Karlsruhe Institute of Technology

ETP-KA/2021-17

Search for the Rare Decay $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ with the Full Event Interpretation at the Belle II Experiment

Laura Frank

Masterthesis

10th December 2021

Institute of Experimental Particle Physics (ETP)

Advisor: Prof. Dr. Torben Ferber
Coadvisor: Prof. Dr. Günter Quast

Editing time: 10th December 2020 – 10th December 2021

Karlsruhe Institute of Technology

ETP-KA/2021-17

Suche nach dem seltenen Zerfall $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ mit der Full Event Interpretation am Belle II Experiment

Laura Frank

Masterarbeit

10. Dezember 2021

Institut für Experimentelle Teilchenphysik (ETP)

Referent:	Prof. Dr. Torben Ferber
Korreferent:	Prof. Dr. Günter Quast

Bearbeitungszeit: 10. Dezember 2020 – 10. Dezember 2021

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 10. Dezember 2021

. . . . . .

(**Laura Frank**)

Contents

1. Introduction

Nowadays, the fundamental description of our nature is given by the Standard Model of particle physics (SM), developed in the early 1970s [1]. The SM comprises both the basic building blocks of matter, the fundamental particles, and three of the four known fundamental forces, which describe the interactions between the fundamental particles. Over the years, this model has been established and probed by many experiments. Despite its great success and remarkable predictions, the SM is still not capable of including the fourth fundamental force, gravity, into the theory. Furthermore, the explanation of Dark Matter (DM) as well as the existence of neutrino oscillation are amongst other challenging problems of the SM.

Thus, the development and extension of the SM has become a main goal of modern physics. Particularly, collider experiments such as the Belle II experiment are powerful way to test the SM and probe possible new physics (NP). The SuperKEKB accelerator in Tsukuba, Japan, is an electron positron collider operating at the center of mass energy of the $\Upsilon(4S)$ resonance. This resonance mainly produces B mesons pairs, therefore SuperKEKB is called a B factory. The Belle II detector is located at the interaction point of the electron and positron beams, and is designed to precisely measure the properties of B mesons. The predecessor Belle experiment (1999-2010) made remarkable discoveries, such as the measurement of CP violation in the B meson system [2].

The rare decay $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ has not been observed yet. The measurement of this decay gives access to important parameters of the B meson. In particular, a parameter called the first inverse moment of the light-cone distribution amplitude, λ_B, is incorporated in the branching fraction of $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ for high energetic photons. This parameter is of great relevance for theoretical predictions for non-leptonic B meson decays, described by quantum chromodynamics (QCD) approaches. Although the experimental techniques and precision have improved in the past, the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ decay is quite challenging. The current limits $\Delta\mathcal{B}(\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma) < 3 \cdot 10^{-6}$ and $\lambda_\mathcal{B} < 240\,\mathrm{MeV}$ at 90% confidence level are reported by the dedicated 2018 analysis of the Belle collaboration [3].

With regard to the forthcoming large data set to be collected by the Belle II experiment, this work resumes and further develops the above mentioned search for the rare decay $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$. In Chapter 2 the fundamental theoretical foundations for this analysis as well as the status of current results is presented. An overview of the experimental setup of Belle II is given in Chapter 3. Chapter 4 describes the software framework and important tools used for this work. The analysis strategy pursued throughout this thesis is explained in

Chapter 5. In Chapter 6 the details of the performed analysis steps are given. The final signal extraction and the underlying methods are discussed in Chapter 7, which furthermore presents a validation of the used fit on real Belle II data. In conclusion, a summary of this work and an outlook is presented in Chapter 8.

2. Theoretical Foundations

This chapter gives a theoretical introduction to the following analysis. The theoretical discussion of the nominal decay of interest $B^+ \to \ell^+ \nu_\ell \gamma$ is presented in Section 2.1. In addition, the decay $B^+ \to \pi^0 \ell^+ \nu_\ell$ is discussed in Section 2.2, as this decay appears as an important background in the analysis of $B^+ \to \ell^+ \nu_\ell \gamma$ decays. Throughout this thesis, all quoted decays of B-mesons imply the charge conjugate processes.

2.1. The Rare Decay $B^+ \to \ell^+ \nu_\ell \gamma$

This analysis focuses on the decay of a charged B meson into a lepton, its respective neutrino, and a photon. Although this radiative leptonic decay $B^+ \to \ell^+ \nu_\ell \gamma$ is often categorised as unwanted background in pure leptonic decays, $B \to \ell^+ \nu_\ell$, its analysis offers great opportunities. In contrast to the pure leptonic case, the helicity suppression is removed from the $B^+ \to \ell^+ \nu_\ell \gamma$ decay due to the emission of the photon. This leads to an increase in the branching fraction for the lighter leptons e^+ and μ^+. Concurrently, the branching fraction is reduced through the addition of the fine structure constant α_{em}.

An important parameter in the decay width of $B^+ \to \ell^+ \nu_\ell \gamma$ decays is the first inverse moment of the light-cone distribution amplitude, λ_B. The light-cone distribution amplitude (LCDA) plays an essential role in the QCD factorisation of non-leptonic B meson decays, which is described in detail in Section 2.3. The following is based on the comprehensive theoretical discussions in [4, 5].

As shown in Figure 2.1, the leading Feynman graph at tree-level is given by the emission of the photon from the light up antiquark. Further higher orders result if the photon is emitted from the heavy bottom quark or the lepton. In the following derivation of the branching fraction, higher order contributions are perceived as radiative and power corrections.

The differential decay width for the $B^+ \to \ell^+ \nu_\ell \gamma$ decay can be calculated through

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{\text{em}} G_F^2 |V_{ub}|^2}{48\pi^2} m_B^4 x_\gamma^3 (1 - x_\gamma)[F_A^2 + F_V^2], \tag{2.1}$$

where α_{em} and G_F are the fine-structure constant and the Fermi coupling constant, quantifying the strength of the electromagnetic and weak coupling, respectively. The matrix element V_{ub} of the Cabibbo–Kobayashi–Maskawa (CKM) matrix indicates the probability of a b quark transforming into an u quark. The B meson mass is denoted with m_B. The

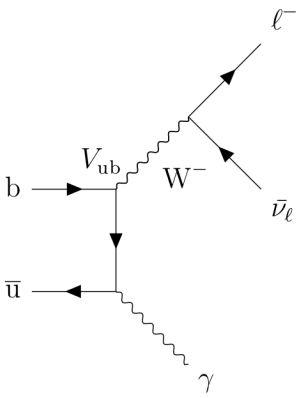


Figure 2.1.: Feynman diagram for the leading order for the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ decay at tree-level. The photon radiates from the up antiquark. Taken from [6].

abbreviation x_γ corresponds to $x_\gamma = 2E_i/m_\mathrm{B}$. The axial and vector form factors are indicated with F_A and F_V, respectively. The QCD calculation of the two form factors assuming photon energies of order $m_\mathrm{B}/2$ gives

$$\begin{aligned} F_\mathrm{V}(E_\gamma) &= \frac{Q_\mathrm{u} m_\mathrm{B} f_\mathrm{B}}{2E_\gamma \lambda_\mathrm{B}(\mu)} R(E_\gamma, \mu) + \left[\xi(E_\gamma) + \frac{Q_\mathrm{b} m_\mathrm{B} f_\mathrm{B}}{2E_\gamma m_\mathrm{b}} + \frac{Q_\mathrm{u} m_\mathrm{B} f_\mathrm{B}}{(2E_\gamma)^2} \right], \\ F_\mathrm{A}(E_\gamma) &= \frac{Q_\mathrm{u} m_\mathrm{B} f_\mathrm{B}}{2E_\gamma \lambda_\mathrm{B}(\mu)} R(E_\gamma, \mu) + \left[\xi(E_\gamma) - \frac{Q_\mathrm{b} m_\mathrm{B} f_\mathrm{B}}{2E_\gamma m_\mathrm{b}} - \frac{Q_\mathrm{u} m_\mathrm{B} f_\mathrm{B}}{(2E_\gamma)^2} + \frac{Q_\ell f_\mathrm{B}}{E_\gamma} \right]. \end{aligned} \tag{2.2}$$

The Q_i refers to the charge of the respective particle i ($i = \mathrm{u}, \mathrm{b}, \ell$), and f_B denotes the decay constant of the B meson. Furthermore, the above-mentioned first inverse moment of the LCDA, $\lambda_\mathrm{B}(\mu)$, appears in the denominator of the fraction, where μ indicates the energy scale. The radiative correction factor $R(E_\gamma, \mu)$ depends on the photon energy calculated at an energy scale μ. Thus, the first term represents the leading order contribution for the photon emission from the light u quark (denoted as the fraction), corrected with $R(E_\gamma, \mu)$ for higher-order radiations. Both F_A and F_V show the same leading power contribution in the heavy quark expansion. The additional terms in brackets correspond to $1/m_\mathrm{b}$ power corrections for higher orders. The symmetry preserving term $\xi(E_\gamma)$ is equal for both form factors, whereas the remaining symmetry breaking terms differ. The term containing Q_b describes the suppressed photon emission from the bottom quark. The Q_u term gives a correction for the emission from the light up quark, since this is suppressed compared to the leading order emission from an up antiquark. Considering F_A, the additional last term with Q_ℓ describes the correction for the emission from the lepton.

Since Equation (2.2) is only valid for the requirement $E_\gamma \sim m_\mathrm{B}/2$, a lower limit for E_γ is considered. For this analysis, $E_\gamma > 1\,\mathrm{GeV}$ is chosen, as highly recommended in [4]. By integrating the differential width in Equation (2.1) over the selected photon energy range and multiplying by the B meson lifetime, τ_B, the resulting partial branching fraction reads

$$\Delta\mathcal{B}(\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma) = \tau_\mathrm{B} \int_\mathrm{Selection} \frac{d\Gamma}{dE_\gamma} dE_\gamma \,. \tag{2.3}$$

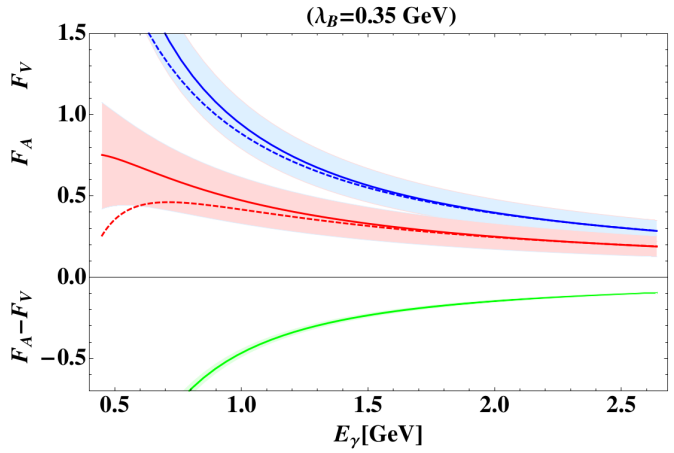


Figure 2.2.: Dependence of the axial and vector form factor, F_A (red) and F_V (blue), as well as their difference, $F_A - F_V$ (green), on the photon energy E_γ. The form factors are calculated for an estimated value of the first inverse moment of the light-cone distribution amplitude, $\lambda_\mathrm{B} = 0.35\,\mathrm{GeV}$. The solid line indicates the resulting form factors assuming $\mu = m_\mathrm{b}$, the dashed line assuming $\mu = 2E_\gamma$. The bands describe the uncertainties. Taken from [4].

The most recent experimental upper limit of the partial branching fraction of $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ decays was achieved by the Belle II collaboration [6]. The result is $\Delta\mathcal{B} < 3 \cdot 10^{-6}$ at 90% confidence level [6]. The prospects of a significant observation of the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ decay at the current Belle II II experiment are promising, especially for the full expected data set of $50\mathrm{ab}^{-1}$ [7]. Theoretical considerations for $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ decays are exploited in detail in [5, 8].

2.2. The Decay $\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell$

Decays, which do not solely produce a lepton and its corresponding neutrino but also one or more hadrons, are called semileptonic. Semileptonic decays are studied intensively, as they play an important role in the determination of the CKM matrix elements V_cb and V_ub. In this work, the semileptonic decay $\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell$ is used as a control channel for the nominal $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ analysis. Reference [9] gives the differential decay width in terms of the four-momentum transfer $q = p_\mathrm{B} - p_\pi$ to the leptonic system, where p_B and p_π indicate the four-momentum of the B meson and the π, respectively. The decay width reads

$$\begin{aligned}\frac{d\Gamma(\mathrm{B} \to \pi \ell \nu_\ell)}{dq^2} = &\frac{G_\mathrm{F}^2}{24\pi^3 m_\mathrm{B}^2 q^4} |V_\mathrm{ub}|^2 (q^2 - m_\ell^2)^2 |\vec{p}_\pi| \\ &\times \left[\left(1 + \frac{m_\ell^2}{2q^2}\right) m_\mathrm{B}^2 |\vec{p}_\pi|^2 |f_+(q^2)|^2 + \frac{3m_\ell^2}{8q^2} (m_\mathrm{B} - m_\pi)^2 |f_0(q^2)|^2 \right],\end{aligned} \tag{2.4}$$

where m_ℓ denotes the lepton and m_π the pion mass, respectively. Two form factors are introduced, the scalar $f_0(q^2)$ and the vector $f_+(q^2)$. The magnitude of the pion momentum

in the B meson rest frame is indicated with $|\vec{p}_\pi|$. In the case of light leptons, terms depending on m_ℓ become negligible, so that Equation (2.4) reduces to

$$\frac{d\Gamma(\mathrm{B} \to \pi\ell\nu_\ell)}{dq^2} = \frac{G_\mathrm{F}^2}{24\pi^3}|V_\mathrm{ub}|^2|f_+(q^2)|^2|\vec{p}_\pi|^3 \ . \tag{2.5}$$

As a consequence, the remaining relevant form factor is solely $f_+(q^2)$. According to [10], the parametrisation

$$z(t, t_0) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}} \tag{2.6}$$

is used to map the variable q^2, where $t_\pm = (m_\mathrm{B} \pm m_\pi)$ and $t_0 = (m_\mathrm{B} + m_\pi)(\sqrt{m_\mathrm{B}} - \sqrt{m_\pi})^2$ are chosen. The form factor $f_+(q^2)$ transforms to the so-called Bourrely-Caprini-Lellouch expansion [11]

$$f_+(q^2) = \frac{1}{1 - q^2(z)/m_{\mathrm{B}^*}^2} \sum_{n=0}^{N_z-1} b_n \left[z^n - (-1)^{n-N_z} \frac{n}{N_z} z^{N_z} \right], \tag{2.7}$$

where N_z indicates the number of expansion parameters and m_{B^*} denotes the B^* meson mass. The expansion parameters are determined by fitting z to the experimental results from Belle and BaBar combined with theoretical lattice-QCD calculation. For a detailed description as well as the resulting form factors and theoretical prediction of V_ub, the reader is again referred to [10]. Latest research results on the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ decay as well as on the determination of V_ub achieved by the Belle II collaboration can be found in [12].

2.3. Non-leptonic B Meson Decays

This section gives an introduction to non-leptonic B meson decays, i.e. decays which do not show any lepton as a final state product. Considering a B meson decay into a pair of light mesons, M_1 and M_2, the amplitude is given by

$$\mathcal{A}(\mathrm{B} \to M_1 M_2) = \frac{G_\mathrm{F}}{\sqrt{2}} \sum_i \lambda_i C_i(\mu) \langle M_1 M_2 | \mathcal{O}_i | \mathrm{B} \rangle (\mu) \ . \tag{2.8}$$

The λ_i indicates the CKM factor, which is multiplied by the coefficient function $C_i(\mu)$ and a matrix element $\langle M_1 M_2 | \mathcal{O}_i | \mathrm{B} \rangle (\mu)$, with a local operator $\mathcal{O}_i$. Since the calculation of this matrix element is non-trivial, the computation is often reduced to approximations. Therefore, the ansatz of QCD factorization (QCDF) is introduced. The QCDF approach aims to convert the matrix element in terms of experimentally measurable form factors, hadronic light-cone distribution amplitudes, and hard scattering functions. In the context of the QCDF scheme in [13], the QCDF is defined as the separation between long-distance (soft) and short-distance (hard) effects. The latter is accessible via pertubative expansions, whereas the soft contribution must be calculated in non-pertubative or experimental determinations. Often, the non-pertubative parameters are easier to calculate, due to a simpler structure compared to the original quantity. In the following, two compositions of the meson pair are

Table 2.1.: Overview of λ_B values, considering the theoretical (QCDF and QCD sum rules) predictions and the experimentally determined limit by the Belle Collaboration.

	λ_B (MeV)
QCDF [14, 15]	≈ 200
QCD sum rules [16]	460 ± 110
Belle [6]	> 240 (90%CL)

distinguished, where the meson pair consists either of a heavy and a light meson or of two light mesons. In the former case of a heavy meson M_1 and a light meson M_2 with masses m_1 and m_2, respectively, the factorised matrix element reads

$$\langle M_1 M_2 | \, \mathcal{O}_i \, | \overline{\mathrm{B}} \rangle = \sum_j F_j^{\mathrm{B}\to M_1}(m_2^2) \int_0^1 du \, T_{ij}^I(u) \, \Phi_{M_2}(u) \ , \tag{2.9}$$

whereas in the latter case, the factorised matrix element is given by

$$\begin{aligned} \langle M_1 M_2 | \, \mathcal{O}_i \, | \mathrm{B} \rangle = & \sum_j F_j^{\mathrm{B}\to M_1}(m_2^2) \int_0^1 du \, T_{ij}^I(u) \, \Phi_{M_2}(u) \\ & + \sum_j F_j^{\mathrm{B}\to M_2}(m_1^2) \int_0^1 dv \, T_{ij}^I(v) \, \Phi_{M_1}(v) \\ & + \int_0^1 d\xi \, du \, dv \, T_i^{II}(\xi, u, v) \, \Phi_{\mathrm{B}}(\xi) \, \Phi_{M_1}(v) \, \Phi_{M_2}(u) \ . \end{aligned} \tag{2.10}$$

The corresponding form factors of the transition $\mathrm{B} \to M_{1,2}$ are denoted with $F_j^{\mathrm{B}\to M_{1,2}}(m_{1,2})$, resepectively. The $T_{ij}^I(u)$ and $T_i^{II}(\xi, u, v)$ indicate hard scattering functions, which depend on the fraction of the meson momenta carried by the light quarks, ξ, u, v. The pertubative calculation of these short-distance functions can be performed by an expansion of the strong coupling constant $\alpha_s(m_\mathrm{b})$. Furthermore, the light-cone distribution amplitude (LCDA) Φ_X is introduced, which describes the quark-antiquark Fock state of meson X, where $X = M_1, M_2, \mathrm{B}$. As already mentioned in Section 2.1, the first inverse moment of the light-cone distribution amplitude of the B meson, λ_B, can be accessed via the branching fraction of the rare decay $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$. The corresponding B meson LCDA, Φ_B, and λ_B are related by

$$\frac{1}{\lambda_B} = \int_0^\infty \frac{d\xi}{\xi} \Phi_\mathrm{B}(\xi) \ . \tag{2.11}$$

Thus, the determination of λ_B reveals insights to the QCDF and probes its theoretical predictions. Yet, the experimental and theoretical predictions of λ_B differ in a wide range. An overview of the predicted values by QCDF, the approach using QCD sum rules, and the experimental lower limit found by the Belle Collaboration, is given in Table 2.1. Since the determination of λ_B via the semileptonic decay $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ is a promising method, this has been subject of several researches at various e^+e^- colliders in the last years [6, 17, 18]. Due to the expected large data set collected by the Belle II experiment, the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ decay could be observed soon. Thus, the prospect of an experimental determination of λ_B is promising. Furthermore, theoretical considerations for QCDF optimisation [5] and other approaches for the calculation of λ_B could improve the result [19].

3. The Belle II Experiment

Since this work is based on the Belle II experiment, this chapter gives a short introduction to the SuperKEKB accelerator in Section 3.1 and the Belle II detector system in Section 3.2.

3.1. SuperKEKB

The SuperKEKB collider is an electron positron collider located at the High Energy Accelerator Research Organisation (KEK) in Tsukuba, Japan. As an upgrade of the former KEKB accelerator, SuperKEKB targets a luminosity 40 times greater than its predecessor. This section is based on [20]. The first collisions were recorded in 2018. A schematic overview of the collider is presented in Figure 3.1. A linear accelerator (linac) with a damping ring for positrons is used to inject the electrons and positrons into respective storage rings. The electrons are stored in the high energy ring (HER) at 7 GeV, while the positrons are kept at 4 GeV in the low energy ring (LER). In total, each ring measures a circumference of 3016 m. Four experimental halls exist, named Oho, Nikko, Tskubua, and Fuji hall. The only collision point of the beam is located at Tsukuba hall. In the case of a regular run, the electrons and positrons are kept at an energy of 7 GeV and 4 GeV, respectively. Therefore, the accelerator operates at the $\Upsilon(4S)$ resonance at a center of mass energy of $\sqrt{s} = 10.58\,\text{GeV}$, which further decays in $> 96\%$ of the time in a B meson pair [21], $e^-e^+ \to \Upsilon(4S) \to B\overline{B}$. Thus, SuperKEKB is a typical B factory. Operations on other resonances ($\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(5S)$) as well as off-resonance runs are possible.

3.2. The Belle II Detector

The Belle II detector consists of several layers, centered around the collision region, to detect as many particles as possible. The sub-detector parts are described in the following, based on [22]. A schematic overview of the detector is shown in Figure 3.2. As can be seen, the detector is asymmetric, due to the different beam energies and the corresponding boost along the beam axis.

