Karlsruhe Institute of Technology

IEKP-KA/2016-6

Analysis of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decays with the Belle II Analysis Software Framework

Master's Thesis of

Felix Metzner

at the Department of Physics of the
Karlsruhe Institute of Technology (KIT)
Institute of Experimental Nuclear Physics (IEKP)

Reviewer:	Prof. Dr. Michael Feindt
Second reviewer:	Prof. Dr. Ulrich Husemann
Advisor:	Dr. Martin Heck
Second advisor:	Dr. Pablo Goldenzweig

June 1st, 2015 – May 31st, 2016

Karlsruher Institut für Technologie
Institut für Experimentelle Kernphysik
Campus Süd
Building 30.23
Wolfgang-Gaede-Str. 1
D-76128 Karlsruhe
Germany

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, June 1, 2016

. . . .

(Felix Metzner)

Contents

1. Introduction

Over the last decades the Standard Model of Particle Physics (SM) proved to be able to explain almost all of the observations made at particle physics experiments at man-made accelerator facilities. The model correctly predicted the components of the elemental particle zoo as well as the interactions among them, and is therefore one of the best-verified theories in physics. The last important confirmation of the model was the discovery of the Higgs boson — the last missing component of the SM particle zoo — by the two LHC experiments ATLAS and CMS in 2012 [1][2]. Despite the apparent beauty and completeness of the theory for the precise description of many measurements at collider experiments, the SM still lacks the capability to explain effects such as neutrino oscillation [3][4][5] or the baryon asymmetry of the universe [6]. Hence, the search for further divergences from the SM at particle accelerators is of particular interest for particle physicists, since such deviations might hint at necessary extensions to the model which could eventually explain the aforementioned observed effects.

Various particle physics experiments pursue different approaches in the search for physics beyond the Standard Model. The high-energy frontier attempts to unveil new physics by pushing towards ever higher energies, for instance at the Large Hadron Collider (LHC) at CERN. The general-purpose experiments ATLAS and CMS benefit from the high center of mass energy and high luminosity of the LHC to search for new heavy particles and deviations from the SM.

In contrast to this strategy stands the approach of the high-precision frontier. In this sector, among other facilities such as the LHCb experiment, are B factories an instrument to achieve new insight into particle physics. B factories operate mainly at the center of mass energy of the $\Upsilon(4S)$ resonance to produce B meson pairs in the clean environment of electron–positron collisions. The former B factory experiments Belle and BaBar [7] accomplished important goals, such as the discovery of CP-Violation in the B-system [8] and constraining the angles of the Unitarity Triangle [9]. The CP-Violation measurement was the initial aim for the construction of the Belle experiment and lead to a Nobel Prize in physics in 2008 for the two japanese physicists M. Kobayashi and T. Masukawa who described the origin of the broken symmetry and motivated the experiment. Furthermore, the evaluation of rare B decays is possible at these facilities. These decays are sensitive to deviations from the SM, as small effects originating from new physics are more likely to be detectable in branching fraction measurements of such rare decays. The potential to search for new physics via rare B decays is one of the motivations for the next generation of B factories which will be commenced with the upgrade of the Belle experiment at the Japanese High Energy Accelerator Research Organization (KEK) to Belle II.

However, to determine a deviation from the SM, accurate theoretical predictions to compare the experimental results with must be available [10]. As some Belle and BaBar studies already exceed the precision of theoretical predictions, the need for improvements on the theoretical side to match the accuracy of the next generation of B factories becomes apparent. The description of the non-trivial hadronic interactions occurring during B meson decays at B factories via Quantum Chromodynamics (QCD) is the bottleneck for the theoretical predictions. QCD calculations of long-distance interactions inside the B meson cannot be calculated perturbatively and thus call for a factorization approach which requires empirical input parameters [11] to reduce the uncertainties of the theoretical predictions. One such parameter can be determined via a branching fraction measurement of the rare leptonic decay $B^+ \to \ell^+ \nu_\ell \gamma$ which is the focus of this work.

The decay $B^+ \to \ell^+ \nu_\ell \gamma$ ($\ell = e, \mu$) was last investigated by *Andreas Heller et al.* [12] who set the most stringent upper limit on the branching fraction of the combined decay $B^+ \to \ell^+ \nu_\ell \gamma$, which, with $\mathcal{B}(B^+ \to \ell^+ \nu_\ell \gamma) < 3.5 \times 10^{-6}$ at 90% confidence level, lies in the region corresponding to theoretical SM predictions [13]. Improvements to algorithms crucial for the analysis of the decay have been made in the scope of the development of the new software framework for the Belle II experiment, motivating the reproduction of the analysis with the enhanced software. Of particular importance is a exclusive B-tagging algorithm reimplemented in the Belle II Analysis Software Framework (BASF2) under the name Full Event Interpretation (FEI) [14][15][16]. Exclusive B-tagging is a method to gain additional information about the signal event by reconstructing the second B meson originating from the $\Upsilon(4S)$ resonance, which accompanies the signal B meson. It allows one to infer the properties of the undetectable neutrino of the process of interest, but has a severe impact on the signal yield of the analysis, as the second B meson can be provided with an efficiency of less than 1 % in the case of the study[1] conducted by *Andreas Heller et al.* [12, p. 47]. The effect of the innovations introduced to the algorithm in the course of the reimplementation for BASF2 might increase the efficiency of the signal selection and thus the chances to measure the branching fraction of the decay $B^+ \to \ell^+ \nu_\ell \gamma$.

The Monte Carlo (MC) study presented herein is the first to employ this improved version of the exclusive B-tagging algorithm FEI up to the point of a signal selection ready for the determination of the signal yield via a fit. The study is based on scaled samples of simulated Belle data corresponding to the full Belle data set of about 771.6×10^6 B meson pair events. Thus, the obtained results can be compared to the ones of the study conducted by *Andreas Heller et al.* [12] for which the same data set was used. A conversion tool called B2BII (short for Belle to Belle II) available in BASF2 allows one to process the Belle data format with the new software. Hence, if the new FEI B-tagging algorithm is proven to be more effective, it might lead to a first observation of the decay based on the already available Belle data.

In the scope of this analysis a signal selection for a branching fraction measurement of the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$ based on Belle MC was performed. The new Belle II software was used to conduct the MC study. Firstly, this allowed for the utilization of the new analysis tools such as the FEI B-tagging algorithm for the analysis, which should be of

---

[1] See also Reference [17] for details on the exclusive B-tagging implementation employed by *Andreas Heller et al.* for the study of $B^+ \to \ell^+ \nu_\ell \gamma$ [12].

benefit for the signal selection efficiency. Secondly, the study poses as validation of the new software and in particular the improvements to the FEI B-tagging algorithm.

A short introduction to the studied decay is provided in Chapter 2. In Chapter 3 an overview of the experiments Belle and Belle II is given, with an emphasis on the improvements introduced by the upgrade. The Belle II analysis software framework used for this analysis is described in Chapter 4. In Chapter 5, the procedure of the analysis of $B^+ \to \ell^+ \nu_\ell \gamma$ based on simulated Belle data corresponding to the full Belle data set is presented. This Chapter concludes with two approaches for the final signal selection and their results in Section 5.7. The latter are evaluated in Chapter 6. Lastly, Chapter 7 summarizes the results of the MC study and gives an outlook for further improvements and next steps.

2. The Decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$

The radiative leptonic decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ offers important insights for the theoretical predictions of a particular subset of B meson decays: non-leptonic B meson-decays into light meson pairs. The emission of the photon probes the first inverse moment λ_B of the light-cone distribution amplitude (LCDA) of the B meson. This parameter is a vital input to QCD factorization schemes for the non-perturbative calculation of aforementioned non-leptonic B meson-decays [11][10]. The leading-order LCDA can be understood as the probability for finding a light quark with a certain momentum fraction inside the B meson, and were first introduced by Reference [18]. It is process independent and plays a similar role for hard exclusive B meson processes as the parton distribution functions play for the inclusive processes of the proton [19, p. 6]. Its first inverse moment λ_B is inversely proportional to the branching fraction of the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ under the assumption that the mass of the charged final state lepton can be neglected compared to the mass of the b quark [13]. Thus, for this analysis the light leptons e and μ are considered as final state particles for the decay — neglecting the heavier and experimentally more difficult τ lepton, which contributes less to the branching fraction of the decay.

Prior studies of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ were able to set upper limits on the branching fraction [20][21], the most stringent one being the Belle analysis this work is based on [12]. In the rest of this chapter a summary of the theoretical background of the parameter λ_B and the role of the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ for its measurement based on reference [13] is provided. Furthermore, the experimental challenges of the analysis are presented in a short introduction of the process $B^+ \rightarrow \ell^+ \nu_\ell \gamma$.

2.1. The Theoretical Relevance

As the precision of the experimental measurements of B decays will improve due to the upgrade to Belle II, an improvement of the SM predictions for B meson interactions would be beneficial for further tests of the model and potential new physics. With QCD the SM provides the theoretical foundation for the calculation of long-distance strong interactions as they occur in the B meson. However, the perturbative ansatz of these calculations fails at low energies due to the strong coupling constant α_s, which is of order one and thus does not allow for a perturbative treatment. This issue is the origin for the biggest uncertainties in the theoretical calculations of B meson decays [10]. These uncertainties already cause theoretical predictions to show higher uncertainties than experimental measurements.

Exclusive non-leptonic weak decays of B mesons into two mesons are one example for decays affected by QCD uncertainties. Such decays can be calculated by applying a factorization scheme which systematically separates the non-perturbative long-distance effects from the short-distance effects [11]. The latter can be treated with a perturbative ansatz. The non-perturbative parts of the factorization can be reduced to parameters which must be obtained empirically.

Reference [11] provides an elaborate derivation of factorization schemes for the cases of the decay into a pair of energetic light mesons with masses M_1 and M_2 and for the decay into one light and one heavy meson ($M_1 \gg M_2$). As a result the factorization of the interaction via the effective weak Hamiltonian H yields for the former case the formula

$$\begin{aligned}\langle M_1 M_2 | H_i | \overline{B} \rangle = \sum_j F_j^{B \to M_1}(m_2^2) \int_0^1 du\, T_{ij}^{\mathrm{I}}(u)\, \Phi_{M_2}(u) \quad + \quad (M_1 \leftrightarrow M_2) \\ + \int_0^1 d\xi\, du\, dv\, T_i^{\mathrm{II}}(\xi, u, v)\, \Phi_B(\xi)\, \Phi_{M_1}(u)\, \Phi_{M_2}(v).\end{aligned} \tag{2.1}$$

For the latter case with one heavy meson M_1 the solution

$$\langle M_1 M_2 | H_i | \overline{B} \rangle = \sum_j F_j^{B \to M_1}(m_2^2) \int_0^1 du\, T_{ij}^{\mathrm{I}}(u)\, \Phi_{M_2}(u) \tag{2.2}$$

is provided. Here the condition of a light meson is only fulfilled for mesons composed of the quark flavors u, d and s and is imposed due to the necessity of at least one high momentum meson. The resulting hadronic matrix elements for the weak interaction can be calculated for different orders i of the power series in α_s of the perturbative part which is reduced to the so-called short-distance scattering kernels T_i^{I} and T_i^{II}. The non-perturbative parts of the calculations are reduced to the $B \to M$ transition form factors $F^{B \to M}$ and the light-cone distribution amplitudes $\Phi_X(u)$ of the quarks within the meson X which depend on the momentum fractions u, v or ξ of the quarks.

Of particular significance is the B meson LCDA Φ_B which represents the interactions of the spectator quark in the process. It is its first inverse momentum which is denoted as λ_B:

$$\lambda_B^{-1} := \int_0^\infty \frac{d\xi}{\xi}\, \Phi_B(\xi), \tag{2.3}$$

where ξ stands for the light-cone projection of the energy of the spectator quark in the B meson. This parameter of the QCD factorization schemes is crucial for the accuracy of theoretical predictions of color-suppressed decay modes. Experimental and theoretical results for its value differ. QCD-sum-rule based estimates predict values of λ_B between 350 and 500 MeV [22], whereas predictions from comparisons with $B \to \pi\pi,\ \pi\rho,\ \rho\rho$ decay data yields lower values of about $\lambda_B \sim 200$ MeV [10]. The latest attempt for an experimental measurement of the parameter via the decay $B^+ \to \ell^+ \nu_\ell \gamma$ by the Belle collaboration was able to set a lower limit for the size of λ_B with $\lambda_B > 238$ MeV at 90% credibility level [23].

The Role of the Decay $B^+ \to \ell^+ \nu_\ell \gamma$ as an access point for an empirical determination of the parameter λ_B becomes apparent in the theoretical calculation of the decay. *Beneke* and *Rohrwild* present in Reference [13] a detailed theoretical summary of the process $B^+ \to \ell^+ \nu_\ell \gamma$. They are the first to include radiative and power corrections to the calculation of the decay's branching fraction up to next-to-leading order.

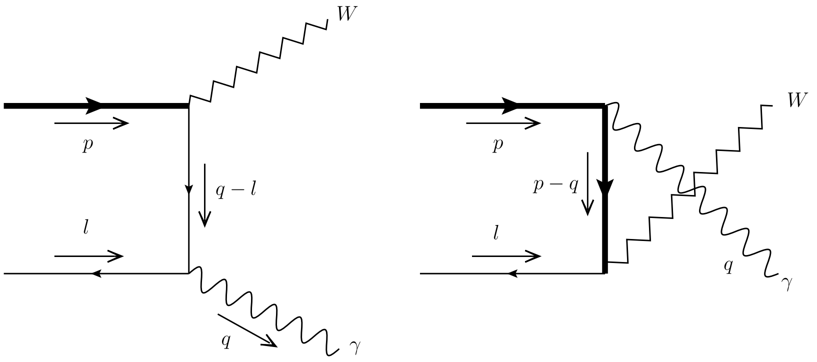


Figure 2.1.: Feynman diagrams for the process $b\bar{u} \to \gamma W^*$, where the photon can either be emitted from the light spectator quark (left) or the heavy b quark (right) [13].

Compared to the non-radiative decay $B^+ \to \ell^+ \nu_\ell$, the inclusion of the photon emission adds the fine-structure constant α_{em} as a further factor to the branching ratio. Despite this branching fraction reducing effect, the addition of the photon lifts the helicity suppression which is present in the non-radiative decay, eventually resulting in a higher branching fraction. The helicity suppression is caused by the parity violating weak vector–axial current which couples only to left-handed particles or right-handed anti-particles and reduces the branching fraction of the process for smaller lepton masses. A photon emission from one of the quarks of the B meson excites the meson into a virtual B_1^+ or B^{*+} state with spin one, thereby allowing the decay into a right-handed charged lepton and the respective right-handed anti-neutrino and vice versa. Thus the helicity suppression present in the non-radiative process is removed and so is the associated reduction of the branching ratio.

To achieve the excited spin one state, the photon has to be emitted from one of the quarks of the B meson. Furthermore a photon emission from the heavy b quark is suppressed as the propagator from the electromagnetic vertex of the photon emission to the vertex of the weak interaction in the tree-level process $b\bar{u} \to \gamma W^*$ (see Figure 2.1) has hard virtuality of the order of $1/(m_b)^2$ compared to the hard-collinear virtuality of order $1/(m_b \Lambda_{\mathrm{QCD}})$ in the case of the radiation from the light u quark[1] (see [13]). Hence, to leading order, the photon is emitted from the spectator u quark of the B meson. Reference [13] takes the higher order contributions from the emissions from the lepton or the b quark as power corrections into account and provides a thorough derivation of the process of interest, which yields the differential decay width

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{\mathrm{em}}\, G_{\mathrm{F}}^2\, |V_{ub}|^2}{48\,\pi^2}\, m_B^4\, (1 - x_\gamma)\, x_\gamma^3 \left[F_A^2 + F_V^2\right], \tag{2.4}$$

where

$$F_V(E_\gamma) = \frac{Q_u\, m_B\, f_B}{2E_\gamma\, \lambda_B(\mu)}\, R(E_\gamma, \mu) + \left[\xi(E_\gamma) + \frac{Q_u\, m_b\, f_B}{(2E_\gamma)^2} + \frac{Q_b\, m_b\, f_B}{2E_\gamma\, m_b}\right] \tag{2.5}$$

---

[1] Λ_{QCD} is the scale parameter of quantum chromodynamics which has a size of about 220 MeV.

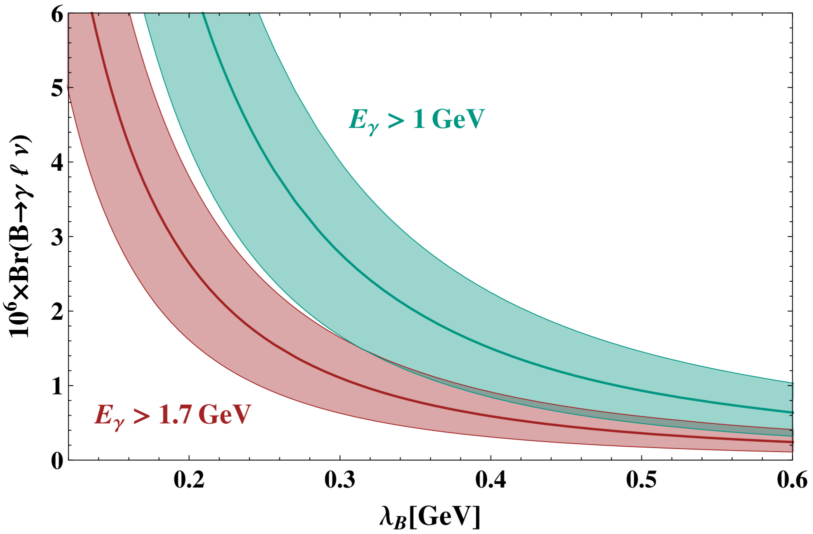


Figure 2.2.: Dependence of the theoretical prediction for the partial branching fraction $\mathcal{B}(B^+ \to \ell^+ \nu_\ell \gamma)$ on the value of the QCD factorization parameter λ_B for two signal photon selection criteria: The theoretical safer cut $E_\gamma > 1.7$ GeV and the cut $E_\gamma > 1.0$ GeV which yields a significantly higher efficiency [13].

and

$$F_A(E_\gamma) = \frac{Q_u\, m_B\, f_B}{2E_\gamma\, \lambda_B(\mu)}\, R(E_\gamma, \mu) + \left[\xi(E_\gamma) - \frac{Q_u\, m_b\, f_B}{(2E_\gamma)^2} - \frac{Q_b\, m_b\, f_B}{2E_\gamma\, m_b} + \frac{Q_\ell\, f_B}{E_\gamma}\right] \tag{2.6}$$

are the vector and axial form factors, respectively. Besides the leading-power contribution and the higher-order radiative correction factor $R(E_\gamma, \mu)$ these form factors contain (in the square brackets): the relative $1/m_b$ power corrections originating from local (Q_u-term) and non-local unknown ($\xi(E_\gamma)$) contributions from the photon emission from the light quark; contributions from the suppressed emission from the b quark (Q_b-term); and the emission from one of the leptons (Q_ℓ-term) which only contributes to the axial form factor. Here m_B and m_b are the masses of the B meson and b quark, respectively, E_γ is the energy of the emitted photon and $x_\gamma = 2E_\gamma/m_B$ its energy relative to the B meson mass. Moreover the differential decay rate depends on the aforementioned fine-structure constant α_{em}, the B meson decay constant f_B, the Fermi constant G_F and the CKM matrix element V_{ub}. $Q_{u,b,\ell}$ denote the electric charge of the up quark, bottom quark and lepton, respectively. Finally, $\lambda_B(\mu)$ is the first inverse momentum of the LCDA, which depends on the hard-collinear energy scale μ.

The partial branching ratio $\Delta\mathcal{B}$ is given by the integral of the differential decay width over the photon energies relevant for the considered selection, multiplied by the lifetime of the B meson τ_B:

$$\Delta\mathcal{B}(B^+ \to \ell^+ \nu_\ell \gamma) = \tau_B \int_{\text{Selection}} dE_\gamma\, \frac{d\Gamma}{dE_\gamma}, \tag{2.7}$$

where photon energies below 1 GeV are considered as unsafe since the factorization approach for the calculation of the form factors F_V and F_A requires the condition $2E_\gamma \sim m_b$.

Hence, only photons with energies above this threshold are considered for this analysis. This allows one to determine the QCD factorization scheme parameter λ_B via the measurement of the partial branching fraction of the decay. Figure 2.2 shows the theoretical relation between the value of λ_B and the respective partial branching ratio for two selection criteria for the signal photon energy.

Further constraints on the energy of the neutrino would enable the experimental examination of the difference between the axial and vector form factor and thereby the impact of the power-suppressed contributions to the decay width (see [13, eq. 3.3]). However, the selection required for this study — the neutrino has to receive the majority of the B meson's energy — reduces the statistics significantly, rendering it not feasible at the moment. Future analyses based on the Belle II data set might want to address this aspect of the decay as well. For now the aim is to accomplish the branching fraction measurement, which might be possible with the aid of the new analysis tools as they could enhance the signal selection efficiency sufficiently. The experimental problems that have to be faced for this task are elaborated on in the following section.

2.2. The Experimental Challenges

Assuming the branching fraction for the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ to be in the range depicted in Figure 2.2 the expected number of events for this process in the full Belle data set is of the order of 10^3. The accurate calculation based on the upper limit of $\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 3.5 \times 10^{-6}$ established by [12][23] predicts a yield of less than

$$\begin{aligned} N_{B^+ \rightarrow \ell^+ \nu_\ell \gamma} &= N_{B\overline{B}} \times 2 \times \mathcal{B}(\Upsilon(4S) \rightarrow B^+ B^-) \times \mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) \\ &= 771.6 \times 10^6 \times 2 \times 0.514 \times 3.5 \times 10^{-6} = 2.8 \times 10^3 \end{aligned} \tag{2.8}$$

where $N_{B\overline{B}}$ is the number of B meson pair events in the full Belle data sample taken on the $\Upsilon(4S)$ resonance as reported in reference [24]. The factors $\mathcal{B}(\Upsilon(4S) \rightarrow B^+ B^-)$ and 2 take the probability for a decay of the resonance into a pair of charged mesons [25] and the fact that both of them can decay via the studied channel into account. One has to keep in mind, though, that the resulting number gets drastically reduced by the efficiency of the signal selection which is not yet considered at this point.

The Tagging of the Companion B Meson

Due to the signal neutrino's practically non-existing interaction with other matter, this product of the signal decay cannot be detected. To retrieve the information about the 3-body signal decay carried by the neutrino, the advantage of the well-known initial state of the $\Upsilon(4S)$ meson produced in e^+e^--collisions is exploited. As illustrated in Figure 2.3, $\Upsilon(4S)$ mesons decay in about 50 % of the cases into two entangled charged B mesons, one of which must decay via the channel $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ for the event to be of interest for this analysis. The second B meson carries information necessary to interpret the full event and thus to infer the properties of the neutrino of the signal B decay via the initial state of the system. Hence, an evaluation of the B_{tag} meson accompanying the signal B_{sig} meson is

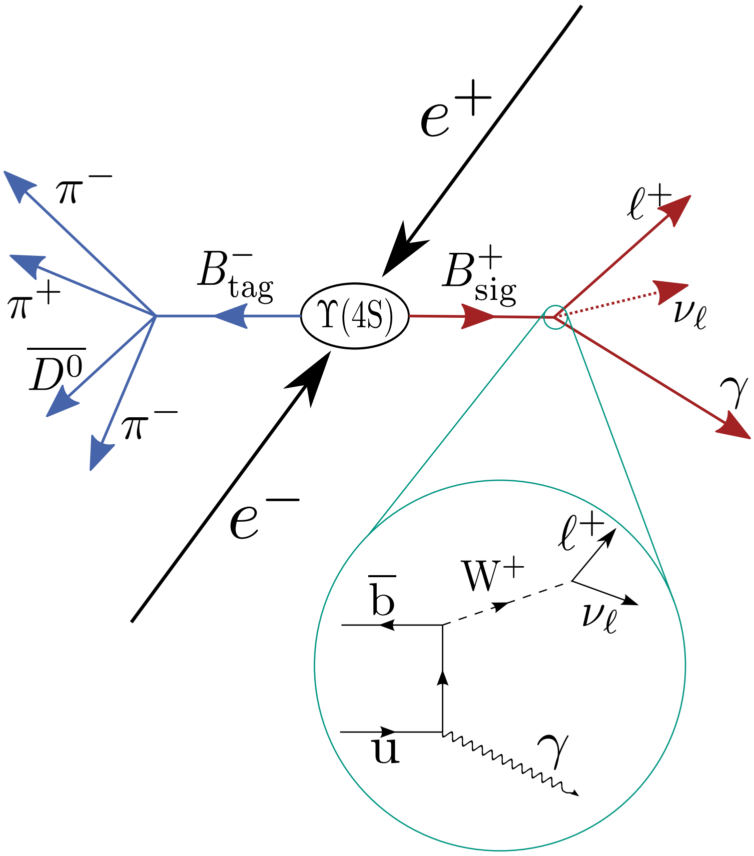


Figure 2.3.: Schematic illustration of the $\Upsilon(4S)$ decay into the signal side B_{sig} — which decays further to the charged lepton, the photon and the corresponding, undetectable neutrino — and the companion B_{tag} meson. The latter decays generically with a hadronic decay into a neutral D meson and three charged pions shown in this case. Also included is a sketch of the leading-order process Feynman diagram of the signal side decay.

conducted — a practice which is referred to as B-tagging. One distinguishes between two common approaches to this task: Inclusive B-tagging and exclusive B-tagging.

Inclusive Tagging is performed on all remaining particles of the event which are unaccounted for after a candidate for the signal-side B_{sig} meson is recombined. The combined 4-momentum of these remaining particles is assigned to the tag-side B_{tag} meson. Under consideration of the initial state of the $\Upsilon(4S)$ system and the recombined signal candidate this allows an estimation of the 4-momentum of the signal-side neutrino.