The innermost detector part is the two-layer pixel detector (PXD), consisting of silicon pixel sensors. At the time of this analysis, the data was recorded with a single layer PXD. A new two-layer PXD will be installed during the first long shutdown planned within the next years. The PXD is surrounded by the silicon vertex detector (SVD), comprising four layers of double-sided silicon strip sensors. Both PXD and SVD sub-detectors together are

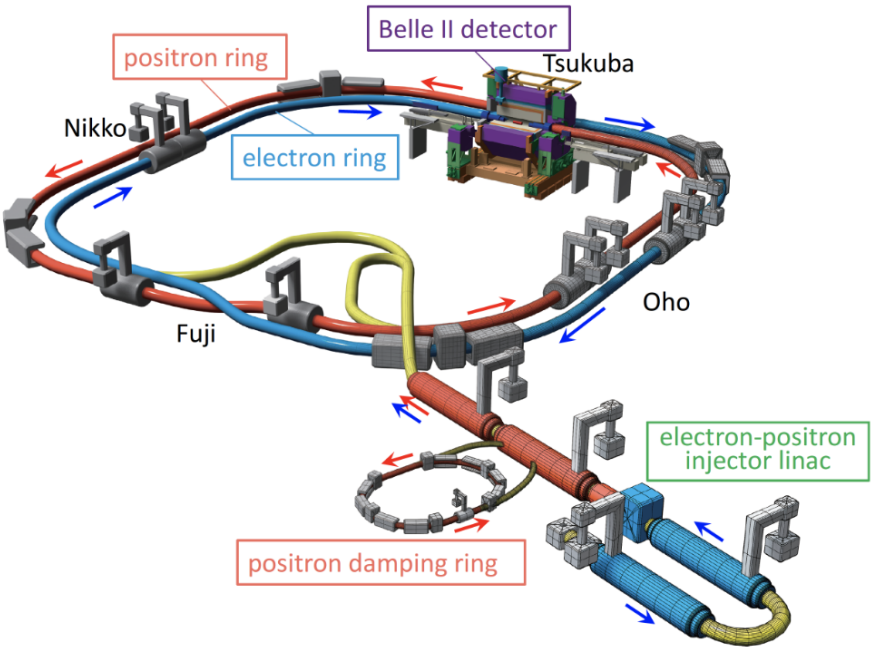


Figure 3.1.: Illustration of the SuperKEKB collider. Starting from the linac with the positron damping ring, the electron and positron storage rings are shown in blue and red, respectively, the beam directions are marked by arrows. All four experimental halls (Oho, Nikko, Tskubua, Fuji) are inscribed. Furthermore, the Belle II detector is presented. Taken from [20].

often referred to as the vertex detector (VXD). The VXD information is used to precisely reconstruct decay vertices. The next layer is the central drift chamber (CDC), filled with a Helium-Methane gas and sense wires. The CDC provides information on the particle trajectories and momenta. As for the identification of charged particles, the Time-Of-Propagation (TOP) in the barrel region as well as the Aerogel Ring-Imaging Cherenkov (ARICH) detector in the forward end-cap use timing and pattern information of Cherenkov effects to discriminate particles. The following electromagnetic calorimeter (ECL) gives precise energy measurements for photons and electrons. Between the ECL and the next layer, a superconducting coil generates a magnetic field with a strength of 1.5 T. Thus, the charged particles inside the magnetic field are forced on curved trajectories, which enables the determination of their momenta. The outermost K_L^0 and muon detector (KLM) detects long living neutral Kaons and muons.

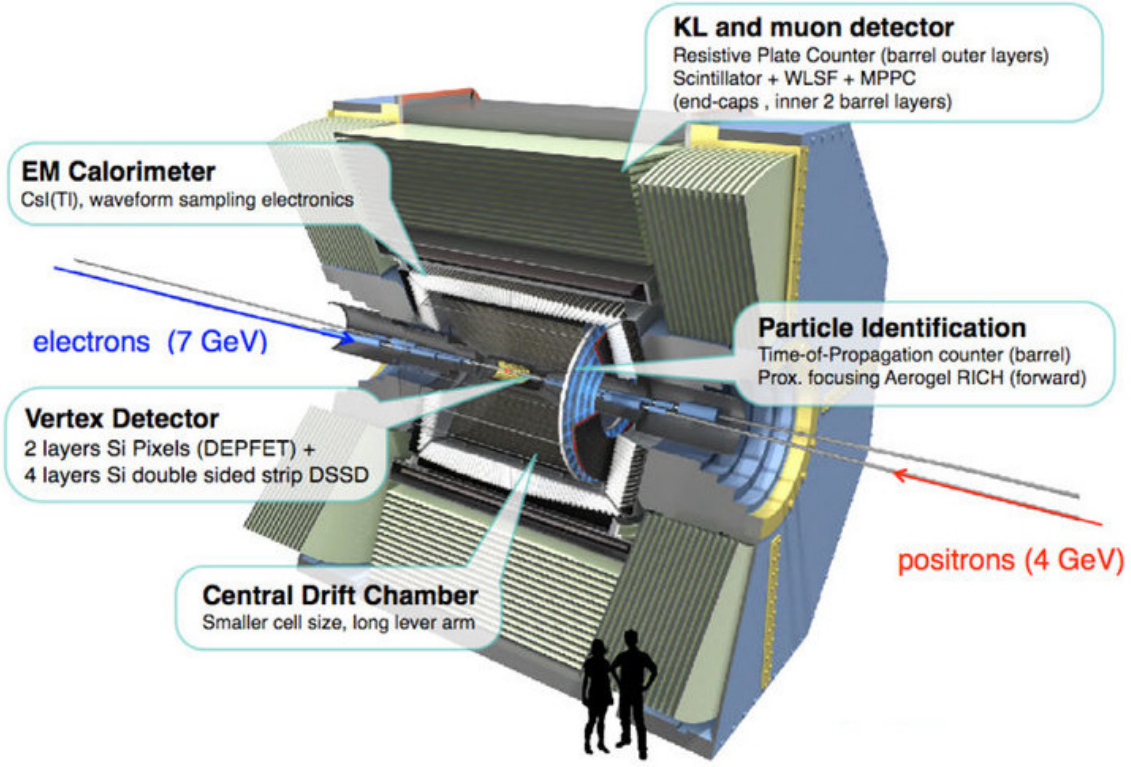


Figure 3.2.: Overview of the Belle II detector. The sub-detector components are indicated. Furthermore, the electron and positron beams with the respective beam energies in a regular run at the $\Upsilon(4S)$ resonance are presented in blue and red. As can be seen, the detector is asymmetric. Taken from [23].

4. Software Framework and Tools

The following sections give an overview of the most important software and tools essentially used in this analysis. More precisely, the Belle II specific software framework is introduced in Section 4.1. Since an analysis should be evolved without any bias with regard to the result, this work is performed on simulated Monte Carlo data, which is described in Section 4.2. A reconstruction method called tagging is presented in Section 4.3. Section 4.4 provides general information on machine learning techniques, which play an essential role for the exclusive tagging algorithm called *Full Event Interpretation*. The introduction of this tagging approach is subject of the last Section 4.5.

4.1. The Belle II Software Framework

Since the Belle II experiment collects an enormous amount of data, a reliable software framework is needed. The `Belle II Analysis Software Framework` (*basf2*) provides tools not only to handle data-taking but also for the generation of simulated data, the reconstruction of tracks and clusters, and high-level offline analysis.

The software architecture is composed of modules, paths, and packages, which are combined in a python based configuration file, the steering file. The *basf2* modules are small processing blocks, written in C++, and perform specific tasks independently. Their functionality is outsourced to shared libraries to ensure an independent and reusable usage of algorithms, such that different modules can access the same code. The steering file is used to configure these modules and add them in a linear chain to a path. The path contains all required modules, which are organized in a strict order. Multiple connected paths in a steering file are possible. [24]

By processing a path, all modules inside the path are executed one after another in a strict linear order. During the processing of a path, all data is solely exchanged with the data storage (DataStore), i.e. a module reads data from and writes the processed data back to the DataStore. [25] Some modules are logically organized in categories, so-called packages, which often include some python scripts to simplify the handling of modules.

4.2. Monte Carlo Simulation and MC Matching

Most analyses are initially studied on simulated data to optimise the process and take advantage of knowing the true information as well as to avoid a biased analysis. A reliable

data simulation in high energy physics (HEP) is provided by Monte Carlo (MC) data. As for Belle II, the MC events are generated using different event generators, such as the `EvtGen` package [26] and `PYTHIA8.2` [27]. Afterwards, the propagation of the generated particles in the detector is simulated, especially the interaction of final state particles with the sub-detectors. This task is performed by the `Geant4` package [28]. Although `Geant4` cannot simulate every detector component perfectly due to a limited precision in the knowledge of the exact hardware composition and unknowns in the underlying physics processes, the output should be comparable to what we expect from real detector measurements.

To connect the generated and simulated events, and consequently exploit the information that a particle is reconstructed correctly or not, so-called MC Truth Matching can be applied.

4.3. B Meson Tagging

In contrast to pile-up collisions at hadron colliders, B-factories offer the great advantage of a well-known initial state of the $\Upsilon(4S)$ resonance. In more than 96 % of the time the $\Upsilon(4S)$ decays in either a neutral or a charged $B\overline{B}$ pair [21]. The B mesons decay via various decay channels and the resulting final state particles are measured as tracks and clusters by the Belle II detector. By combining the final state particles hierarchically to intermediate states, the decay chains of the initial B mesons can be reconstructed.

In particular, one B meson, called B_{sig}, is reconstructed in the desired signal decay. A challenging component, especially in rare decays such as $B^+ \to \ell^+ \nu_\ell \gamma$, are undetectable neutrinos as final state particles. However, one can deploy the additional information provided by the well-known four-momentum of the initial $\Upsilon(4S)$, and thus be able to interpret the full event by considering the accompanying B meson, the so-called B_{tag}. More precisely, this method is referred to as tagging. Considering that a set of final state particles is assigned to the signal-side by reconstructing the B_{sig}, all remaining tracks and clusters are called the rest-of-event (ROE). An illustration of the tagging concept is presented in Figure 4.1.

The ROE should be compatible with all particles from the tag-side, if the event is correctly reconstructed. Therefore, combining the B_{tag} meson and the B_{sig} meson no additional tracks and clusters should appear in the event, otherwise the consequently wrongly reconstructed event is rejected as intrusive background. This requirement is known as the completeness constraint [30]. In general, the tagging performance can be evaluated with three properties:

- **Tagging efficiency** quotes the fraction of tagged B mesons relative to the total number of B mesons.
- **Tag-side efficiency** gives the fraction of correctly reconstructed tagged B mesons relative to the total number of B mesons.
- **Tag-side purity** states the fraction of correctly reconstructed B mesons relative to the number of tagged B mesons, therefore indicating the tagging quality.

As for the B-tagging methods, two different approaches are commonly used, namely inclusive and exclusive tagging. In case of inclusive tagging no specific B_{tag} decay mode is required

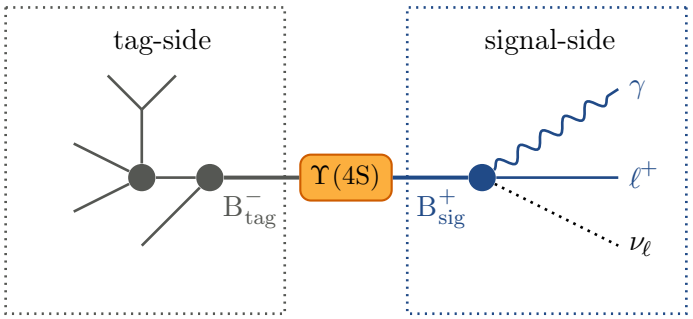


Figure 4.1.: Illustration of the tagging method, applied for the $\Upsilon(4S)$ resonance decaying into a B meson pari. The $B^+ \to \ell^+ \nu_\ell \gamma$ decay is chosen as the signal B meson, B^+_{sig}. The corresponding particles are shown on the right side and to the signal side. The neutrino is marked with a dashed line, since it is not detectable. The second B meson on the left, B^-_{tag}, refers to the tag-side. Adapted from [29].

but rather all four-momenta remaining after the signal-side selection are combined without explicitly reconstructing a B_{tag} decay chain. Consequently, for inclusively tagged events the tagging algorithm always provides a valid B_{tag} in any case. Despite the resulting high tagging efficiency (theoretically up to 100%), this method suffers from an immense impurity [29].

In contrast to inclusive tagging, the exclusive approach requires the reconstruction of certain B_{tag} decay modes. The tag-side efficiency is highly constrained by the limited number of reconstruction decay modes in comparison to the huge multiplicity and the large amount of possible decay channels. This leads to a low efficiency of a few percent, which is furthermore worsened by the possibly incorrect track and cluster reconstruction. Hence, decay channels with a high branching fraction are preferred as exclusive B_{tag} modes. Although the exclusive tagging algorithm has a low efficiency, it provides the advantage of a high purity.

Depending on the signal channel of interest, either a hadronic or a semileptonic approach for the exclusive tagging method is chosen. The former uses purely hadronic decays for the tag-side reconstruction, while the latter considers semileptonic modes containing a lepton-neutrino pair.

Conceptually, the good tag-side purity for the hadronic case is advantageous, since the reconstruction without any undetectable neutrinos is quite precise. Particularly B_{sig} decays including a neutrino, such as $B^+ \to \ell^+ \nu_\ell \gamma$, benefit from a well-known hadronic tag-side, as the information loss is minimized. Nevertheless, the small hadronic branching fractions decrease the number of tagged events and therefore limit the tag-side efficiency.

Conversely, semileptonic tagging provides a lower purity due to missing kinematic information, but the abundance of semileptonic decays and their higher branching fractions compared to hadronic decays enhances the efficiency. An overview of the different approaches is presented in Figure 4.2, where the three methods are ranked according to their efficiency and purity.

Given these points, one has to consider the previously discussed pros and cons with regard to the signal model and the studied variables of interest.

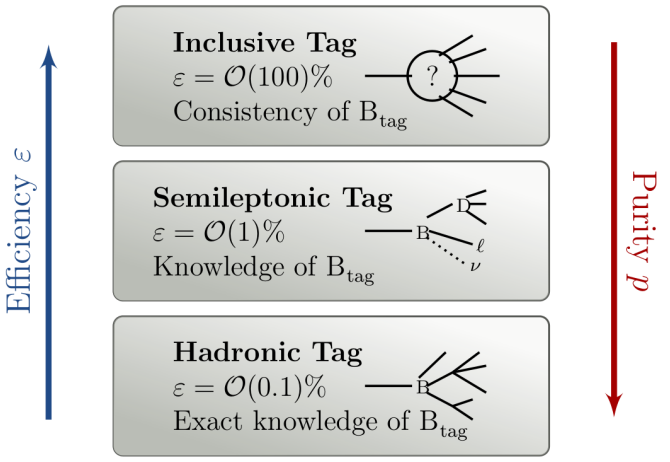


Figure 4.2.: Overview of the three different tagging approaches and their advantages and disadvantages. Taken from [29].

4.4. Multivariate Classifier and Boosted Decision Trees

Many tasks and challenges in HEP find excellent application for machine learning (ML) techniques. More precisely, the vast amount of not only experimental but also simulated data gives both necessity and advantage of using data-driven algorithms. Especially for this work, multivariate analysis (MVA) methods play an important role to reconstruct event candidates and separate signal from background events. The following gives an introduction to MVA classifiers and is based on [31–33].

Given a certain data set, a multivariate classifier considers multiple significant features $\vec{x} = \{x_1, ..., x_d\}$ and their correlations to distinguish between signal and background, which is indicated by a class label y set to $y = 1$ and $y = 0$, respectively. The class label can be interpreted as a target information, notably provided either by Monte Carlo simulations or by data-driven methods.

Conceptually, the classification process is divided in two phases:

- **The fitting phase** prepares the statistical model of the classifier by adjusting its internal parameters. The goal of this task is that the model automatically learns from experience and knowledge. In general, the parameters are trained on a provided training data sample.
- **The inference phase** is intended for the application of the learned model to an independent validation or test data-set. Ideally, the classifier is able to deduce the class labels of each event in the new sample. The capability of a good prediction is known as generalisation [31].

Furthermore, regarding the provision of the target information during the fitting phase, three different machine learning techniques exist. If not only the input variables, but also the target information are handed to the classifier, the process is called supervised learning. In the case of unsupervised learning, the classifier does not receive any target information

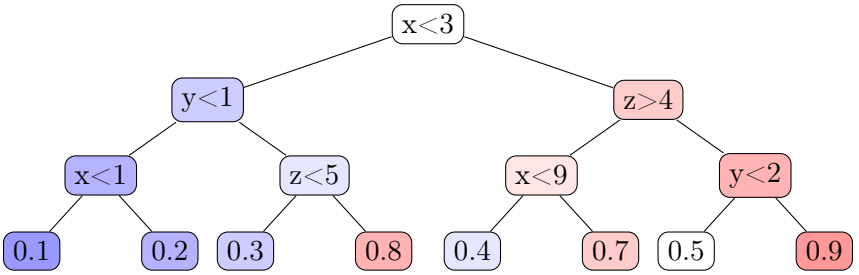


Figure 4.3.: Schematic overview of a DT with three layers, starting from the top and ending in the terminal nodes at the bottom. At each node, a binary decision is made. Adapted from [33]

and therefore is not explicitly taught. This approach is widely used in clustering and density estimation problems. Another technique is reinforced learning, which is a more interactive method based on rewards for desired behaviours and penalties for undesired performances.

Depending on the model complexity, mainly influenced by the number of degrees of freedom, the classification performance suffers from different kind of errors. On the one hand, in case of a too complex and large model or a too small training set compared to the internal parameters, the risk of overfitting occurs; although the training data is perfectly fitted, the generalisation error of the test sample will be large. On the other hand, underfitting is a consequence of a too simple model. This model is often biased and not able to correctly fit new data.

Since classification problems, such as the distinction between correctly reconstructed $B^+ \to \ell^+ \nu_\ell \gamma$ signal events and misidentified background events in this analysis, are most conveniently handled by supervised learning, the following will focus on the supervised approach. As for HEP, a further advantage on the subject of supervised learning is given by the fact that training data can easily be generated by simulation. More detailed information on the other techniques can be found in [32, 34].

Widely used algorithms in supervised learning are tree-based models, especially Decision Trees (DT). In its simplest form, a DT successively splits (or explicitly does not split) nodes by making binary decisions for a data-point along a sequence of selections. A visualisation of a DT is shown in Figure 4.3. Hence, a tree-like structure with a depth limited by the number of consecutive cuts is built. The procedure is executed for each data-point in the sample. The output is a signal probability of each point. Unfortunately, a simple DT classification often suffers from statistical fluctuations and, in consequence, overfitting.

A more stable and accurate way of prediction is provided by boosting. The concept of a boosted decision tree (BDT) relies on the gathered output of several similar classifiers. Even though these classifiers can be simple DTs and therefore "weak" learners with a rather imprecise classification, the "committee" of all trees can perform significantly better and minimize the overall classification error. In particular, the single classifiers are trained sequentially on a weighted form of the training set, in which each data point is assigned a weighting coefficient resulting from the performance of the previous weak classifier. Consequently, challenging data points are re-evaluated with a higher count and gain impact [31].

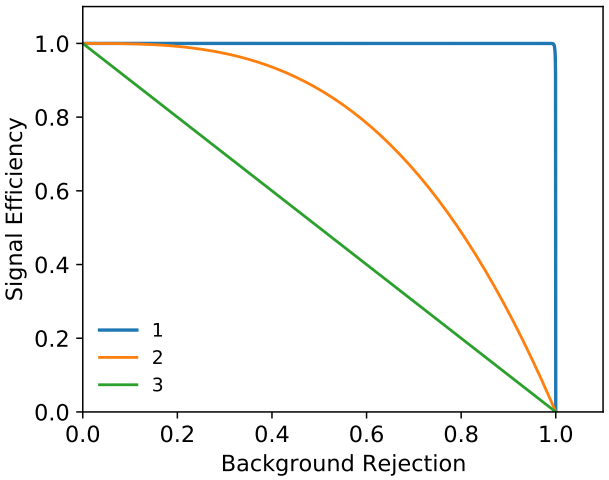


Figure 4.4.: Schematic ROC curves. Three different curves for the signal efficiency plotted over the background rejection are given. The blue ROC curves indicates an almost perfect classifier, whereas the green ROC curve corresponds to a classifier without any separation skill. The orange ROC curve represents a classifier with discrimination power, but no perfect separation.

Furthermore, the classification can be optimised by gradient descent. Gradient descent is an optimisation algorithm, which aims for the local minimum of a certain function, e.g. the loss-function. The loss-function indicates the performance of the model's coefficients when it comes to fit the given data. Given that the weights of the function are adjusted incrementally by moving the weight vector towards the steepest decrease in the error function, the loss is minimised. An even more powerful approach is a Stochastic Gradient Boosted Decision Tree (SGBDT) [35]. In this case, a certain level of randomness is added by using randomly drawn subsamples of the full training sample. This procedure strengthens the algorithm against overfitting.

As mentioned in Section 4.1, the analysis framework for Belle II is *basf2*, consisting of a variety of tools and functions. Regarding multivariate methods, an `mva` package is part of the *basf2* framework. The package can not only handle the required tasks for the fitting, but also the inference on test data as well as the evaluation of the results. As backend, different MVA frameworks are accessible. Here, the multivariate classifier `FastBDT` is chosen. `FastBDT` provides a fast fitting and application performance in terms of CPU time and a robust prediction quality. Furthermore, `FastBDT` supports negative weights and missing values and returns an estimation of the feature importance. The implementation and validation of `FastBDT` is further described in [33].

A commonly used indicator of a classifier is the Receiver Operating Characteristic (ROC) curve. With regard to this analysis, the ROC curve is constructed as the signal efficiency plotted against the background rejection. The corresponding Area Under Curve (AUC) characterises the classifier performance, where a AUC score close to one (0.5) indicates a reliable (random) separation of signal and background. Exemplary ROC curves are shown in Figure 4.4.

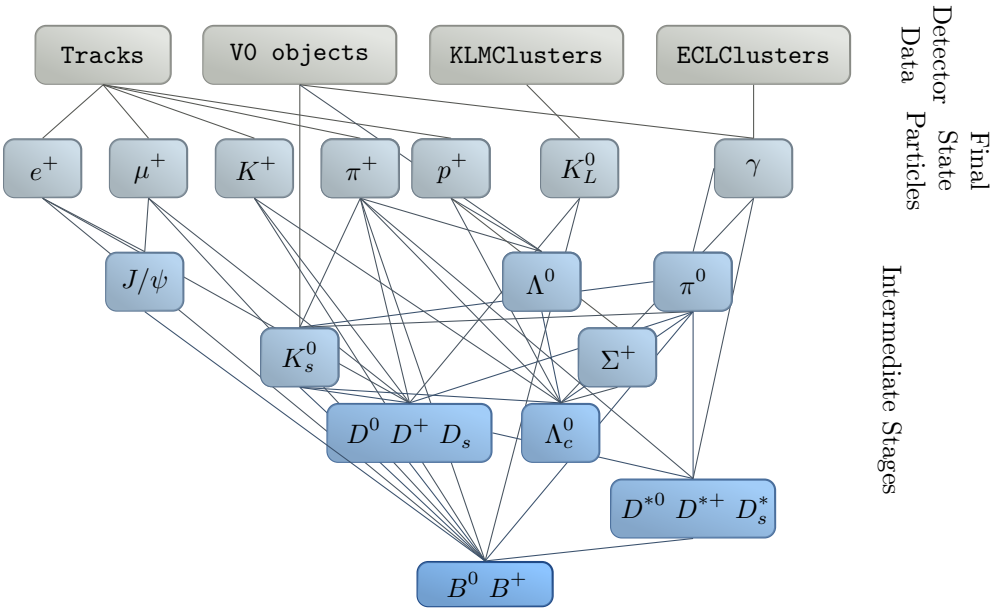


Figure 4.5.: Schematic illustration of the hierarchical structure of the FEI. In a first step, all detector information is collected (tracks, displaced vertices V0, clusters). Secondly, final state particles ($e^+, \mu^+, K^+, \pi^+, K_L^0, \gamma$) are built as combination of the gathered detector data. In further stages, intermediate particles are formed and combined (J/ψ, π^0, Λ, Σ), up to the terminal $B\overline{B}$ stage. Adapted from [29].

In conclusion, the use of multivariate classifiers is a powerful method in high energy physics and the mva package is extensively used throughout this thesis. Furthermore, the mva package is extensively used at Belle II in many offline and online reconstruction tasks, such as particle identification and tagging algorithms.

4.5. The Full Event Interpretation

In an ideal world, the previously described tagging method is able to fully reconstruct the entire event and hence provide an exact interpretation of all information gained by detected (or simluated) tracks and clusters. The so-called Full Event Interpretation (FEI) is an exclusive tagging algorithm based on machine learning techniques [29]. The FEI was originally developed for Belle II and is therefore included in *basf2* and provides an exclusive tagging approach, either reconstructing hadronic or semileptonic B_{tag} candidates. A schematic overview of the FEI is shown in Figure 4.5. Following a hierarchical structure, the FEI algorithm covers six stages. The basis is formed by reconstructed tracks, ECL clusters, and KLM clusters, as well as V0 objects, which are track pairs with a common vertex. At first, this collected detector data is assigned to final state particles (e^+, μ^+, K^+, π^+, p, K_L^0, γ). In the next step, the final state particle candidates form intermediate particles, in particular J/ψ, π^0, and Λ. Subsequently, further intermediate stages are created in three additional categorical stages by using the intermediate particles of the preceding

stages. This includes K_S^0 and Σ^+, various D and Λ_c^+, and various D^* particles. Finally, the B candidates are reconstructed, either being correctly or wrongly combined, and consequently categorised as signal or background, respectively.

At each stage, a multivariate classifier based on `FastBDT` (see Section 4.4) is trained for each final state or intermediate particle. As real-valued output of the classifier, a signal probability is assigned. Every classifier considers the probability estimates of the corresponding previous particles in the hierarchy as input features. Kinematic and vertex fit information are regarded as additional input features. The network of multivariate classifiers results in a final probability estimate $\mathcal{P}_{\mathrm{FEI}}$ of the B meson candidate. Since multiple decay channels for each intermediate state particle exist, no unique decay chain is given but rather various B_{tag} candidates are provided. By using roughly 100 explicit decay channels, the `FEI` correspondingly leads to more than 10000 decay chains. On the subject of combinatorics, the number of tracks and clusters factorises the range of all possible combinations of B meson candidates.

To reduce the amount of incorrectly reconstructed candidates and hence scale down computing time, one can apply loose pre-selections (before the multivariate classification) on fast accessible quantities, e.g. the photon energy or the beam constrained mass for hadronic B mesons. The beam constrained mass M_{bc} is defined as

$$M_{bc} = \sqrt{E_{beam}^2 - \mathbf{p}_B^2}, \tag{4.1}$$

and E_{beam} indicates the beam energy and $\mathbf{p}_B$ denotes the momentum of the B_{sig}. Additionally, the remaining candidates are ranked with respect to this quantity. After the execution of all computational expensive tasks including the multivariate classification, a loose fixed selection (post-selection) on the signal probability reduces the number of possible candidates. Furthermore, the candidates which survived the selection criteria are ranked according to their signal probability. As a tighter post-selection, only a certain number of candidates over all decay channels is considered, forming a best candidate selection.

In general, the `FEI` is trained on MC data. Considering the training method, different approaches are possible:

Generic `FEI` uses generic MC events as training data, with no regard to any signal-side selection. After training the generic `FEI` independently, it is applicable to any signal decay of interest. Usually, the generic `FEI` is trained centrally in each MC production campaign by the collaboration.

Specific `FEI` depends on a certain signal-side. Initially, the visible B_{sig} daughters are reconstructed, leaving the ROE as training space for the specific `FEI`. Since the training is performed on a signal-adjusted background, this method is advantageous for constraining non-trivial background.

Obviously the `FEI` is applied on the $\Upsilon(4S)$ resonance as preferred for B-factories, but can be modified with respect to other resonances, e.g. $\Upsilon(5S)$ as discussed and applied in [36].

5. Analysis Strategy

This chapter describes the pursued strategy chosen for this $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ analysis. An overview of the analysis steps can be found in Section 5.1. In addition, the variable used for the evaluation of the signal extraction is presented in Section 5.2.

5.1. Overview

The goal of this analysis is to measure the partial branching fraction $\Delta\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma)$ with a photon energy of $E_\gamma > 1\,\mathrm{GeV}$. In addition, the first inverse moment of the light-cone distribution amplitude, described in Chapter 2, can be determined by using the calculated $\Delta\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma)$. To achieve this goal, an analysis strategy is pursued. The milestones of the strategy are visualised in Figure 5.1. First of all, the tag-side is reconstructed with a generic hadronic `FEI`, described in Section 4.5. Afterwards, the signal-side $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ is reconstructed by combining final state particles. After the combination of selected signal- and tag-side candidates, the background is suppressed. The background consists of falsely reconstructed events, which initially originate from decay channels other than $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$, that are mistakenly reconstructed as such. The contributing background components in this analysis are grouped in $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ events, $\mathrm{B}^+ to \rho^0\ell^+\nu_\ell$ events, other resonant and non-resonant $\mathrm{B}^+ \to X_\mathrm{u}\ell^+\nu_\ell$ events, $\mathrm{B}^+ \to X_\mathrm{c}\ell^+\nu_\ell$ events, continuum events, events occurring through indirect leptons, and a remaining component, which corresponds to other $\mathrm{B\overline{B}}$ processes not mentioned before. Aiming for a high signal retention, the unwanted background is rejected by applying selection criteria (Section 6.2) and special background suppression tasks. More precisely, this analysis includes special suppression tasks for continuum background events, as well as for so-called peaking background events. The first one arises from e^+e^- collisions resulting in a quark-antiquark pair instead of the desired $\Upsilon(4\mathrm{S})$ resonance. The latter is an analysis-specific background, whose events show similar characteristics as the ones expected for the signal decay. In this $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ analysis, the peaking background mainly consists of the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ decay. Both continuum and peaking background are described in detail in Section 6.3.

Since this analysis is developed on Monte Carlo data, some corrections are applied to ensure a more precise agreement between real data and MC data. As mentioned above, the peaking background is dominated by the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ background. Although the peaking background is reduced during the background suppression task, a fraction of $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ decays remains. Therefore, the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ channel is used as a control

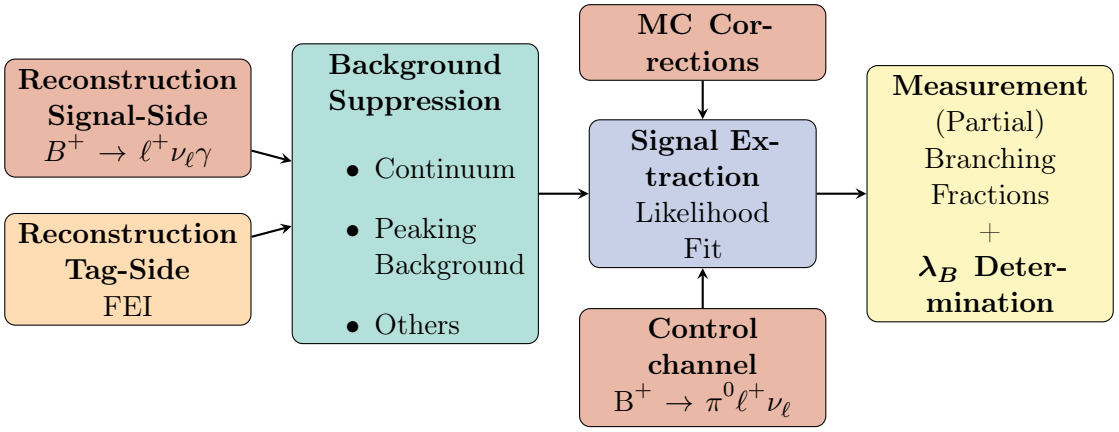


Figure 5.1.: Schematic overview of the analysis strategy.

channel to constrain the background in the nominal $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ channel. Thus, a separate $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ selection is performed in addition to the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ selection.

To determine the partial branching fraction of the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ decays as well as the branching fraction of the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ decays, the distribution of the so-called squared missing mass M^2_{miss} is analysed. The definition and meaning of M^2_{miss} are discussed in the following Section 5.2. The M^2_{miss} distribution is fitted with a binned maximum likelihood fit, which is described in detail in Chapter 7. Both the distribution templates of the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ and the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ selection are considered to perform a simultaneous fit. The resulting signal events extracted with the fit are taken as input for the calculation of the partial branching fraction $\Delta\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma)$, to finally determine λ_{B}. As a crosscheck, the branching fraction $\mathcal{B}(\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell)$ can be measured with a binned maximum likelihood fit to test the stability of the analysis.

In this thesis, the above-mentioned steps are evaluated and performed. Even though the limited scope of this thesis does not allow for a full signal extraction on real Belle II detector data, but only simulations, this work was able to perform a validation of the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ control channel with a data set corresponding to an integrated luminosity of $\mathcal{L} = 63.25\,\mathrm{fb}^{-1}$.

5.2. Missing Mass

The fitting variable chosen for this work is the squared missing mass M^2_{miss}. Here, the missing mass corresponds to the squared neutrino mass and is defined as

$$\begin{aligned} M^2_{\mathrm{miss}} &= (p_{\mathrm{B_{sig}}} - p_\ell - p_\gamma)^2 \\ &= \left(\begin{pmatrix} \frac{E_{\mathrm{CMS}}}{2c} \\ -\vec{p}_{\mathrm{B_{tag}}} \end{pmatrix} - p_\ell - p_\gamma \right)^2 , \end{aligned} \tag{5.1}$$

where the four-momenta of the $\mathrm{B_{sig}}$, the lepton, and the photon, are indicated with $p_{\mathrm{B_{sig}}}$, p_ℓ, and p_γ, respectively. The replacement of $p_{\mathrm{B_{sig}}}$ with half of the center-of-mass (CMS) energy E_{CMS} and the opposite $\mathrm{B_{tag}}$ momentum $\vec{p}_{\mathrm{B_{tag}}}$ is justified by the back-to-back

production of the B_{sig} and B_{tag} in the CMS frame. For correctly reconstructed $B^+ \to \ell^+ \nu_\ell \gamma$ signal events, the M^2_{miss} distribution is expected to peak around zero due to the massless neutrino. The same conclusions can be drawn for $B^+ \to \pi^0 \ell^+ \nu_\ell$ events. Thus, mistakenly reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ events mimic the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ signal and are called peaking background, as mentioned in the previous section. The distribution of the squared missing mass M^2_{miss} forms the basis of the signal extraction with the binned maximum likelihood fit.

6. Analysis

This chapter describes the details of the analysis procedure. More precisely, a discussion of the used data samples is given in Section 6.1. The event reconstruction and considered selections for the $B^+ \to \ell^+ \nu_\ell \gamma$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ are specified in Section 6.2. To further reduce unwanted background, two suppression tasks are introduced in Section 6.3. A validation of the control channel $B^+ \to \pi^0 \ell^+ \nu_\ell$ is obtained in Section 6.4. In addition, the performance of the cluster reconstruction is indicated in Section 6.5. To correct discrepancies between MC simulated and real data, some corrections are applied on MC, which is described in Section 6.6.

6.1. Data Samples

The Belle II experiment is still collecting data, aiming for a total amount of $50\,\mathrm{ab}^{-1}$ within the next years. Currently, the data status records around $246\,\mathrm{fb}^{-1}$ on the $\Upsilon(4S)$ resonance. Since this analysis is evaluated on MC studies, a large simulation sample is required. In particular, the following official Belle II MC data sets are used:

Generic events, containing various decays with b $\to$ u,s,d,c transitions. The generic sample includes both charged and mixed samples, corresponding to $\Upsilon(4S) \to B^+B^-$ and $\Upsilon(4S) \to B^0\overline{B}^0$, respectively. The decays are produced with `EvtGen` based on a decay table, containing all possible decays with the corresponding branching ratios as well as the used underlying model.

Continuum events describe background events occurring through e^+e^- collisions resulting in a quark anti-quark pair $q\bar{q}$ (with q =u,d,s,c) instead of an $\Upsilon(4S)$ resonance. The hadronisation of these light quarks leads to a dominant background. In contrast to the generic samples, the continuum events are generated using `KKMC` [37] followed by `PYTHIA8.2` [27], therefore being model-dependent and rely on the underlying parton fragmentation model.

$B^+ \to \ell^+ \nu_\ell \gamma$ events are added to thoroughly study the signal decay of interest. Both final states, $B^+ \to e^+ \nu_e \gamma$, and $B^+ \to \mu^+ \nu_\mu \gamma$, are included equally in the sample. The sample is produced with `EvtGen` and an estimated partial branching fraction of $5.0 \cdot^{-6}$ is used to weight the sample.

$b \to u \ell \nu_\ell$ events consider semileptonic decays with b $\to$ u transitions, produced with `EvtGen`. In particular, a distinction of charged $B^+ \to X_u \ell \nu_\ell$ and mixed $B^0 \to X_u \ell \nu$

Table 6.1.: Overview of the used MC samples and the number of events included in each sample.

Sample	Number of events ($\cdot 10^6$)
Generic charged	54
$u\bar{u}$	160.5
$d\bar{d}$	40.1
$s\bar{s}$	38.3.5
$c\bar{c}$	132.9
Resonant $B^+ \to X_u \ell \nu_\ell$	50
Non-resonant $B^+ \to X_u \ell \nu_\ell$	50
$B^+ \to e^+ \nu_e \gamma$	40
$B^+ \to \mu^+ \nu_\mu \gamma$	40

is made, where X_u indicates a hadron. Furthermore, each sample differentiates between resonant ($X_u = \pi, \rho, \eta, \eta', \omega$) and non-resonant decays. These samples are used to calculate the hybrid weights, described below. In addition, the control channel $B^+ \to \pi^0 \ell^+ \nu_\ell$, which is used to constrain the background in the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ analysis, is included in the resonant $B^+ \to X_u \ell \nu_\ell$ sample. The assumed branching fraction of the $B^+ \to \pi^0 \ell^+ \nu_\ell$ decay is $1.56 \cdot 10^{-4}$.

An overview of the used samples and their corresponding number of events used for this analysis can be found in Table 6.1. The combination of all above-mentioned samples requires a weighting of the samples according to a hybrid weight, following [38]. The hybrid weights are calculated per event in 3-dimensional bins, taking into account three variables: E_ℓ^B (the lepton momentum with regard to the B meson frame), q^2 (the square of the four-momentum transferred), and M_X (the hadron mass). A form factor re-weighting tool called `eFFORT` [39] is used to determine the hybrid weights. The number of total hybrid events per bin, H_i, is determined by adding the number of resonant events, R_i, to the number of non-resonant events, I_i, which are weighted according to a scaling weight w_i

$$H_i = R_i + w_i I_i \quad . \tag{6.1}$$

The weights are chosen such that $R_i + w_i I_i = I_i$ in each bin i. A more detailed and comprehensive discussion of hybrid weights is provided by [39]. In total, the used MC events are weighted so that they correspond to a total integrated luminosity of $400\,\mathrm{fb}^{-1}$. Since the additionally added $b \to u \ell \nu_\ell$ decays are concurrently included in the generic sample, the corresponding generic events are replaced by the equivalent amount of the respective dedicated MC.

6.2. Event Reconstruction

This section evaluates the selection of events as well as a Bremsstrahlung correction, which is applied to the electrons in the reconstruction. Both signal-side and tag-side selections are considered, as well as criteria on the resulting full $\Upsilon(4S)$ event.

Table 6.2.: Overview of the event-based pre-selections, applied before running the FEI.

Variable	Event-based selection
E_{vis}	$> 4\,\text{GeV}$
$E_{\text{ECL}}^{\text{extra}}$	$\in [2, 7]\,\text{GeV}$
n_{tracks}	≥ 3
n_{cluster}	≥ 3
d_0	$< 0.5\,\text{cm}$
$\lvert z_0 \rvert$	$< 2\,\text{cm}$
p_{T}	$> 0.1\,\text{GeV}$
θ	$\in [17^\circ, 150^\circ]$
E	$> 0.1\,\text{GeV}$

6.2.1. Event Selection

This analysis uses the tagging algorithm FEI, as described in Section 4.5, to reconstruct the full $\Upsilon(4S)$ event. Since the application of the FEI is computationally intensive, preprocessed data sets are provided centrally. These so-called skims are available for data and MC, both for hadronic and semileptonic tagging.

To reduce the size of processing data, some loose selection criteria are applied to the original data before applying the FEI, aiming for rejection of background events without significantly affecting the number of signal events. In particular, so-called clean tracks are ensured by requiring a lower limit on the transverse momentum $p_t > 0.1\,\text{GeV}$ as well as a close vicinity to the interaction point (IP), i.e. $d_0 < 0.5$ and $|z_0| < 2\,\text{cm}$, indicating the distance to the point of closest approach (POCA) in the $r - \phi$ plane and the POCA z-coordinate, respectively. Similarly, clean ECL clusters are filtered with an energy threshold of $E > 0.1\,\text{GeV}$ and within a polar angle acceptance of $17^\circ < \theta < 150^\circ$. The number of cleaned tracks n_{tracks} and ECL clusters n_{cluster} have to be greater than three, respectively. In total, the deposited energy in the ECL ranges over $2\,\text{GeV} < E_{\text{ECL}} < 7\,\text{GeV}$. Furthermore, events with a visible energy (in the CMS frame) lower than 4 GeV are rejected.
Besides the aforementioned pre-selections, further built-in selections occur during the FEI training. These vary according to the hadronic or semileptonic FEI approach and are applied on the B_{tag}, therefore representing candidate-based criteria. In particular, the selection criteria for a hadronic FEI, such as considered for this analysis, include a lower limit on the beam constrained mass, $M_{bc} > 5.24$. Moreover, a maximum absolute energy difference $|\Delta E| < 0.2$, and a signal probability $\mathcal{P}_{\text{FEI}} > 0.001$ are required. The applied event-based pre-selections are summarised in Table 6.2, and a list of the B_{tag} selection criteria applied during the FEI training can be found in Table 6.3. This analysis uses the official hadronic FEI skim as input.

Besides the skim selection criteria, further constraints on the ROE ensure an appropriate usage of information. This is done by applying a so-called ROE mask, defining a set of track-based and ECL-based selections. The underlying ROE mask used in this analysis passes a selection on the transverse momentum, $p_{\text{t}} > 0.2\,\text{GeV}$, as well as selections on the impact parameter, $dr < 2\,\text{cm}$ and $|dz| < 4\,\text{cm}$, for all tracks in the ROE, where dr denotes the transverse and dz the z-direction distance to the interaction point. In addition, a CDC

Table 6.3.: Overview of the B_{tag} selection criteria applied during the training of the `FEI`.

Variable	B_{tag} candidate-based selection
M_{bc}	$> 5.24\,\text{GeV}$
$\lvert\Delta E\rvert$	$< 0.2\,\text{GeV}$
$\mathcal{P}_{FEI}$	> 0.001

Table 6.4.: Overview of the used ROE mask criteria.

Variable	selection
p_t	$> 0.2\,\text{GeV}$
dr	$< 2\,\text{cm}$
$\lvert dz\rvert$	$< 4\,\text{cm}$
$clusterNHits$	> 1.5
θ	$\in [17^\circ, 150^\circ]$
$E_{forward}$	$> 0.08\,\text{GeV}$
E_{barrel}	$> 0.03\,\text{GeV}$
$E_{backward}$	$> 0.06\,\text{GeV}$

angular acceptance for θ is required, $17^\circ < \theta < 150^\circ$. Concerning the ECL cluster, a lower limit of 1.5 is set the number of all crystals in an ECL cluster, $clusterNHits$. Actually, $clusterNHits$ does not return the number of crystals but the sum of the crystal weights, in the case of overlapping clusters. Furthermore, selections on the photon energy deposited in different cluster regions are applied to reduce the number of background clusters. The regional energy depositions requirements are $E_{forward} > 0.08\,\text{GeV}$ for the forward end-cap, $E_{barrel} > 0.03\,\text{GeV}$ for the barrel region, and $E_{backward} > 0.06\,\text{GeV}$ for the backward end-cap energy.