Exclusive Tagging is a more sophisticated strategy for which the explicit recombination of the companion B_{tag} meson in either fully hadronic or semileptonic decay channels is attempted. In this approach the intermediate resonances of the tag-side decay process are resolved, thereby retrieving additional information about the event which results in a better resolution of the 4-momentum determination of the tag-side and thus the 4-momentum of the neutrino. One must distinguish the two cases, as semileptonic B-tagging is affected by additional neutrinos on the tag-side, thereby degrading the 4-momentum resolution. Because of the additional neutrino, semileptonic B-tagging also allows only for loose selection criteria on the tag-side B_{tag} meson. However, the semileptonic decay channels have a higher branching fraction.

The exclusive tagging technique requires the explicit analysis of a certain set of decay channels for the tag-side process and, thus, has a reduced efficiency compared to the inclusive tagging. Despite the lower efficiency of the exclusive tagging it offers further benefits. The exclusive interpretation of the event up to the point of the recombination of the $\Upsilon(4S)$ allows for additional tests of the signal-side. If the tag- and signal-side are correctly identified, no further particles unassociated to these two B mesons should be present. If no valid tag-side candidate is found which includes all the remaining particles, it is more likely that the signal-side is not correctly recombined. These additional information allow one to discriminate background events to a much better extent than with inclusive tagging, resulting in a higher signal purity.

Both tagging techniques affect the signal yield severely, with an efficiency of the order of 1 % for the inclusive approach and less than 1 % for exclusive tagging. In case of the decay $B^+ \rightarrow \ell^+\nu_\ell\gamma$, analyzed with exclusive B-tagging, this reduces the number of expected events to the order of 10. The performance of the applied tagging technique is a crucial part of the experimental evaluation of the decay. A more thorough illustration of the exclusive tagging approach and especially an algorithm dedicated to this task can be found in Section 4.3.

The Peaking Background

A further obstacle for the analysis is the signal-like background represented by the channels $B^+ \rightarrow X_u\ell^+\nu_\ell$, where X_u represents a light neutral meson containing an u quark. This light meson can decay into two photons. If one of those is not correctly taken into account, the process can be misinterpreted as the signal decay of interest. These signal mimicking processes show properties very similar to the decay $B^+ \rightarrow \ell^+\nu_\ell\gamma$ and cannot be fully suppressed. If the second photon is not detected (e.g. due to the restricted acceptance of the detector), or if the two photons cannot be separated by the resolution of the detector, such events pose an irreducible background to the signal decay.

Of particular significance are the processes with $X_u = \pi^0$ and $X_u = \eta$ which dominate the signal region, owing to their high branching ratios compared to the signal decay ($\mathcal{B}(B^+ \rightarrow \pi^0\ell^+\nu_\ell) \approx 7.8 \times 10^{-5}$ and $\mathcal{B}(B^+ \rightarrow \eta\ell^+\nu_\ell) \approx 3.8 \times 10^{-5}$, [25]). Furthermore, the low masses of these particles make it harder to separate the two photons and resolve the light mesons. For this study a final classifier dedicated to distinguishing these peaking backgrounds from the signal decay has been developed to reduce their impact on the potential branching fraction measurement (see also Section 5.6).

3. The Experiments Belle and Belle II

The analysis of $B^+ \to \ell^+ \nu_\ell \gamma$ presented in this thesis is performed on Belle MC using the Belle II software framework. Hence the experimental setups of both projects are introduced. Both experiments use the tunnel of the former TRISTAN electron–proton collider which is situated at the Japanese High Energy Accelerator Research Organization KEK in Tsukuba. The corresponding electron–positron colliders dedicated for B meson physics are KEKB and its successor SuperKEKB. The former delivered an integrated luminosity of about 1040 fb^{-1} during its run time from 1999 until 2010.

Directly after the shut down of KEKB the work on the upgrade to SuperKEKB started. First positron and electron beams were stored in the new machine in February 2016 [26]. The storage ring features one interaction region (IR) where the electron–positron-collisions take place. The detector is located at the IR enclosing the interaction point (IP) to a solid angle close to 4π.

Based on the more detailed References [27] and [7] for the Belle experiment and [28] for the Belle II experiment the two detectors and colliders are presented in the following Sections 3.1 and 3.2, respectively.

3.1. The Belle Experiment

The Belle detector observed and recorded about 770 million $B\overline{B}$ pair events. This was made possible by the record instantaneous luminosity of the two-ring, asymmetric-energy electron–positron collider KEKB, which reached a maximum of $2.1 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ in 2009 [29]. As an electron–positron collider, KEKB offered clean events with a well-known center of mass energy, since the colliding particles do not have a substructure. With the chosen center of mass energy of $E_{\mathrm{CMS}} = \sqrt{s} = 10.58\,\mathrm{GeV}$ the KEKB collider operated at the $\Upsilon(4S)$ resonance and therefore just above the threshold for the $B\overline{B}$ meson pair production of interest. Hence the entangled B mesons — to which the $\Upsilon(4S)$ decays with a branching fraction of $\mathcal{B}(\Upsilon(4S) \to B\overline{B}) > 96\%$ — are produced without additional particles. For the life time measurement of the B mesons a known Lorentz boost is introduced to the collision products, thus allowing for a time measurement via the offset of the B meson decay vertices. The boost is realized by using different energies for the electron and positron beams (see Table 3.1). As depicted in Figure 3.1, KEKB features two separate storage rings: the high energy ring HER for the electrons and the low energy ring LER for the positrons with their respective energies $E_{\mathrm{HER}} = 8.0\,\mathrm{GeV}$ and $E_{\mathrm{LER}} = 3.5\,\mathrm{GeV}$. Due to this a Lorentz boost of

$$\beta\gamma = \frac{E_{\mathrm{HER}} - E_{\mathrm{LER}}}{E_{\mathrm{CMS}}} = 0.425 \tag{3.1}$$

in the direction of the HER is achieved. The two beams cross at an angle of 22 mrad to allow for a simple separation of electrons and positrons once they pass the IP. Additionally,

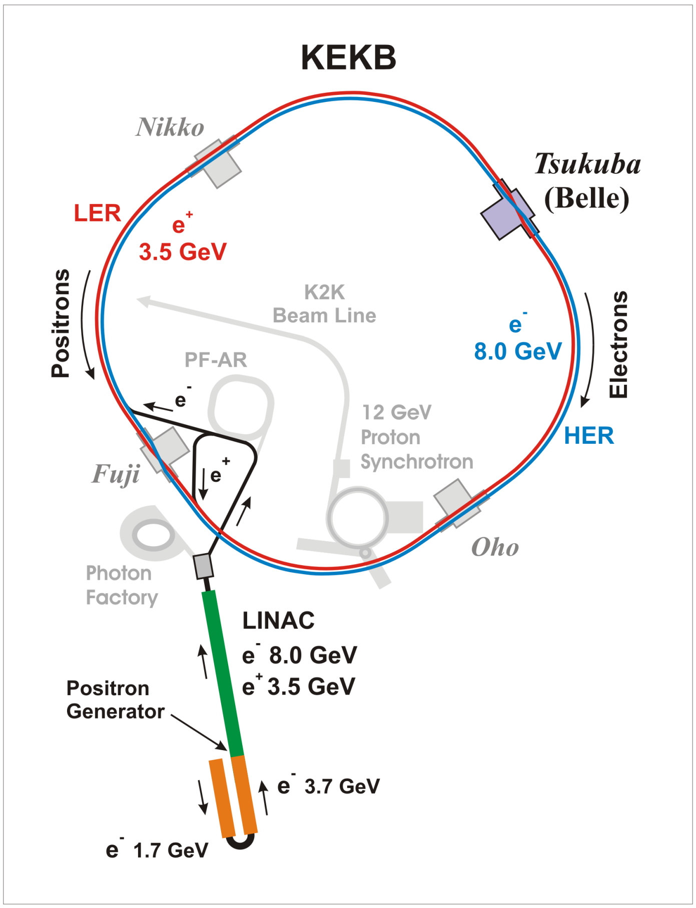


Figure 3.1.: Schematic view of the KEKB collider with its two rings: LER in red and HER in blue. The positrons for the LER are produced at the positron generator by shooting electrons on a tantalum target. Both electrons and positrons are accelerated to their respective energies in the linear accelerator (LINAC) shown in green. The location of the Belle detector is shown at the Tsukuba interaction region [30].

the area of interaction between to two beams is increased due to the angle and thereby the luminosity. So-called crab cavities tilt the beam bunches at the IP to ensure a frontal collision of the bunches despite the crossing-angle [7, p. 6, p. 45]. This feature of the KEKB collider — installed in January 2007 — made the high instantaneous luminosity possible.

The asymmetry of the particle beam energies is reflected in the asymmetry of the detector which can be seen in Figure 3.2. To harvest a maximal amount of information about the physical processes taking place at the IP, the Belle detector was designed to identify and measure energy and momentum of the resulting decay products. Thus the detector consists of several specialized sub-detectors which enclose the IP in 360° of the azimuth angle ϕ and from 17° to 150° of the polar angle θ.

The Silicon Vertex Detector (SVD) was the innermost sub-detector used for vertex detection and installed directly on the beryllium beam pipe at the IP. It consisted of four layers of double-sided silicon-strip detectors with an angular acceptance of $17° < \theta < 150°$ and layer radii of 20 mm, 44 mm, 70 mm and 88 mm. The SVD was used for the measurement of the vertices of decaying particles and delivered additional information for particle tracking, which was particular valuable for low momentum particles.

The Central Drift Chamber (CDC) was Belle's main track reconstruction detector and with an inner radius of 80 mm directly connected to the SVD. Only a thin wall separated the two

]

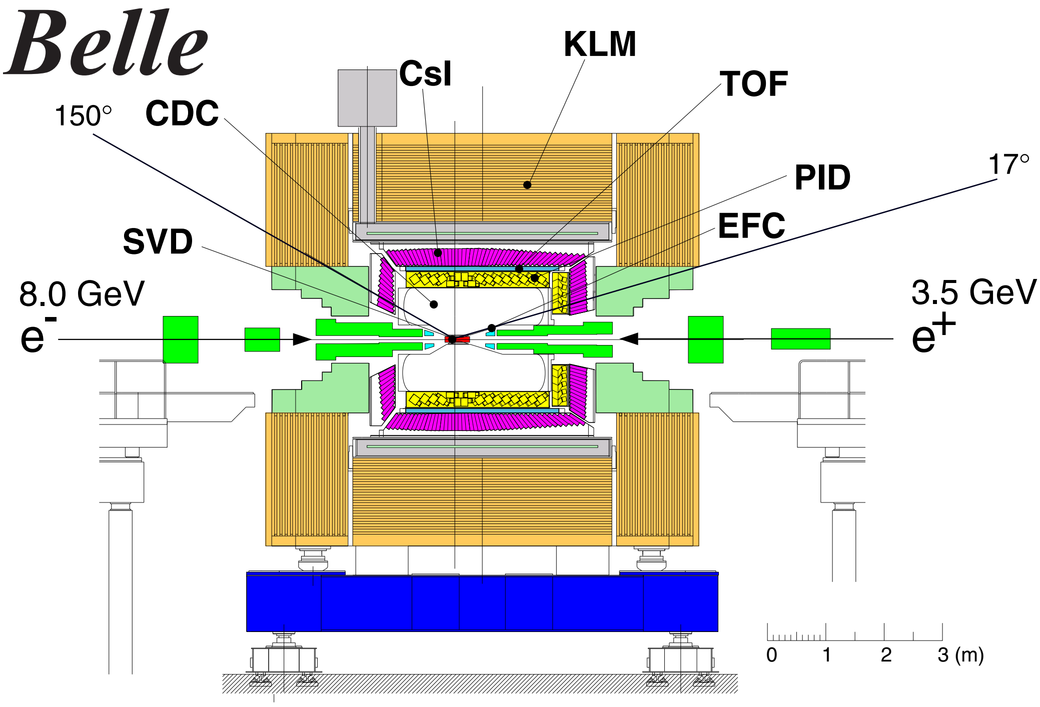


Figure 3.2.: Cross-section of the original Belle detector. Here, the ECL is marked as CsI, as it consists of CsI(Tl) crystals, and the ACC (yellow) is labeled as PID [27].

sub-detectors to minimize multiple scattering. The volume enclosed by the outer wall with a radius of 880 mm covered a polar angle from 17° to 150° and contained a mixture of 50 % ethane gas and 50 % helium. Charged particles passing through the volume ionized the helium atoms, thus produce free electrons which are accelerated in the artificial electric field maintained between the sense and field wires installed in the chamber and induced a sharp peak collected by the read-out electronics. The wires were arranged parallel to the beam line in 8 400 almost square drift cells with the gold-plated tungsten sense wire in the middle surrounded by the aluminum field wires. Each cell had a maximum drift distance between 8 and 10 mm. Combined they formed 50 cylindrical layers which were either axial layers with respect to the z-axis or slightly tilted stereo layers. The latter are necessary to gain information about the z-position of the particle tracks. This setup allowed for the reconstruction of the tracks of charged particles and hence the determination of their momenta based on the curvature of the track in the homogenous 1.5 Tesla magnetic field of the superconducting NbTi/Cu solenoid surrounding the inner sub-detectors. Additionally, the energy loss dE/dx of the particles could be measured and used as input for the particle identification.

The Aerogel Cherenkov Counter (ACC) was one of the sub-detectors dedicated to particle identification (PID) and located in the barrel and at the forward endcap, adjoining the CDC, as can be seen in Figure 3.2 where it is labeled “PID”. Particles with a velocity higher than the speed of light in the aerogel blocks caused Cherenkov radiation which was detected with fine mesh-type photo multiplier tubes. In combination with the momentum measurement based on the tracking detectors this information serves as discriminator for different mass hypotheses, an thus as input for the particle identification likelihoods. This method is only applicable for particles with a velocity above the threshold specified by the refractive indices of the dielectric aerogel. Slower particles do not cause Cherenkov radiation, which is also an important information for the likelihood calculation.

The Time-Of-Flight detector (TOF) offered PID information for slower particles for which a time measurement with a resolution of 100 ps for the travel time between IP and the sub-detector was possible. For faster particles this resolution is not sufficient. Again this measurement is used to discriminate between different mass hypotheses for the detected particles via their velocity and the momentum information. The TOF was installed in the barrel region of the Belle detector and thus covered only an acceptance of 33° to 121° in θ.

The Electromagnetic Calorimeter (ECL) — composed of 8 736 Thallium-doped CsI scintillator crystals with silicon photo diodes attached to their ends — was the outermost of the sub-detectors placed inside of the solenoid. This is sensible since the ECL is designated to detect electro-magnetic showers caused by either electrons or photons due to pair production or Bremsstrahlung which should occur ideally only in the optical material of the ECL. As the momenta of these particles vary strongly depending on their generation process, the ECL crystals have to provide a sufficient number of interaction lengths for high momentum particles as well as a good energy resolution for low momenta. Each crystal points almost directly to the IP, however a small divergence is maintained to avoid gaps between the optical elements, guaranteeing a hermetic coverage. Photons are detected as showers in the ECL without an associated particle track. The angular resolution obtained for photons depends on their energy and improves from about 13 mrad to 3 mrad for increasing energies [7, p. 31]. This angular resolution allows for a separation of photon pairs originating from the decay of uncharged pions with a π^0 mass resolution of about 4.5 MeV/c^2. The energy of photons is determined based on the amount of secondary photons recorded by the ECL's photo diodes. The resolution of this energy measurement varies between 4 % at 100 MeV and about 1.6 % at 8 GeV [7, p. 31]. Furthermore, the ECL offers the possibility to identify electrons or positrons in combination with their tracks, distinguishing them from charged hadrons or muons.

The Extreme Forward Calorimeter (EFC) was an addition to the ECL which was located close to the beam pipe to extend the coverage of the ECL with an acceptance in the polar angle of 6.2° to 11.6° in forward, and 163.1° to 171.5° in backward direction. Both parts were composed of 160 radiation-hard bismuth-germanate crystals and attached to cryostats of the compensation solenoid magnets of the KEKB collider ring as indicated in Figure 3.2. The sub-detector also acted as beam and luminosity monitor as well as as shield from beam backgrounds for the tracking detectors in addition to its role as calorimeter.

The K_L^0 and Muon Detection System (KLM) was installed in the iron return yoke of the superconducting solenoid. 14 layers of 4.7 cm thick iron plates equipped with glass-electrode resistive plate counters provided 3.9 interaction lengths of material for hadron interactions. The active elements of the KLM detect hadronic showers and non-showering charged particles such as muons which can pass through the whole detector. Hadronic showers are clusters of secondary particles originating from strong interactions of charged and neutral hadrons in the iron. Neutral particles — in particular K_L^0 — show no associated tracks. The KLM thus allowed for the identification of such particles and provided a direction measurement assuming the IP is their origin. However, it was not able to determine their energy due to large statistical fluctuations caused by the low number of secondary particles produced in the showers. Furthermore, a distinction between charged pions and non-showering muons, which have a similar mass, was possible as the weakly interacting muons penetrated on average more layers of the KLM. Hence muons can be identified via a combination of the momentum measurement of the tracking detectors, the velocity measurement provided by TOF and ACC and an observation of a non-showering particle in the outer layers of the KLM.

3.2. The Belle II Experiment

The objective of the Belle II experiment is the accumulation of a data set of $B\overline{B}$ pair events exceeding the Belle data set by a factor of 50 [28][29]. To achieve this goal the e^+e^--collider at KEK has been extensively upgraded. The modified machine is called SuperKEKB and has a design instantaneous luminosity of about $8.0 \times 10^{35} \mathrm{cm}^{-2}\mathrm{s}^{-1}$ [28, sec. 2]. To attain this increase in luminosity by a factor of 40, the nano-beam scheme, which is further elaborated on in [28, sec. 2], is utilized. In principle the colliding beams are focused by additional final focusing quadrupoles near the IP, thereby obtaining a beam size of $\sigma_y^* = 0.08\ \mu\mathrm{m}$. Compared to KEKB this results in a reduction of the vertical beta function at the IP $\beta_{y\pm}^*$ for both beams by a factor of almost 20. An overview of the important machine parameters of SuperKEKB is presented in Table 3.1 where the numbers are compared to the ones of KEKB. The effect of this reduction on the instantaneous luminosity can be estimated via equation 3.2 where the assumption of flat beams and equal horizontal and vertical beam sizes for both, the e^+ and the e^- beam at the IP is made [28, sec. 2].

$$L = \frac{\gamma_\pm}{2er_e}\left(\frac{I_\pm \xi_{y\pm}}{\beta_{y\pm}^*}\right)\left(\frac{R_L}{R_{\xi_y}}\right) \tag{3.2}$$

As the luminosity is indirectly proportional to the vertical beta functions it is enhanced accordingly. In addition, the luminosity depends on the relativistic Lorentz factors $\gamma_\pm$ of the beams, the electron charge e and its classical radius r_e; the quotients of the reduction factors for the luminosity R_L and the vertical beam-beam parameter R_{ξ_y}; the vertical beam-beam parameters $\xi_{y\pm}$ themselves and the beam total currents $I_\pm$, where ± either denotes the electron beam (−, HER) or the positron beam (+, LER). Moreover, an increase of the beam currents by a factor of about two results in an overall 40 times higher instantaneous luminosity compared to KEKB.

The hardware modifications necessary to obtain these design goals included a redesign of the IP and in particular a larger crossing angle of 83 mrad to accommodate the final focusing quadrupoles. To produce the low emittance electrons and positrons necessary for the nano-beam scheme, the injector complex was upgraded with a new radio frequency gun for the production of low emittance electrons and a new tungsten positron target. Since the positrons are produced as high emittance secondary particles from e^- collisions

Table 3.1.: Comparison of achieved KEKB machine parameters and design values of SuperKEKB [28, sec. 2].

Machine Parameter	KEKB		SuperKEKB	
	HER	LER	HER	LER
Beam current (A)	1.64	1.19	3.60	2.61
Energy (GeV)	8.0	3.5	7.0	4.0
β_y^* (mm)	5.9	5.9	0.27	0.41
Crossing angle (mrad)	22		83	
Beam emittance (µm)	300	2100	20	10
Beam lifetime (min)	200	150	10	10
Luminosity (10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	2.11		80	

with this target, the addition of a positron damping ring — which is displayed in Figure 3.3 — was necessary for the dramatic reduction of their emittance compared to KEKB (see Table 3.1).

As a side effect of the small vertical beta function at the IP and the higher beam current the rate of intra-bunch scattering — the so called Touschek effect [31] — increases, resulting in a lower beam lifetime. Furthermore, the amount of beam-induced background is increased significantly due to the Touschek effect and other sources, such as beam-gas scattering and synchrotron radiation. The impact of the Touschek effect is reduced by exploiting its proportionality to E_{Beam}^{-3} and changing the energies of LER and HER to 4.0 and 7.0 GeV, respectively. Thus, the Lorentz boost is reduced to $\beta\gamma = 0.284$, but the same center of mass energy is maintained.

To cope with the expected increase of beam-background and to adapt technical innovations, the Belle II detector also receives major upgrades. Nonetheless, some parts of the old detector are inherited almost as they are: The **superconducting solenoid** and its iron return yoke are preserved. Only the instrumentation of the yoke for the **KLM** receives minor improvements, such as the replacement of the resistive plate counters by plastic scintillator strips in the endcaps and the inner two layers of the barrel to compensate for the higher background rates in these regions [28, sec. 10][32]. Likewise, most of the **ECL** crystals are reused — only the crystals in the endcap regions might be replaced with pure CsI crystals which offer a shorter scintillation time [33]. Again this region is especially affected by the increased beam-background. Furthermore, the electronics of the ECL will be substituted with a faster readout with waveform sampling, thus augmenting the performance of the whole ECL under the new conditions [33].

Figure 3.4 shows a side by side-comparison of the detectors Belle and Belle II in which the major innovations to the detector are highlighted in color. Besides the afore-mentioned new beryllium beam pipe with a radius of 12 mm, the quadrupoles and the considered upgrade of the ECL, the tracking detectors and the PIDs are most affected by the upgrade. These sub-detectors are discussed slightly more thoroughly below.

The VerteX Detectors (VXD), which are the two-layered PiXel Detector (PXD) and the four-layered Silicon Vertex Detector (SVD), are responsible for the measurement of vertices

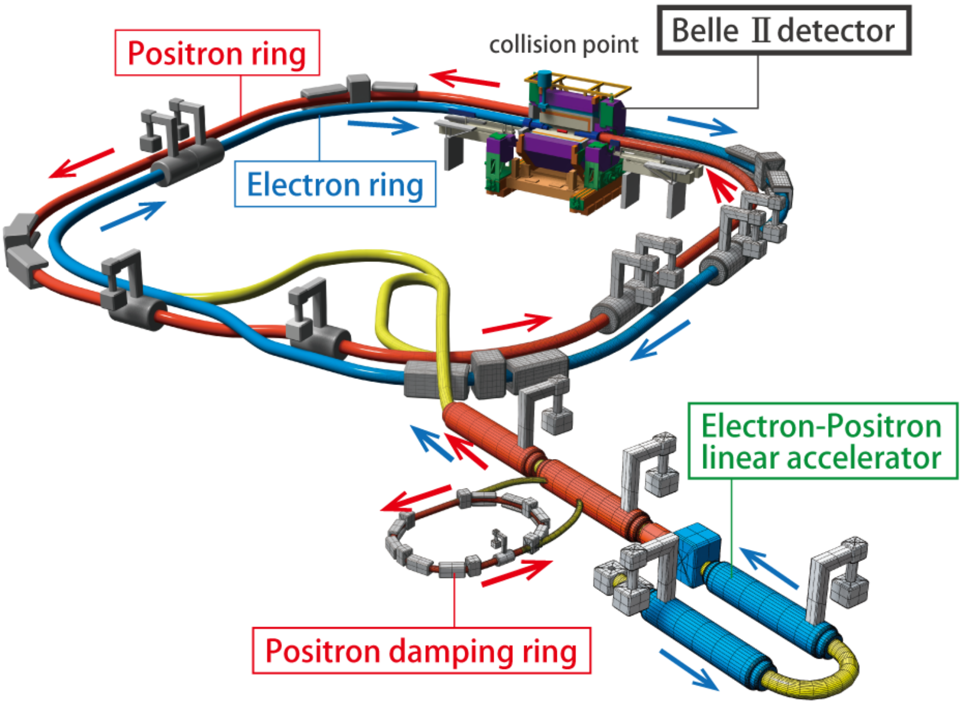


Figure 3.3.: Sketch of the SuperKEKB collider with its high energy electron ring (HER, blue) and low energy positron ring (LER, red), as well as the added positron damping ring and the Belle II detector [26].

and tracking of charged particles with low momenta. They replace the SVD2 and are installed closest to the IP — the innermost PXD sensor layer with a radius of 14 mm sits directly on the beam line. Combined with the second layer with a radius of 22 mm the PXD has over eight million depleted field effect transistor pixels (DEPFET) which guarantee a low occupancy of the detector despite the the higher background rate. Moreover, the DEPFET technology allows for a very thin design of the sensor elements, thereby reducing multiple scattering in these parts of the detector.

The PXD is accompanied by the Silicon Vertex Detector (SVD) surrounding it with radii between 38 and 140 mm. To cover the same acceptance of 17° to 150° in θ as the rest of the PXD, the three outermost layers of the SVD are tilted towards the beam line in the forward region with an angle of about 15 degrees [28, p. 142]. Thus, the amount of material necessary for the double-sided-readout silicon strip detector elements is reduced. The readout-strips on the two sides of the elements are oriented approximately perpendicular to each other in order to provide complete positional information about the hits produced by charged tracks. A faster readout time compared to the PXD makes the elements also less prone to beam-background. The SVD acts as a connection between PXD and CDC to extrapolate reconstructed tracks to their respective vertex. Furthermore, it provides the only tracking information for charged particles which do not reach the CDC due to low momentum and assists the CDC in the other cases.