6.2.2. Bremsstrahlung Correction

Charged particles emit electromagnetic radiation while decelerating, an effect known as Bremsstrahlung radiation. As for Belle II, this occurs especially for the light electrons during their propagation through the detector components. Consequently, the electrons suffer from energy loss due to the emission of Bremsstrahlung photons. To balance this effect, a Bremsstrahlung correction on electrons can be applied. The most energetic Bremsstrahlung photons are radiated around the initial momentum direction of the electron. Thus, photons found in a cone of 5° around the momentum are considered as possible candidates arising from Bremsstrahlung radiation. More precisely, the cone is placed at the point of the closest approach with respect to the interaction point. To recover the initial energy of the electron, the four-momenta of the photons are added to the electron. For each electron, daughters are assigned. If one or more Bremsstrahlung photon candidates are found, the daughters consist of the initial electron and the corresponding Bremsstrahlung photons, otherwise, the only daughter is the uncorrected original electron. This provides the advantage that Bremsstrahlung photons cannot be allocated twice, since they already occur in the daughter list of the corrected electron.

6.2.3. Signal-Side Reconstruction

The foundation of a correct signal-side reconstruction is based on a careful selection of the recombined particles, especially for rare decays such as $B^+ \to \ell^+ \nu_\ell \gamma$. In this section, the selection criteria for both signal decays, $B^+ \to \ell^+ \nu_\ell \gamma$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$, are discussed. The selection criteria are based on the $B^+ \to \ell^+ \nu_\ell \gamma$ Belle analysis of 2018 [6] and the $B^+ \to \pi^0 \ell^+ \nu_\ell$ Belle study performed in 2013 [9]. At first, the final state particles are collected. Both leptons, e^+ and μ^+, are assigned with a particle identification probability ($particleID$), representing a likelihood of the particle's own hypothesis, taking into account all detector information. A lower limit of $particleID > 0.9$ is set for both lepton types. In addition, the lepton momentum p_ℓ is required to measure at least 400 MeV. Regarding the lepton tracks, the proximity to the interaction point is ensured by $dr < 2\,\text{cm}$ and $|dz| < 5\,\text{cm}$. Furthermore, the leptons are supposed to be within the angular range of CDC acceptance, restricting the polar angle to $17^\circ < \theta < 150^\circ$. The aforementioned criteria on the final state electrons and muons are applied in both signal channels, $B^+ \to \ell^+ \nu_\ell \gamma$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$. In case of several lepton candidates per event, the one providing the highest particle identification probability is kept.

As elaborated in Chapter 2, the signal photon in the $B^+ \to \ell^+ \nu_\ell \gamma$ channel is required to have an energy $E_\gamma > 1\,\text{GeV}$ due to theoretical considerations on the form factors. Although theory prefers an even tighter selection criterion, the limit of 1 GeV is chosen to ensure enough statistics. Referring to $clusterNHits$, a lower limit of 1.5 is set. To differentiate between photons and hadrons detected in the ECL, a characteristic called $clusterE9E21$ is investigated. The $clusterE9E21$ gives the ratio of energies in the inner $3x3$ crystals of a cluster, $E9$, and $5x5$ crystals around the central crystal in the cluster without taking the four crystals in the corner into account. Lower values of $clusterE9E21$ are expected for hadrons, whereas photons tend to give ratios close to 1.0, the maximal value. These selections are only applied to the photon in the $B^+ \to \ell^+ \nu_\ell \gamma$ channel.

In case of the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel, some selection criteria are dedicated to the π^0 and the corresponding daughter photons. The π^0 mass M_π is allocated to be in the range $[0.12, 0.145]\,\text{GeV}$. Similar to the ROE selection criteria, the regional energy depositions in the ECL are required to be $E_{\text{forward}} > 0.08\,\text{GeV}$ for the forward end-cap, $E_{\text{barrel}} > 0.03\,\text{GeV}$ for the barrel region, and $E_{\text{backward}} > 0.06\,\text{GeV}$ for the backward end-cap energy. As for the photons in the $B^+ \to \ell^+ \nu_\ell \gamma$ channel, the $clusterNHits > 1.5$ selection is applied to the π^0 daughter photons as well. Moreover, the opening angle of the two daughter photons, $\psi_{\gamma\gamma}$, is restricted by requiring $\cos\psi_{\gamma\gamma} > 0.25$. By adding the four-vectors of the reconstructed π^0 and the lepton in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel one can form a pseudoparticle Y. The absolute cosine of the angle between the flight-direction of the nominal B_{sig} and the pseudoparticle Y should fulfill the selection $|\cos\theta_{\text{BY}}| < 3$, if only a neutrino is missing in the reconstruction. A criterion of $|\cos\theta_{\text{BY}}| < 1$ would be sufficiently reasonable, but the selection is loosened to $|\cos\theta_{\text{BY}}| < 3$ for resolution effects. Given that the missing four-momentum of this event is defined by

$$p_{\text{miss}} = (E_{\text{miss}}, \vec{p}_{\text{miss}}) = p_{B_{\text{sig}}} - p_Y, \tag{6.2}$$

a lower limit of $E_{\text{miss}} > 0.3\,\text{GeV}$ is required.

After reconstructing the final state particles and combining the selected photons to a π^0, the B_{sig} candidates are built by combining particles that passed the above-mentioned criteria,

Table 6.5.: Overview of the selection criteria applied during the $B^+ \to \ell^+ \nu_\ell \gamma$ reconstruction.

Particle	Variable	selection
Lepton	$particleID$	> 0.9
	p_ℓ	$> 0.4\,\text{GeV}$
	dr	$< 5\,\text{cm}$
	$\lvert dz \rvert$	$< 2\,\text{cm}$
	θ	$\in [17^\circ, 150^\circ]$
Photon	E_γ	$> 1\,\text{GeV}$
	$clusterNHits$	> 1.5
	$E9E21$	> 0.9
B_{sig}	$cos\,\theta_{\nu\gamma}$	> -0.9

Table 6.6.: Overview of the selection criteria applied during the $B^+ \to \pi^0 \ell^+ \nu_\ell$ reconstruction.

Particle	Variable	selection
Lepton	$particleID$	> 0.9
	p_ℓ	$> 0.4\,\text{GeV}$
	dr	$< 5\,\text{cm}$
	$\lvert dz \rvert$	$< 2\,\text{cm}$
	θ	$\in [17^\circ, 150^\circ]$
π^0 daughter photons	$clusterTheta$	$\in [17^\circ, 150^\circ]$
	$clusterNHits$	> 1.5
	E_{forward}	$> 0.08\,\text{GeV}$
	E_{barrel}	$> 0.03\,\text{GeV}$
	E_{backward}	$> 0.06\,\text{GeV}$
π^0	M	$\in [0.120, 0.145]\,\text{GeV}$
	$E_{residual}$	$< 0.6\,\text{GeV}$
	$\cos\psi_{\gamma\gamma}$	> 0.25

Table 6.7.: Selections applied to the $\Upsilon(4S)$ reconstruction.

Variable	Selection
M	$\in [7.5, 10.5]\,\mathrm{GeV}$
$E_{\mathrm{ECL}}^{\mathrm{extra}}$	$< 0.9\,\mathrm{GeV}$

to reconstruct the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ and $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ decays, respectively. In addition to the initial selection criteria on the final state particles, further selections are considered for the $\mathrm{B}_{\mathrm{sig}}$ candidates. Regarding the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ channel, the angle between the signal photon and the missing momentum (i.e. the undetectable neutrino) is restricted by $\cos\theta_{\gamma\nu} > -0.9$ to reject continuum events [40].

6.2.4. Tag-Side Reconstruction

As previously discussed in Section 4.5, the `FEI` is a powerful tagging algorithm to reconstruct the $\mathrm{B}_{\mathrm{tag}}$, and is therefore used for this analysis. In particular, the generic hadronic `FEI` is chosen. A complete listing of the possible $\mathrm{B}_{\mathrm{tag}}$ channels used for the `FEI` training can be found in [29]. In addition to the selections in Section 6.2.1, the $\mathrm{B}_{\mathrm{tag}}$ in the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ channel requires $\mathcal{P}_{\mathrm{FEI}} > 0.01$.

6.2.5. $\Upsilon(4S)$ Reconstruction

In a final step, the $\Upsilon(4S)$ event is reconstructed as a combination of the signal-side and tag-side, where the $\mathrm{B}_{\mathrm{sig}}$ and $\mathrm{B}_{\mathrm{tag}}$ are reconstructed according to the selection criteria of the preceding sections. Additionally, the $\Upsilon(4S)$ event is restricted with various further selections to decrease the background retention. Considering the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ selection, the invariant mass of the $\Upsilon(4S)$ is limited to the range $[7.5, 10.5]\,\mathrm{GeV}$. The extra energy in the ECL, which is not associated to a particle in the reconstruction, is constrained by an upper limit, $E_{\mathrm{ECL}}^{\mathrm{extra}} < 0.9\,\mathrm{GeV}$. A similar selection is applied to the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ channel, where the extra energy from neutral ECL clusters not used for the reconstruction is restricted to $E_{\mathrm{ECL}}^{\mathrm{extra}} < 0.6$. Since a clean ROE is preferred, only events without extra tracks are considered, hence the completeness constraint (see Section 4.3) is used for both reconstruction channels.

6.3. Background Suppression

As mentioned in Section 5.1, mistakenly reconstructed events occurring from other processes appear in the than initial analysis of the decay of interest. Two of these background processes, continuum and peaking background, are discussed in this section. A classification of continuum and signal events is performed for the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ as well as for the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ reconstruction. The classification for the peaking background suppression is only evaluated for the nominal $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ channel. In the end, the best cut values of the resulting classification indicators is found.

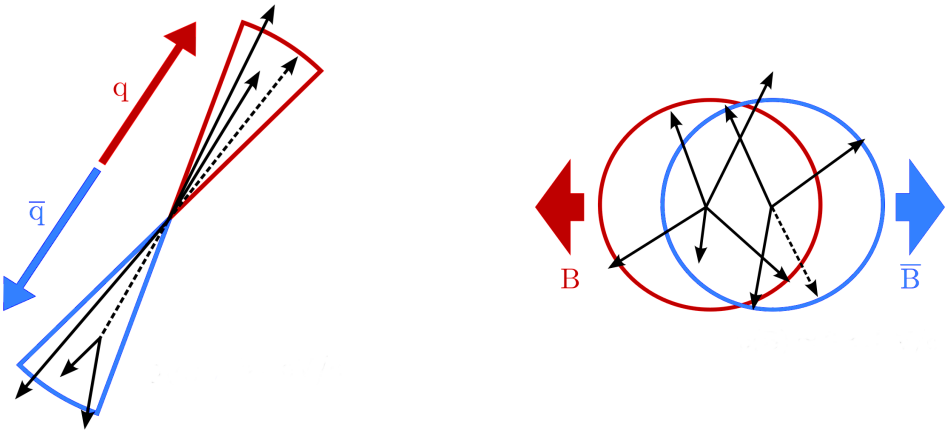


Figure 6.1.: Event topology of a jet-like continuum (left) and an isotropic B$\overline{\text{B}}$ event (right), respectively. Adapted from [41].

6.3.1. Continuum Suppression

The dominant combinatorial background of e^+e^- collisions is constituted by continuum events, i.e. the production of $u\bar{u}, d\bar{d}, s\bar{s}$ or $c\bar{c}$ pairs. These quark-antiquark pairs randomly produce final state particles, which can mimic signal trajectories. Especially analyses of rare decays such as $\text{B}^+ \to \ell^+\nu_\ell\gamma$, which do not decay via $b \to c$ transitions, suffer from the presence of continuum background. Thus, advanced techniques to suppress continuum events have been employed, exploiting the difference in the event shape of B$\overline{\text{B}}$ decays compared to continuum events. The event shape is mainly influenced by the angular distribution of the particle momenta. In the case of continuum events, the event fragments into two back-to-back jets of light hadrons, since the light quark and antiquark carry a large momentum. A schematic continuum event is illustrated on the left in Figure 6.1. In contrast, a B$\overline{\text{B}}$ pair proceeding from an $\Upsilon(4\text{S})$ event decays approximately at rest in the $\Upsilon(4\text{S})$ frame. Henceforth, an isotropic distribution is expected, as can be seen on the right in Figure 6.1.

A description of the different event shapes is provided by various variables, explained in the following.

Thrust

The thrust axis $\vec{T}$ describes the unit vector in the direction of the maximised projection of N particles momenta $\vec{p_i}$, $i = 1, ..., N$. Given the thrust axis $\vec{T}$ and the momenta $\vec{p_i}$, the so-called thrust T can be calculated as

$$T = \frac{\sum_i^N |\vec{T}\vec{p_i}|}{\sum_i^N |\vec{p_i}|}. \tag{6.3}$$

The thrust angle θ_T, defined as the angle between the thrust of all signal B daughter particles and the thrust of all ROE particles, provides powerful information. The decay products of B$\overline{\text{B}}$ events and consequently their thrust axes are randomly distributed. Accordingly, the cosine of the corresponding thrust angle $|\cos\theta_T|$ is uniform. In

contrast, for continuum events $|\cos\theta_T|$ peaks at large θ_T values, since the jet-like event shape results in directional thrusts. Alternatively, similar deductions can be made by looking at the angle between the signal B thrust and the beam axis z, $|\cos\theta_{T,z}|$, instead of T_{ROE}.

CLEO Cones

The naming of these parameters stems from the CLEO collaboration, which developed the cones [42]. In total, 9 CLEO cones are defined, concentrically distributed in steps of 10° around the thrust axis. For each cone, the absolute momenta of all particles within the corresponding angular region are summed. The cones are available not only for all final state particles in the event, but also only for ROE particles.

Fox-Wolfram Moments

The Fox-Wolfram moments were introduced for e^+e^- annihilation [43] and provide a description of the momentum and energy flow in the event containing N particles with momenta p_i $(i = 1, ..., N)$. The definition is given by

$$H_k = \sum_{i,j}^{N} |\vec{p_i}||\vec{p_j}| P_l(\cos\theta_{i,j}), \tag{6.4}$$

where $\theta_{i,j}$ denotes the angle between $\vec{p_i}$ and $\vec{p_j}$ and P_l refers to the l-th order Legendre polynomial. In general, a normalisation of the Fox-Wolfram moments $R_l = H_l/H_0$ is used, since $H_0 = 1$ for vanishing particle masses. In case of jet-like collimated events, R_l approaches zero (one) for odd (even) numbers of l, consequently providing discrimination power. In particular, R_2 is a powerful and commonly used variable to discriminate between $\mathrm{B\overline{B}}$ and continuum events, and is thus included in the continuum suppression in this work.

Kakuno-Super-Fox-Wolfram Moments

A modified version of the aforementioned Fox-Wolfram moments are the so-called Kakuno-Super-Fox-Wolfram moments (KSFW), developed by the Belle collaboration. In comparison to Fox-Wolfram moments, the KSFW signal-background discrimination for charmless two-body decays and three-body decays is more reliable. There are a total of 17 KSFW moment variables; this analysis uses only 16 of them, as one KSFW moment is correlated to the fitting variable.

For further, more comprehensive descriptions see [44].

The set of all above-mentioned variables is used as feature input to train a multivariate classifier with the `FastBDT` algorithm (see Section 4.4). Subsequently, the output of the classifier, $\mathcal{P}_{\mathrm{CS}}$, is an indicator of facing a continuum event, where zero (one) corresponds to an event most likely originating from continuum (signal). Initially, the classifier is trained on a dedicated training sample, containing both signal and continuum events. In a second step, the classifier is applied to an independent, smaller test sample. As can be seen on the left in Figure 6.2, the application on a test sample performs a distinct separation of $\mathcal{P}_{\mathrm{CS}}$ for continuum and signal events, peaking around zero and one, respectively. The performance of the classifier is shown on the right in Figure 6.2. The AUC of the test

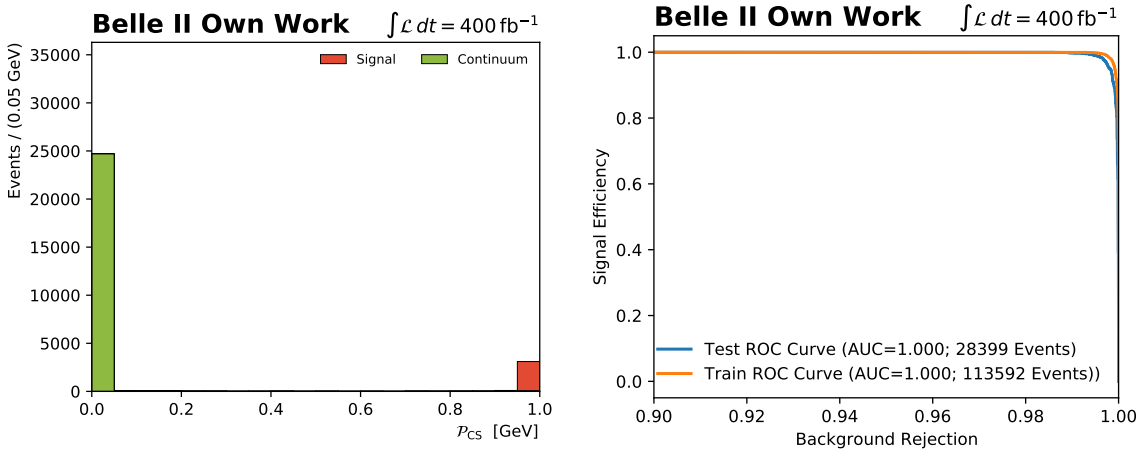


Figure 6.2.: Left: Stacked distribution of the continuum suppression classifier output $\mathcal{P}_{\mathrm{CS}}$, calculated for a test sample. Signal events show output values peaking around one, whereas continuum background events are gathered around zero. Right: ROC curve for the continuum suppression classifier. The classifier is applied on both train (orange) and test (blue) samples. The AUC score and the number of sample events can be found in the lower-left corner for the train and test sample, respectively.

sample scores AUC = 1.00, within the precision of rounding, indicating an almost perfect separation. Therefore, the signal efficiency as well as the background rejection are reasonable. Overtraining is ruled out, since the classifier output of the train and test sample show comparable distributions. Continuum suppression is applied to both the $B^+ \to \ell^+\nu_\ell\gamma$ and the $B^+ \to \pi^0\ell^+\nu_\ell$ selection. The identification of the best selection value on the respective $\mathcal{P}_{\mathrm{CS}}$ is discussed in Section 6.3.3. The feature importance calculated with `FastBDT` can be found in Appendix A.

6.3.2. Peaking Background Suppression

In addition to the above-described continuum suppression, another important background must be taken into account, namely the peaking background, as already mentioned in Chapter 5. These events mimic signal properties, which leads to a distortion of the signal extraction. As for this $B^+ \to \ell^+\nu_\ell\gamma$ analysis, the peaking background occurs due to semileptonic $B^+ \to X_u\ell\nu_\ell$ decays, since the light meson X_u decays in a photon pair. If only one of the two photons is detected, it can mistakenly be reconstructed as the signal photon. This misidentification may occur if for example both photons are emitted approximately in the same direction. The ECL clustering algorithm then might identify those as a single photon. Alternatively, only one of both photons may be identified as a signal photon, while the other is either not detected at all or mistakenly interpreted as a tag-side or Bremsstrahlung photon. The dominating fraction of $B^+ \to X_u\ell\nu_\ell$ decays for $B^+ \to \ell^+\nu_\ell\gamma$ reconstructions is caused by the $B^+ \to \pi^0\ell^+\nu_\ell$ channel, thus this work focuses on the suppression of $B^+ \to \pi^0\ell^+\nu_\ell$ background events.

Similar to the continuum suppression described in the preceding Section 6.3.1, a multivariate

classifier for the peaking background suppression is used. This task is only applied to the $B^+ \to \ell^+ \nu_\ell \gamma$ selection. The variables listed below are taken as features for the multivariate peaking background classifier.

π^0 Mass Veto To probe whether a signal photon originates from a neutral pion, the signal-side photon is combined with any photon of the ROE to form a π^0 candidate. If the resultant π^0 candidate shows a mass $M_{\gamma_{sig},\gamma_{ROE}}$ close to the nominal π^0 mass $M_{\pi^0} = 0.134\,\text{GeV}$, this so-called π^0 mass veto is applied to the signal photon. In particular, the range $M_{\gamma_{sig},\gamma_{ROE}} \in [0.11, 0.16]\,\text{GeV}$ is chosen. In case no corresponding ROE photon candidate is found, the $M_{\gamma_{sig},\gamma_{ROE}}$ is allocated to infinity, to enhance the information of the missing value as recommended in the implementation of `FastBDT` [33]. In addition, some selection criteria are applied to the ROE photon, in particular $E_{\text{forward}} > 0.08\,\text{GeV}$ for the forward end-cap, $E_{\text{barrel}} > 0.03\,\text{GeV}$ for the barrel region, and $E_{\text{backward}} > 0.06\,\text{GeV}$ for the backward end-cap energy, as well as $clusterNHits > 1.5$.

$\theta_{\ell,\gamma}$

The lepton-photon opening angle $\theta_{\ell,\gamma}$ is taken into account, since those particles are often produced back-to-back in $B^+ \to \ell^+ \nu_\ell \gamma$ decays.

Beam Background Probability

To further exclude beam background photons, *basf2* provides an MVA classifier already implemented in the online analysis, *beamBackgroundSuppression*. The output of the classifier indicates whether a true photon cluster or rather a beam background cluster is found. The used training features can be found in Appendix B.

π^0-η-Veto Probability

Another MVA classifier trained online is *writePi0EtaVeto*, which calculates a probability for the signal photon to originate from a π^0 or η as output. The used training features can be found in Appendix B.

clusterNHits

The weighted sum of all crystals, in which energy was deposited, of an ECL cluster. Since nearby clusters can overlap, the crystals are weighted according to the energy splitting between the clusters. In case no overlapping occurs, this equals the number of all crystals in the cluster.

clusterE9E21

The energy in the inner $3x3$ crystals, $E9$, is related to the energy of the $5x5$ crystals around the central crystal, excluding the four crystals in the corner. The ratio is called E9E21.

clusterLAT

The lateral energy distribution of the ECL cluster tends to be broader for two merged photons originating from a π^0 than for an incident photon.

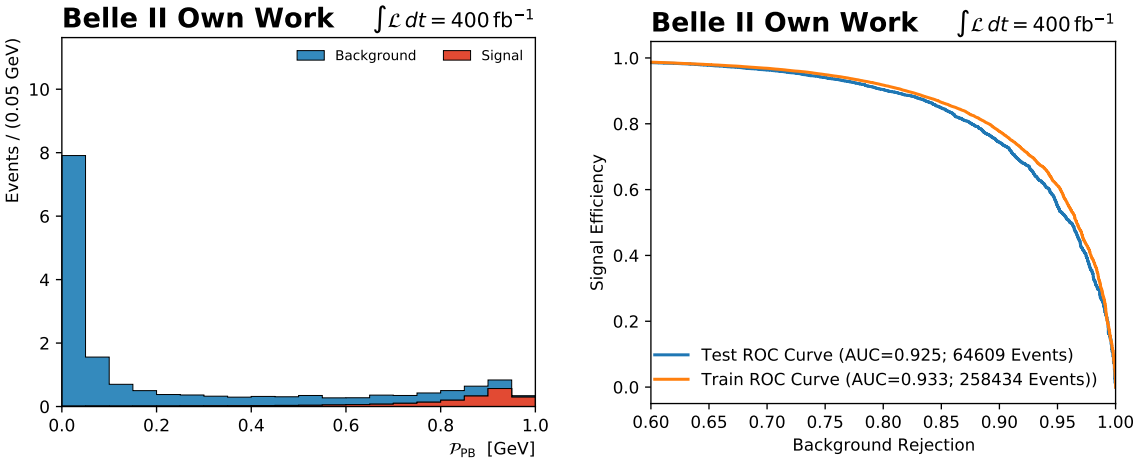


Figure 6.3.: Left: Stacked distribution of the peaking background classifier output $\mathcal{P}_{\mathrm{PB}}$, calculated for a test sample. Signal events (background events) show output values peaking around one (zero). Right: ROC curve for the peaking background classifier applied on the train (orange) and test (blue) samples. The respective AUC scores and the number of sample events can be found in the lower-left.

$\theta_{\gamma,\nu}$

This variable describes the angle between the signal photon and the missing neutrino $\theta_{\gamma,\nu}$ with respect to the $\mathrm{B}_{\mathrm{sig}}$ rest frame. Although these particles are produced back-to-back in the $W^{\pm}$ rest frame considering a $\mathrm{B}^{+} \to \ell^{+}\nu_{\ell}\gamma$ decay, the emission of the photon causes a boost of the lepton and neutrino.

$E_{\mathrm{ECL}}^{\mathrm{extra}}$

As already described for the event selection (Section 6.2.5), $E_{mathECL}^{mathextra}$ is the remaining energy in the ECL, which has not been assigned to any particle.

Energy Asymmetry

The energy asymmetry for a particle with i daughters is given by

$$A_E = \frac{\prod_i E_i}{\sum_i E_i}. \tag{6.5}$$

As for the $\mathrm{B}_{\mathrm{sig}}$, the above formula reveals the asymmetry in the distribution of the lepton and photon candidate.

A train and test sample are defined, using the $\mathrm{B}^{+} \to \ell^{+}\nu_{\ell}\gamma$ signal and $\mathrm{B}^{+} \to \pi^{0}\ell^{+}\nu_{\ell}$ background sample. Both the train and the test samples are stratified by signal and background events. Similar to $\mathcal{P}_{\mathrm{CS}}$, high (low) values of $\mathcal{P}_{\mathrm{PB}}$ indicate the probability of facing a signal (background) event. The resulting classifier output, the peaking background probability $\mathcal{P}_{\mathrm{PB}}$, is evaluated in Figure 6.3. On the left side, the output of the classifier applied on a test sample is shown, whereas the right side presents the ROC curve with an AUC score of AUC = 0.925 for the test sample. The feature importance calculated with

`FastBDT` can be found in Appendix B. The investigation of the best selection value for $\mathcal{P}_{\mathrm{PB}}$, with respect to a low background retention and a high efficiency, is estimated in the following section.

6.3.3. $\mathcal{P}_{\mathrm{CS}}$ And $\mathcal{P}_{\mathrm{PB}}$ Selection Optimisation

Aiming for a background rejection as high as possible and concurrently a moderate signal loss, reasonable selections on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$ are required. Therefore, a figure of merit is used to indicate the performance of the classifiers. Particularly for small expected signals, as for the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ analysis, the so-called Punzi's figure of merit (PFOM) [45] provides a reasonable indication. This figure is defined as

$$\mathrm{PFOM} = \frac{\epsilon}{\frac{\sigma}{2} + \sqrt{B}}, \tag{6.6}$$

where ϵ indicates the reconstruction efficiency, σ denotes the desired significance and B gives the count of background events. For this analysis, a significance of $\sigma = 3$ is considered. Regarding the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ selection, both the continuum suppression classifier and the peaking background identifier are simultaneously optimised by maximising the PFOM. In particular, a module called `scipy.optimize` with a Powell algorithm is used [46]. Besides the mathematical determination, a crosscheck with a 2-dimensional grid search with a stepsize of $n_{\mathrm{steps}} = 100$ is performed. Both optimisation processes are carried out in an optimisation window $M^2_{\mathrm{miss}} \in [-0.5, 0.5]\,\mathrm{GeV}^2$. The achieved results from `scipy` measure $\mathcal{P}_{\mathrm{CS}} = 0.979$ and $\mathcal{P}_{\mathrm{PB}} = 0.437$, which is in good agreement with the grid search, as can be seen in Figure 6.4. The PFOM and the signal retention as a function of the classifier output are shown in Figure 6.5 for continuum suppression and peaking background classification, respectively.

Figure 6.6 shows the distribution of M^2_{miss} before (left) and after (right) the selections on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$ are applied. Further background components are distinguished. More precisely, the distinct components are $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ (signal), $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$, $\mathrm{B}^+ \to \rho^0\ell^+\nu_\ell$, nonresonant $\mathrm{B}^+ \to \mathrm{X_u}\ell\nu_\ell$, other resonant $\mathrm{B}^+ \to \mathrm{X_u}\ell\nu_\ell$, $\mathrm{B}^+ \to \mathrm{X_c}\ell\nu_\ell$, continuum, indirect lepton, and other $\mathrm{B\overline{B}}$ processes. The signal shape in the left plot is enhanced by a factor of 20. Both the continuum and $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ component are significantly reduced after the background suppression selections. Further background events in other components are rejected due to the selection criteria. Nevertheless, a fraction of the peaking $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ background still survives the selections. Therefore, as mentioned in Chapter 5, the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ control sample is used to constrain the remaining background in the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ selection. The procedure is in detail described in Chapter 7. Since the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ selection only considers continuum suppression, a selection value of $\mathcal{P}_{\mathrm{CS}} = 0.8$ is chosen. Table 6.8 gives an overview of the obtained signal retention and background rejection within the optimisation window for the selection criteria of both channels.

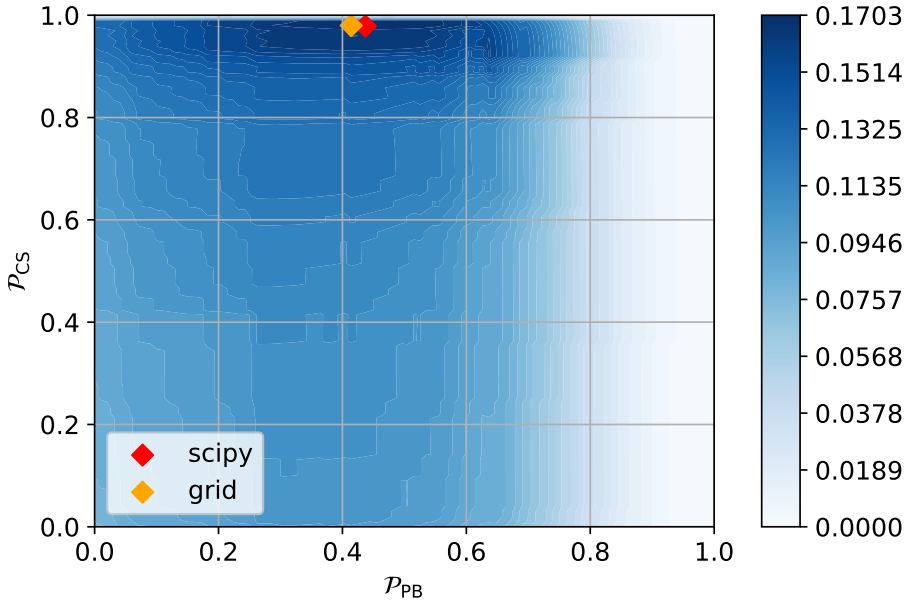
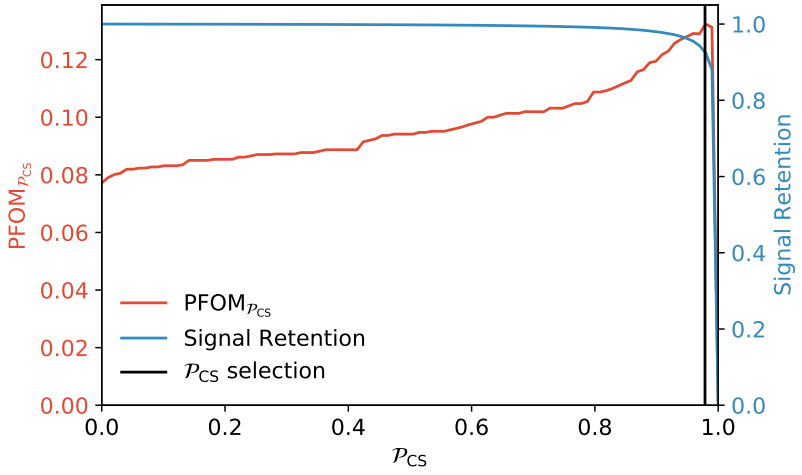


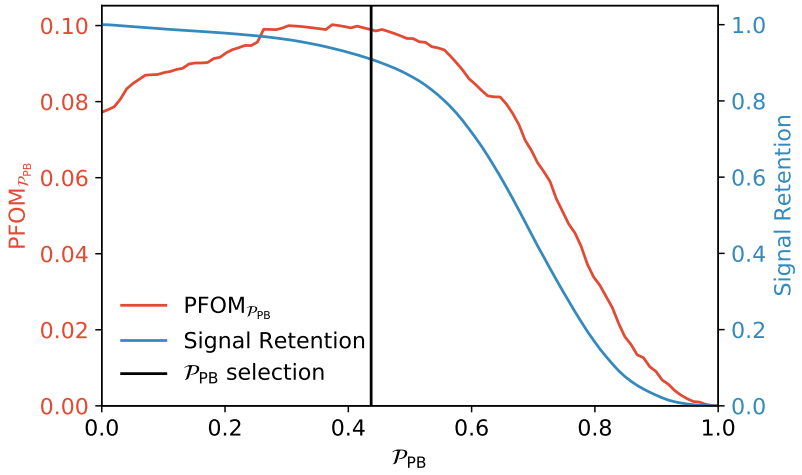
Figure 6.4.: Visualisation of the two-dimensional grid-search for the best selection values for $\mathcal{P}_{\text{CS}}$ and $\mathcal{P}_{\text{PB}}$, plotted on the y-axis and x-axis, respectively. The color indicates the magnitude of the PFOM, where light (dark) blue is assigned to the lowest (highest) value. The best selection value resulting from the `scipy` calculation is marked in red, the result obtained by the grid-search in orange.

Table 6.8.: Overview of the retained signal and the rejected background events for the nominal $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ and the control $\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell$ channel within the optimisation window $M^2_{\text{miss}} \in [-0.5, 0.5]\,\text{GeV}$. The results are given in relation to the number of events before the selection criteria on $\mathcal{P}_{\text{CS}}$ and $\mathcal{P}_{\text{PB}}$ are applied.

	Signal Retention	Background Rejection
$\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$	0.84 %	0.41 %
$\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell$	0.98 %	0.30 %

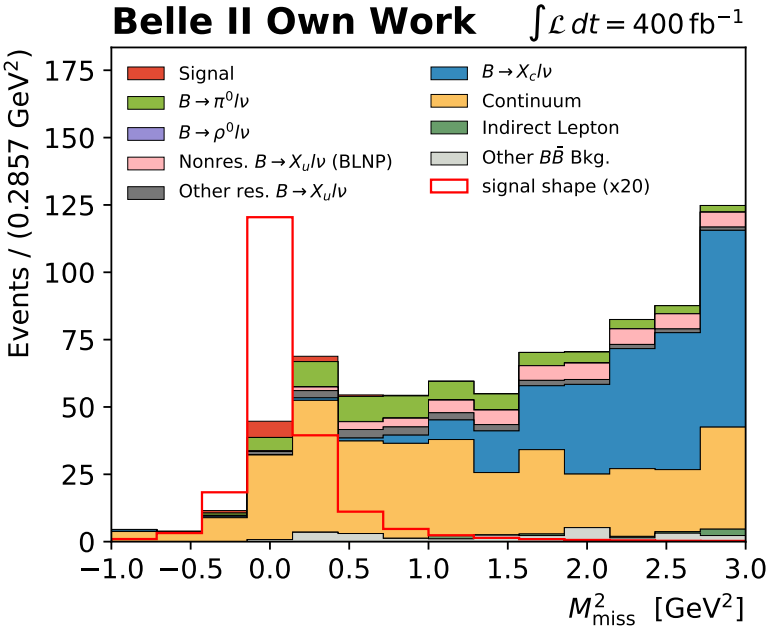


(a) Continuum Suppression Classifier.

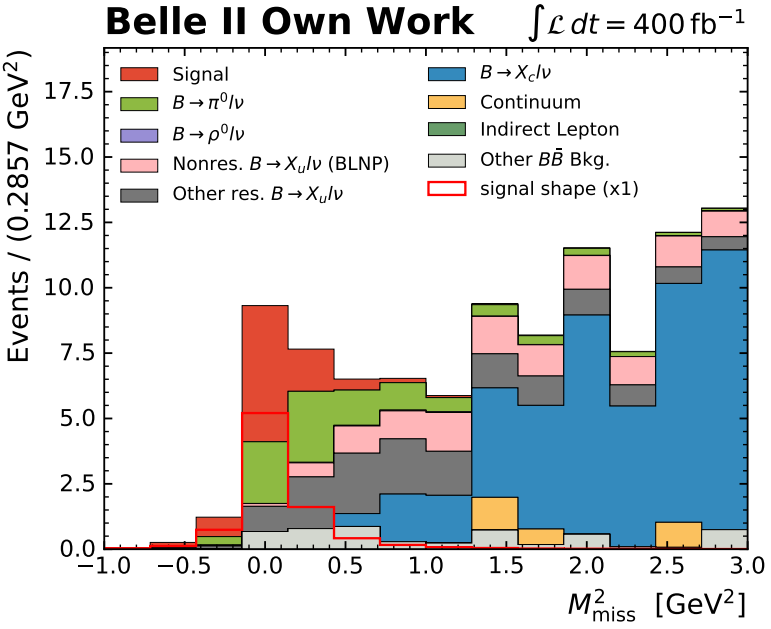


(b) Peaking Background Classifier.

Figure 6.5.: The resulting PFOM (red) is plotted against the respective classifier output for the continuum suppression and peaking background classifier. In addition, the signal retention (blue) is shown. The black vertical lines mark the corresponding calculated selection value, estimated with the `scipy` method.

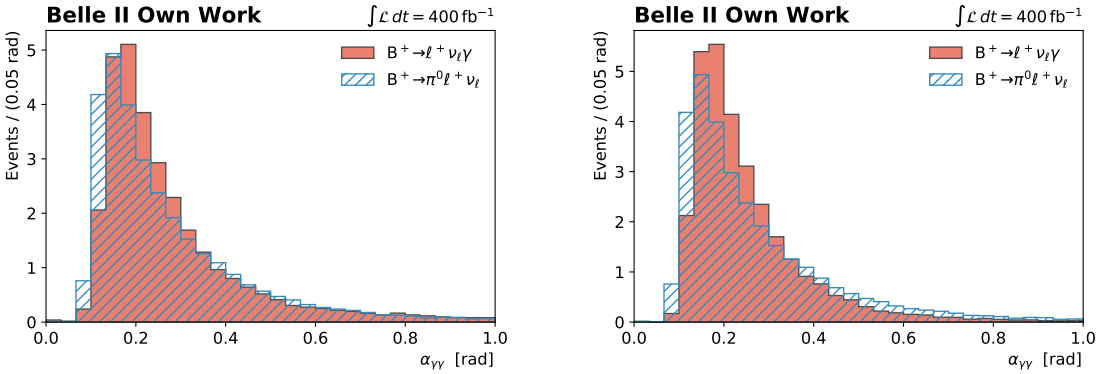


(a) After selection criteria on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$ are applied.



(b) Before selection criteria on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$ are applied.

Figure 6.6.: Distribution of M^2_{miss} for the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ selection before (upper plot) and after (lower plot) the selections on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$ are applied. In addition to the stacked components, the signal shape is drawn. For better visibility, the signal shape in the upper plot is enhanced by a factor of 20.



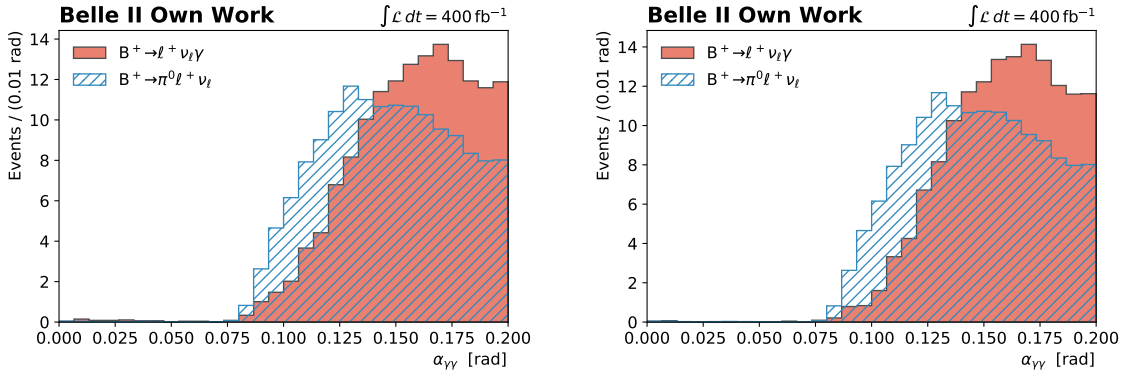
(a) Before selection on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$.

(b) After selection on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$.

Figure 6.7.: Distribution of the angle between the two daughter photons of the π^0, occurring in the background of the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ reconstruction (blue) and the control $B^+ \to \pi^0 \ell^+ \nu_\ell$ selection (orange). The angle is calculated with regard to the true generator information. Both distributions, before (left) and after (right) the selections applied to $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$, are shown.

6.4. Validation $B^+ \to \pi^0 \ell^+ \nu_\ell$ Control Channel

As already mentioned in Chapter 5, the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel is employed to constrain the remaining peaking background in the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ channel. To verify whether the events in the nominal selection and those in the control channel are comparable, the angular distribution of the two π^0 daughter photons is examined. Since the investigation is performed on MC data, the true origin of the photons is known. Therefore, the angle between the two initially generated π^0 photons, $\alpha_{\gamma\gamma}$ can be calculated. As can be seen in Figure 6.7, the distributions for $\alpha_{\gamma\gamma}$ for the background π^0 candidates in the nominal signal $B^+ \to \ell^+ \nu_\ell \gamma$ selection compared to the $\alpha_{\gamma\gamma}$ distribution for the π^0 candidates in the control $B^+ \to \pi^0 \ell^+ \nu_\ell$ selection are in reasonable agreement. Both cases, before and after the selections on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$, discussed in the preceding section, are shown in the left and right plot, respectively. The shapes of the distributions are comparable, which justifies the use of the control sample to constrain the background. The critical region of small angles ($\alpha_{\gamma\gamma} \in [0, 0.2]\,\mathrm{rad}$) is enlarged in Figure 6.8. Photons emitted in approximately the same direction are hard to distinguish as separate photons.



(a) Before selection on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$.

(b) After selection on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$.

Figure 6.8.: Distribution of the angle between the two daughter photons of the π^0, occurring in the background of the nominal $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ reconstruction (blue) and the control $\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell$ selection (orange) in the section $[0, 0.2]$ rad. The angle is calculated with regard to the true generator information. Both distributions, before (left) and after (right) the selections applied to $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$, are shown.

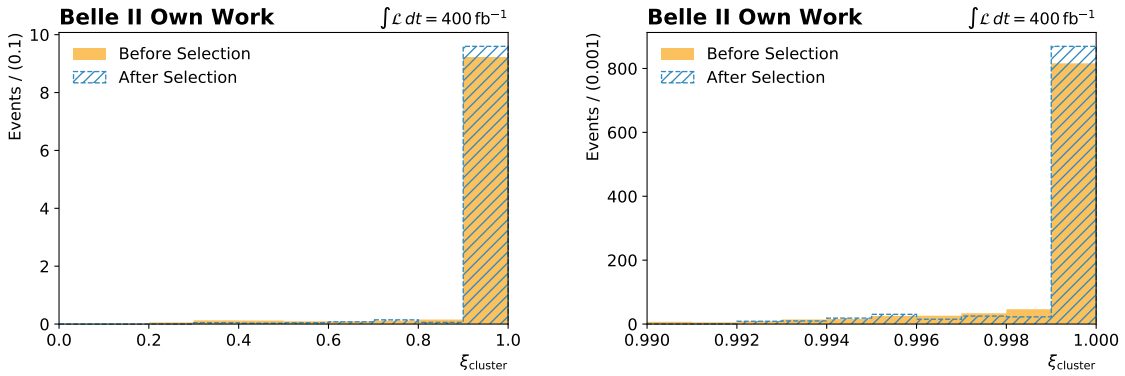
6.5. Cluster Reconstruction Performance

The performance of the ECL cluster reconstruction is tested for the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ selection, to investigate the distinction between two merged photons originating from light mesons and a single photon. In particular, the cluster weights, and consequently the energy depositions in the cluster, are analysed. On the one hand, the variable *clusterTotalMCMatchWeight* describes the sum of the weighted energies of the cluster crystals of all related MC particles. On the other hand, the variable *ClusterBestMCMatchWeight* gives the energy fraction of the MC particle with the highest weight in the cluster. Since charged particles are matched to MC according to their track, the *ClusterBestMCMatchWeight* can deviate from the deposited energy in the ECL cluster.