The Central Drift Chamber (CDC) in Belle II is larger than its predecessor — featuring an inner and outer radius of 160 mm and 1 130 mm, respectively. Similar to the setup of Belle, its 14 336 almost square drift cells are arranged in 56 layers which are grouped into 9 superlayers, alternating between axial and stereo orientation. A denser arrangement of the wires of the innermost superlayer allows for the reduction of the occupancy of this region affected the most by the higher background rate. Additionally, new readout electronics utilizing ASIC chips reduce the deadtime of the sense wires of the whole system. Thus, the CDC should be able to provide accurate information for track finding and the identification of charged particles under the new conditions. Combining these improvements on the hardware side with the considerably enhanced tracking algorithms implemented in BASF2, which employ new techniques such as multivariate classifiers and the integration of local track finding algorithms (see References [34][35]), a significantly higher tracking efficiency can be achieved.

The Particle Identification Detectors (PID) are the subjects of a complete redesign. The voluminous ACC and the TOF are replaced by a Time-Of-Propagation counter (TOP) in the barrel and a proximity-focusing Aerogel Ring-Imaging CHerenkov detector (ARICH) in the forward endcap region (see Figure 3.4). This system offers not only an improved discrimination between kaons and pions, but also requires less space and material, which allows for the larger CDC and a better energy measurement in the ECL.

Composed of 16 synthetic fused, 2 cm thick silica radiator bars with photo-multipliers at their ends, the TOP measures the time of impact as well as the location where the charged particles enter the optical component. A Cherenkov image is reconstructed from these information and the Cherenkov photons which propagate in the quartz bars towards the photo-multiplier in a specific time depending on their Cherenkov angle ϕ. Hence, the TOP offers a time-of-flight measurement as well as a velocity measurement based on the equation for the Cherenkov angle $\cos(\phi) = 1/(n\beta)$, where n is the refractive index of the optical component and β is the particle velocity relative to the speed of light in the vacuum. Both measurements are combined into a likelihood for the different final state particle hypotheses.

The ARICH detects Cherenkov photons produced in its aerogel radiator which consists of two layers with different refractive indices — the one of the inner layer being smaller than the other. Position sensitive photon detectors reproduce the ring projection of the Cherenkov cone, which is focused due to the decreasing refractive indices. Again, the Cherenkov angle can be determined from the image, thereby providing the velocity of the observed particle.

As in the case of the Belle spectrometer, the PIDs allow for the discrimination between different mass hypotheses in combination with the information of the tracking detectors. In the backward-region not covered by the PIDs, this task is required to be fulfilled by the remaining sub-detectors.

Further information about the method of operation of the different sub-detectors of Belle II can be obtained from [28].

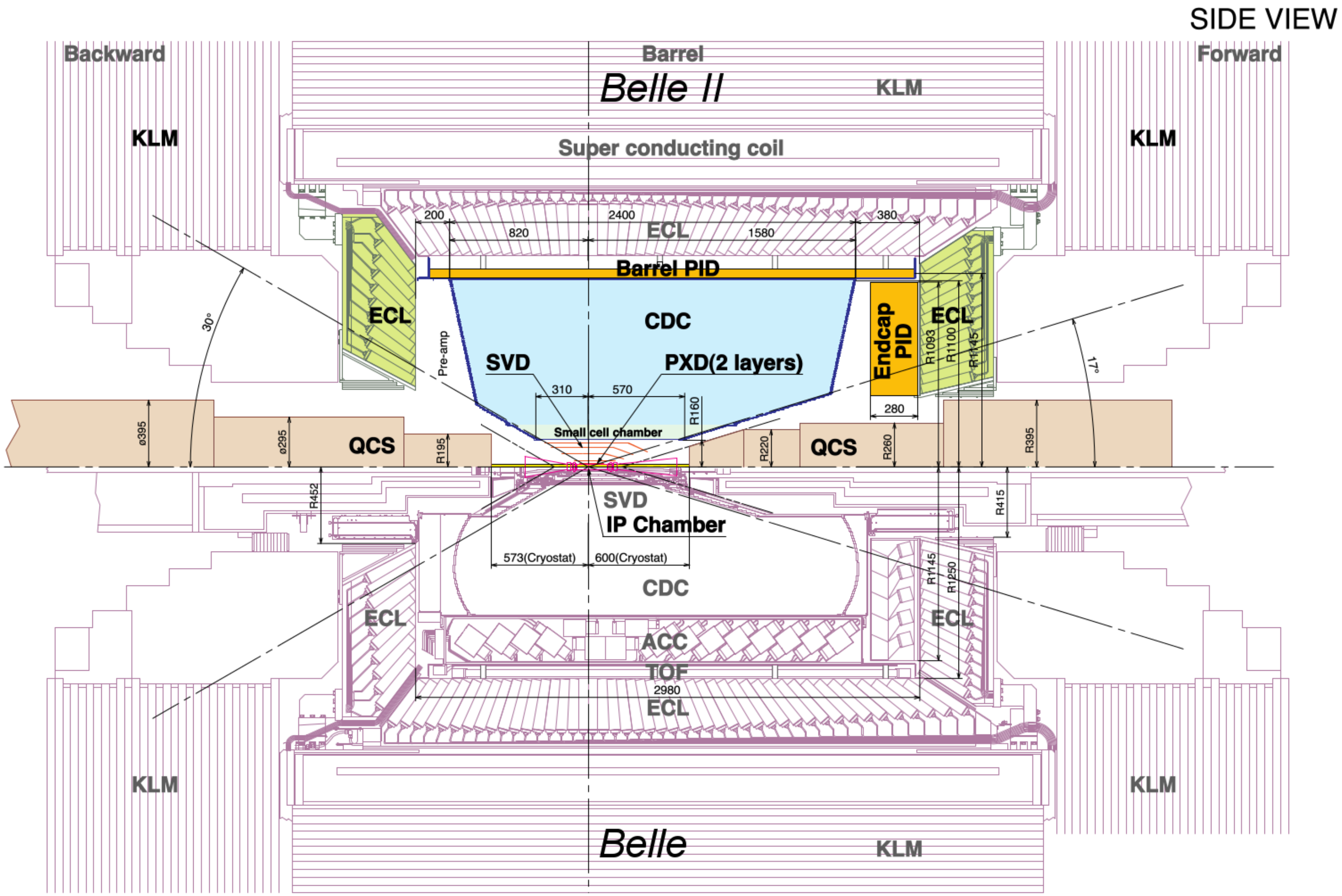


Figure 3.4.: Comparison of Belle (lower half) and Belle II (upper half); the most significant changes are marked in color [28].

4. The Software Framework

As interface between the act of data-taking at the detector site and the final analysis of the gathered data, the software framework of the Belle II experiment has to provide the tools for online accumulation of data and monitoring as well as for the offline examination of the recorded data. Initially the proposal to reuse an updated version of Belle's analysis framework BASF to provide these features for the upgraded experiment was pursued [36]. A second software framework which should utilize ideas of other major high energy particle physics experiments such as ATLAS or LHCb was developed in parallel and eventually chosen over the reworked version of BASF. The new software — which is called Belle II Analysis Software Framework (BASF2) — is written in C++ and features a Python 3 based user interface which offers a consistent and simple way to use the software. The work flow for different tasks such as online data acquisition, MC generation, detector simulation, tracking of charged particles or data analysis can be defined in so-called steering files written in Python 3. In these steering files, a path consisting of chained modules — small processing units designed to perform a specific task on event-by-event basis — is defined, which usually starts and ends with a data input and output module, respectively (see Figure 4.1). With the aid of the `boost::python` library of the boost package for C++ [37] these modules access the BASF2 libraries on the C++ level of the software where the actual operations on the data are performed. The exchange of the data-objects between the different modules occurs via the common `DataStore`, which allows one to store arbitrary event-based or persistent objects and the relations between them.

Certain libraries such as the MC generation or the detector simulation rely on additional external libraries, e.g. EvtGen [38] or Geant4 [39] in these cases. Further external packages include CERN's particle physics data analysis framework ROOT for data handling [40] or the googletest unit-testing framework [41].

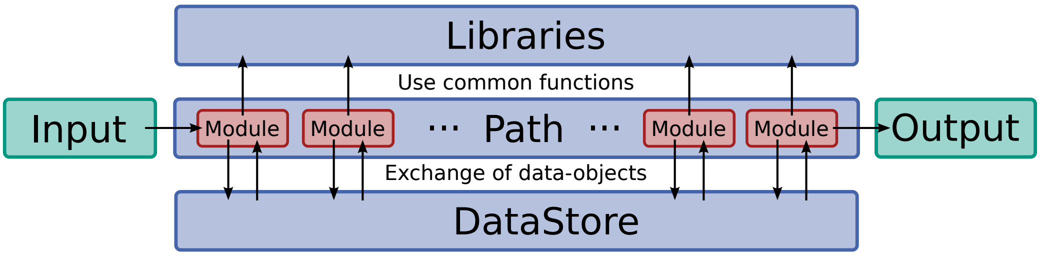


Figure 4.1.: Schematic of the execution of modules chained into a path as it is done via a Python steering file. The modules utilize functionalities defined in the C++ libraries of BASF2. They exchange data-objects via a common data storage and process these on an event-by-event basis. The initial (final) data input (output) is handled by modules designated to this task [15].

A general overview of the Belle II analysis software framework can be found in [42]. References [15], [43] and in particular [16] provide a deeper insight into the structure and functionality of the software and include some examples.

In the scope of this work the new software framework was validated with respect to its usability and performance under the circumstances of a realistic analysis. Particular focus has been lain on the analysis tools provided by BASF2, which will be elaborated on in the following Section 4.1. As preprocessed Belle MC is used for the analysis presented herein, the BASF2 tracking, detector simulation and event simulation was not used within the scope of this work. Instead the Belle-to-Belle II conversion tool B2BII (see Section 4.2) was employed and tested with regard to its correctness and applicability. A vast part of the validation was dedicated to the performance of one particular high-level analysis tool included in BASF2 — the Full Event Interpretation (see Section 4.3). It was tested and revised extensively for the purposes of this work and still offers room for studies and improvement, as it is progressively developed further.

The application of the Belle II software in actual use case scenarios offered insights into the functionality of the software framework, which cannot be obtained from simple tests during the development process or via unit-tests. As a result of this work many software bugs were revealed and repaired. Moreover, an expansion of the analysis tools regarding general functionalities and variables necessary for an analysis was accomplished.

4.1. The Analysis Package

The analysis package offers a collection of tools to extract physically relevant information from reconstructed particle tracks, particle identification (PID) likelihoods and calorimeter clusters which are stored in so-called mini Data Summary Tables — mDST for short. These data objects summarize the physical process observed by the detector and are retrieved from the detector data by other packages of BASF2, e.g. the one dedicated to the finding and fitting of charged tracks. The mDST files also contain essential information about the beam conditions of the accelerator, MC information, and experiment, run and event numbers[1].

The analysis tools provided by the analysis package allow the user to exploit this information with the aim to interpret the events and identify the physical processes they contain. Under the assumption of certain particle hypotheses, candidates for final state particles are generated from the mDST data objects and gathered in a particle list for each event. One can impose analysis specific conditions on these candidates to restrict the number of candidates to the ones relevant to the process of interest. E.g. a cut on the energy of photon candidates, which are reconstructed from ECL clusters without an associated charged track, can be applied.

---

[1] The Belle and Belle II experiments divide the time of data acquisition into a number of periods which are called *experiments*. These in turn are subdivided by changes in the accelerator or detector conditions into *runs* for which these conditions are assumed to be stable. The event number is a unique index for an event of a certain run.

The Particle Combiner Module allows one to form combinations of members of two or more of those particle lists to recombine intermediate particles with higher masses. The module considers the charge of the recombined particles and sets relations between the respective particle candidates which are necessary for subsequent calculations of the properties of the recombined particle candidate. These relations ensure the usage of the correct particle hypotheses for the calculations and avoid double-counting of a specific observation under different hypotheses. With these conditions for the recombined particle in mind, further restrictive demands can be imposed on the new higher level particle candidate. For instance a cut on the mass M of the intermediate particle candidate or its divergence dM from the nominal mass of the particle hypothesis can be applied.

By consecutively employing the particle combiner module ever higher level particles can be recombined up to the point of the recombination of the $\Upsilon(4S)$ meson. In this fashion a complete interpretation of an event is possible.

The Monte Carlo Matching Module enables the user to check for the correctness of the recombined particle candidates if the MC information of the event is available. The module can accept certain exceptions, such as the lack of neutrinos or photons originating from final state radiation. Besides a boolean value which represents the correctness of the recombination, an error index is provided, which allows one to evaluate the mistakes made for an incorrect recombination.

The Common Set of Variables defined on the C++ level of the analysis package allow the user to access the physical properties which can be calculated from the recombined particles based on the obtained interpretation of an event. This includes the 4-momenta p of final state particles, which can be retrieved from track fit results; particle masses M, which can depend either on the assumed particle hypothesis for final state particles, or the sum of the recombined particles for an intermediate particle; or more complex variables, such as the squared missing mass m_{miss}^2 which is calculated from two recombined B mesons as described in Section 5.2. As this set of variables is defined within the analysis package, it is equal for every user and should provide all necessary variables for a common analysis. Additional variables can be implemented on the C++ level and made accessible for other users. The variables can be used for selective operations to identify the decay processes of interest for an analysis and of course for the output of physical quantities.

A Interface to Multivariate Classifier Methods is also provided by the analysis package. As perpendicular cuts are not the best strategy to achieve an optimal selection of signal events, the utilization of multivariate classifiers to combine several weakly separating quantities to a strong discriminator is a common practice in data analysis. Thus, support for a wide range of multivariate classifier methods is supplied by the analysis package. These include the popular ROOT Toolkit for MultiVariate data Analysis (TMVA) [44], the commercial application NeuroBayes [45][46], and a speed optimized boosted decision tree implementation called FastBDT which was developed specifically for the FEI [15]. This allows the user to train, configure and apply a variety of classification methods and pick the most suitable one. Due to the common interface provided by the analysis package, these classifiers can be shared and reproduced with ease, thus maintaining a collaboration wide compatibility.

High-Level Analysis Tools which are commonly employed in analyses are already available in the analysis package and can be adapted to fit the needs for a specific decay study. These algorithms utilize the multivariate analysis methods and the other modules of the analysis package to conduct complex tasks such as

> **Continuum Suppression:** A common background for all analysis are the $e^+e^- \to q\bar{q}$ $(q = u, d, s, c)$ continuum events. To distinguish these events from the $\Upsilon(4S)$ events of interest various sets of variables describing the difference in the event topology of the two processes have been developed at different experiments. As described to more detail in the upcoming Section 5.4, these weakly separating continuum suppression variables are combined with the help of a multivariate classifier to discriminate continuum background.
>
> **Flavor Tagging:** For some analyses the knowledge about the flavor of the companion B meson is of interest to infer the initial flavor of the signal B meson. The flavor tagging algorithm combines information about the companion B meson in multiple stages of classifiers to provide this information for the user.
>
> **Exclusive B-Tagging** requires a complete evaluation of the decay process of the tag-side B meson of the event (Section 2.2). This process is automatized to a high degree in form of the Full Event Interpretation (FEI) algorithm which utilizes a complex network of multivariate classifiers and other modules of the analysis package, such as the particle combiner. The algorithm and its application are described in the Sections 4.3 and 5.5, respectively.

Like in the case of the multivariate classifiers, the results of these high-level analysis tools can be shared and reproduced within the collaboration as they are all based on the common foundation provided by the analysis package.

Decay Vertex Reconstruction is a task which is performed by an external package adapted to the needs of the Belle 2 experiment: The vertex reconstruction toolkit RAVE (Reconstruction of vertices in Abstract, Versatile Environments), which was originally developed for the International Linear Collider and was made available as an independent package [47]. This vertex fitter generates useful positional information about decay processes necessary for the measurement of time-dependent CP-violation, as input for the continuum suppression and for the event interpretation in general (e.g. FEI).

The analysis package provides a comprehensive and solid base for studies with the Belle II software. As described in the following Section 4.2, a conversion of the Belle data format to the new Belle II format allows one to employ the tools of the analysis package for studies conducted on Belle data and MC, thus profiting from the improved algorithms. In the scope of this work the analysis tools have been validated and improved in close cooperation with the developers of the software based on such a study on converted Belle MC.

4.2. The Belle-to-Belle II Conversion

Retaining the usability of the Belle data within the Belle II software grants the possibility to conduct Belle analyses with the new tools and, furthermore, to validate the viability of the framework. Thus, the conversion of Belle data stored in the PANTHER table format [48] into the ROOT file format which is accessible for BASF2 is beneficial and enables Belle II user who are unfamiliar with the BASF software to access to the 1040 fb^{-1} of Belle data. Additionally, this guarantees the preservation of the Belle data in a fashion that does not rely on the Belle software which eventually will be outdated and incompatible with modern computer systems.

The functionalities necessary for this conversion from the Belle format to the Belle II format are provided in the B2BII (short for *Belle-to-Belle II*) package of BASF2. Belle files are read by the `B2BIIMdstInput` module of this package. Subsequently, the conversion process is performed by the `B2BIIConvertMdst` module and the output can be validated with the help of the module `B2BIIMCParticlesMonitor`, which compares the result with the respective Belle counterparts if MC data is processed. Additionally, the module `B2BIIFixMdst` applies a general filter which comprises cuts imposed on recorded Belle data to discard unwanted QED background and contaminations originating from $\tau^+\tau^-$, $\gamma\gamma$ and beam gas processes. This filter is called HadronB(J) [49][50] and must also be applied on MC data to ensure conditions similar to the ones of recorded data.

The conversion is performed on the mDST format level which contains an already preprocessed representation of the detector data. This means, for instance, that the reconstructed tracks are converted instead of the hit information from the SVD and CDC. Hence, the improved track finding and fitting algorithms of BASF2 cannot be utilized for the Belle data and one relies on the BASF tracking. The B2BII conversion retrieves the following data objects from Belle mDST files and assigns them to the corresponding BASF2 objects which can be accessed by the analysis tools:

- Fitted tracks
- ECL clusters
- Particle identification (PID) likelihoods
- V-shaped tracks (two charged tracks originating from a π^0 or K_s^0 decay)
- Photon, π^0 and K_s^0 lists
- Beam parameters
- MC information inherited from EvtGen

The PID likelihoods are available as likelihood-ratios in the case of kaons, charged pions and protons [51] and as separate variables for electrons [52] and muons [53]. Furthermore, the information from ACC and TOF are mapped to the corresponding new sub-detectors ARICH and TOP. Not (yet) converted are the information from the KLM sub-detector as well as EFC clusters.

The beam parameters, which comprise the run-dependent beam energies of the high- and low-energy ring as well as the resulting center of mass energy and the beam angle, are read from a Belle database server or, in case of privately produced MC, retrieved directly from the Belle mDST file. They are required to calculate important quantities, such as the beam-constrained mass m_{bc} of a B meson, which is calculated under the assumption that the energy of the B meson corresponds to $E_{\mathrm{CMS}}/2$, or the difference ΔE between the determined energy of the B meson and $E_{\mathrm{CMS}}/2$.

The B2BII input module `B2BIIMdstInput` can be used as a regular BASF2 file input module at the beginning of a module path in a Python steering file. Belle mDST files can be loaded directly by providing the file name, or via URLs provided by the Belle File Search Engine[2] [54]. Depending on the task, the module-chain following the B2BII conversion within an analysis path can vary. One can distinguish between three methods to utilize the B2BII conversion:

On-the-Fly Conversion: As the `B2BIIMdstInput` module behaves — apart from the additional conversion task — like the standard `RootInput` module, which loads the Belle II ROOT files, one can perform normal analysis operations directly after the conversion. Hence, extra storage space is not required, as the Belle mDST files can be used directly. The conversion is a fast process and can be neglected compared to complex analysis operations such as the FEI. However, this method has a drawback, as it requires one to process all Belle mDST files every time the signal selection process defined in the Python steering file changes. For small tests and feasibility studies on a limited amount of data this might be insignificant and offers a fast solution. In the case of a full scale analysis, however, the number of files can increase significantly, thus rendering this approach less optimal.

Convert and Save: The conversion of the whole Belle data set and the available MC is easily possible. The converted file is saved in the ROOT format and the beam parameters are stored in a dedicated local database. Once converted, the data can be loaded via the `RootInput` module. Due to the higher compression of ROOT files and the HadronB(J) filter applied during the conversion, a converted file requires only about 55 % of the storage space of its PANTHER table counterpart.

Convert and Skim: The convert-and-save method does not make use of the ability to perform analysis tasks on the already loaded file. A conversion and a subsequent skim before the file is saved in the Belle II ROOT format represents the obvious middle course of the two aforementioned methods that includes the customary process of skimming into the steering file. A skim applies loose pre-cuts and other selective operations on the data to discard the events which are not of interest for the analysis at hand. This allows for a considerable reduction of the necessary storage space and the number of events that have to be processed in following independent analysis tasks.

More technical details about the B2BII conversion process can be found in Reference [16, Sec. 8.1]

---

[2] This option is only available for Belle data and MC produced by the Belle collaboration. The latter is referred to as *public MC* from hereon, whereas MC specifically produced for this study is referred to as *private MC*.

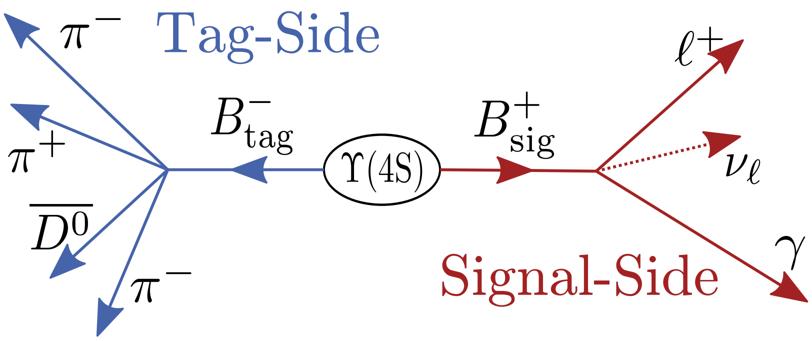


Figure 4.2.: Inclusion of arbitrarily decaying tag-side B_{tag} mesons for a full interpretation of an event featuring a B_{sig} meson decaying via the channel of interest.

4.3. The Full Event Interpretation

The Full Event Interpretation (FEI) is an implementation of the exclusive approach to B-tagging (see also Section 2.2) available as a high-level analysis tool in BASF2. The technical details of this algorithm can be found in References [15], [16] and [14]. Inspired by the original Belle implementation [17][55], the reimplementation of the algorithm for BASF2 faces the complex task of exclusive B-tagging with a sophisticated ansatz to reduce the computational effort and improve the efficiency of the tagging. As the algorithm plays a central role for the study presented herein, the key features of this ansatz and the innovations introduced with the reimplementation for BASF2 shall be summarize based on the aforementioned References. Furthermore, the most recent innovations to the algorithm are presented — referring to the status of the FEI version 3.0 “Cyberman”.

The Recombination of the Tag-Side B Meson offers additional information about the B_{sig} meson decaying via the transition of interest as they both originate from the well-known initial state of the $\Upsilon(4S)$ resonance. Recombining both the signal-side B_{sig} meson and the companion B_{tag} meson essentially is an interpretation of the complete event as illustrated in Figure 4.2. Hence, no additional tracks and ECL clusters are expected to be present in the case of the combination being correct. This allows for a rejection of wrongly recombined events and thereby mainly unwanted background of the signal decay. Furthermore, the full interpretation of the event is particularly advantageous for analyses with one or more neutrinos in the final state, as knowledge of the 4-momentum of explicitly recombined tag-side B meson allows one to infer the properties of the undetectable neutrinos. Resolving the intermediate resonances of the tag-side decay increases the resolution of the 4-momentum determination and thereby the resolution of quantities such as the missing mass and missing momentum which correspond to the neutrino’s properties. In the case of this study for which the signal decay features only one neutrino, this results in an improvement of the resolution of the squared missing mass m^2_{miss} associated with the neutrino — the quantity which will eventually be used to determine the signal yield.

Fully Hadronic or Semileptonic decay channels can be considered for the recombination of the B_{tag} meson. The latter entails the involvement of further neutrinos on the tag-side which affect the resolution of the B_{tag} meson’s 4-momentum negatively and thus results in a more diffuse missing mass and missing momentum of the event. However, the semileptonic channels feature high branching fractions which result in higher B-tagging efficiencies. The pros and cons of both options have to be considered with respect to the demands of an analysis. For either of the two options, the evaluation of many exclusive decay channels is necessary.

The Efficiency of the exclusive approach to B-tagging depends on the total number N of considered channels as well as the branching fraction $\mathcal{B}_i$ and recombination efficiency ϵ_i for each of these individual exclusive B decay channels:

$$\epsilon_{\text{tot}} = \sum_{i}^{N} \epsilon_i \times \mathcal{B}_i. \tag{4.1}$$

The aim of an exclusive tagging algorithm is to maximize the total efficiency ϵ_{tot} with which a valid B_{tag} candidate can be provided, whilst keeping the amount of computing effort at a reasonable level. As the branching fractions are constants defined by the laws of nature, an enhancement of this efficiency can only be achieved by increasing either the number of considered channels or the recombination efficiency of them. For analyses depending on quantities derived from the tag-side B meson, such as the missing mass, this efficiency severely affects the signal selection efficiency.

The Hierarchical Approach to the problem — introduced by Reference [17] — reduces the effort required when the number of considered channels is increased. Instead of addressing each exclusive decay separately, the hierarchical approach reduces these processes to a series of general tasks like the identification of stable final state particles and the recombination of heavier unstable particles as illustrated in Figure 4.3. For each of these tasks — e.g. the identification of photons, the recombination of a π^0 from two photons or the eventual recombination of the charged or uncharged tag-side B meson — a multivariate classifier is trained. The output provide by these classifiers can be interpreted as probability for the correctness of the product of the specific task, for instance the recombined π^0, and are reused as input for subsequent higher-level classifiers. Thus, it serves as input for the verification of the recombination of particles of higher levels and eventually as a quantity representing the correctness of the B_{tag} candidate provided by the algorithm. As the classifiers of higher stages depend on the output of the lower ones, the network of classifiers defined by the algorithm is trained in several stages, corresponding to the stages of the hierarchy, until the final B meson candidates are recombined.

These tasks are applicable in different combinations to reproduce multiple exclusive decay channels, thus reducing the evaluation of over 1 000 exclusive hadronic B meson decay channels to $\mathcal{O}(100)$ multivariate classifiers [14]. Hence, the number of required classification tasks is decreased by the application of the hierarchical ansatz, thereby reducing both the human and computational effort necessary to conduct the evaluation of many channels. This allows for an expansion of the list of considered decay channels, thus contributing to the total efficiency of the exclusive tagging algorithm.

For the BASF2 implementation of this algorithm — the FEI — the modules of the analysis package are utilized in a sophisticated fashion to conducted these tasks. The user can configure the algorithm with regard to the considered decay channels, decay specific pre-

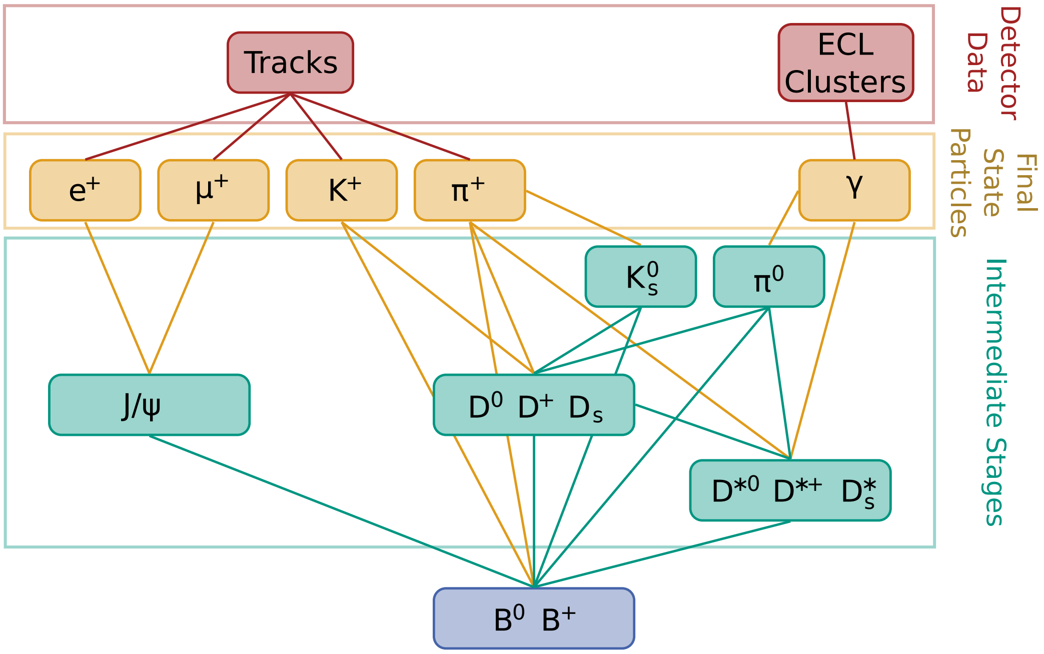


Figure 4.3.: Illustration of the hierarchical approach to the complete recombination of tag-side B_{tag} mesons, beginning with the identification of final state particles from detector information and via common intermediate states [15].

and post cuts and the used multivariate classifiers as well as their settings. Furthermore, the FEI allows one to restrict the evaluation of the tag-side B decay to fully hadronic or semileptonic decay channels, as well as to consider both modes.

The FEI also introduces the following two innovations to the algorithm.

The Selection Technique imposed on the particle candidates which emerge during the different stages of the hierarchical algorithm is a critical part of the algorithm. If the selection criteria are chosen to be too tight, the number of surviving particles is reduced, thereby decreasing the amount of training data available for subsequent stages as well as the overall efficiency of the algorithm. Overly loose selection criteria increase the combinatorics and thus the computational effort. The initial implementation of the FEI utilized a histogram-based automatic pre-cut determination which selected optimal cuts for each particle candidate recombined via a specific decay channel and for each final state particle candidate [16][14].

With FEI version 3.0 “Cyberman” the approach to this particle candidate selection was changed, discarding the automated cut-based approach and replacing it with a best candidate selection. After a loose cut rejects unphysical candidates, the 10 best candidates (in some cases more or less, e.g. 20 for photons, 5 for muons) are selected and used for subsequent stages. The quantity on which the loose cut and the best candidate selection is performed on is chosen for each particle type specifically to provide the best background discrimination (see also [15, Sec. 5.3.1]). This approach favors the survival of at least one candidate per signal event. The other side of the coin, however, is that this approach also boosts the survival of combinatorial background. An adequate user-specific post-selection is required to reduce the background again and thus obtain a reasonable purity for correctly recombined tag-side B mesons.

A Signal Specific Approach to the training and application of the exclusive B-tagging algorithm was introduced during the reimplementation of the algorithm for BASF2. This means that the FEI allows for a pre-selection of events which contain the relevant signal B decay before the recombination of the tag-side B meson is attempted. This further extends the versatility of the FEI and allows the algorithm to learn the specific event topology typical for the signal decay. Hence, the FEI should be able to differentiate between a companion B meson that is independent from the signal side, and the case of a mis-identified signal B_{sig} meson for which missing or additional particles are expected for the tag-side.

For this application option of the FEI the network of multivariate classifiers is trained on signal and generic MC, the latter one representing background in this case. Only events in which a signal candidate can be found are used as input for the FEI. The hierarchical recombination strategy is then applied on the reduced set of particles left after the signal candidate is isolated — the so-called *rest of the event*. This approach reduces the amount of combinatorics for the recombination and, furthermore, the training data to events which contain relevant signal candidates, thus encouraging the learning of the characteristics of the signal decay at hand.

This signal specific way of utilizing the tagging algorithm is referred to as *specific* FEI, whereas a FEI trained solely on generic MC without a signal-specific pre-selection is called a *generic* FEI.

This Study is the first to incorporate a FEI utilizing the best candidate selection approach and also the first to employ a signal-specific FEI up to the point of the final signal selection. The analysis is performed on B2BII converted Belle MC and thus does not profit from the enhanced track finding and fitting algorithms of BASF2 which would improve the performance of the FEI considerably.

Nonetheless, this allows for a validation of the FEI based on the decay $B^+ \to \ell^+ \nu_\ell \gamma$ and provides insights into the performance of the algorithm. The results obtained herein can be compared with the ones of Reference [12] where the Belle *Full Reconstruction* algorithm [17] for exclusive hadronic B-tagging was applied. Moreover, the effects of the innovations to the algorithm are of interest. The specific proceedings used for the analysis of $B^+ \to \ell^+ \nu_\ell \gamma$ presented herein are illustrated in the following chapter.

5. The Analysis

The goal of this Monte Carlo study of the decay $B^+ \to \ell^+ \nu_\ell \gamma$ ($\ell = e, \mu$) is to achieve a signal selection similar to or preferably better than the one presented in [12] with the analysis tools provided by the new software framework for the Belle II experiment. As this rare leptonic decay features an undetectable neutrino, additional information about the decay of the signal-side is retrieved from the remaining particles of the event. Combined, these particles make up the tag-side of the event as has been illustrated in Section 2.2. The interpretation of the full event via the recombination of the initial $\Upsilon(4S)$ meson thus provides additional information to suppress unwanted background and allows one to infer the properties of the signal-side neutrino as well as to determine the missing mass m_{miss} of the event. For this purpose the tag-side is recombined in fully hadronic decay channels by employing the FEI algorithm in a signal decay specific mode. This analysis makes first use of this signal specific exclusive tagging algorithm. All results presented herein were obtained using the Revision #26252 of the Belle II Analysis Software Framework from March 11th, 2016 and the FEI version 3.0 “Cyberman”.

Before a branching fraction measurement on actual data is possible, a analysis procedure has to be developed and verified on simulated data. A signal selection technique has to be prepared and optimized to achieve the maximal possible signal significance for a subsequent fit which will determine the signal yield on actual data. The fit procedure and the evaluation of systematic uncertainties of the analysis are also studied on MC before the analysis procedure is deemed ready for the application on detector data. The study presented herein focuses on the signal selection for the decay $B^+ \to \ell^+ \nu_\ell \gamma$ based on the new Belle II software and simulated Belle data. Fit procedures and systematic uncertainties are not discussed.

The preparation of the MC samples used for the evaluation of the signal selection process presented in this study is described in the following Section 5.1. An important variable for the analysis — the squared missing mass m^2_{miss} — is introduced in Section 5.2. It is calculated from the information obtained from the tag-side B meson and serves as basis for a simple counting experiment to evaluate the quality of the signal selection presented herein.

The actual signal selection process is split into several steps, beginning with the illustration of a pre-selection of signal candidates in Section 5.3 for eventual multivariate classifier trainings and the signal specific FEI algorithm. The training and application of this algorithm for exclusive B-tagging is described in Section 5.5. An important part of the signal selection process is the treatment of unwanted background. The Sections 5.4 and 5.6 are dedicated to the suppression of two specific types of background, namely continuum background and signal mimicking background, respectively. For the continuum suppression a strategy common for Belle and Belle II studies is applied. The approach to the discrimination of signal-like background was motivated by [12], but includes some changes which have been introduced in the scope of the adaption to BASF2. In the concluding Section 5.7 the process of the final signal selection is described. Two options for this final selection are presented. The first option follows the idea of a fit of the m^2_{miss} distribution in bins of the output of the signal-like background suppression classifier as it has been done by in [12]. The second option utilizes the separation power of the same classifier to prepare a m^2_{miss} distribution for a one-dimensional fit. Both options are compared to the signal selection obtained by *Andreas Heller et al.* [12].

5.1. The Monte Carlo Samples

The analytical procedure and the results for the MC study of the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$ presented in this chapter are based on simulated Belle data. The MC events are scaled to correspond to the expected yields for the full Belle data set, which was recorded in the time span from 1999 to 2010 at the $\Upsilon(4S)$ resonance and contains approximately 771.6×10^6 B meson pair events. In addition to the MC samples provided by the Belle Collaboration, private MC samples for the signal decay channels and signal mimicking backgrounds were produced for which the scaling factors have to be calculated separately. The considered peaking backgrounds are decays of the type $B^+ \to X_u\, \ell^+ \nu_\ell$, where X_u represents a light neutral meson that could decay into two photons. One of those photons could be misinterpreted as the signal photon, what is particularly often the case if the neutral meson is either a π^0 or an η. Hence, for these two cases MC samples with higher statistics were produced, using the latest measurements of the respective branching fractions.

For all privately produced MC events the second B meson of the event decays generically, meaning that the b quark decays via the weak interaction $b \to cW$ into a charm quark and a charged W boson. This decay type poses the majority of the B meson decays recorded at Belle. An event in which both B mesons decay via these channels is referred to as *generic* event. Another background that has to be considered are *continuum* events which occur independently from the center of mass energy when the colliding electrons and positrons do not result in the bottomonium resonance $\Upsilon(4S)$: $e^+e^- \to q\bar{q}$ $(q = u, d, s, c)$. For both continuum and generic background, as well as the remaining $B^+ \to X_u\, \ell^+ \nu_\ell$ channels, all available MC samples provided by the Belle Collaboration were used.

A full overview of the MC data used to obtain the presented results can be found in Table 5.1. The table also lists the scaling factors necessary to maintain the correct proportion between the evaluated MC samples. For the privately produced MC these weights are calculated with the formula

$$\text{weight} = \frac{N_{B\bar{B}} \times 2 \times \mathcal{B}(\Upsilon(4S) \to B^+B^-) \times \mathcal{B}(\text{MC type})}{\text{Number of generated MC events}} \tag{5.1}$$

which can easily be derived from Equation 2.8. In addition to the number of generated MC events, the formula also requires a prior knowledge of the expected branching fraction of the respective channel. The branching fraction of the signal decay is assumed to be $\mathcal{B}(\mathrm{B}^+ \to \ell^+ \nu_\ell \gamma) = 5.0 \times 10^{-6}$ to ensure a comparability to the results of [12]. The values for the peaking backgrounds are taken from [25]: $\mathcal{B}(\mathrm{B}^+ \to \pi^0 \ell^+ \nu_\ell) = 7.8 \times 10^{-5}$ and $\mathcal{B}(\mathrm{B}^+ \to \eta\, \ell^+ \nu_\ell) = 3.8 \times 10^{-5}$.

Table 5.1.: Overview of used MC for the signal channels and the relevant background and the respective scaling factors. For the privately produced MC samples the number of events is listed directly; the public MC samples are available as so-called *streams*, each of which containing an amount of the respective decay type corresponding to the number of events expected in the Belle data set.

Channel		Number of Events	Scaling Factor
Signal (Electron)	$B^+ \to e^+ \nu_e \gamma$	10×10^6	3.966×10^{-4}
Signal (Muon)	$B^+ \to \mu^+ \nu_\mu \gamma$	10×10^6	3.966×10^{-4}
Generic Bkg.	$b \to cW$	10 Streams	1/10
Continuum Bkg.	$e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$	6 Streams	1/6
Peaking Bkg.	$B^+ \to X_u \ell^+ \nu_\ell$	20 Streams	1/20
↳ $\pi^0 + e$	$B^+ \to \pi^0 e^+ \nu_e$	10×10^6	6.187×10^{-3}
↳ $\pi^0 + \mu$	$B^+ \to \pi^0 \mu^+ \nu_\mu$	10×10^6	6.187×10^{-3}
↳ $\eta + e$	$B^+ \to \eta\, e^+ \nu_e$	10×10^6	3.014×10^{-3}
↳ $\eta + \mu$	$B^+ \to \eta\, \mu^+ \nu_\mu$	10×10^6	3.014×10^{-3}

All MC data were produced with the Belle analysis framework BASF and converted to the Belle II format with the B2BII package of BASF2.

5.2. The Missing Mass

As the neutrino of the signal decay cannot be detected directly, the full event has to be evaluated to recover the properties of this massless particle. In Chapter 2.2 some strategies were introduced to gain information about the second B_{tag} meson of the event, once a candidate for the signal side B_{sig} meson is found. An important quantity that can be retrieved from this additional Information is the squared missing mass of the event which can be calculated via the equation

$$\begin{aligned} m^2_{\text{miss}} &= (p_{B_{\text{sig}}} - p_\ell - p_\gamma)^2 \\ &= ((E_{\text{CMS}}/(2c), -\vec{p}_{B_{\text{tag}}}) - p_\ell - p_\gamma)^2 \end{aligned} \tag{5.2}$$

for the decay at hand, where p_ℓ and p_γ are the 4-momenta of the selected lepton and photon candidates, respectively. The 4-momentum of the signal B_{sig} meson can be calculated due to the well-known initial state of the $\Upsilon(4S)$ produced in the electron–positron collision. As the two B mesons are produced in a back-to-back decay of the $\Upsilon(4S)$, the momentum of the signal-side can be estimated with the negative momentum of the tag-side B meson in the rest frame of the $\Upsilon(4S)$ meson. The resolution of the missing mass can be improved by using the center of mass energy E_{CMS} instead of the energy of the tag-side for its calculation, thereby including the knowledge about the initial state.

Due to the neutrino being massless, the squared missing mass is expected to peak at zero for a correctly interpreted event, whereas a deviation is expected in the case of wrongly recombined events. However, even for correctly recombined events a variance originating from the momentum resolution of the detector can be observed. As demonstrated in [12], these effects are dominated by the contribution of the photon, for which the momentum is determined from the energy deposit in the ECL.

In case of wrongly recombined events, the squared missing mass deviates more distinctly from zero. Thus, the variable provides good separation power to distinguish between signal and most background candidates. It is an ideal candidate to obtain a measurement for the number of observed signal events via a fit as in the case of [12] or a counting experiment as it will be attempted in this study. Variables dependent on the missing mass, e.g. the highly correlated quantity ΔE of the signal-side, which represents the difference between the expected and the calculated energy of a signal B candidate, are therefore excluded as input for multivariate classifiers.

5.3. The Pre-Selection

An initial signal selection is performed as part of the skim conducted subsequently to the B2BII conversion by combining B_{sig} candidates from

- ECL clusters with an energy above 1.0 GeV and no associated charged track and
- a charged track with either an electron or muon PID larger than 0.4, respectively.

Here PID refers to the particle identification (PID) likelihood of the electrons and muons recorded by the Belle detector, which are described in [52] and [53], respectively. Additionally, the selected charged tracks are required to originate no farther than 2.0 cm in the x-y-plain and 4 cm in z-direction away from the interaction point specified by the beam parameters. The reconstructed mass of the charged B meson candidate resulting from this particle combination is expected to be in the interval between 1.0 GeV and 6.0 GeV, as a majority of the B meson's energy can be passed to the undetectable neutrino not considered here. This loose cut captures the whole mass range obtained for actual signal events, but does cut away unwanted continuum background.

The energy of the photon candidate is required to be larger than 1.0 GeV to fulfill the demands imposed by the theoretical treatment of the studied decay, as has been illustrated in Chapter 2. This cut also conforms the selection of the nominal analysis in the References [12] and [23].

Furthermore the cuts of the HadronB(J) skim [49][50] are applied as part of the B2BII conversion process (see also Section 4.2). These cuts are chosen to reduce unwanted QED background, as well as τ pair, photon pair and beam gas background while maintaining a high efficiency of 99.1 % for hadronic B meson pair events. In the case of the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$ the HadronB(J) skim has a reduced efficiency of about 90 % for $\ell = e$ and a slightly higher efficiency for the decay into a muon. This is due to the neutrino, which can carry a large amount of the released energy, and due to the lower number of tracks for these rare event types.

The combined efficiency of this first selection is shown in Table 5.2 and reflects the recombination efficiency of the signal decay. One observes, that for all decays the electron case can be recombined with a sightly higher efficiency than the muon case, which is to

Table 5.2.: Efficiencies of pre-selection for the signal channels and the main peaking backgrounds.

Channel	Events after Skim	Efficiency
$B^+ \to e^+ \nu_e \gamma$	4716943	47.17 %
$B^+ \to \mu^+ \nu_\mu \gamma$	4662097	46.62 %
$B^+ \to \pi^0 e^+ \nu_e$	4348901	43.49 %
$B^+ \to \pi^0 \mu^+ \nu_\mu$	4220244	42.20 %
$B^+ \to \eta\ e^+ \nu_e$	2855218	28.55 %
$B^+ \to \eta\ \mu^+ \nu_\mu$	2709306	27.09 %

be expected, as the muon can only be identified conclusively if it reaches the KLM. This initial selection serves as basis for all following analysis methods. The trainings of all employed multivariate classifiers was conducted on MC data which passed this skim.

5.4. The Continuum Suppression

A large fraction of the background events originate from $q\overline{q}$ continuum processes ever present at e^+e^- collider experiments. In these processes pairs of the lighter u, d, s or c quarks are produced instead of the bottomonium resonance Υ(4S). Compared to the B meson pairs produced in $B\overline{B}$ events, the light quarks of continuum events receive a considerably larger momentum due to the higher amount of released energy. As the quarks are produced back-to-back in the boosted reference frame of the asymmetrically colliding electrons and positrons and carry a high momentum of almost half of the center of mass energy, the event has a distinct jet-like shape. In contrast to this the Υ(4S) events, in which the B mesons are produced nearly at rest, show a more spherical event topology, as is illustrated in Figure 5.1. On average, the products of hadronic B decays acquire a much smaller transverse momentum. In case of the leptonic signal decay $B^+ \to \ell^+ \nu_\ell \gamma$ discussed herein, this criteria is not fulfilled. However, combined with the tag-side B meson, a signal event should provide enough information to be distinguishable from continuum.

As continuum events pose a special type of background affecting all analyses conducted on the Belle and Belle II data, a common strategy is recommended which is discussed in Reference [7, sec. 9.5]. Several weak discriminating variables reflecting the features of the event topology have been studied and improved over time. In this analysis, a gradient-boosted decision tree is employed to combine these variables to a powerful discriminator to discard the continuum background. The following overview based on Reference [7, sec. 9.5] categorizes the relevant quantities and provides a short description.

Thrust Variables are based on the definition of thrust established by the high energy experiments at the PEP and PETRA accelerators in the 1980s to describe jets. The thrust axis is defined as the unit vector $\vec{n}$ for which the sum of the projections of the considered

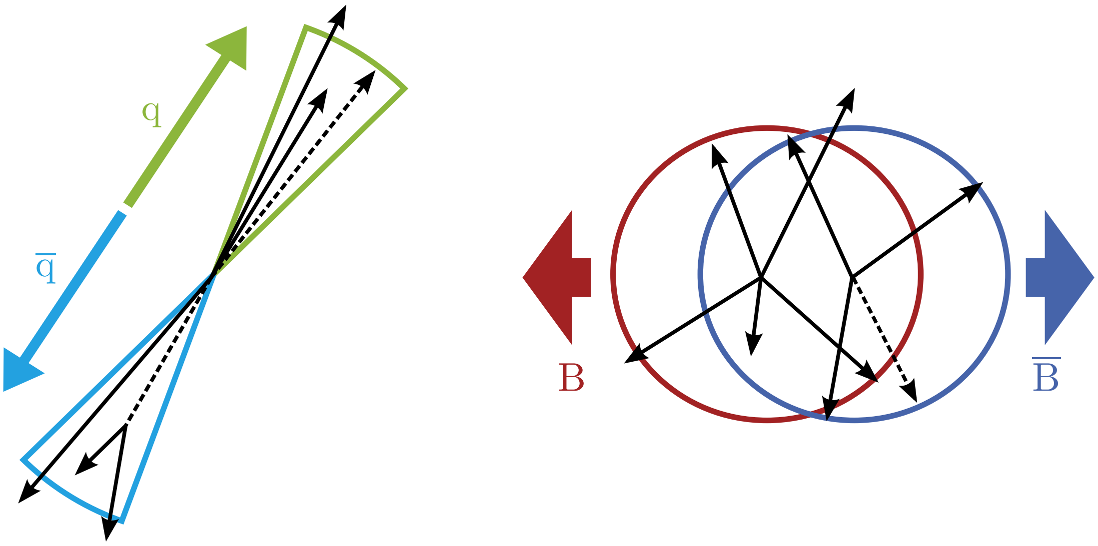


Figure 5.1.: Illustration of the differences in the event shape for the jet-like $e^+e^- \to q\overline{q}$ ($q = u, d, s, c$) continuum events displayed on the left and $B\overline{B}$ events on the right. The latter result in a more spherical event topology, which allows one to distinguish between the two event types [56].

particle momenta on itself is maximal:

$$\vec{n} = \max_{|\vec{n}=1|} \left(\sum_{i}^{N} \vec{n} \cdot \vec{p}_i \right). \tag{5.3}$$

Based on this axis, the scalar thrust T is defined as

$$T = \max_{|\vec{n}|=1} \frac{\sum_i^N \vec{n} \cdot \vec{p}_i}{\sum_i^N \vec{p}_i}. \tag{5.4}$$

These are used to determine the jet-likeness of the to be evaluated signal B_{sig} and the tag-side made up of the remaining particles. To distinguish between $B\overline{B}$ and continuum events, the cosine of the angle between the thrust axis of the B_{sig} meson and the combined remaining particles of the event $cos(\theta_{B_{\mathrm{sig}}B_{\mathrm{tag}}\mathrm{Thrusts}})$ is calculated. As a spherical event topology is expected, the absolute value of this quantity should show an uniform distribution, whereas the jet-like topology of continuum events should result in a peak at 1, due to the back-to-back configuration of the quarks. Additional discriminating power is obtained from the magnitudes of the thrusts T of both B mesons. In particular the tag-side should show a low thrust, because of the spherical topology. Furthermore the cosine of the angle between the thrust axis of the B_{sig} meson and the beam axis is considered.

Fox-Wolfram Moments (FWM) are more advanced event shape variables introduced by *Fox* and *Wolfram* in 1978 [57]. They describe the phase-space distribution of momentum and energy flow for a collection of N particles. The ℓ-th order FWM H_ℓ is calculated via the equation

$$H_\ell = \sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| \; P_\ell(cos\theta_{ij}), \tag{5.5}$$

where $|\vec{p}_{i,j}|$ is the magnitude of the momentum of the i-th or j-th particle, respectively, and θ_{ij} the angle between them. $P_\ell(cos\theta_{ij})$ is the ℓ-th order Legendre polynomial determined with the cosine of this angle. In the limit of vanishing invariant particle masses the energy-momentum conservation requires the 0th order FWM to assume the value $H_0 = 1$. Thus, the normalized FWM ratios $R_\ell = H_\ell/H_0$ provide separation power, as they take values close to 0 for odd ℓ and close to 1 for even ℓ for events with strongly collimated jets. As input for the classifier the ratio of the 2nd and 0th FWM $R_2 = H_2/H_0$ is used.