By building the ratio of *ClusterBestMCMatchWeight* and *clusterTotalMCMatchWeight*, the result should be approximately one for single photons. In the case of two or even several merged photons, the ratio scales down to lower values. To identify whether the photons reconstructed in the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ channel are initial photons or merged photons, this ratio is evaluated, as can be seen in Figure 6.9. Significantly, most clusters originate from single photons, recognisable due to the clear peak close to one. The distributions are normalised, thus the rising peak after the background suppression compared to the pre-selections version indicates an enhancement in the single-photon purity of the clusters.

6.6. MC Corrections

To balance differences between MC and data, certain corrections are applied on the reconstructed MC events. More precisely, this includes the efficiency of the lepton identification as well as the probability of pions being misidentified as leptons. The correction factors for the respective leptons are based on a study of the corresponding lepton track polar angle θ



(a) Before selection on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$.

(b) After selection on $\mathcal{P}_{\mathrm{CS}}$ and $\mathcal{P}_{\mathrm{PB}}$.

Figure 6.9.: Distribution of ratio Ξ_{cluster} for the $\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma$ selection performed before (left) and after (right) the application of the continuum suppression and peaking background suppression selection criteria. The distributions are normalised.

and its lab frame momentum p for well-known processes. The study is performed by the Belle II collaboration [47]; this work uses the resulting recommended correction factors. In addition, the discrepancies in the hadronic tag-side reconstruction efficiency between MC and data are considered by a `FEI` calibration factor of 0.63, which is derived in a separate study [12]. As for the $\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell$ reconstruction, a further scaling factor $\mathrm{SF}_\pi = 0.945$ is applied [12], which is obtained by an independent study of $\eta \to 3\pi^0$ decays.

7. Signal Extraction

As mentioned before, the partial branching fraction is determined by extracting the fraction of the desired signal process contributing to a given data set. This contribution, called signal yield, is extracted from the M^2_{miss} distribution using a binned maximum likelihood method. This chapter describes the general procedure of a maximum likelihood fit in Section 7.1. To state the goodness of the fit result, the significance is introduced in Section 7.2. As this work uses templates of different contributing processes to fit the missing mass distribution, this template method is explained in Section 7.3. The determination of the partial branching fraction with use of the result from the template fit is given in Section 7.4. The fitting setup for the simultaneous fit of the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ and control $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel is specified in Section 7.5. Considering the validation of the fit, a fit on so-called Asimov data is performed in Section 7.7. In addition, the $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fraction is determined with data obtained by the Belle II experiment in Section 6.4, which provides a further fit validation.

7.1. Maximum Likelihood Method

A commonly used technique for parameter estimation is the maximum likelihood fit. The probability distribution of an observable $\boldsymbol{x}$ can be described by $f(x, \boldsymbol{\theta})$, the so-called probability density function (pdf), where $\boldsymbol{\theta}$ denotes a set of parameters. These parameters are unknown, whereas the functional shape of the pdf is known. The pdf returns how likely x describes an observed data set x_0. An estimation of the parameters $\boldsymbol{\theta}$ is provided by the maximum likelihood method. The likelihood function $L(\boldsymbol{\theta})$ is defined as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i|\boldsymbol{\theta}), \tag{7.1}$$

where x_i denotes the i-th measurement of the observable. The maximum likelihood estimators for the parameters θ_i for $i = 1, ..., M$ are then calculated as

$$\frac{\partial L}{\partial \theta_i} = 0 \ . \tag{7.2}$$

Therefore, the estimates maximise the likelihood function, assuming that L is differentiable in θ_i. More conveniently, the negative log-likelihood function

$$-\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log f(x_i|\boldsymbol{\theta}) \tag{7.3}$$

is minimised instead of maximising Equation (7.2). Since the logarithm transforms the product in L into a sum and possible exponentials into simple factors, the log-likelihood is computational easier to calculate. For each event x_i, the log-likelihood is calculated according to Equation (7.3), which is computationally intensive for large data samples. Therefore, a binned maximum likelihood function provides a more efficient method, defined as

$$-\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N_{\text{bins}}} n_i \cdot \log f(y_i|\boldsymbol{\theta}) \quad . \tag{7.4}$$

In that case, a histogram with N_{bins} bins is used, where n_i events are counted in a bin i with the bin center y_i. The computation time is reduced, as it scales with N_{bins} instead of the number of events. Although this method might be less precise compared to the unbinned one, the effect is negligible for small bin sizes [48].

The expected number of events in the $i-$th bin, $\nu_i(\boldsymbol{\theta})$, can be estimated by assuming a Poisson distribution

$$\nu_i(\boldsymbol{\theta}) = n_{\text{tot}} \int_{y_i^{\text{min}}}^{y_i^{\text{max}}} f(y_i|\boldsymbol{\theta}) dy, \tag{7.5}$$

where n_{tot} indicates the total number of observed entries, and y_i^{min} and y_i^{max} are the respective lower and upper bin limit. If n_{tot} is considered as a random variable, it can be estimated by a Poisson distribution with mean ν_{tot}. As consequence, $L(\boldsymbol{\theta})$ then becomes the product of the Poisson probabilities to find n_{tot}, called extended likelihood function. In the case of binned data, the binned extended log-likelihood function is given by

$$\log L(\nu_{\text{tot}}, \boldsymbol{\theta}) = -\nu_{\text{tot}} + \sum_{i=1}^{N_{\text{bins}}} n_i \log \nu_i(\boldsymbol{\theta}) \,. \tag{7.6}$$

If no dependence between ν_{tot} and $\boldsymbol{\theta}$ exists, the estimator is given by $\hat{\nu}_{\text{tot}} = n_{\text{tot}}$. A more detailed derivation can be found in [48].

7.2. Significance

The significance of a measurement indicates its goodness of fit. This section is based on [49], where detailed descriptions and comprehensive deviations of the following equations can be found. This work uses a profile likelihood ratio to calculate the significance for the result. The profile likelihood for a specific ϕ is defined as

$$\lambda(\phi) = \frac{L(\phi, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\phi}, \hat{\boldsymbol{\theta}})}, \tag{7.7}$$

where $\hat{\hat{\boldsymbol{\theta}}}$ maximises L given the fixed ϕ. The numerator shows the maximised L for $\hat{\phi}$ and $\hat{\boldsymbol{\theta}}$. The underlying test statistic is given by

$$q_0 = \begin{cases} -2\log(\lambda(0)) & \hat{\phi} \geq 0 \\ 0 & \hat{\phi} < 0 \end{cases} , \tag{7.8}$$

where $\lambda(0)$ denotes the the profile likelihood ratio from Equation (7.7). Equation (7.8) is only valid, if a positive ϕ for the signal process is expected. This ensures, that the hypothesis of not finding a signal is rejected. The resulting significance is estimated as

$$Z = \sqrt{-2\log\lambda(0)}, \tag{7.9}$$

and is commonly given in units of standard deviations σ.

7.3. Template Method

As mentioned above, the binned maximum likelihood method provides an estimation. Since the underlying pdfs are not known, the so-called template method [50] is used. Different processes contribute to the distribution of an observable, the corresponding shapes of their partial distributions can be used as templates. In particular, the templates are inferred from the histograms based on MC simulations. Each initial template is normalised. The normalisation constant of each template is considered as a floating model parameter in the fit. By fitting all model parameters to the data, the inference of the summed templates and the data is optimised.

Given a histogram with N_{bins} for a variable $\boldsymbol{x}$, the n_i entries of each bin i fluctuate around an expectation value ν_i. The fluctuations can be described by a Poisson distribution, independent for each bin. Taking into account the M different processes, all j templates for $j = 1, ..., M$ can be extracted by evaluating the probability density for the respective process. Correspondingly, each process shows a certain yield n_j. The likelihood is

$$L(\boldsymbol{n}) = \prod_{i=1}^{N_{\text{bins}}} \frac{\nu_i(\boldsymbol{n})^{m_i}}{m_i!} e^{\nu_i(\boldsymbol{n})}, \tag{7.10}$$

where m_i and $\nu_i(\boldsymbol{n})$ indicate the number of measured and expected events, respectively. The set of process yields is denoted by $\boldsymbol{n}$. The expected entries ν_i for bin i in the histogram are assumed to be

$$\nu_i(\boldsymbol{n}) = \sum_{j=1}^{M} \nu_{ij}(n_j), \tag{7.11}$$

where the expected number of events in bin i contributed by process j is indicated by ν_{ij}. The calculation of ν_{ij} is given by

$$\nu_{ij} = n_j f_{ij}(h_j) \quad . \tag{7.12}$$

The f_{ij} represents the fraction of the number of events in bin i contributed by process j, h_{ij}, relative to the total number of entries in the histogram

$$f_{ij} = \frac{h_{ij}}{\sum h_{ij}} \quad . \tag{7.13}$$

The interpretation of occurring systematic uncertainties is performed by including so-called nuisance parameters δ^{sys}. These systematic uncertainties are caused by the event generation, simulation, and reconstruction. In particular, the process rate and the template shape

can be affected by systematic uncertainties. Assuming R different sources, the systematic uncertainties affecting the process rate n_j can be integrated into the likelihood function by the transformation

$$n_j \to n_j \sum_{k}^{N_{j,\text{rate}}} (1 + \epsilon_{jk}\delta_{jk}^{\text{sys}}) \ , \tag{7.14}$$

where ϵ_{jk} is the relative uncertainty corresponding to the n_j and δ_{jk}^{sys} indicates the nuisance parameter for a process j arising from a source k. Here, a Gaussian distribution with mean value of zero and a standard deviation of one is assumed for the nuisance parameters. Changes on the template shape consequently influence the fraction given by Equation (7.13). Therefore, the fraction f_{ij} is substituted according to

$$f_{ij} \to f'_{ij} = \frac{h_{ij}(1 + \epsilon_{ij}\delta_{ij}^{\text{sys}})}{\sum h_{ij}(1 + \epsilon_{ij}\delta_{ij}^{\text{sys}})} \ , \tag{7.15}$$

where the relative uncertainty ϵ_{ij} and the nuisance parameter δ_{ij}^{sys} correspond to bin i and template j. Since this analysis requires a simultaneous fit of the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ selection and the $\mathrm{B}^+ \to \pi^0\ell^+\nu_\ell$ selection, different channels k ($k = 1, ..., N_{\text{channels}}$) are incorporated into the likelihood function in Equation (7.10), resulting in

$$L(\boldsymbol{n}) = \prod_{k=1}^{N_{\text{channels}}} \prod_{i=1}^{N_{\text{bins}}} \frac{\nu_{ik}(\boldsymbol{n})^{m_{ik}}}{m_{ik}!} e^{-\nu_{ij}(\boldsymbol{n})} \ , \tag{7.16}$$

where the expected number of events in bin i for channel l, ν_{il}, is given by

$$\nu_{ik} = \sum_{j=1}^{M} n_j f_{ikj} \epsilon_{kj} \quad . \tag{7.17}$$

As stated above, Equation (7.17) shows a dependence on the fraction ϵ_{kj} of process j in channel k, connecting the process yields n_j. Equation (7.16) forms the basis of the minimisation of the negative log-likelihood, the corresponding set of parameters consists of the process yield n_j, the nuisance parameters $\boldsymbol{\theta}$, and the fraction ϵ_{kj}.

7.4. Branching Fraction Determination

The resulting signal yields extracted from the M_{miss}^2 distribution are utilised to determine the partial branching fraction of the $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ decay for photon energies $\gamma > 1\,\text{GeV}$. The partial branching fraction $\Delta\mathcal{B}$ for $\mathrm{B}^+ \to \mathrm{e}^+\nu_\mathrm{e}\gamma$ and $\mathrm{B}^+ \to \mu^+\nu_\mu\gamma$ is defined as

$$\Delta\mathcal{B}(\mathrm{B}^+ \to \ell^+\nu_\ell\gamma) = \frac{N_{\text{sig}}}{2 \cdot \epsilon_\gamma \cdot N_{\mathrm{B\overline{B}}} \cdot \mathcal{B}(\Upsilon(4\mathrm{S}) \to \mathrm{B}^+\mathrm{B}^-)} \ , \tag{7.18}$$

where N_{sig} denotes the extracted signal yield and ϵ_γ is the reconstruction efficiency, corresponding to the ratio of reconstructed $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ signal events to the number of produced $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ signal MC events. The number of $\mathrm{B\overline{B}}$ pairs recorded in the data set is indicated by $N_{\mathrm{B\overline{B}}}$ and the branching fraction of the $\Upsilon(4\mathrm{S})$ decaying into a pair of charged B mesons

in the denominator is given by $\mathcal{B}(\Upsilon(4S) \to B^+B^-) = 0.514$ [21]. The factor of 2 in the denominator occurs due to the two B mesons in the $\Upsilon(4S)$ decay. In the case of a combined partial branching fraction for both lepton flavors, this factor turns into 4.

Considering the $B^+ \to \pi^0 \ell^+ \nu_\ell$ reconstruction, the branching fraction is given by

$$\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell) = \frac{N_\pi}{2 \cdot \epsilon_\pi \cdot N_{B\overline{B}} \cdot \mathcal{B}(\Upsilon(4S) \to B^+B^-)\cdot} \,, \tag{7.19}$$

where N_π denotes the extracted $B^+ \to \pi^0 \ell^+ \nu_\ell$ yield and ϵ_π the reconstruction efficiency for $B^+ \to \pi^0 \ell^+ \nu_\ell$ events. The MC corrections described in Section 6.6 are incorporated in the reconstruction efficiencies ϵ_γ and ϵ_π.

7.5. Fitting Setup

This section describes the used templates for the $B^+ \to \ell^+ \nu_\ell \gamma$ and the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel, respectively.

7.5.1. $B^+ \to \ell^+ \nu_\ell \gamma$ Templates

Considering the $B^+ \to \ell^+ \nu_\ell \gamma$ selection, three templates are extracted from the missing mass distribution of the MC data.

$B^+ \to \ell^+ \nu_\ell \gamma$ Signal
This template describes the nominal signal component.

$B^+ \to \pi^0 \ell^+ \nu_\ell$ Background
The template of the peaking background, i.e. mistakenly reconstructed $B^+ \to \pi^0 \ell^+ \nu_\ell$ events in the $B^+ \to \ell^+ \nu_\ell \gamma$ reconstruction. This template is supposed to be constrained by the corresponding process in the control channel.

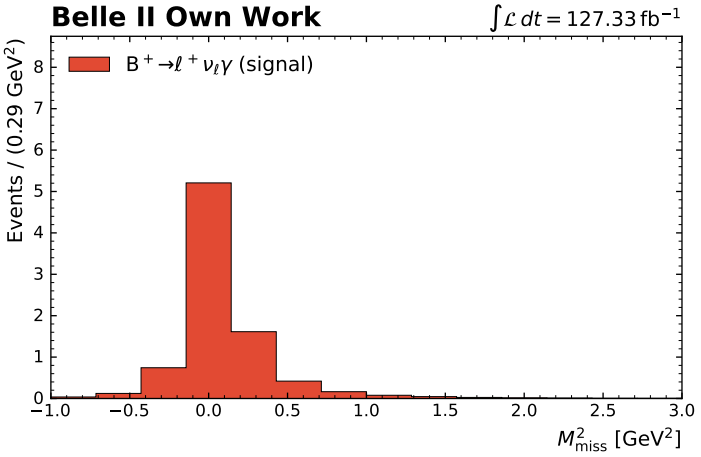
Remaining Background
This template contains all events, which are neither $B^+ \to \ell^+ \nu_\ell \gamma$ nor $B^+ \to \pi^0 \ell^+ \nu_\ell$ events, but appear in the $B^+ \to \ell^+ \nu_\ell \gamma$ reconstruction.

The weights are adjusted, so that the underlying M^2_{miss} distribution of all templates corresponds to a recorded data luminosity of $\mathcal{L} = 127.33\,\text{fb}^{-1}$. The corresponding templates are shown in Figure 7.1. The distinct templates for each lepton flavor ($\ell = e, \mu$) can be found in Appendix C.

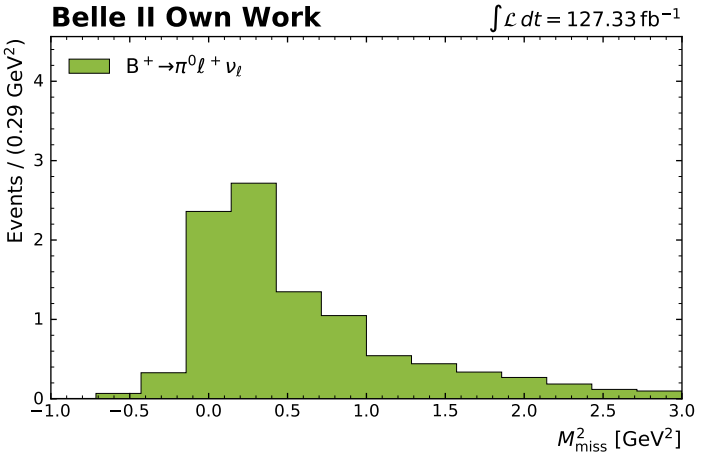
7.5.2. $B^+ \to \pi^0 \ell^+ \nu_\ell$ Templates

In the case of the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel, two processes templates are considered.

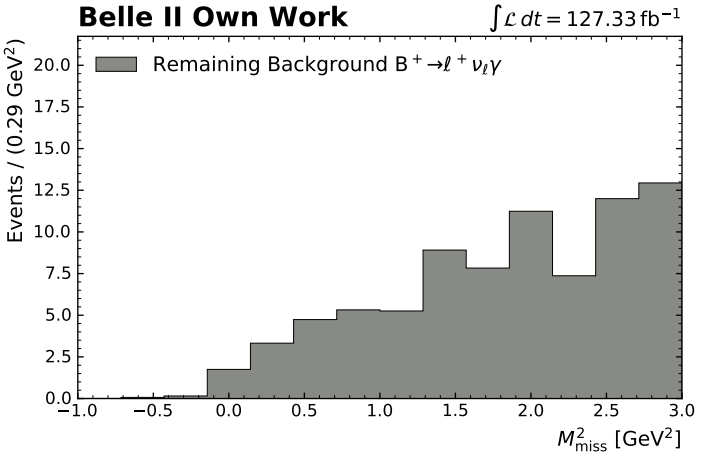
$B^+ \to \pi^0 \ell^+ \nu_\ell$ Signal
The template functions as the signal process in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel. Concurrently, this template is connected to the $B^+ \to \pi^0 \ell^+ \nu_\ell$ background template in the $B^+ \to \ell^+ \nu_\ell \gamma$ channel, and is used to constrain this background.



(a) $B^+ \to \ell^+ \nu_\ell \gamma$ signal template.



(b) $B^+ \to \pi^0 \ell^+ \nu_\ell$ background template.



(c) Remaining background template.

Figure 7.1.: The resulting M^2_{miss} template histograms evaluated on MC for the $B^+ \to \ell^+ \nu_\ell \gamma$ reconstruction channel.

Remaining Background

The remaining background template corresponds to the distribution of all events, which do not originate from $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays but are mistakenly reconstructed as such.

Again, the weights are scaled according to the recorded luminosity of $\mathcal{L} = 127.33\,\text{fb}^{-1}$. The resulting templates are presented in Figure 7.2. The distinct templates for each lepton flavor ($\ell = e, \mu$) can be found in Appendix C.

7.6. Simultaneous Fit

As stated in the previous chapters, this work uses a simultaneous fit to constrain the $B^+ \to \pi^0 \ell^+ \nu_\ell$ background in the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ analysis. The simultaneous fit considers both reconstruction channels and connects the $B^+ \to \pi^0 \ell^+ \nu_\ell$ process of both channels. Therefore, according to Section 7.5.1 amd Section 7.5.2 the resulting four template components of the simultaneous fit are listed below.

- $B^+ \to \ell^+ \nu_\ell \gamma$ (nominal signal)
- $B^+ \to \pi^0 \ell^+ \nu_\ell$
- Remaining Background in the $B^+ \to \ell^+ \nu_\ell \gamma$ reconstruction
- Remaining Background in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ reconstruction

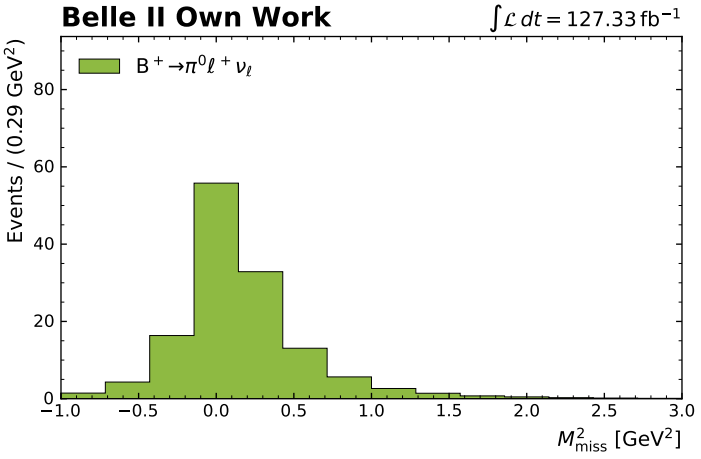
Since the $B^+ \to \pi^0 \ell^+ \nu_\ell$ process is present in both channels, the contributions to the total $B^+ \to \pi^0 \ell^+ \nu_\ell$ yield of the respective templates is determined. In the following, this ratio is referred to as ϵ_{yield}.