Further high-level variables are constructed from FWMs in the form of two Fisher discriminants. The two Fisher discriminants take the affiliation of considered tracks to either the signal B_{sig} meson or the tag-side into account. They are referred to as Super Fox-Wolfram moments (SFWM) and Kakuno Super Fox-Wolfram moments (KSFWM) and have been developed specifically for the event topologies observed at the B factories. A thorough description of their definition can be found in Reference [7, sec. 9.5], where the determination of the coefficients of the Fisher discriminants is elaborated on. For the continuum suppression classifier training performed for this analysis a total of 16 of these $B\overline{B}$ event optimized KSFWM variables were employed.

CLEO Cones were established by the CLEO collaboration in 1996 [58] in context of charmless B decays and provide additional discrimination power. As a set of multivariate discriminant variables they describe the momentum flow around the thrust axis of the signal B meson, binned in 9 cones with opening angles increasing in steps of 10 degrees. In the case of continuum events featuring collimated jets this momentum flow should be higher for the innermost cones.

Flavor Tagging Variables can provide additional information to help distinguish between continuum background and the $B\overline{B}$ events of interest. The flavor tagger algorithm [59] evaluates the tag-side of an event to determine the flavor of the B_{tag} meson. Several multivariate classifiers combine particle information to provide an output variable in the range between −1 and 1 to distinguish anti-particle and particle, where −1 indicates a high probability for a $\overline{B}$ and vice versa. Despite its original purpose as an important input for time-dependent CP-asymmetry studies, this quantity has been found useful to discriminate continuum background. In the case of a continuum event, the flavor tagger should not be able to determine a flavor for the tag-side with a high certainty, which will be reflected in its output. Hence the flavor tagger output is also used in this analysis as an input variable for the continuum suppression classifier.

Another variable which can be associated with the measurement of time-dependent CP-asymmetry is Δz — the distance between the fitted vertices of the signal-side B meson and the tag-side. As the B mesons travel a small distance before decaying into the observable final state particles, the distance between the fitted vertices should be larger than in the case of the falsely recombined B mesons in continuum events.

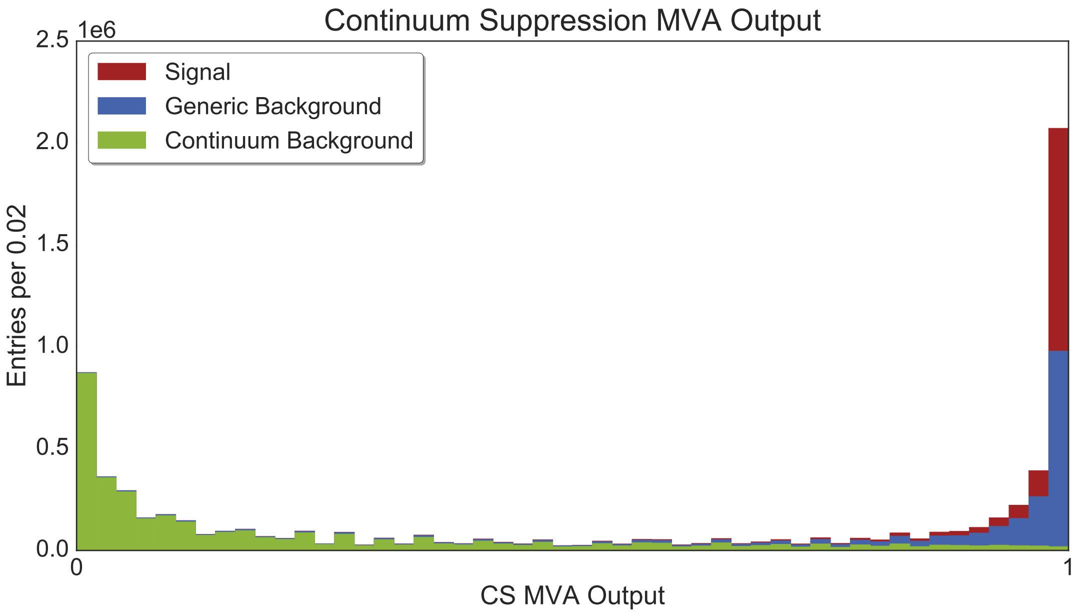


Figure 5.2.: Validation of the discrimination power of the continuum suppression classifier based on an unscaled MC sample with signal $B^+ \to \ell^+ \nu_\ell \gamma$ and continuum $e^+e^- \to q\bar{q}$ $(q = u, d, s, c)$ events, as well as generic $b \to cW$ events.

The Combination of all these weakly separating variables to a strong discriminator is achieved with the aid of a fast gradient-boosted decision tree (FastBDT). For the FastBDT a forest of 100 trees is generated, allowing a maximal depth of 3 layers, the shrinkage set to the value 0.10 and the number of attempts to find the optimal cut set to 8. The multivariate classifier is trained in a signal specific manner on 10 million independent signal MC events and an equal amount of simulated continuum events from 1 stream to identify non-continuum events. The input sample is randomly split into two independent subsets, one of which is used for the learning process, the other one for the validation of the training result. It is important to note that, even though the discriminator is trained specifically for the signal decay, the target for the learning process is to distinguish between continuum and non-continuum events. Thus the MC truth based variable `isNotContinuumEvent` is used as target variable. A training performed with the target variable `isSignal` yields unsatisfactory results, as in this case wrongly recombined signal events are assigned to the background category, thereby blurring the true properties of the background.

As a figure of merit for the performance of the resulting discriminator the area under the curve of the receiver operator characteristic shall be stated, which, with a value of 95.8 %, speaks for the applied procedure. The result of the check for over-training performed by TMVA can be seen in Figure A.1 in Appendix A. The output of the classifier can be interpreted as probability for an event to be of the category *non-continuum*. To validate the continuum suppression, the classifier was applied on a further independent and unscaled sample, which also contained generic background. The resulting distribution is shown in Figure 5.2. It is clearly visible, that the quantity provided by the trained gradient-boosted decision tree allows for a good suppression of continuum background. The signal events as well as generic background peak at the value 1, thus indicating that a general separation between $B\overline{B}$ events and continuum events was obtained.

5.5. The Signal-Specific Hadronic B-Tag Selection

To gain information about the neutrino of the signal B decay, an evaluation of the remaining B meson of the tag-side is performed. For this task the exclusive tagging algorithm FEI, which was illustrated in Section 4.3, is used in a signal specific manner. This means that the signal decay is recombined first, thereby restricting the search for a valid tag-side B candidate to the remaining particles in the event. Due to this selection prior to the tag-side recombination, the FEI algorithm can be trained to consider the characteristics of the signal decay at hand, such as the low number of charged tracks in the case of the decay $B^+ \to \ell^+ \nu_\ell \gamma$. Additionally, the combinatorics is reduced, as the algorithm is trained solely on the remaining particles. For the interpretation of the decay of the tag-side B_{tag} meson only fully hadronic decay channels are considered.

The Monte Carlo Samples chosen for the training of the hierarchical exclusive tagging algorithm reflect the signal specific approach applied in this analysis. The classifier network of the FEI is trained on two MC samples consisting of 100×10^6 events of each of the two signal channels $B^+ \to \ell^+ \nu_\ell \gamma$ with either an electron or a muon as product. From those all correctly identified signal events which survived the initial selection process described in Section 5.3 are used for the training. Additionally, about 200×10^6 generic background events are included into the training sample. The signal pre-selection is also performed on this part of the training sample, yielding only wrongly recombined signal B mesons, of course. These generic events provide additional training data for the classifiers of final state and intermediate particles during the hierarchical training process. They can also lead to correctly recombined tag candidates, but are more likely to contain additional charged tracks, which are associated with neither the tag candidate, nor the in this case false signal candidate.

A Preparation of the Rest of the Event — meaning the remaining particles unassociated to the signal candidate — prior to the training and application of the FEI tagging algorithm is necessary to improve its quality and reduce computation time. In the scope of this preparation a filter is imposed on all charged tracks and ECL clusters remaining after the signal pre-selection. The filter rejects events which do not show the potential to provide a valid tag-side candidate after the signal pre-selection by briefly analyzing some features of the sum of all remaining particles, such as

- the beam-constraint mass m_{bc} resulting from the sum of all remaining particles, which must be larger than 4.82 GeV; and

- the energy difference ΔE of the sum of all remaining particles with respect to the expected energy of the tag-side B_{tag} meson resulting from the clean initial state, which must be larger than -380 MeV;

thereby further reducing the combinatorics for the tag-side recombination. Furthermore, a cleansing of the event is performed by the application of cuts on

- the point of closest approach of charged tracks to the IP, requiring a maximal distance of 0.5 mm in the r-ϕ-plane and 1.0 mm in the z-direction;
- the E9E25 ratio of ECL clusters, discarding such ones which show a ratio lower than 0.905; and
- the cluster timing of ECL clusters, rejecting events with a time measurement difference larger than 50 ns with respect to the event time

to reject particles originating from beam background or other effects which do not originate from the interaction point of the electron–positron collision.

The Training of the network of multivariate classifiers comprised by the FEI is conducted on the training samples obtained from the above described selection. Again gradient-boosted decision trees (FastBDT) with the configuration chosen as described in the previous chapter are used. The network of FastBDTs is trained to recombine B_{tag} candidates in the fashion determined by the hierarchical algorithm described in Section 4.3. Each classifier is dedicated to a specific particle recombination, evaluating the properties of the recombined daughters and the resulting particle of higher mass. The result of this evaluation is the output of the classifier, which can be interpreted as a probability for the correctness of the particle combination. This probability is reused as input for higher level classifiers which use the produced particle as daughter for further recombinations.

On each particle combination a loose cut depending on the mother particle is imposed, thus rejecting non-physical combinations. To select the most suitable particle candidates from the ones surviving this cut a best candidate selection is performed. The variable on which the best candidate selection is based on is chosen individually for each particle type in the hierarchical approach. For final state particles either the corresponding PID likelihood or the energy E is used. For each particle the 10 most suitable candidates[1] are chosen for each event and fed into the classifiers of the next higher level for the recombination of particles with higher masses. The criteria used for the selection of the candidates of the intermediate particles recombined on higher levels is either

- the absolute value of the deviation from the nominal mass $|dM|$,
- the absolute value of the difference between the released energy of the decay and the nominal energy of the particle $|dQ|$, or
- the product of the classifier output of the daughter particles which are recombined to form the intermediate resonance.

The criteria used for the selection of the final B_{tag} candidates which are eventually recombined on the highest level of the hierarchical algorithm is the output of the respective classifier. In the last step up to 10 B_{tag} candidates are selected for each event.

---

[1] Depending on the particle type, this value varies between 5 (e.g. for leptons) and 20 (e.g. for photons), but is mostly set to 10.

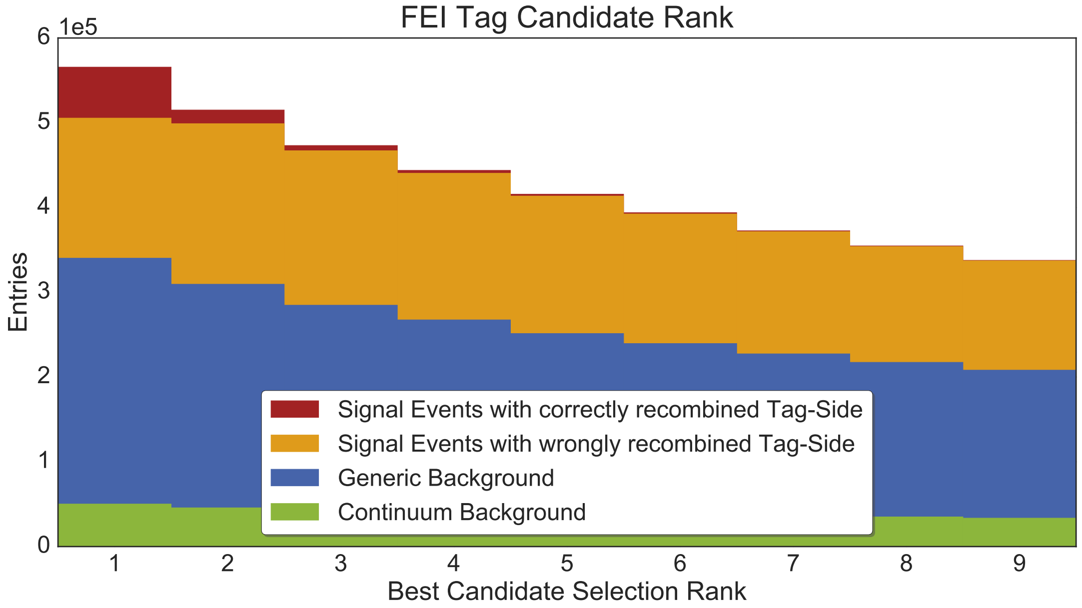


Figure 5.3.: Result of the best candidate selection of the tag candidates provided by the FEI. The Histogram produced from an arbitrary amount of signal and background events shows the number of candidates provided for an event ranked by the probability output of the FEI.

Due to this best candidate selection approach, the selection of the particle candidates is performed on an event basis. Only particle candidates which do not survive the aforementioned particle specific loose cut are rejected. The tagging efficiency of the signal specific FEI algorithm — meaning the efficiency with which at least one tag-side candidate for a signal event can be provided — is about 10 %. However, this quantity does not take the correctness of the B_{tag} candidate into account.

In the majority of the events the maximal number of 10 candidates is provided. This is illustrated in form of the best candidate selection result shown in Figure 5.3. The displayed stacked histogram of the best candidate rank for the charged B mesons provided as tag candidates by the FEI algorithm is obtained from an unscaled MC sample. One can see, that the algorithm does provide multiple candidates not only for signal events, but also for generic and continuum background. For signal events, the correctly recombined tag-side B meson does not always rank highest, meaning that another tag-side candidate achieved a higher probability for being correctly interpreted by the FEI. This is the case for about 35 % of the events for which the FEI can provide the correct tag-side B_{tag} meson.

Only one of the tag-side B meson candidates is chosen for the further evaluation of the event, as only one of them can be correct. A selection of this candidate based on the FEI probability output at this point would result in the rejection of these 35 % of the correct tag-sides. Thus, a further filter in form of the post-processing described in the following is applied beforehand.

The Post-Processing of the candidates provided by the FEI algorithm is imposed to reduce the amount of background which survived the tagging process. Furthermore, the chances

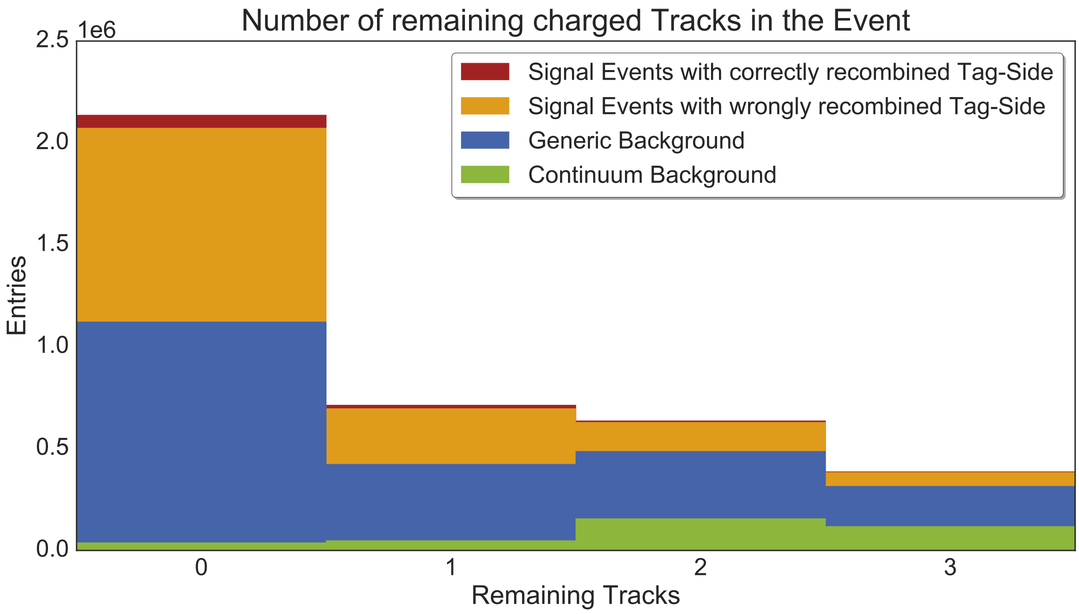


Figure 5.4.: This histogram shows fully interpreted $\Upsilon(4S)$ events retrieved from an unscaled MC sample which signal events as well as generic and continuum background. The signal events are further subcategorised into correctly recombined $\Upsilon(4S)$ and events for which the tag-side is wrongly recombined.

for the correctly recombined tag-side candidate to survive the best candidate selection process should be increased by this post-processing, because incorrectly tagged, but higher-ranking candidates might be rejected by a further filter.

For this selection, properties of the $\Upsilon(4S)$ can be included, as both the tag-side B_{tag} meson and the signal B_{sig} meson are now available and can be combined. First of all, events which do not allow for the recombination of a physically sensible $\Upsilon(4S)$ candidate can be discarded at this point. Subsequently, perpendicular cuts can be applied to reject events which do not show the qualities of a signal event. An important variable for this selection is the number of remaining tracks in the event not associated to the recombined $\Upsilon(4S)$ meson. As mentioned before, in the case of a correctly interpreted event no additional charged tracks are expected. The distribution of this quantity after the recombination, obtained from an unscaled MC sample, is displayed in Figure 5.4.

In the course of the recombination of the signal and tag-side to form the $\Upsilon(4S)$ meson, the **signal-side best candidate selection** is conducted as the first step of the post-processing. For the selection of the most suitable B_{sig} candidate the energy of the photon has proven to be an efficient criteria. The correct signal candidate in a signal event is chosen with an efficiency of 86.04 % if the candidate with the highest photon energy is selected.

After the recombination of the $\Upsilon(4S)$ from the selected signal-side and the tag candidates, the **perpendicular cuts** listed in Table 5.3 are imposed. The first cut is applied on the output of the continuum suppression classifier introduced in Section 5.4. Every candidate for which the probability of originating not from a continuum event, as determined by the multivariate classifier, is below a threshold of 0.2 is discarded.

Table 5.3.: This table shows the signal efficiencies and background rejection rate (both in percent) of the cuts applied to the $\Upsilon(4S)$ candidates prior to the final tag-side selection. The three main background sources are listed separately to show the effect of the cuts on each category.

Cut on	Signal Efficiency	Background Rejection for		
		Generic Bkg	Continuum Bkg	Peaking Bkg
CS Output	99.57	11.14	66.34	2.74
$\Upsilon(4S)$ Mass	99.16	50.91	55.38	20.89
ΔE of B_{tag}	94.49	31.93	34.71	32.75
m_{bc} of B_{tag}	98.79	75.23	87.88	47.12
remaining Tracks	88.11	66.89	66.56	23.91
missing Mass m_{miss}	98.14	71.70	69.82	24.03
Combined	79.70	99.31	99.88	84.18

Cuts on the mass of the recombined $\Upsilon(4S)$ candidate and the missing mass m_{miss} are now possible. Demanding these variables to be in the intervals 7.5 GeV $\leq M \leq$ 10.5 GeV and -2.0 GeV $\leq m_{\text{miss}} \leq$ 4.0 GeV, respectively, results in the rejection of unphysical events.

Furthermore, cuts on the B_{tag} candidate are imposed to discard tag-side candidates with unphysical properties. As the tag-side is recombined considering only fully hadronic decay channels, the candidates should show the features of a charged B meson with no missing energy. Hence, the conditions -150 MeV $\leq \Delta E \leq$ 100 MeV and 5.27 GeV $\leq m_{\text{bc}} \leq$ 5.29 GeV are required to be fulfilled by tag-side candidates for them to be valid.

Lastly, the aforementioned cut on the remaining tracks in the event after the full recombination of the $\Upsilon(4S)$ is imposed by selecting only the candidates for which maximally one additional charged track remains. Despite the undertaken attempts to remove additional charged tracks originating e.g. from beam background, a considerable fraction of correctly interpreted $\Upsilon(4S)$ events still feature an unaccounted track, as can be seen in Figure 5.4. These candidates shall not be rejected at this point, wherefore the cut allows for one remaining charged track.

The resulting signal efficiencies and the background rejection rates of all cuts are listed in Table 5.3. After these cuts the tag candidate is chosen based on the FEI probability. This final **tag-side best candidate selection** results in 71.60 % of the surviving signal events in the correctly recombined $\Upsilon(4S)$ candidates.

The Result of this Full Event Interpretation is a selection yield of 0.78 % of the initial 20×10^6 signal events. This includes the HadronB(J) cuts and the recombination efficiency of the signal-side. However, a considerable amount of background still remains, resulting in a signal purity of 0.28 % for the scaled sample.

The majority of the tag candidates provided by the algorithm are not correctly recombined, as can already be seen in the Figures 5.4 and 5.3. To improve the purity of the sample, a cut on the FEI probability output shown in Figure 5.5 would usually be imposed. As this output reflects the probability of the tag-side recombination being correct, the signal

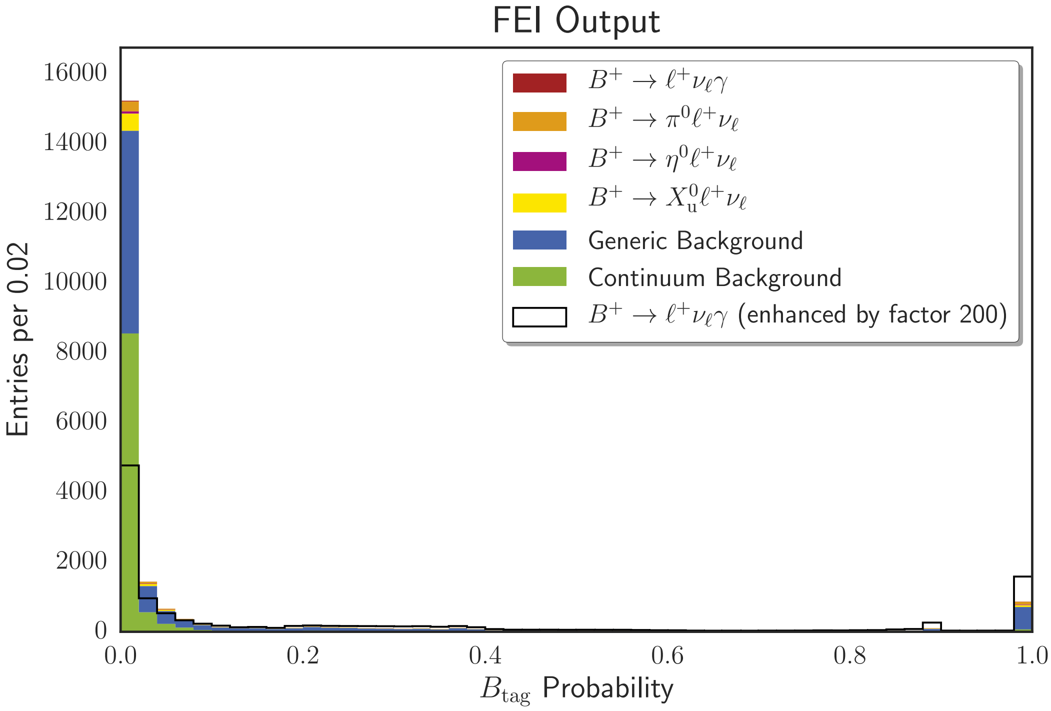


Figure 5.5.: The displayed distribution of the FEI output, which corresponds to the probability for the tag-side being correctly identified, illustrates the background rejection power of the quantity. However, a cut on the variable would also discard a large fraction of signal events for which the tag-side process could not be resolved correctly. To illustrate this, the shape of the signal distribution, enhanced by a factor of 200, is shown in addition to the stacked histogram obtained from the scaled MC sample described in Section 5.1.

candidates with wrongly recombined tag-sides will be discarded by such a cut (see Figure 5.6). Thus, the rejection of background events based on the FEI output has an extreme impact on the signal efficiency as can be estimated from the efficiency-over-purity plot shown in Figure 5.7. A further fine-tuning of the signal selection process to enhance the purity of the sample is described in the following sections.

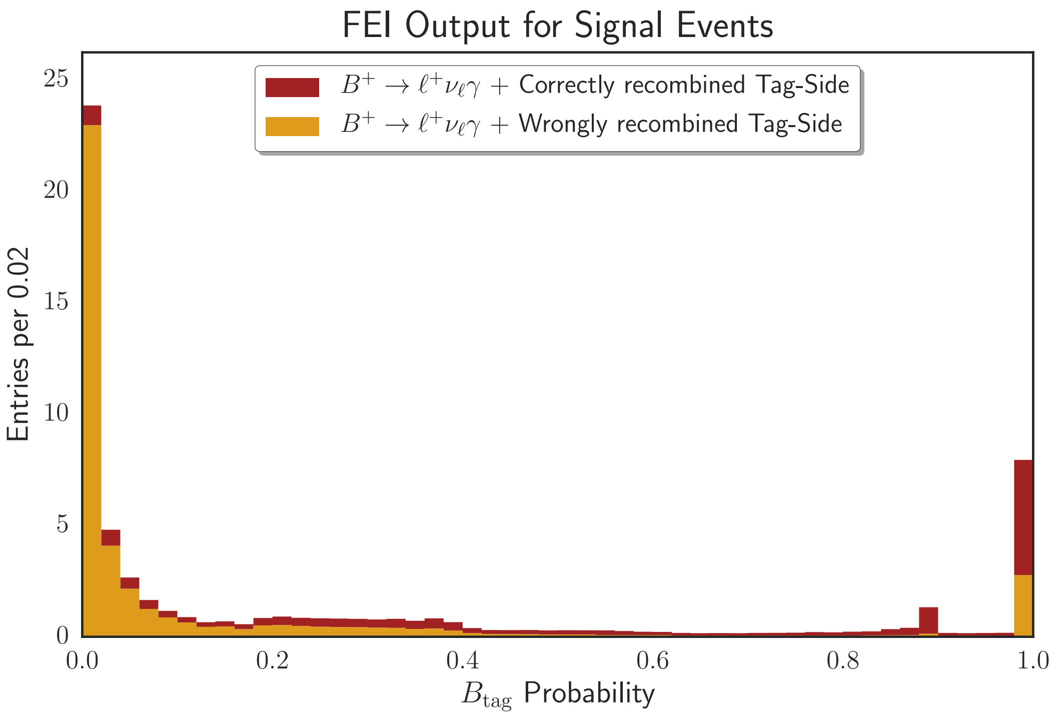


Figure 5.6.: Distribution of the FEI output for signal events with a correctly or wrongly recombined tag-side; shown in the form of a stacked histogram.