7.7. Fit on Asimov Data

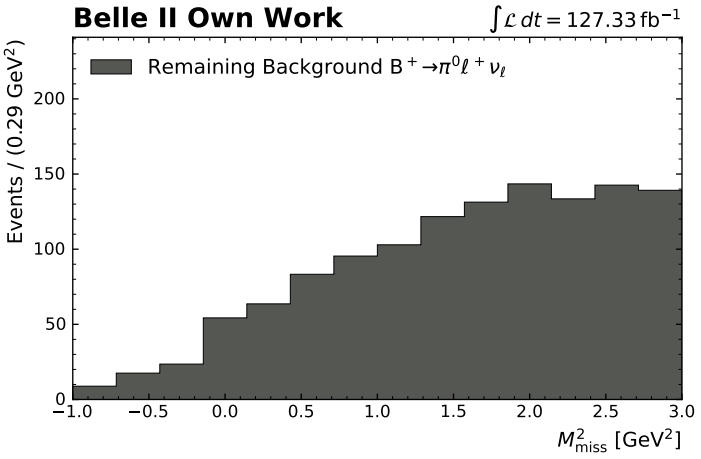
To validate the fit before real data is revealed, a fit on so-called Asimov data [49] is performed. This artificial data set is based on the MC expectations, thus the number of measured events in a bin is set to the number of the summed template events in that bin. The corresponding errors on the Asimov data points are assumed to be the standard Poisson deviation. For this validation, a luminosity of the currently recorded data from Belle II, $\mathcal{L} = 127.33\,\text{fb}^{-1}$, is chosen to simulate the Asimov data. Figure 7.3 presents the stacked templates corresponding to the $B^+ \to \ell^+ \nu_\ell \gamma$ channel (left) and the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel (right) for the M^2_{miss} distribution with Asimov data. As can be seen, the Asimov data exactly matches the bin entries of the template fit results, which is given by definition. According to Equation (7.18), the partial branching fraction obtained from the Asimov data set can be calculated to validate the underlying template fit for the $B^+ \to \ell^+ \nu_\ell \gamma$ channel. The measurement gives

$$\Delta\mathcal{B}_{\text{Asimov}}(B^+ \to \ell^+ \nu_\ell \gamma) = (10.00 \pm 4.67) \cdot 10^{-6}\ , \tag{7.20}$$

with a significance of $1.29\,\sigma$. The error indicates the propagated statistical uncertainties of the $B^+ \to \ell^+ \nu_\ell \gamma$ yield. The result is equal to the assumed theoretical branching fraction of both lepton flavors, which was used as input for the MC sample production. Thus, closure

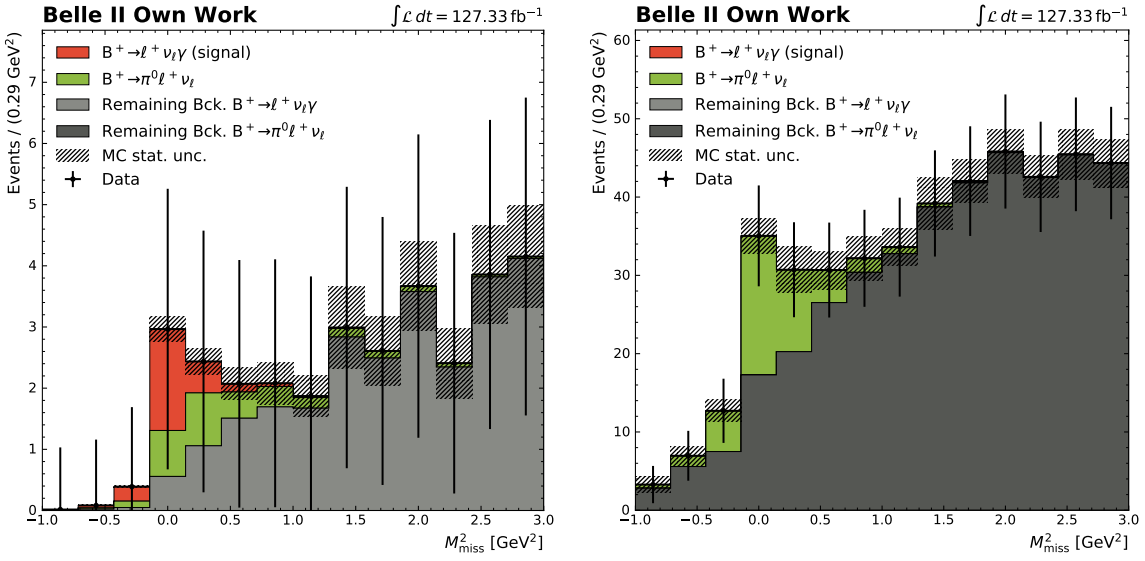


(a) $B^+ \to \pi^0 \ell^+ \nu_\ell$ signal template.



(b) Remaining Background template.

Figure 7.2.: The resulting M^2_{miss} template histograms evaluated on MC for the $B^+ \to \pi^0 \ell^+ \nu_\ell$ reconstruction channel.



(a) $B^+ \to \ell^+ \nu_\ell \gamma$ channel.

(b) $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel.

Figure 7.3.: The resulting M^2_{miss} template fit on the Asimov data set. The nominal $B^+ \to \ell^+ \nu_\ell \gamma$ channel is shown on the left, whereas the control channel $B^+ \to \pi^0 \ell^+ \nu_\ell$ is presented on the right. Each fit is performed with the four template components as described in Section 7.6. The respective remaining background component appears in the $B^+ \to \ell^+ \nu_\ell \gamma$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel, indicated with "Remaining Bck. $B^+ \to \ell^+ \nu_\ell \gamma$ " and "Remaining Bck. $B^+ \to \pi^0 \ell^+ \nu_\ell$ ".

Table 7.1.: Overview of the extracted yields and the fraction of yield events in the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel, ϵ_{yield} for both $B^+ \to \ell^+ \nu_\ell \gamma$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel, resulting from the Asimov fit.

	Asimov Fit Results
N_{sig}	2.71 ± 2.53
N_π	46.22 ± 12.90
ϵ_{yield}	0.93

Table 7.2.: Overview of the determined (partial) branching fraction for the $B^+ \to \ell^+ \nu_\ell \gamma$ channel and the $B^+ \to \pi^0 \ell^+ \nu_\ell$ channel as well as the assumed branching fraction used for MC production.

	Asimov Fit Result	Assumed Value
$\Delta\mathcal{B}(B^+ \to \ell^+ \nu_\ell \gamma)$	$(10.00 \pm 4.67) \cdot 10^{-6}$	$10.00 \cdot 10^{-6}$
$\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell)$	$(1.56 \pm 0.44) \cdot 10^{-4}$	$1.56 \cdot 10^{-4}$

of the fit is confirmed. Similar to the above, the branching fraction for the $B^+ \to \pi^0 \ell^+ \nu_\ell$ process is determined according to Equation (7.19). Since the respective resulting yield contains the contributions from both channels, the yield is weighted with the corresponding fraction, ϵ_{yield}, as mentioned in Section 7.6. The result

$$\mathcal{B}_{\text{Asimov}}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (1.56 \pm 0.44) \cdot 10^{-4} \,, \tag{7.21}$$

with a significance of $3.97\,\sigma$. The error indicates the propagated statistical uncertainties of the $B^+ \to \pi^0 \ell^+ \nu_\ell$ yield. The result is equal to the assumed branching fraction used for MC production of both flavors. The distinct Asimov fit for each lepton flavor ($\ell = e, \mu$) can be found in Appendix C.

7.8. Validation On $B^+ \to \pi^0 \ell^+ \nu_\ell$ Channel

A further validation of the fit is performed by the evaluation of the control channel on real data collected by the Belle II experiment. More precisely, a sub-sample corresponding to an integrated luminosity of $\mathcal{L} = 63.25\,\text{fb}^{-1}$ is used. According to Equation (7.19), the simultaneous fit gives a branching fraction for $B^+ \to \pi^0 \ell^+ \nu_\ell$ of

$$\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (1.17 \pm 0.65) \cdot 10^{-4} \,, \tag{7.22}$$

with a significance of $1.92\,\sigma$. The error refers to the propagated statistical uncertainties of the $B^+ \to \pi^0 \ell^+ \nu_\ell$ yield. Both lepton flavors are considered. Within the uncertainties, the result compares with the branching fraction reported by Belle II in the dedicated 2021 analysis [12], which is in agreement with the current world average $\mathcal{B}_{\text{PDG}}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (7.80 \pm 0.27) \cdot 10^{-5}$ [21].

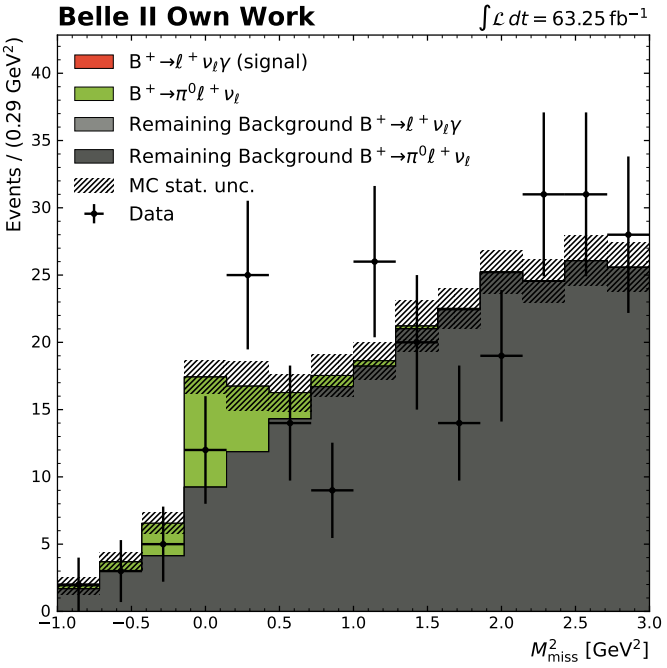


Figure 7.4.: The resulting template fit on the data set corresponding to an integrated luminosity of $\mathcal{L} = 63.25\,\mathrm{fb}^{-1}$.

8. Conclusion

This thesis presents the analysis of the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$ for high energetic photons with $E_\gamma > 1\,\text{GeV}$ (Section 2.1). The study is performed on simulated data, and the control channel $B^+ \to \pi^0 \ell^+ \nu_\ell$ (Section 2.2) is validated with an integrated luminosity of $\mathcal{L} = 63.25\,\text{fb}^{-1}$ of real data collected by the Belle II experiment (Chapter 3).

The decay $B^+ \to \ell^+ \nu_\ell \gamma$ provides access to the first inverse moment λ_B of the light-cone distribution amplitude of the B meson, which is an important input to the QCD factorization scheme used in non-leptonic B decay amplitudes (Section 2.3). Although the decay $B^+ \to \ell^+ \nu_\ell \gamma$ has not been observed at any collider experiment yet due to its low branching fraction, advanced analysis methods as well as experimental improvements provide new opportunities at Belle II. In anticipation of the forthcoming large data set to be collected by the Belle II experiment, this work aims to improve on previous searches for $B^+ \to \ell^+ \nu_\ell \gamma$ and analyses on λ_B. The goal of this work is to determine the partial branching fraction of $B^+ \to \ell^+ \nu_\ell \gamma$ decays, $\Delta\mathcal{B}(B^+ \to \ell^+ \nu_\ell \gamma)$, as well as the $B^+ \to \pi^0 \ell^+ \nu_\ell$ branching fraction $\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell)$. A comprehensive discussion of systematic uncertainties remains for other studies. Starting from the discussions and results of this thesis, further studies could continue the search for the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$.

To achieve the goal of determining the branching fractions, an evaluation of the distribution of the missing mass squared (Section 5.2), M^2_{miss}, of $B^+ \to \ell^+ \nu_\ell \gamma$ decays is performed. Since $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays show similar characteristics as $B^+ \to \ell^+ \nu_\ell \gamma$ events in the M^2_{miss} distribution, a pure extraction of $B^+ \to \ell^+ \nu_\ell \gamma$ events is challenging. Thus, in addition to the nominal study on $B^+ \to \ell^+ \nu_\ell \gamma$ decays, a concurrent analysis of $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays is performed. This control channel serves to constrain the $B^+ \to \pi^0 \ell^+ \nu_\ell$ background in the nominal $B^+ \to \ell^+ \nu_\ell \gamma$ analysis.

Event candidates are reconstructed by using the generic hadronic Full Event Interpretation (Section 4.5) for the tag-side reconstruction and a following signal-side selection (Section 6.2). The selection criteria for each reconstruction stage are adapted to the current recommendations of the Belle II collaboration. Since background events in the reconstruction cause problems for following analysis steps, further background suppression tasks are performed. In particular, continuum events are suppressed by a multivariate classifier (Section 6.3.1). As for the $B^+ \to \ell^+ \nu_\ell \gamma$ anaylsis, $B^+ \to \pi^0 \ell^+ \nu_\ell$ events, referred to as peaking background, are rejected with a second multivariate classifier (Section 6.3.2). The selections applied on both classifiers are simultaneously optimised via maximisation of Punzi's figure of merit (Section 6.3.3) within an optimisation window of $M^2_{\text{miss}} \in [-0.5, 0.5]\,\text{GeV}$. Furthermore, a

verification of the control channel $B^+ \to \pi^0 \ell^+ \nu_\ell$ is studied by comparison of the angular distributions between the two daughter photons of the π^0 (Section 6.4). The performance of the electromagnetic calorimeter is tested to indicate the distinction between single and merged photons (Section 6.5).

As for the extraction of the respective yield for $B^+ \to \ell^+ \nu_\ell \gamma$ and $B^+ \to \pi^0 \ell^+ \nu_\ell$, a template method with a binned maximum likelihood fit is performed (Section 7.3). More precisely, the templates are obtained from the resulting M^2_{miss} distribution of simulated data. Each template is obtained from the underlying distribution of simulated data for the respective channel (Section 7.5.1, Section 7.5.2). As a precondition for the simultaneous fit, both $B^+ \to \pi^0 \ell^+ \nu_\ell$ templates obtained from the nominal and control channel are connected. In total, four components are chosen to form the basis of the simultaneous fit (Section 7.6). A first validation of the distinct templates and the simultaneous fit is ensured by the application on the simulated Asimov data set (Section 7.7). Closure between the fit result and the initial branching fraction used for simulation is given for both channels.

Finally, the simultaneous fit is performed on a data sub-set corresponding to an integrated luminosity of $\mathcal{L} = 63.25\,\text{fb}^{-1}$ obtained by the Belle II experiment (Section 7.8). Since the complete interpretation of the data fit is beyond the scope of this thesis, only the $B^+ \to \pi^0 \ell^+ \nu_\ell$ control channel is evaluated. The extracted branching fraction of $\mathcal{B}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (1.17 \pm 0.65) \cdot 10^{-4}$ is obtained. Within statistical uncertainties, this compares with current results by Belle II [12], which are in agreement with the world average $\mathcal{B}_{\text{PDG}}(B^+ \to \pi^0 \ell^+ \nu_\ell) = (7.80 \pm 0.27) \cdot 10^{-5}$. Thus, a validation of the fit is given by the control channel.

A. Continuum Suppression

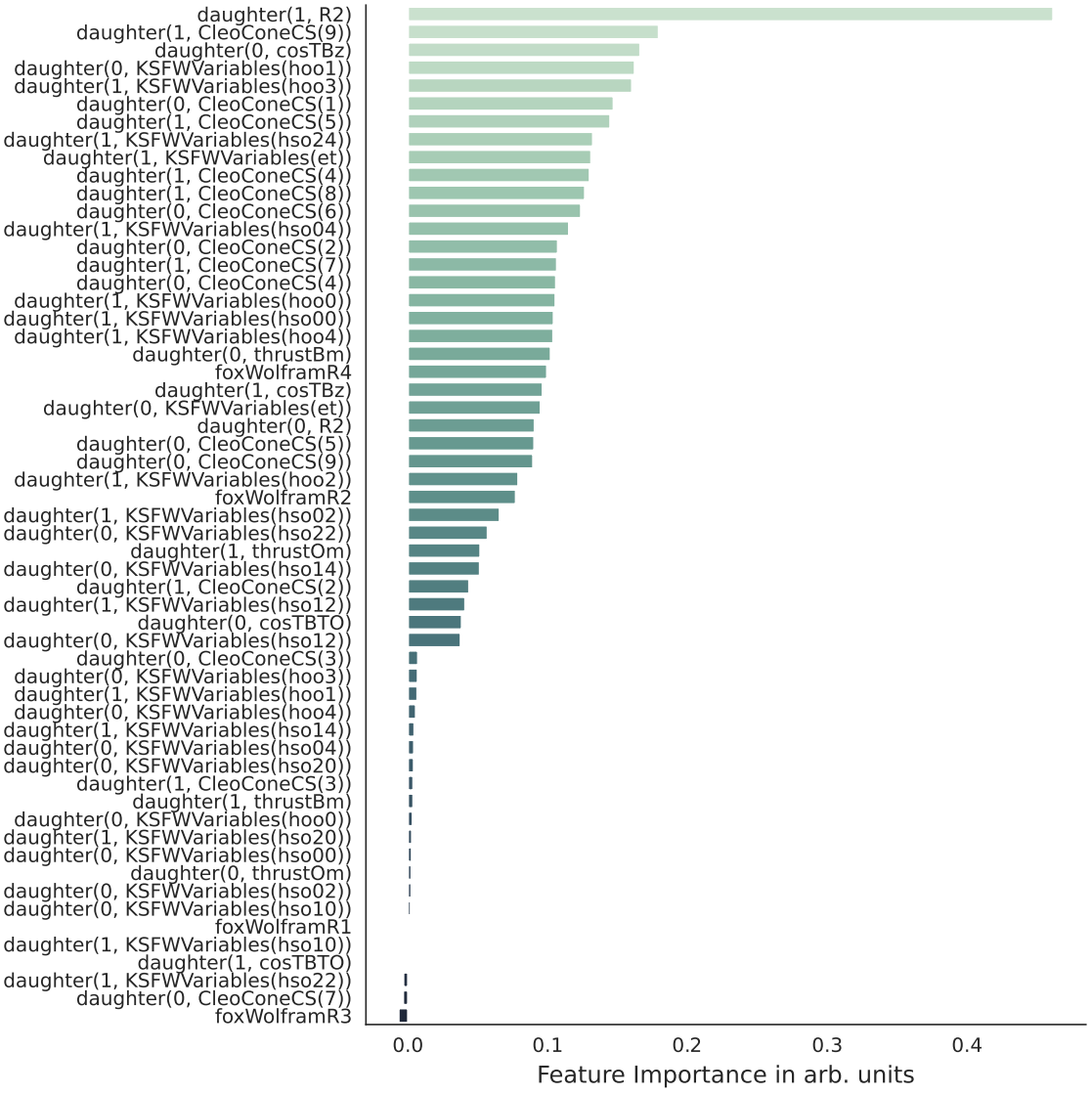


Figure A.1.: Feature importance for the continuum suppression classifier based on the `FastBDT` calculation.

B. Peaking Background Suppression

B.1. Features *BeamBackgroundSuppression* MVA

The *BeamBackgroundSuppression* MVA uses MC samples of signal photons and beam background photons. The features used for training are listed below, the order is given by decreasing significance.

- *clusterTiming*
- *clusterPulseShapeDiscriminationMVA*
- *clusterE*
- *clusterTheta*
- *clusterZernikeMVA*
- *clusterLAT*
- *clusterSecondMoment*

B.2. Features *writePi0EtaVeto* MVA

The *writePi0EtaVeto* MVA is trained with the following variables. As for the hard photon, a lower limit of $E_\gamma > 1.4\,\text{GeV}$ in the CMS frame is set.

- M (for π^0, η candidates)
- *daughter(1,E)*
- *daughter(1,clusterTheta)*
- *daughter(1,minC2TDist)*
- *daughter(1,clusterZernikeMVA)*
- *daughter(1,clusterNHits)*
- *daughter(1,clusterE9E21)*
- *cosHelicityAngleMomentum*

Furthermore, the standard mode is chosen, which provides loose energy selection but no selections on *clusterNHits* for the soft photon.