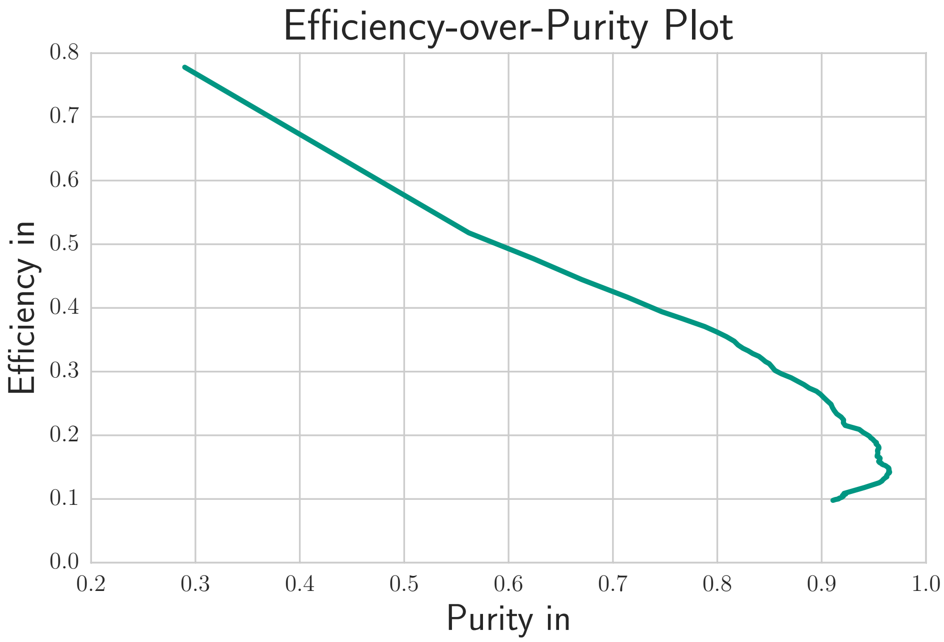


Figure 5.7.: Efficiency-over-purity plot for a cut on the FEI tag-side probability output determined on the scaled MC samples. Due to the signal specific approach, the efficiency and purity displayed here correspond to the total signal selection efficiency.

5.6. The Peaking Background

Signal mimicking decays of the type $B^+ \to X_u\, \ell^+ \nu_\ell$ are particularly hard to discriminate from actual signal events. Especially the cases in which the role of the light meson X_u is taken on by a π^0 or an η dominate the signal region. These light mesons decay into a pair of photons, one of which can be misidentified as the signal photon. The other photon might be assigned to the tag side, rejected as supposed beam background, or it could simply exit the detector beyond its acceptance. Due to its strong resemblance to the signal decay, a total suppression of this background type is impossible. Nonetheless, a veto for these decay channels was employed in an attempt to reject as many false candidates as possible.

For this task all signal candidates are evaluated in a two-step procedure to identify signal photons originating from a π^0 or an η:

> In a **first step** potential π^0 or η candidates are recombined from the signal photon and the photons available on the remaining tag-side.
>
> The **second step** of the procedure evaluates the most suited of those candidates with respect to the signal side lepton for different energy thresholds for the second photon.

For the First Step two discriminators are trained to evaluate the π^0 and η candidates. The FastBDTs which are used for this task, learned to separate true π^0 or η candidates, respectively, from arbitrary combinations of photons which happen to yield an invariant mass close to the nominal value of the respective particles. The variables used as input for the discriminators are typical for the identification of neutral pions:

- The combination's absolute deviation from the nominal particle mass: dM
- Transverse momentum of the two photons and of their combination: $p_{\mathrm{t}_1}, p_{\mathrm{t}_2}, p_{\mathrm{t}}$
- z-component of the two photons and of their combination: p_{z_1}, p_{z_2}, p_z
- Energy of the two photons and of their combination: E_1, E_2 ,E
- Cosine of the angle between the two photons: $\cos(\theta_{12})$
- Cosine of the polar angle for both photon trajectories: $\cos(\theta_1)$, $\cos(\theta_2)$
- Azimuthal angle of both photon trajectories: ϕ_1, ϕ_2
- Energy asymmetry: $(E_1 \times E_2)/(E_1 + E_2)$
- Cluster region and timing of the ECL clusters associated with the two photons
- Number of hits of the ECL clusters associated with the two photons
- E9 over E25 ratio of the ECL clusters associated with the two photons

As the multivariate classifiers are trained solely on signal events and the specific kind of signal mimicking background, this π^0 and η identification is customized for the task at hand. The configuration of the FastBDTs is chosen similar to the one used for the continuum suppression training, setting the forest size to 100, the maximal depth to 3 and the shrinkage to 0.10. The target is the MC truth of the π^0 or η candidate being correctly recombined. The training sample of 10 million events of each of the decays $B^+ \to X_u\, \ell^+ \nu_\ell$ and $B^+ \to \ell^+ \nu_\ell \gamma$ ($\ell = e, \mu$, resulting in a total of 40 million events for each of the two classifier trainings) is randomly split into two subsets for training and validation. The resulting area under the curve of the receiver operator characteristic of 92.6 % (92.8 % for the η identification) speaks for the presented method.

In the Second Step the π^0 and η candidates are prepared using energy thresholds between 100 MeV and 400 MeV for the second photon taken from the tag-side. Based on the mass difference to the nominal π^0 or η mass, respectively, the best candidate for each energy threshold is fed into a further multivariate classifier. The partitioning of the candidates based on the energy of the second photon serves the purpose of reducing the number of arbitrary combinations with invariant masses which are incidentally close to the nominal value of the searched-for particle. Reference [12, sec. 4.3] further elaborates on this problem and motivated the approach presented herein.

In addition to the classifier output of the first stage, information from the remaining lepton and neutrino of the signal-side are used to discriminate between actual signal events and the $B^+ \to X_u\, \ell^+ \nu_\ell$ background in a final classifier. Among this information are the polar and azimuthal angle of the lepton trajectory, the energy of the lepton, as well as its electron and muon PID likelihood. Moreover, the angles between the lepton, the photon and the missing momentum in the reference frame of the $\Upsilon(4S)$ meson are considered. The missing momentum is calculated from the information provided by the FEI exclusive tagging algorithm in a manner similar to the determination of the missing mass illustrated in Section 5.2. Only the directional information about the neutrino is used to avoid a dependence on the missing mass. Further input variables are

- the number of all tracks in the event,
- the number of remaining tracks after the recombination of the $\Upsilon(4S)$,
- and the remaining energy in the ECL not considered for this recombination.

These variables are combined in a gradient-boosted decision tree (FastBDT) which is trained on all considered decay types listed in Table 5.1. One stream of generic and continuum background was used, as well as 10 streams of the rare $B^+ \to X_u\, \ell^+ \nu_\ell$ background. Of the privately produced decay channels 10×10^6 additional independent events were generated for each channel and used for the training. The inclusion of generic and continuum background for the training of this final classifier should allow for a reduction of these background types based on the features of the signal decay. As target for this training served the correctness of the recombination of the signal-side B_{sig} meson which is obtained from the MC truth. Thereby, the focus of the classification task is set on the signal-side B decay and the effect of a wrong tag-side recombination is reduced. The same configuration of the FastBDT as stated above is used.

Table 5.4.: Overview of the signal efficiency and background rejection of the loose cuts on the outputs of the **signal selection** classifier and **veto** classifier applied in this order. All listed values are in percent.

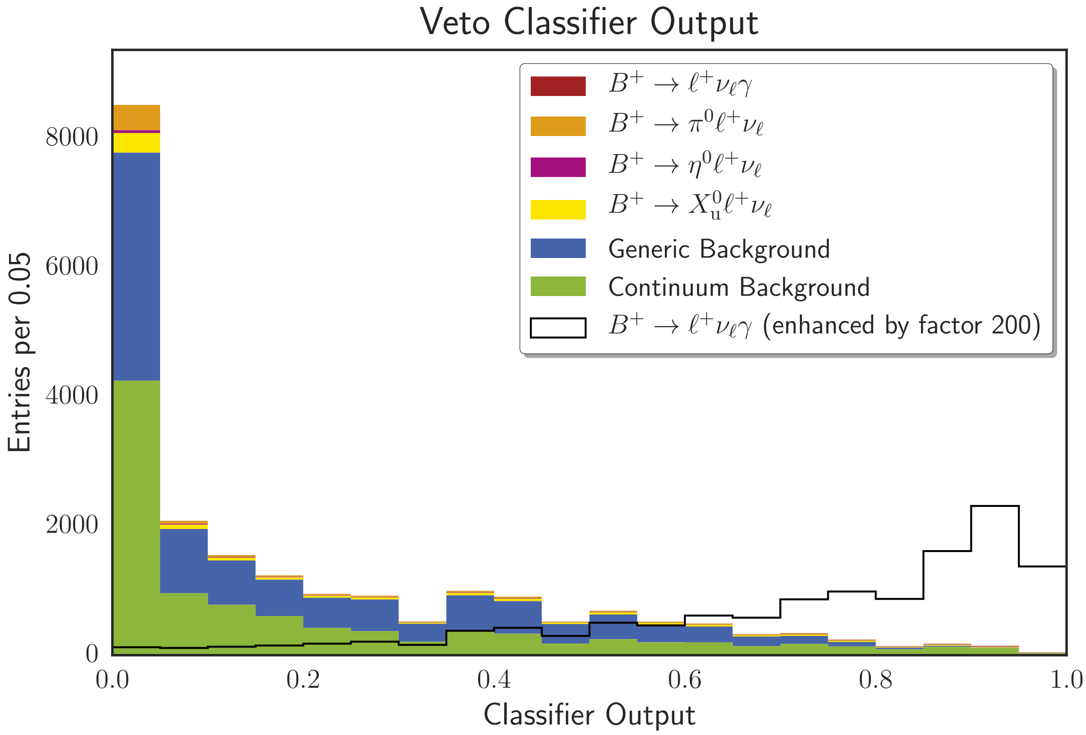
Cut on	Signal Efficiency	Background Rejection for Generic Bkg	 Continuum Bkg	 Peaking Bkg
Signal Selection	96.74	66.82	87.52	58.30
Veto	99.57	3.38	1.01	6.26
Combined	96.32	67.94	87.65	60.91

To utilize the separation power bundled in the output of the continuum suppression classifier described in Section 5.4 and the knowledge about the tag-side available in the form of the FEI output, a second version of the final classifier is trained. The inclusion of these variables should help to separate the generic and continuum background and might enable one to retrieve the signal candidates with wrongly recombined tag-sides.

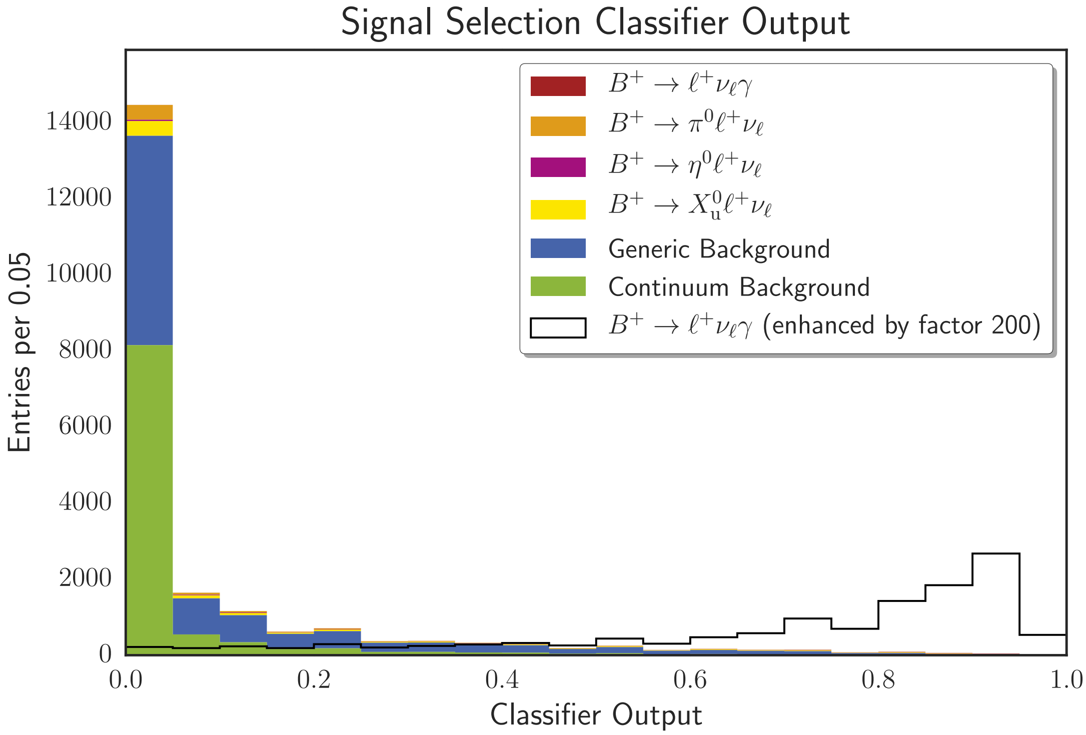
The validation of the training on an independent sample yields areas under the receiver operator characteristic curve of 98.6 % and 96.3 % for the training with and without the additional information of the continuum suppression and FEI output, respectively. Hence, the inclusion of the output of the FEI and continuum suppression classifiers improves the performance of the final classifier significantly. The classifier obtained from the training including the FEI and continuum suppression information shall be referred to as the *signal selection classifier* from here on, whereas the discriminator which does not include the additional information is referred to as the *veto classifier.* An overview of the training validation performed by TMVA for the two classifiers can be found in the appendix of this text in form of the over-training check plot, the cut efficiency evaluation and the ROC curve displayed in the Figures A.2 and A.3 for the veto classifier and the signal selection classifier, respectively.

Applying these classifiers on the scaled MC samples described in Section 5.1 results in distributions of their outputs as shown in Figure 5.8 for both the veto and signal selection classifier. The influence of the additional input variables is clearly visible in Figure 5.8b, especially for the continuum background. In both cases a peak for the signal mimicking background types at zero can be observed. The presented approach also achieves a distribution for signal events which tends to output values of 1.0 for both trained variants. This is an improvement compared to the previous study of the decay $B^+ \to \ell^+ \nu_\ell \gamma$ as presented in Reference [12, sec. 4.3] where a similarly strategy resulted in a uniform distribution of the network output.

To discard the obvious background a cut on both classifier outputs is applied, rejecting events with outputs below 0.1. The signal efficiency and background rejection of both cuts are listed in Table 5.4. This allows a clearer look at the distribution of the background types in the relevant region of the outputs, which is shown in Figure 5.9. The remaining events from the signal mimicking channels $B^+ \to X_u \ell^+ \nu_\ell$ are almost uniformly distributed. A large fraction of them, comparable to the number of signal events, still remains as irreducible background in the bins in which the signal distribution peaks.

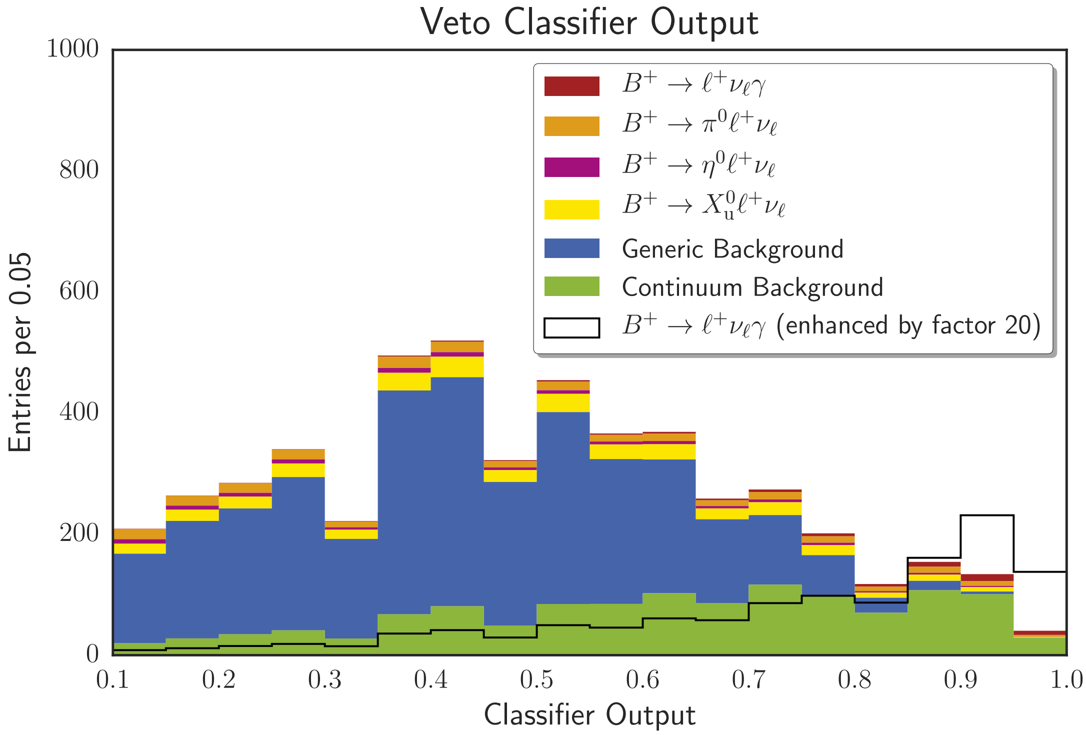


(a) Veto classifier

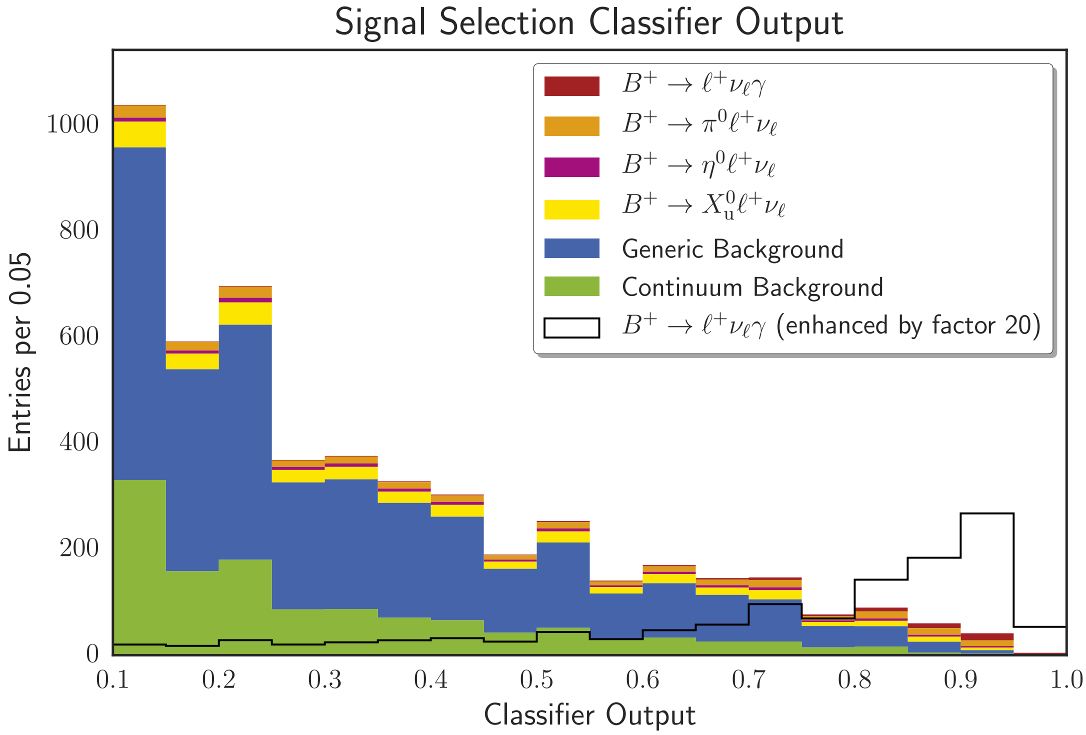


(b) Signal Selection Classifier

Figure 5.8.: Stacked histogram of the outputs of the veto classifier 5.8a and the signal selection classifier 5.8b obtained from the scaled MC sample described in Section 5.1. The shape of the signal distribution enhanced by a factor of 200 is displayed as black line.



(a) Veto classifier



(b) Signal Selection Classifier

Figure 5.9.: Distribution of the veto 5.9a and signal selection classifier 5.9b outputs after the simultaneous application of the loose cuts on both quantities to discard obvious background. The signal shape (enhanced by a factor of 20) is shown in addition to the stacked histogram.

5.7. The Signal Selection

Based on the results of the two final classifiers for the signal selection presented in the previous section, different approaches for the final signal selection are possible. One could

1. pursue a strategy similar to the one presented in Reference [12] and aim for a pseudo 2D fit by fitting the missing mass in bins of the signal selection or veto classifier output, respectively, after applying perpendicular cuts to discard background, or
2. utilize the improved final classifiers by performing cuts on their output and subsequently determine the signal yield with an one-dimensional fit on the missing mass.

To compare the results accomplished herein to the ones of Reference [12], a short overview of the results of the **first option** is given in the following Section 5.7.1. As the outputs of both of the improved final classifiers do peak at 1.0 for signal candidates, the fit in bins of these quantities might not be necessary anymore. Hence, the **second option** is introduced in Section 5.7.2, in which a cut on the quantities is performed to reject background. This results in a significant improvement of the signal yield.

As the FEI version 3.0 "Cyberman", which introduced the best candidate selection approach for the selection of the candidates, was finalized only shortly before the due date of this work, an optimization of the cuts applied in both approaches was not possible. Moreover, a validation of both approaches regarding the optimization of their final signal significance could not be conducted, due to the same reason.

The results for both approaches presented in the following are obtained from the same scaled MC sample described in Section 5.1. This sample is mostly independent from the samples used for the training of the classifiers, with the exception of a subset of the streams of public MC[2]. The decay channels which are available with higher statistics in the form of privately produced MC samples are removed from the public MC samples to maintain the correct branching fractions given by the scaling factors described in Section 5.1 for the scaled sample.

5.7.1. The Preparation for a Fit in Bins of the Classifier Output

As this approach does not allow for a tight cut on the output of neither the signal selection classifier, nor the veto classifier, perpendicular cuts on other variables are necessary to reduce the background. The variables used for this selection and the cuts imposed on them are listed in Table 5.5. Similar to the approach of [12], a cut on the probability output of the exclusive tagging algorithm is applied. Due to the high fraction of signal events with a wrongly recombined tag-side, this cut has a severe effect on the signal efficiency.

A loose cut on the veto classifier output is applied to reject obvious $B^+ \to X_u \ell^+ \nu_\ell$ background. To discard continuum background, a tight cut on the continuum suppression

---

[2] As stated in the previous sections was a subset of 1 stream of generic MC used for the FEI training; 1 stream of continuum MC was for the continuum suppression classifier training; and 1 stream of each generic and continuum MC as well as 10 streams of $B^+ \to X_u \ell^+ \nu_\ell$ MC were used for the training of the final classifiers.