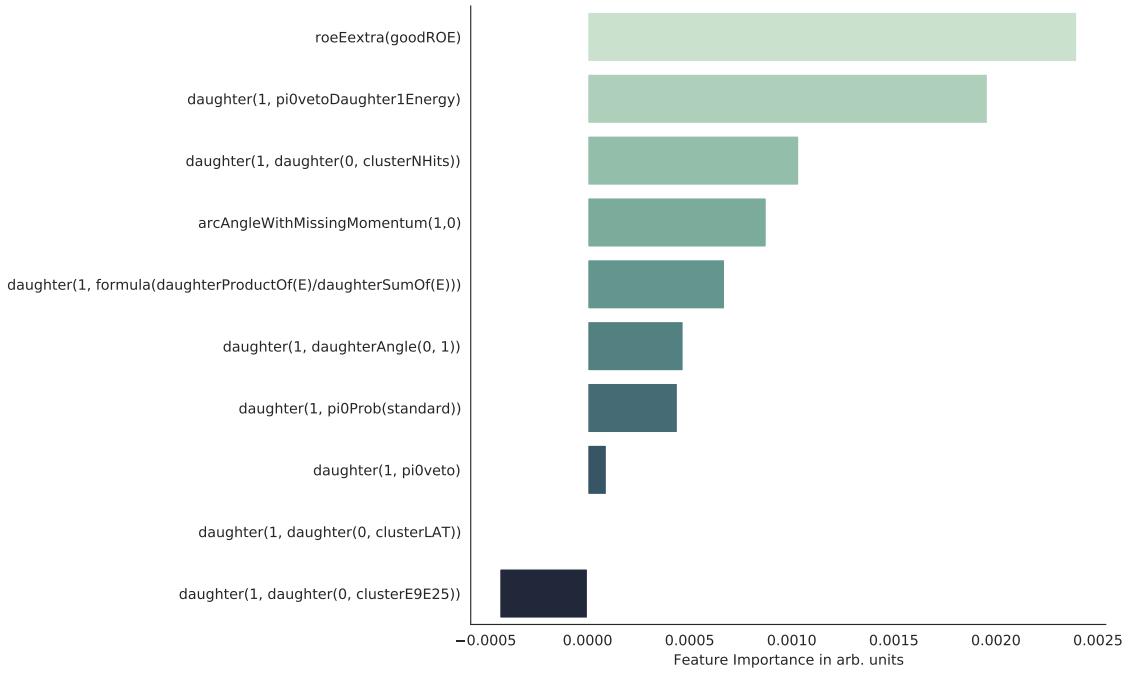
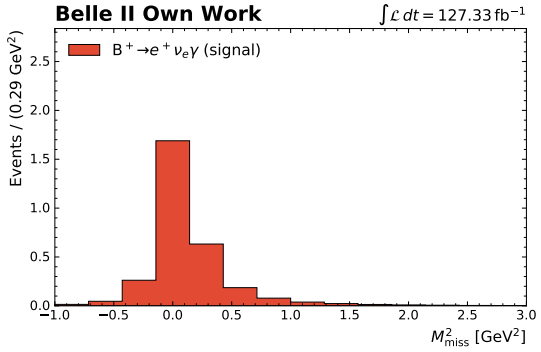


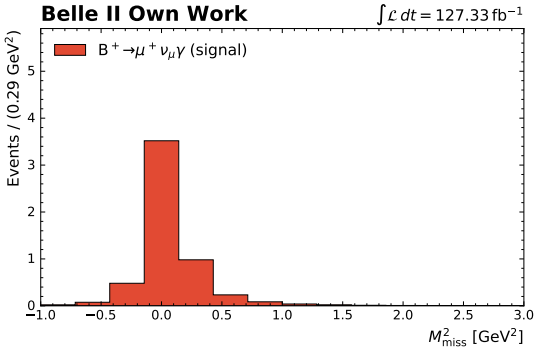
Figure B.1.: Feature importance for the peaking background classifier based on the `FastBDT` calculation.

C. Lepton Flavor Dependent Template Fits

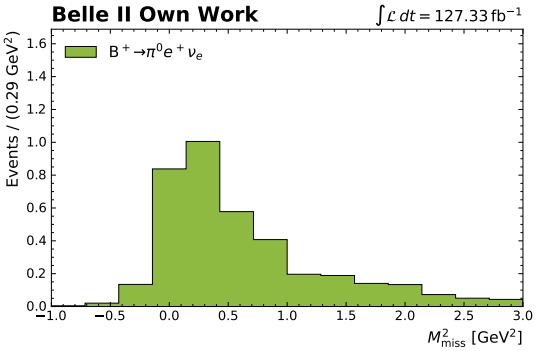
The following gives an overview of the used templates and Asimov fits for the distinct lepton flavor selection, $B^+ \to e^+\nu_e\gamma$ and $B^+ \to \pi^0 e^+\nu_e$, and $B^+ \to \mu^+\nu_\mu\gamma$ and $B^+ \to \pi^0\mu^+\nu_\mu$.



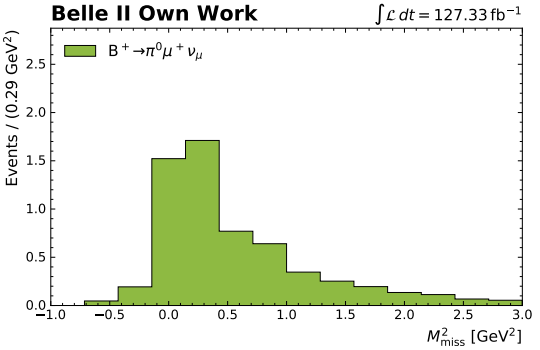
(a) $B^+ \to e^+ \nu_e \gamma$ signal template.



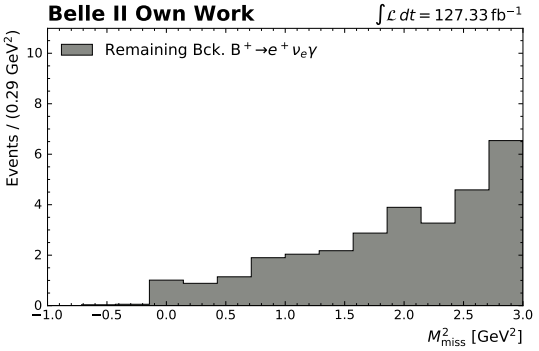
(b) $B^+ \to \mu^+ \nu_\mu \gamma$ signal template.



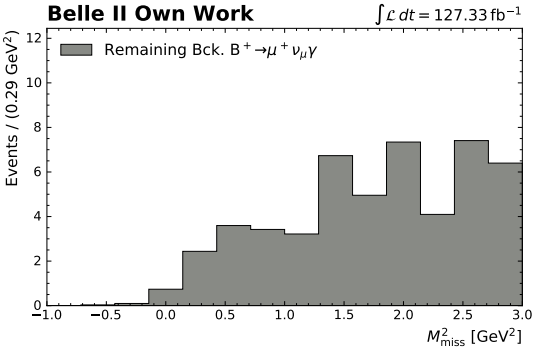
(c) $B^+ \to \pi^0 \ell^+ \nu_\ell$ background template.



(d) $B^+ \to \pi^0 \ell^+ \nu_\ell$ background template.

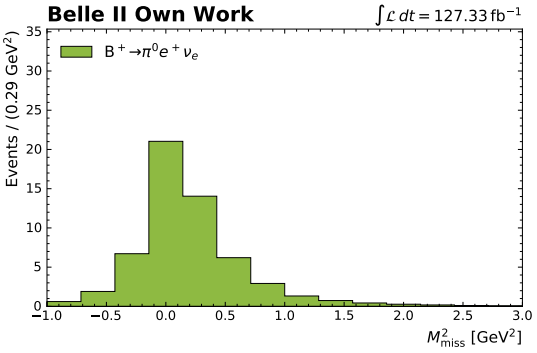


(e) Remaining background template.

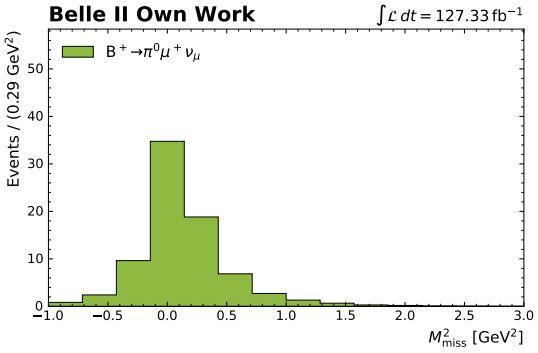


(f) Remaining background template.

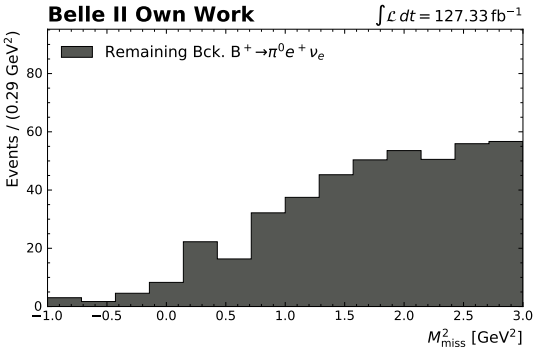
Figure C.1.: The resulting M^2_{miss} template histograms evaluated on MC for the $B^+ \to e^+ \nu_e \gamma$ (left) and $B^+ \to \mu^+ \nu_\mu \gamma$ (right) reconstruction channel.



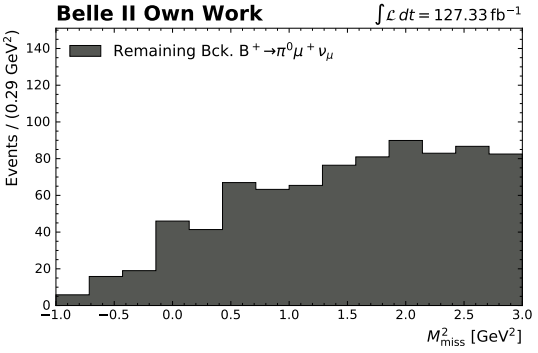
(a) $\mathrm{B}^+ \to \pi^0 e^+ \nu_e$ signal template.



(b) $\mathrm{B}^+ \to \pi^0 \mu^+ \nu_\mu$ signal template.

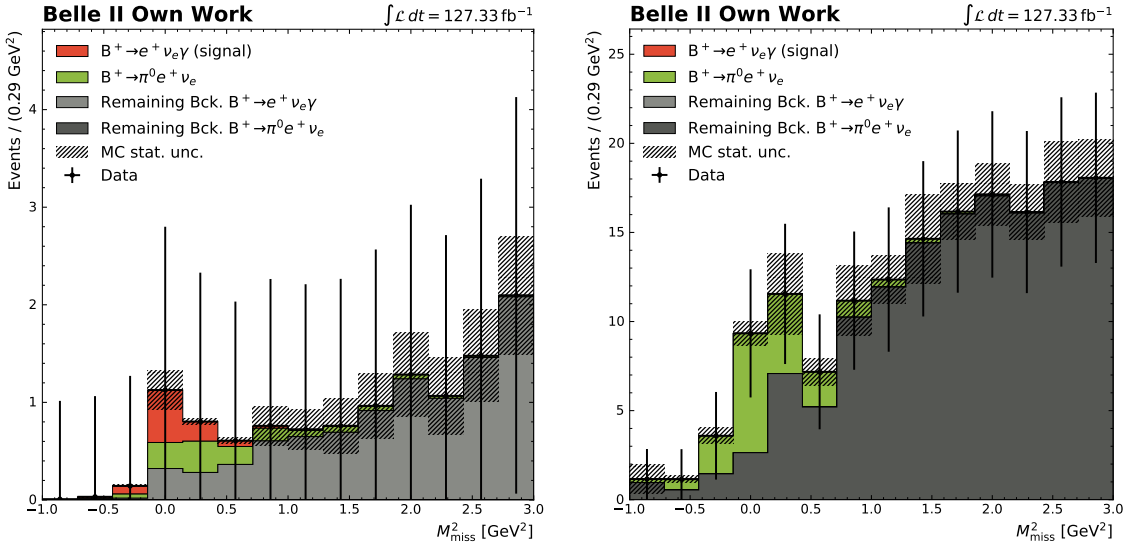


(c) $\mathrm{B}^+ \to \pi^0 e^+ \nu_e$ background template.



(d) $\mathrm{B}^+ \to \pi^0 \mu^+ \nu_\mu$ background template.

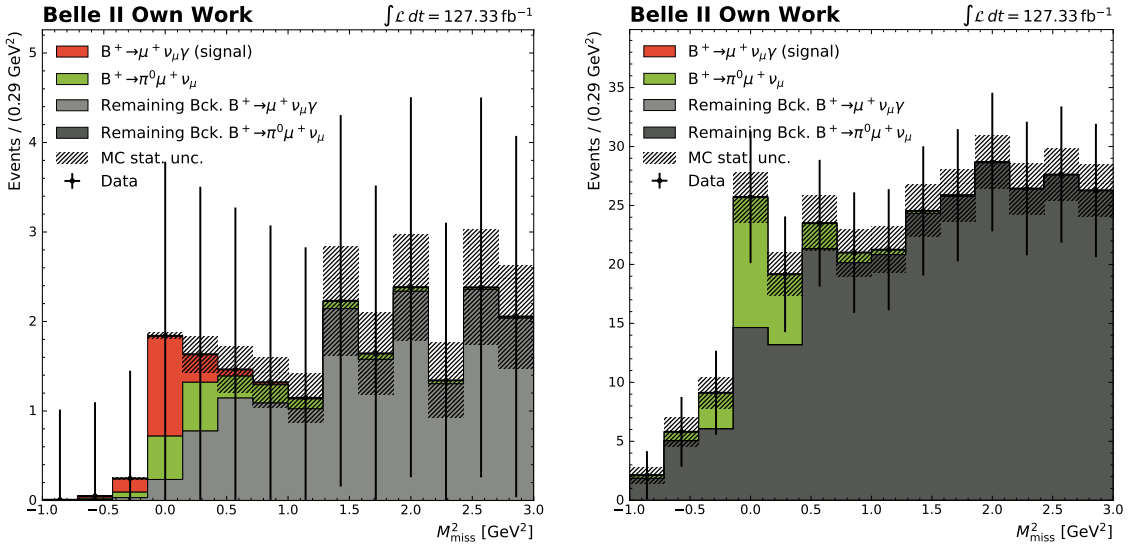
Figure C.2.: The resulting M^2_{miss} template histograms evaluated on MC for the $\mathrm{B}^+ \to \pi^0 e^+ \nu_e$ (left) and $\mathrm{B}^+ \to \pi^0 \mu^+ \nu_\mu$ (right) reconstruction channel.



(a) $B^+ \to e^+ \nu_e \gamma$ channel.

(b) $B^+ \to \pi^0 e^+ \nu_e$ channel.

Figure C.3.: The resulting M^2_{miss} template fit on the Asimov data set. The nominal $B^+ \to e^+ \nu_e \gamma$ channel is shown on the left, whereas the control channel $B^+ \to \pi^0 e^+ \nu_e$ is presented on the right. Each fit is performed with the four template components as described in Section 7.6. The respective remaining background component appears in the $B^+ \to e^+ \nu_e \gamma$ and $B^+ \to \pi^0 e^+ \nu_e$ channel, indicated with "Remaining Bck. $B^+ \to e^+ \nu_e \gamma$ " and "Remaining Bck. $B^+ \to \pi^0 e^+ \nu_e$ ".



(a) $B^+ \to \mu^+ \nu_\mu \gamma$ channel.

(b) $B^+ \to \pi^0 \mu^+ \nu_\mu$ channel.

Figure C.4.: The resulting M^2_{miss} template fit on the Asimov data set. The nominal $B^+ \to \mu^+ \nu_\mu \gamma$ channel is shown on the left, whereas the control channel $B^+ \to \pi^0 \mu^+ \nu_\mu$ is presented on the right. Each fit is performed with the four template components as described in Section 7.6. The respective remaining background component appears in the $B^+ \to \mu^+ \nu_\mu \gamma$ and $B^+ \to \pi^0 \mu^+ \nu_\mu$ channel, indicated with "Remaining Bck. $B^+ \to \mu^+ \nu_\mu \gamma$ " and "Remaining Bck. $B^+ \to \pi^0 \mu^+ \nu_\mu$ ".

Bibliography

[1] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

[2] K. Abe et al. Observation of large *CP* violation in the neutral B meson system. *Phys. Rev. Lett.*, 87:091802, Aug 2001.

[3] M. Gelb et al. Search for the rare decay of $b^+ \to \ell^+ \nu\gamma$ with improved hadronic tagging. *Physical Review D*, 98(11), Dec 2018.

[4] M. Beneke and J. Rohrwild. B meson distribution amplitude from $B \to \gamma\ell\nu$. *The European Physical Journal C*, 71(12):1818, December 2011. arXiv: 1110.3228.

[5] M. Beneke, V. M. Braun, Y. Ji, and Y. Wei. Radiative leptonic decay $B \to \gamma\ell\nu_\ell$ with subleading power corrections. *Journal of High Energy Physics*, 2018(7):154, July 2018. arXiv: 1804.04962.

[6] M. Gelb. *Search for the Rare Decay* $\mathrm{B}^+ \to \ell^+\nu_\ell\gamma$ *with the Full Event Interpretation at the Belle Experiment*. PhD Thesis, Karlsruhe Institute of Technology (KIT), 2018.

[7] P. Gambino et al. Challenges in semileptonic B decays. *The European Physical Journal C*, 80(10), October 2020. Publisher: Springer Science and Business Media LLC.

[8] Y. Shen, Y. Wei, X. Zhao, and S. Zhou. Revisiting radiative leptonic B decay. *Chinese Physics C*, 44(12):123106, November 2020. Publisher: IOP Publishing.

[9] A. Sibidanov et al. Study of Exclusive $B \to X_u\ell\nu$ Decays and Extraction of $|V_{\mathrm{ub}}|$ using Full Reconstruction Tagging at the Belle Experiment. *Physical Review D*, 88(3), Aug 2013.

[10] J. A. Bailey et al. $|V_{ub}|$ from $B \to \pi\ell\nu$ decays and (2+1)-flavor lattice QCD. *Physical Review D*, 92(1), July 2015. Publisher: American Physical Society (APS).

[11] C. Bourrely, L. Lellouch, and I. Caprini. Model-independent description of $B \to \pi\ell\nu$ decays and a determination of $|V_{ub}|$. *Physical Review D*, 79(1):013008, January 2009.

[12] F. Abudinén et al. Exclusive $B \to X_u\ell\nu_\ell$ Decays with Hadronic Full-event-interpretation Tagging in 62.8 fb^{-1} of Belle II Data. *arXiv:2111.00710 [hep-ex]*, November 2021. arXiv: 2111.00710.

[13] 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(1-2):313–418, December 2000. arXiv: hep-ph/0006124.

[14] M. Beneke and M. Neubert. QCD factorization for $B \to PP$ and $B \to PV$ decays. *Nuclear Physics B*, 675(1-2):333–415, December 2003. arXiv: hep-ph/0308039.

[15] M. Beneke, T. Huber, and X. Li. NNLO vertex corrections to non-leptonic B decays: Tree amplitudes. *Nuclear Physics B*, 832(1-2):109–151, June 2010. arXiv: 0911.3655.

[16] V. M. Braun, D. Y. Ivanov, and G. P. Korchemsky. The B-Meson Distribution Amplitude in QCD. *Physical Review D*, 69(3):034014, February 2004. arXiv: hep-ph/0309330.

[17] A. Heller. *Search for $B^+ \to \ell^+\nu\gamma$ decays with hadronic tagging using the full Belle data sample*. PhD Thesis, Karlsruher Institut für Technologie (KIT), 2015.

[18] B. Aubert et al. Model-independent search for the decay $B^+ \to \ell^+\nu\gamma$. *Physical Review D*, 80(11):111105, December 2009.

[19] Y. Wang and Y. Shen. Subleading-power corrections to the radiative leptonic $B \to \gamma\ell\nu$ decay in QCD. *Journal of High Energy Physics*, 2018(5), May 2018. Publisher: Springer Science and Business Media LLC.

[20] K. Akai, K. Furukawa, and H. Koiso. SuperKEKB collider. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 907:188–199, Nov 2018.

[21] Particle Data Group, P. Zyla, et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8):083C01, August 2020.

[22] T. Abe et al. Belle II Technical Design Report, 2010.

[23] D. Matvienko. The Belle II experiment: status and physics program. *EPJ Web of Conferences*, 191:02010, January 2018.

[24] A. Moll. The Software Framework of the Belle II Experiment. *Journal of Physics: Conference Series*, 331(3):032024, December 2011.

[25] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun. The Belle II Core Software: Belle II Framework Software Group. *Computing and Software for Big Science*, 3(1):1, December 2019.

[26] A. Ryd and D. Lange. A Monte Carlo Generator for B-Physics. March 2004.

[27] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, 2015.

[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, July 2003.

[29] T. Keck. *Machine Learning at the Belle II Experiment: The Full Event Interpretation and Its Validation on Belle Data*. Springer Theses. Springer International Publishing, Cham, 2018.

[30] T. Keck et al. The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment. *Computing and Software for Big Science*, 3(1):6, December 2019. arXiv: 1807.08680.

[31] C. M. Bishop. *Pattern recognition and machine learning*. Information science and statistics. Springer, New York, 2006.

[32] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2009.

[33] T. Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Computing and Software for Big Science*, 1(1):2, November 2017.

[34] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. Adaptive computation and machine learning series. The MIT Press, Cambridge, Massachusetts, second edition edition, 2018.

[35] J. Friedman. Stochastic gradient boosting.

[36] M. Bauer. Extending the Full Event Interpretation to the υ (5S) Resonance. Master's thesis, Karlsruhe Institute of Technology (KIT), 2020.

[37] S. Jadach, B. F. L. Ward, and Z. Was. The precision Monte Carlo event generator KK for two-fermion final states in e^+e collisions. *Computer Physics Communications*, 130(3):260–325, 2000.

[38] C. Ramirez, J. F. Donoghue, and G. Burdman. Semileptonic $b \to u$ decay. *Physical Review D*, 41(5):1496–1503, March 1990. Publisher: American Physical Society.

[39] M. T. Prim. Search for $B \to \mu\nu_\mu$ at Belle, 2019.

[40] A. Heller. *Search for $B^+ \to \ell^+\nu\gamma$ decays with hadronic tagging using the full Belle data sample*. PhD Thesis, Karlsruhe Institute of Technology (KIT), 2015.

[41] M. Röhrken. *Time-Dependent CP Violation Measurements*. Springer Theses. Springer International Publishing, Cham, 2014.

[42] D. M. Asner et al. Search for exclusive charmless hadronic B decays. *Physical Review D*, 53(3):1039–1050, Feb 1996.

[43] G. C. Fox and S. Wolfram. Observables for the Analysis of Event Shapes in e^+e Annihilation and Other Processes. *Physical Review Letters*, 41(23):1581–1585, December 1978.

[44] A. Bevan et al. The Physics of the B Factories. *European Physical Journal C*, November 2014.

[45] G. Punzi. Sensitivity of searches for new signals and its optimization. *arXiv:physics/0308063*, December 2003. arXiv: physics/0308063.

[46] P. Virtanen et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.

[47] LeptonID group and Belle II Collaboration. Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2021. March 2021.

[48] G. Cowan. *Statistical Data Analysis*. Oxford University Press, 1998. Google-Books-ID: JhL2FiF3hO4C.

[49] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2):1554, February 2011. arXiv: 1007.1727.

[50] C. Hensel and K. Kröninger. *Statistical Methods Commonly Used in High Energy Physics*, pages 329–356. 06 2013.