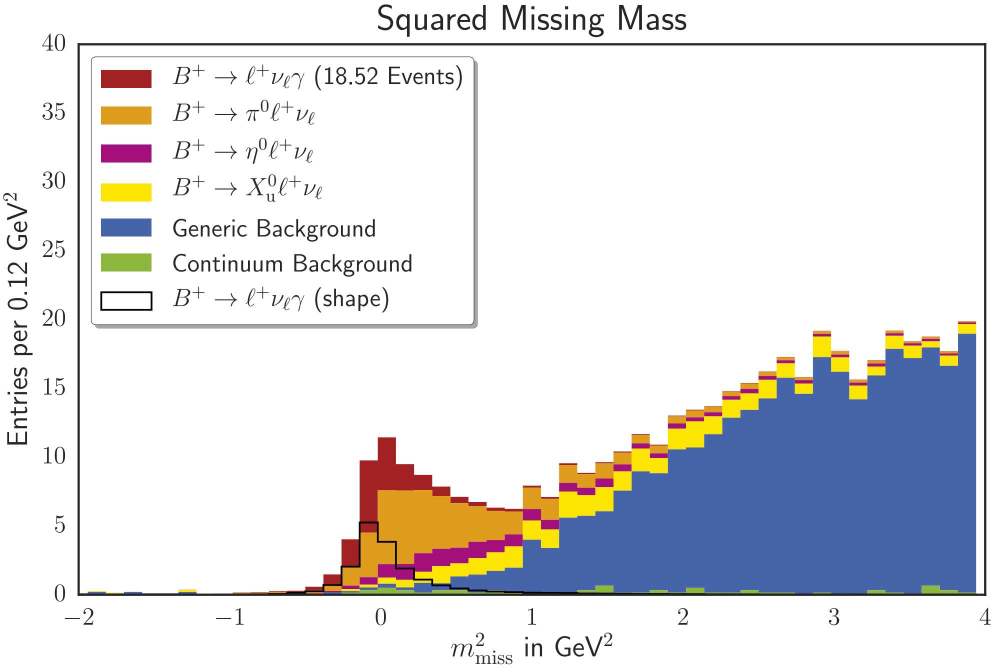
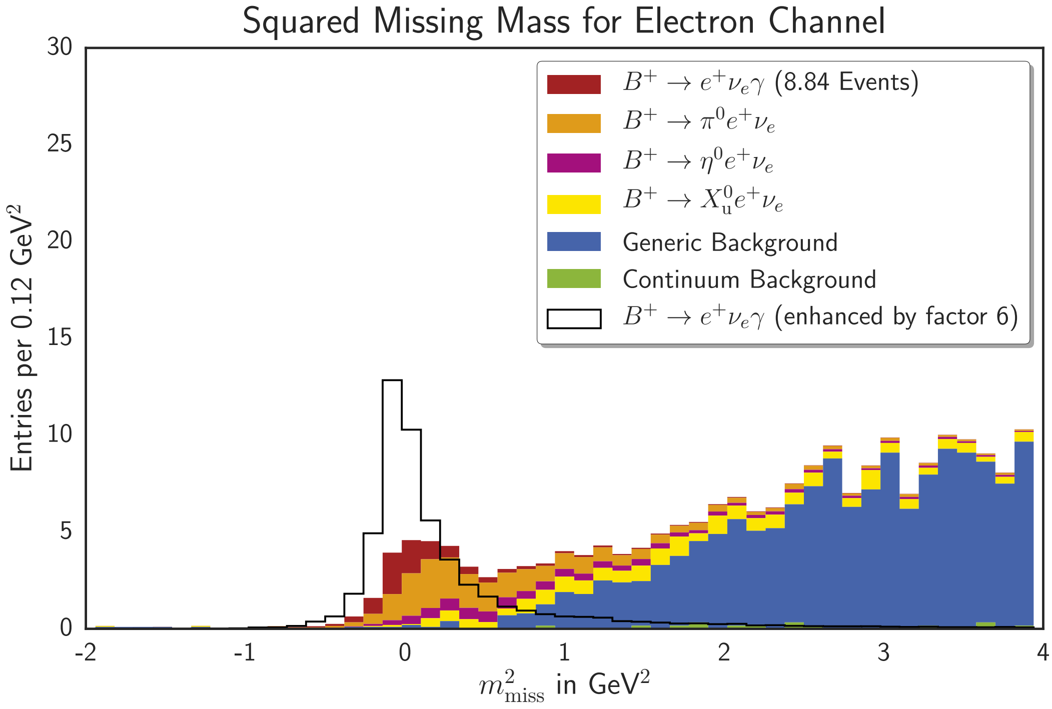


Figure 5.10.: Stacked histogram of the distribution of the squared missing mass m^2_{miss}. Additionally, the shape of the signal distribution is displayed as black line.

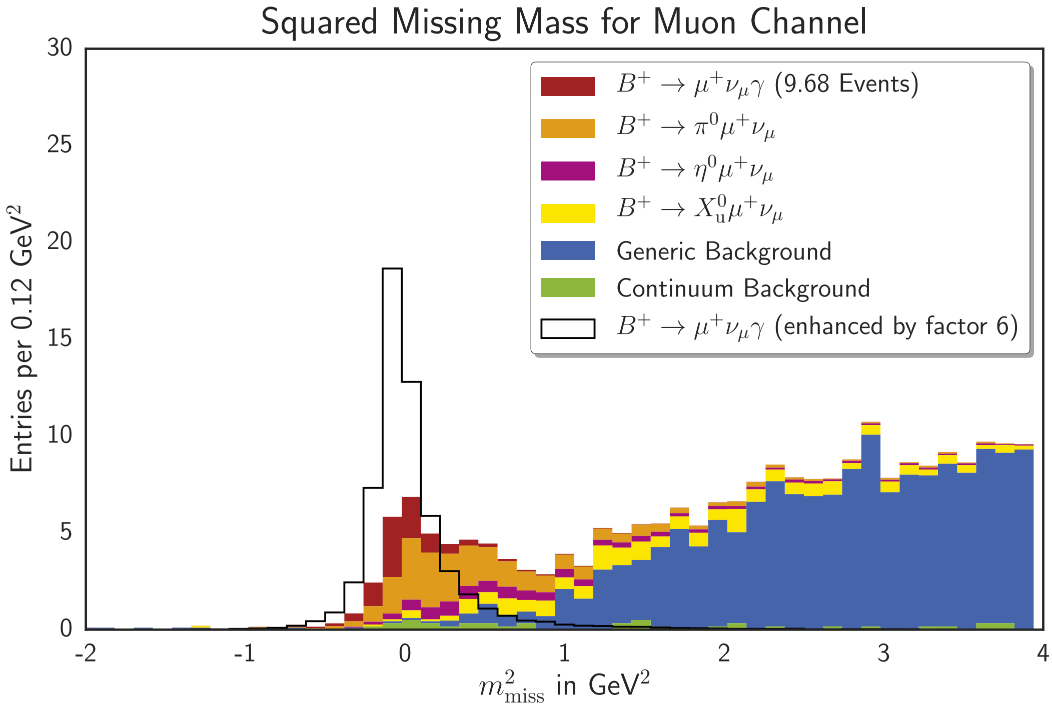
output is performed. Moreover, the aforementioned cut on the number of remaining tracks in the event after the recombination of the $\Upsilon(4S)$ meson is applied, despite the fact that this cut does affect the signal efficiency. This cut and the cuts on the remaining variables listed in Table 5.5 are also motivated by the approach presented in Reference [12].

The resulting squared missing mass distribution for the combined decay $B^+ \to \ell^+\nu_\ell\gamma$ ($\ell = e, \mu$) is shown in Figure 5.10 as stacked histogram obtained from the scaled MC sample. Separate histograms for the electron channel and muon channel are displayed in the Figures 5.11a and 5.11b, respectively. The binning of the displayed histograms is chosen to fit the one used in the References [12] and [23], to ensure a comparability of the results obtained from the MC samples. For the same reason, the factors by which the signal shape is enhanced in the Figures 5.11a and 5.11b and the branching fraction used for the calculation of the scaling factors in Section 5.1 are chosen as they are.

The signal yields obtained from this approach are stated in Table 5.7 together with the corresponding values from Reference [12]. The latter are yields obtained from a fit on MC for which the uncertainties are neglected here, whereas the values from the approach presented herein are the number of remaining MC events after the application of the above described signal selection process. Hence, the small improvement of the presented approach compared to the values of Reference [12, page 67], which is also listed in Table 5.7, has to be taken with a grain of salt. Due to the change in the distribution of the output of the signal selection classifier illustrated in Section 5.6 compared to the one of Reference [12], the available fitting algorithm from named Reference could not be employed. A more thorough evaluation of the results is given in the following Chapter 6.



(a) Electron Channel



(b) Muon Channel

Figure 5.11.: The two stacked histograms show the distribution of the squared missing mass separately for the electron channel (top) and the muon channel (bottom). Additionally, the shape of the distribution for signal events enhanced by a factor of 6 which corresponds to a branching fraction of 30×10^{-6} is plotted as black line.

Table 5.5.: Signal efficiencies and background rejections, both in percent, for all **cuts** imposed to prepare the MC sample for a fit in bins of the signal selection classifier output. The listed values do not include the pre-selection efficiency and the tag-side selection efficiency, which are stated in the Sections 5.3 and 5.5, respectively.

Variable	Cut	Signal Eff.	Background Rejection for		
			Gen. Bkg	Con. Bkg	Peaking Bkg
FEI Output	>0.1	44.70	75.32	95.69	67.10
Continuum Suppression	>0.8	93.95	27.76	80.92	12.39
Remaining Tracks	=0	79.53	43.99	51.02	27.23
Veto Classifier	>0.3	93.78	64.89	70.42	67.82
$\cos(\theta_{\nu,\gamma})$	>-0.9	99.60	0.92	7.04	1.02
Extra Energy	<4.5 GeV	99.32	1.31	3.03	0.74
E9 over E29 ratio	>0.9	96.87	6.29	25.00	3.99
Combined		30.01	96.79	99.92	93.63

Table 5.6.: **Total signal efficiencies** for this selection approach; including pre-selection and tag-side selection.

	Electron Channel $B^+ \to e^+\nu_e\gamma$	Muon Channel $B^+ \to \mu^+\nu_\mu\gamma$	Combined $B^+ \to \ell^+\nu_\ell\gamma$
Total Efficiency	0.22 %	0.24 %	0.23 %

Table 5.7.: **Signal yields** of the signal selection for the preparation of a fit in bins of the final classifier output as described in Section 5.7.1. The values are compared to the MC fit results of the nominal analysis (without uncertainties) presented in Reference [12]. Additionally, the number of background events observed in the signal region are stated for this approach.

	Electron Channel $B^+ \to e^+\nu_e\gamma$	Muon Channel $B^+ \to \mu^+\nu_\mu\gamma$	Combined $B^+ \to \ell^+\nu_\ell\gamma$
Signal Yield for this Approach	8.8	9.7	18.5
Background Events	203	220	423
Signal yield of Reference [12]	8.0	8.7	16.5
Improvement	10 %	11 %	12 %

5.7.2. The Utilization of the Improved Classifiers

As the final classifiers presented in Section 5.6 deliver output distributions which peak for signal events, the idea of a fit in bins of these outputs might be abandoned. A cut on the classifier outputs could yield better results, and thus this approach will be pursued in the following. Since the signal selection classifier already includes the information from the FEI output, a cut on the FEI output is not used. Thus, the signal events which are combined with wrongly recombined tag-side B_{tag} mesons are not directly discarded, as it was the case in the approach presented above. Also the cut on the number of remaining tracks after the recombination of the $\Upsilon(4S)$ meson is not imposed in this approach, thereby allowing for one remaining track (a cut applied during the tag-side selection already restricted the number of remaining charged tracks to 1).

Instead, tighter cuts on the signal selection and the veto classifier output are applied in addition to further cuts, all of which are listed in Table 5.8 together with the corresponding efficiencies. To further discard continuum background a very tight cut on the continuum suppression classifier output is imposed by demanding a value larger than 0.99.

The signal efficiency achieved with this approach exceeds the one of the approach described in the previous Section 5.7.1 significantly, whilst maintaining a similar background rejection. The total signal efficiencies which include the previous steps of the pre-selection and the tag-side selection described in the Sections 5.3 and 5.5, respectively, are shown in Table 5.9. The signal yields obtained via this approach are listed in Table 5.10, where they again are compared to the MC values of Reference [12]. Despite neglecting the uncertainties of the latter, the improvement of the approach presented herein is obvious. The signal yield increases by a factor of about 2.

The corresponding histograms of the squared missing mass obtained from the same independent MC sample as the one used in the previous section are shown in Figure 5.12 for the combined analysis, and in the Figures 5.13a and 5.13b separately for the electron and muon channel, respectively. Again, the bin size and the branching fractions are chosen to obtain results comparable to the ones of Reference [12].

The signal region is inevitably dominated by the contribution of the signal mimicking background; the channel $B^+ \to \pi^0\, \ell^+ \nu_\ell$ being the most important one. This background type still remains irreducible with this approach. A further evaluation of these results is provided in Chapter 6.

Table 5.8.: Signal efficiencies and background rejections, both in percent, for all **cuts** imposed in the scope of the approach utilizing the improved classifiers. The listed values do not include the pre-selection efficiency and the tag-side selection efficiency, which are stated in the Sections 5.3 and 5.5, respectively.

Variable	Cut	Signal Eff.	Background Rejection for Gen. Bkg	 Con. Bkg	 Peaking Bkg
Signal Selection Class.	>0.4	85.08	90.27	96.85	80.74
Veto Classifier	>0.4	97.80	8.28	1.12	9.14
Continuum Suppression	>0.99	68.36	65.31	95.81	40.81
$\cos(\theta_{\nu,\gamma})$	>-0.9	99.74	0.43	5.13	0.71
E9 over E29 ratio	>0.9	97.42	7.02	27.03	3.55
Combined		55.27	97.13	99.91	90.08

Table 5.9.: **Total signal efficiencies** for this selection approach; including pre-selection and tag-side selection.

	Electron Channel $B^+ \to e^+\nu_e\gamma$	Muon Channel $B^+ \to \mu^+\nu_\mu\gamma$	Combined $B^+ \to \ell^+\nu_\ell\gamma$
Total Efficiency	0.41 %	0.45 %	0.43 %

Table 5.10.: **Signal yields** of the signal selection utilizing the improved classifiers as described in Section 5.7.2. The values are compared to the MC fit results of the nominal analysis (without uncertainties) presented in Reference [12]. Additionally, the number of background events observed in the signal region are stated for this approach.

	Electron Channel $B^+ \to e^+\nu_e\gamma$	Muon Channel $B^+ \to \mu^+\nu_\mu\gamma$	Combined $B^+ \to \ell^+\nu_\ell\gamma$
Signal Yield for this Approach	16.4	17.7	34.1
Background Events	227	222	449
Signal yield of Reference [12]	8.0	8.7	16.5
Improvement	105 %	103 %	107 %

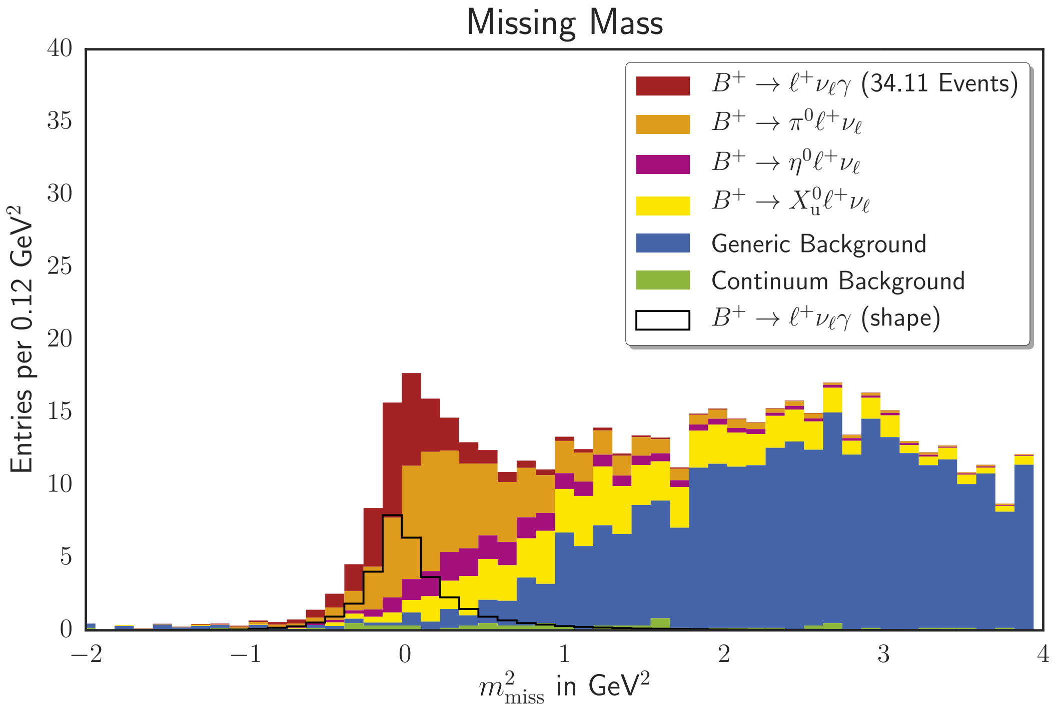
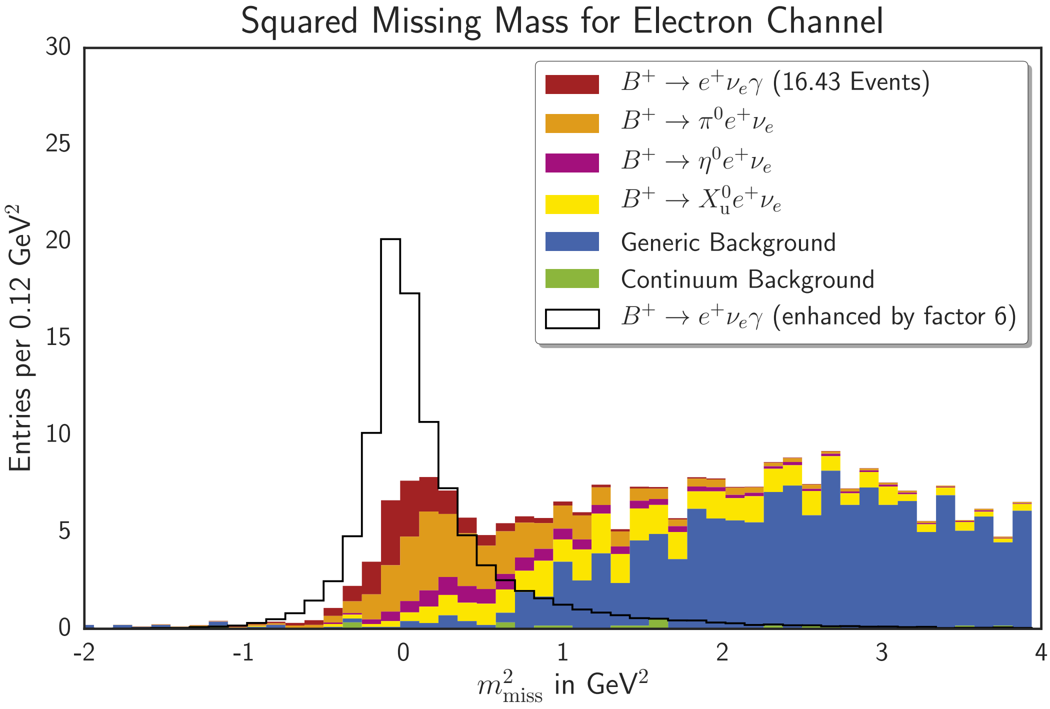
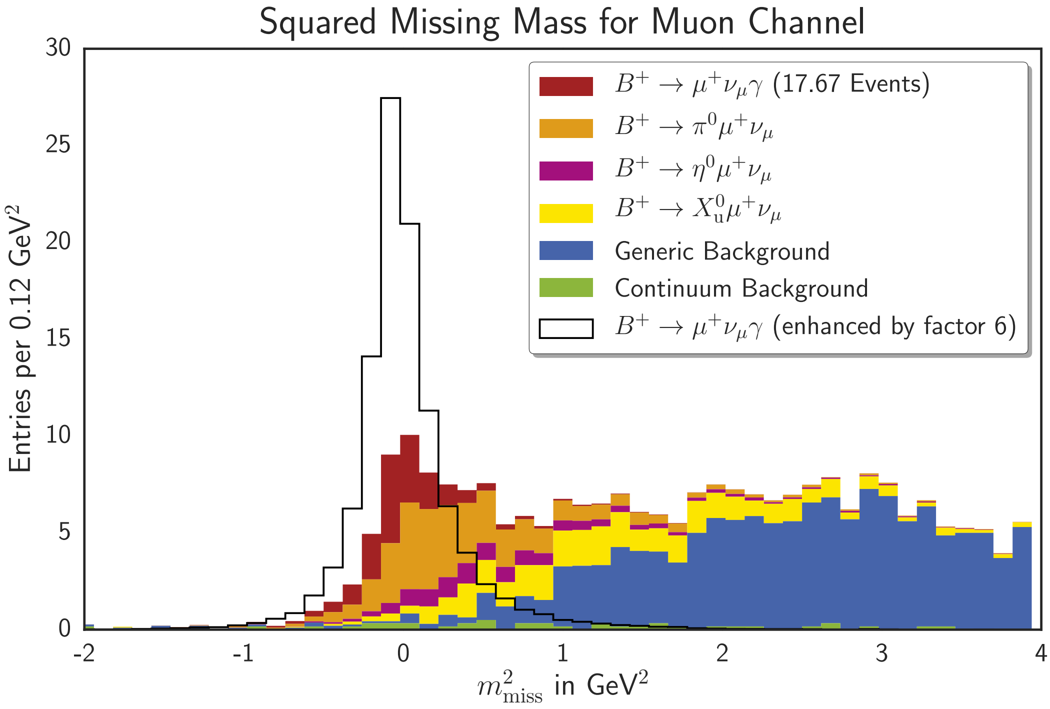


Figure 5.12.: Stacked histogram of the distribution of the squared missing mass m^2_{miss}. Additionally, the shape of the signal distribution is displayed as black line.



(a) Electron Channel



(b) Muon Channel

Figure 5.13.: The two stacked histograms show the distribution of the squared missing mass separately for the electron channel (top) and the muon channel (bottom). Additionally, the shape of the distribution for the signal events enhanced by a factor of 6 which corresponds to a branching fraction of 30×10^{-6} is plotted as black line.

6. Evaluation of the Results

In this Monte Carlo study of the rare decay $B^+ \to \ell^+ \nu_\ell \gamma$ the FEI tagging algorithm was employed in a signal specific mode to provide tag-side candidates. The version of this algorithm, which is presented in this text for the first time, utilizes a new approach for the selection of these candidates based on a best candidate selection. This approach increases the efficiency with which tag candidates are provided, as it principally favors the survival of at least one candidate. The correct method to handle the candidates provided by this algorithm has yet to be studied, as the enhancement in efficiency is mostly due to wrongly recombined tag-side. A first conservative approach via signal decay specific perpendicular cuts for the tag-side selection was presented herein.

The signal selection classifier introduced in Section 5.6, which combines signal-side variables and the outputs of the FEI and the continuum suppression, provides good background rejection power, as can be seen in the Figures 5.8b and 5.9b. This discriminator can either be used

1. as second variable in a 2D fit together with the squared missing mass, or
2. as variable to cut on for a 1D fit in the squared missing mass distribution.

For both approaches a signal selection was conducted in Section 5.7. In option 1 no cut on the signal selection classifier output was imposed, whereas in option 2 this quantity was used to discard a large amount of background.

Instead, a cut on the FEI output was applied in **option 1** to reduce the amount of combinatorial background, thereby rejecting a large fraction of signal events with wrongly recombined tag-sides. This results in a total signal selection efficiency of **0.23 %** for the combined decay $B^+ \to \ell^+ \nu_\ell \gamma$ which yields an amount of signal events comparable to the signal selection of Reference [12].

Refraining from the perpendicular cut on the FEI output and instead relying on the discrimination power of the signal selection classifier — as it was done for **option 2** — partially retrieves the signal events with wrongly recombined tag-sides. This enhances the signal selection efficiency to **0.41 %** whilst maintaining a similar background rejection. Plotting the missing mass distribution for this approach for signal events only (Figure 6.1), differentiating between the ones with correctly and with wrongly recombined tag-side, shows the effect on the distribution of this quantity when the wrongly recombined tag-sides are included. The width of the peak increases only slightly due to the events with wrongly recombined tag-sides — an effect which has to be studied further to be quantified. Among the errors occurring during the recombination of the tag-side are

- tracks and ECL clusters originating from beam-background or wrong track reconstruction,

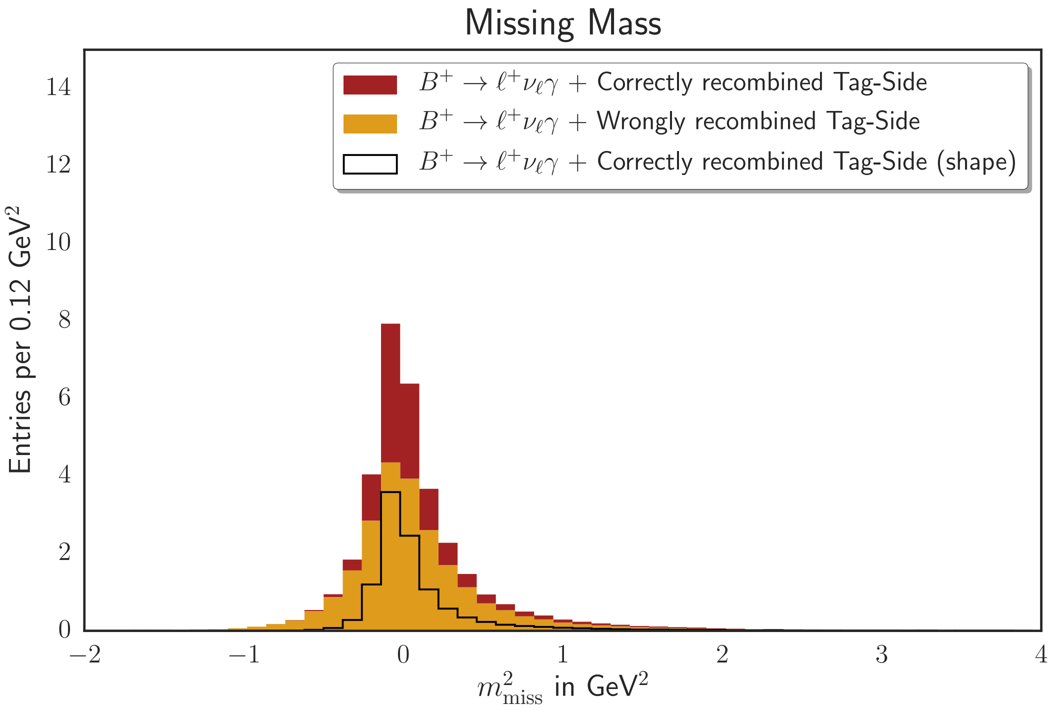


Figure 6.1.: Squared missing mass distribution for signal events of the decay $B^+ \to \ell^+\nu_\ell\gamma$ which are combined with either correctly or wrongly recombined tag-sides. The stacked histogram is generated from signal events from the scaled MC samples which survived the signal selection process presented in Section 5.7.2. Only a slight increase in width of the peak due to the events with wrongly recombined tag-sides can be observed.

- recombination of a wrong intermediate particles, and
- misidentified final state particles.

The additional events included in the approach of option 2 effectively represent inclusive tag-side candidates which are included into the exclusive tagging algorithm by means of the best candidate selection approach. Due to the signal specific manner in which the FEI is trained and applied, these candidates underwent a selection process similar to the one which would be imposed in the scope of inclusive tagging. For these events, the FEI algorithm was not able to resolve the intermediate resonances that occurred during the decay of the tag sides correctly. Nonetheless, these events contribute to the analysis at hand. A further optimization of the way this tagging algorithm is utilized is necessary to fully evaluate the potential of this approach. Independent from this possibility, an optimization of the cuts imposed as part of the signal selection presented in this thesis should lead to a higher selection efficiency. For this a figure of merit as discussed in Reference [60] can be employed.

7. Conclusion

The FEI Tagging Algorithm, as presented herein, shows the potential to combine the inclusive and exclusive tagging approaches. The best candidate selection approach, which is utilized during the hierarchical recombination process of the companion B meson, allows for the survival of events for which the intermediate resonances of the B decay cannot be fully resolved. The validity of these incorrectly recombined tag-sides is ensured by the signal specific pre- and post-processing, which is similar to a selection imposed in the scope of inclusive B-tagging. Further optimizations of the algorithm are necessary to fully exploit the advantages of both tagging methods. A proper evaluation of the errors occurring during the exclusive tag-side recombination is necessary to understand the possibilities of this combined approach. Special cases have to be considered — for instance the lack of a charged track can lead to unphysical results in which a charged B meson is recombined instead of a neutral B meson, or vice versa. Furthermore, the amount of background surviving the selection process of the algorithm due to the best candidate selection approach needs to be handled to achieve an acceptable signal purity.

In the scope of this work an unoptimized approach was applied, resulting in an enhancement of the signal selection efficiency by a factor of about 2. Rejecting the *inclusively* tagged candidates posed by the events with a wrongly recombined tag-side leads to a reduced signal selection efficiency comparable to the one obtained in the study of the decay conducted by *A. Heller et al.* [12].

For the Analysis of the Decay $B^+ \to \ell^+ \nu_\ell \gamma$ the signal selection presented herein provides a good basis. Further steps have to be undertaken to reach the stage of a branching fraction measurement on Belle data with the help of the FEI algorithm and the Belle II software. First of all the cuts used in the signal selection should be optimized by maximizing a figure of merit. This may give a further improvement of the selection efficiency. A fitting process has to be established to obtain the signal yield via a 1D fit on the squared missing mass distribution or a 2D fit including the final signal selection classifier output. The signal significances of both fit methods have to be evaluated to chose the most suitable approach.

An important part of the preparation for the application of the FEI algorithm — and therefore the signal selection approach illustrated herein — on actual data is the evaluation of the performance difference between MC and data. Notwithstanding the fact that this has not yet been evaluated for the FEI algorithm in general, the signal specific application of the algorithm presented herein calls for a signal specific evaluation of this effect. Correction factors for a possible data – Monte Carlo discrepancy of the FEI algorithm have to be determined on an independent decay channel. Possible candidates are the decays $B^+ \to D^0\,\pi^+$, which offers a very pure sample, or $B^+ \to D^{*0}\,(\to D^0\,\gamma)\,\ell^+$ which has the advantage of high statistics.

These tasks exceed the scope of this Master's Thesis and are work for the future. **Building upon the FEI algorithm and the improved signal selection process, may yet lead to a measurement of the branching fraction of the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ and thus the first inverse moment of the light-cone distribution amplitude λ_B with the existing Belle data set in the Belle II software framework.** For a study of the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ based on Belle II data, further improvements due to the advanced tracking algorithms, the new hardware and of course the increased luminosity are expected.

Bibliography

[1] **ATLAS** Collaboration, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.* **B716** (2012) 1–29, `arXiv:1207.7214 [hep-ex]`.

[2] **CMS** Collaboration, S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.* **B716** (2012) 30–61, `arXiv:1207.7235 [hep-ex]`.

[3] **Super-Kamiokande** Collaboration, Y. Fukuda *et al.*, “Evidence for Oscillation of Atmospheric Neutrinos,” *Phys. Rev. Lett.* **81** (1998) 1562–1567, `arXiv:hep-ex/9807003 [hep-ex]`.

[4] **SNO** Collaboration, Q. R. Ahmad *et al.*, “Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.* **87** (2001) 071301, `arXiv:nucl-ex/0106015 [nucl-ex]`.

[5] **SNO** Collaboration, Q. R. Ahmad *et al.*, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.* **89** (2002) 011301, `arXiv:nucl-ex/0204008 [nucl-ex]`.

[6] A. D. Sakharov, “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe,” *Soviet Physics Uspekhi* **34** no. 5, (1991) 392. `http://stacks.iop.org/0038-5670/34/i=5/a=A08`.

[7] **Belle, BaBar** Collaboration, A. J. Bevan *et al.*, “The Physics of the B Factories,” *Eur. Phys. J.* **C74** (2014) 3026, `arXiv:1406.6311 [hep-ex]`.

[8] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction,” *Prog. Theor. Phys.* **49** (1973) 652–657.

[9] **Particle Data Group** Collaboration, K. A. Olive *et al.*, “Review of Particle Physics,” *Chin. Phys.* **C38** (2014) 090001. Section 12. The CKM Quark-Mixing Matrix.

[10] T. Feldmann, “Non-Leptonic Heavy Meson Decays - Theory Status,” in *12th Conference on Flavor Physics and CP Violation (FPCP 2014) Marseille, France, May 26-30, 2014*. 2014. `arXiv:1408.0300 [hep-ph]`. `https://inspirehep.net/record/1309602/files/arXiv:1408.0300.pdf`.

[11] M. Beneke, G. Buchalla, M. Neubert, and C. Sachrajda, “QCD factorization for exclusive non-leptonic B-meson decays: general arguments and the case of

heavy–light final states," *Nuclear Physics B* **591** no. 1–2, (2000) 313 – 418. `http://www.sciencedirect.com/science/article/pii/S0550321300005599`.

[12] 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. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48743`.

[13] M. Beneke and J. Rohrwild, "B meson distribution amplitude from $B \to \gamma \ell \nu$," *Eur. Phys. J.* **C71** (2011) 1818, `arXiv:1110.3228 [hep-ph]`.

[14] C. Pulvermacher, T. Keck, M. Feindt, M. Heck, and T. Kuhr, "An automated framework for hierarchical reconstruction of B mesons at the Belle II experiment," *Journal of Physics: Conference Series (ACAT 2014)* **608** no. 1, (2015) 012048.

[15] T. Keck, "The Full Event Interpretation for Belle II," Master Thesis, Karlsruhe Institut of Technology (KIT), 2014. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48602`.

[16] C. Pulvermacher, *Analysis Software and Full Event Interpretation for the Belle II Experiment.* PhD Thesis, Karlsruhe Institute of Technology (KIT), 2015. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48741`.

[17] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander, and A. Zupanc, "A Hierarchical NeuroBayes-based Algorithm for Full Reconstruction of B Mesons at B Factories," *Nucl. Instrum. Meth.* **A654** (2011) 432–440, `arXiv:1102.3876 [hep-ex]`.

[18] V. M. Braun and I. B. Filyanov, "Conformal invariance and pion wave functions of nonleading twist," *Zeitschrift für Physik C Particles and Fields* **48** no. 2, (1990) 239–247. `http://dx.doi.org/10.1007/BF01554472`.

[19] Y. Grossman, M. König, and M. Neubert, "Exclusive Radiative Decays of W and Z Bosons in QCD Factorization," *JHEP* **04** (2015) 101, `arXiv:1501.06569 [hep-ph]`.

[20] **BaBar** Collaboration, B. Aubert *et al.*, "Search for the radiative leptonic decay $B^+ \to \gamma \ell^+ \nu_l$," `arXiv:0704.1478 [hep-ex]`. `http://arxiv.org/abs/0704.1478`.

[21] **BaBar** Collaboration, B. Aubert *et al.*, "Model-independent search for the decay $B^+ \to l^+ \nu_l \gamma$," *Phys. Rev. D* **80** (Dec, 2009) 111105. `http://link.aps.org/doi/10.1103/PhysRevD.80.111105`.

[22] G. Bell and V. Pilipp, "$B^- \to \pi^- \pi^0 / \rho^- \rho^0$ to NNLO in QCD factorization," *Phys. Rev.* **D80** (2009) 054024, `arXiv:0907.1016 [hep-ph]`.

[23] **Belle** Collaboration, A. Heller *et al.*, "Search for $B^+ \to \ell^+ \nu_\ell \gamma$ decays with hadronic tagging using the full Belle data sample," *Phys. Rev. D* **91** (Jun, 2015) 112009, `arXiv:1504.05831 [hep-ex]`. `http://link.aps.org/doi/10.1103/PhysRevD.91.112009`.

[24] K. Nishimura, “"Number of B events in Hadron B(J)".” `http://belle.kek.jp/secured/nbb/nbb.html`. Belle internal, last accessed April 27th 2016.

[25] **Particle Data Group** Collaboration, K. A. Olive *et al.*, “Review of Particle Physics,” *Chin. Phys.* **C38** (2014) 090001.

[26] K. Akai, H. Koiso, and S. Okada, “KEK Press Release (March 2nd 2016): "First turns and successful storage of beams in the SuperKEKB electron and positron rings".” `https://www.kek.jp/en/NewsRoom/Release/20160302163000/`. Last accessed March 29th 2016.

[27] A. Abashian *et al.*, “The Belle detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **479** no. 1, (2002) 117 – 232. `http://www.sciencedirect.com/science/article/pii/S0168900201020137`. Detectors for Asymmetric B-factories.

[28] **Belle II** Collaboration, T. Abe *et al.*, “Belle II Technical Design Report,” `arXiv:1011.0352 [physics.ins-det]`.

[29] C. Schwanda, “SuperKEKB machine and Belle II detector status,” *Nuclear Physics B - Proceedings Supplements* **209** no. 1, (2010) 70 – 72. `http://www.sciencedirect.com/science/article/pii/S0920563210004482`. Proceedings of the Third Workshop on Theory, Phenomenology and Experiments in Heavy Flavour Physics.

[30] Krib, “Schematic layout of the KEKB particle accelerator at KEK in Tsukuba (Japan),” Aug, 2013. `https://commons.wikimedia.org/wiki/File:KEKB.png`. Last accessed May 19th 2016.

[31] C. Bernardini, G. F. Corazza, G. Di Giugno, G. Ghigo, J. Haissinski, P. Marin, R. Querzoli, and B. Touschek, “Lifetime and Beam Size in a Storage Ring,” *Phys. Rev. Lett.* **10** (May, 1963) 407–409.

[32] L. Piilonen, “B-KLM Summary Talk.” 11th B2GM, Mar, 2012. `http://kds.kek.jp/contributionDisplay.py?sessionId=29&contribId=17&confId=8895`. Belle II internal Talk.

[33] V. Aulchenko *et al.*, “Electromagnetic calorimeter for Belle II,” *Journal of Physics: Conference Series* **587** no. 1, (2015) 012045. `http://stacks.iop.org/1742-6596/587/i=1/a=012045`.

[34] O. Frost, “A Local Tracking Algorithm for the Central Drift Chamber of Belle II,” Diploma Thesis, Karlsruhe Institute of Technology (KIT), 2013. `http://ekp-invenio.physik.uni-karlsruhe.de/record/48172`.

[35] N. Braun, “Momentum Estimation of Slow Pions and Improvements on the Track Finding in the Central Drift Chamber for the Belle II Experiment,” Master Thesis, Karlsruhe Institute of Technology (KIT), Karlsruhe, 2015. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48740`.

[36] S. Mineo, R. Itoh, N. Katayama, and S. Lee, “Distributed parallel processing analysis framework for Belle II and Hyper Suprime-Cam,” *PoS (ACAT2010)* **026** (2010) . `https://inspirehep.net/record/924965/files/ACAT2010_026.pdf`.

[37] “Boost C++ Libraries.” `http://www.boost.org`.

[38] D. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[39] S. Agostinelli *et al.*, “GEANT4: A Simulation toolkit,” *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[40] R. Brun and F. Rademakers, “ROOT – an object oriented data analysis framework,” *Nucl. Instrum. Meth.* **A389** no. 1, (1997) 81–86.

[41] “Google C++ Testing Framework.” `https://code.google.com/p/googletest/`.

[42] A. Moll, “The software framework of the Belle II experiment,” *Journal of Physics: Conference Series (CHEP 2010)* **331** (2011) 032024.

[43] M. Staric, “Physics Analysis Software Framework for Belle II,” Apr, 2015. `http://d2comp.kek.jp/record/295/`. Belle II internal, last accessed April 28th 2016.

[44] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS* **ACAT** (2007) 040, `arXiv:physics/0703039`.

[45] M. Feindt and U. Kerzel, “The NeuroBayes neural network package,” *Nucl. Instrum. Meth.* **A559** no. 1, (2006) 190–194.

[46] M. Feindt, “A Neural Bayesian Estimator for Conditional Probability Densities,” *ArXiv Physics e-prints* (Feb., 2004) , `arXiv:physics/0402093`.

[47] W. Waltenberger, “RAVE" A Detector-Independent Toolkit to Reconstruct Vertices,” *IEEE Transactions on Nuclear Science* **58** no. 2, (April, 2011) 434–444.

[48] I. Adachi, R. Itoh, N. Katayama, T. Tsukamoto, T. Hibino, M. Yokoyama, L. Hinz, and F. Ronga, “Computing system for the Belle experiment,” *eConf* **C0303241** (2003) MODT010, `arXiv:physics/0306120 [physics]`.

[49] B. Casey, “HadronB.” Belle internal Note 390, Jan, 2001. `https://belle.kek.jp/secured/belle_note/gn390/bn390_012901.ps.gz`.

[50] I. Adachi, "Selection Criteria in Classifying Beam Data." `http://belle.kek.jp/secured/runinfo/dst/skimmed_file/SkimmedFile/e000007/selection.html`. Belle internal.

[51] G. Majumder, "Proton identification efficiency and fake rates from K/π." Belle internal Note 670, November, 2003. `https://belle.kek.jp/secured/belle_note/gn670/note670.ps.gz`.

[52] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, "Electron identification in Belle," *Nucl. Instrum. Meth.* **A485** no. 3, (2002) 490–503.

[53] A. Abashian *et al.*, "Muon identification in the Belle experiment at KEKB," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **491** no. 1–2, (2002) 69 – 82. `http://www.sciencedirect.com/science/article/pii/S0168900202011646`.

[54] "Belle File Search Engine; Version 4.0 (Belle internal)." `http://bweb3.cc.kek.jp/index_m.php`.

[55] S. Neubauer, *Search for* $B \to K^{(*)} \nu\bar{\nu}$ *Decays Using a New Probabilistic Full Reconstruction Method.* PhD Thesis, Karlsruhe Institute of Technology (KIT), 2011. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48215`.

[56] M. Röhrken, *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment.* PhD Thesis, Karlsruhe Institute of Technology (KIT), 2012. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48212`.

[57] G. C. Fox and S. Wolfram, "Observables for the analysis of event shapes in e^+e^- annihilation and other processes," *Phys. Rev. Lett.* **41** (Dec, 1978) 1581–1585.

[58] D. M. Asner *et al.*, "Search for exclusive charmless hadronic B decays," *Phys. Rev. D* **53** (Feb, 1996) 1039–1050. `http://link.aps.org/doi/10.1103/PhysRevD.53.1039`.

[59] M. Gelb, "Neutral B Meson Flavor Tagging for Belle II," Master Thesis, Karlsruhe Institut of Technology (KIT), 2015. `https://ekp-invenio.physik.uni-karlsruhe.de/record/48719`.

[60] G. Punzi, "Sensitivity of searches for new signals and its optimization," *eConf* **C030908** (2003) MODT002, `arXiv:physics/0308063 [physics]`. [,79(2003)].

Appendix

A. Classifier Training Validation

The following figures show the results of the automatic classifier training validation performed by TMVA for a selection of the FastBDT trainings performed in the scope of this thesis.

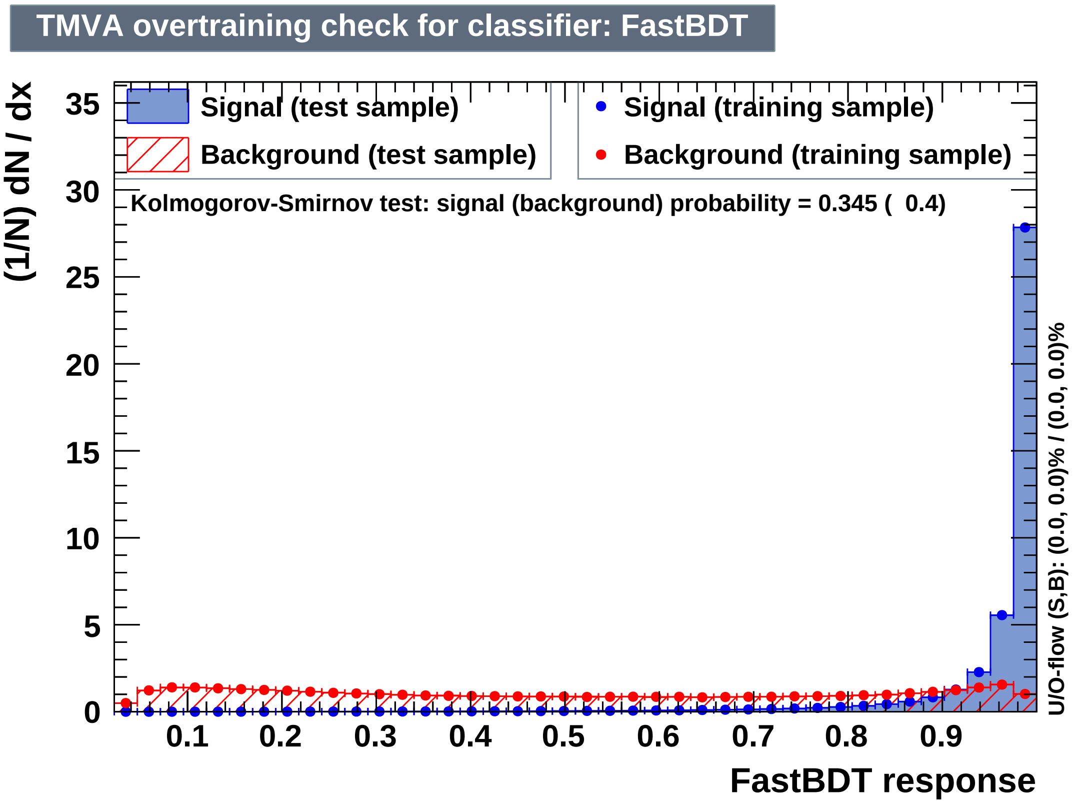
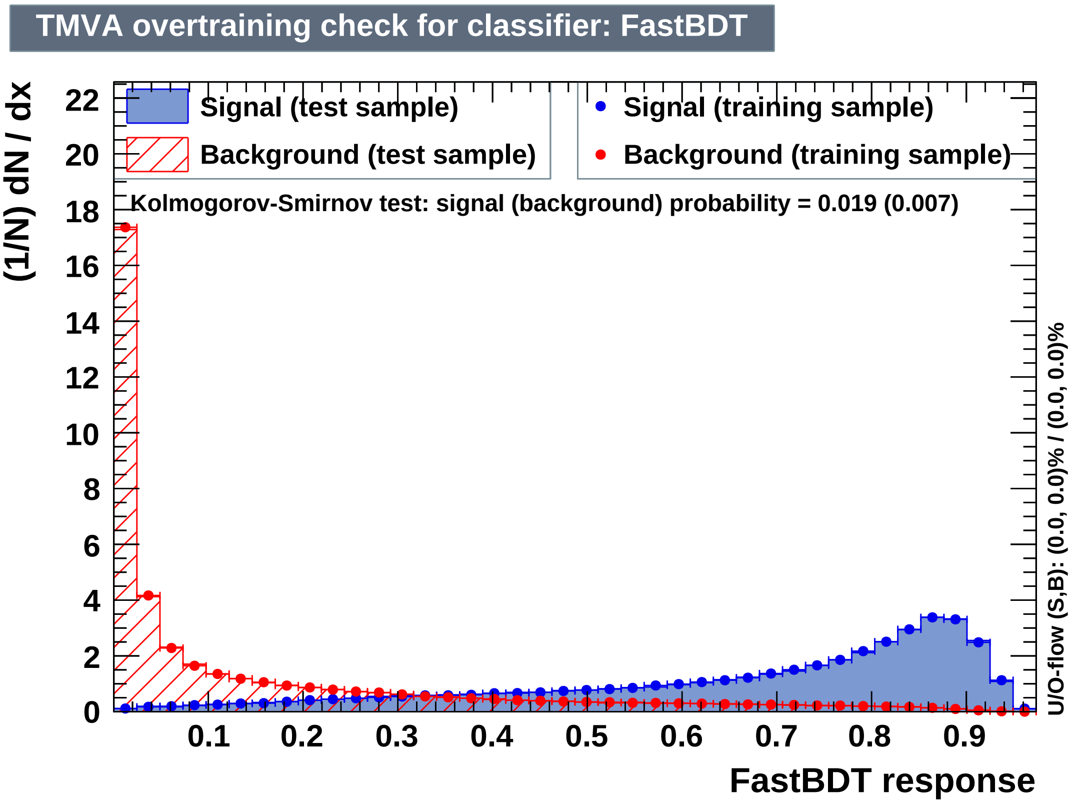
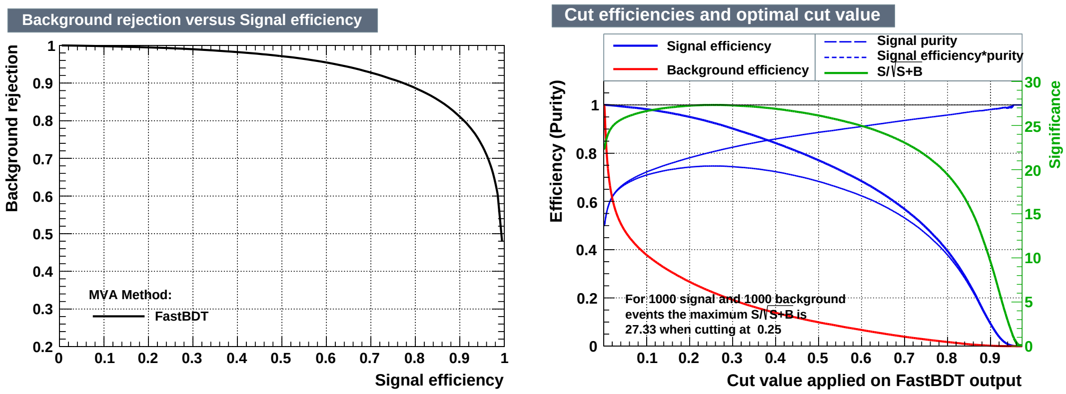


Figure A.1.: The displayed over-training check plot for the **continuum suppression classifier** was automatically generated as part of the TMVA classifier training validation. It superimposes the output of the classifier obtained from the training sample and the independent validation sample. As the result for the independent sample agrees with the test sample, one can assume that no over-training of the classifier occurred.



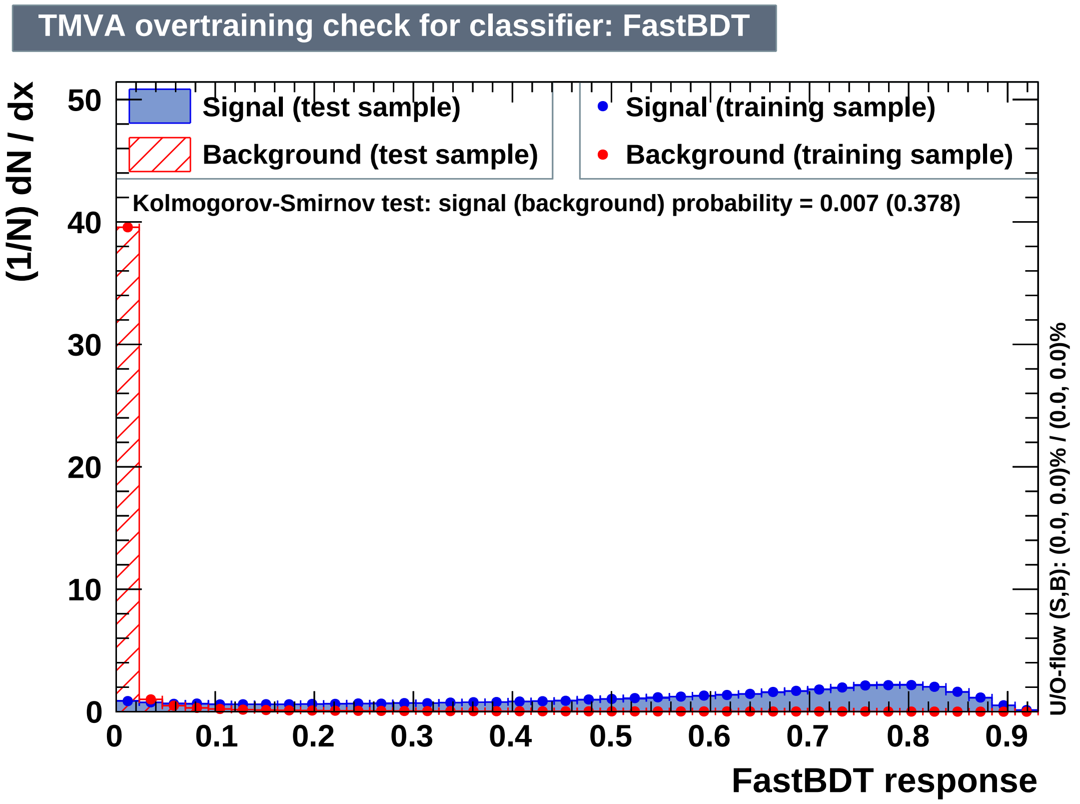
(a) TMVA over-training check plot



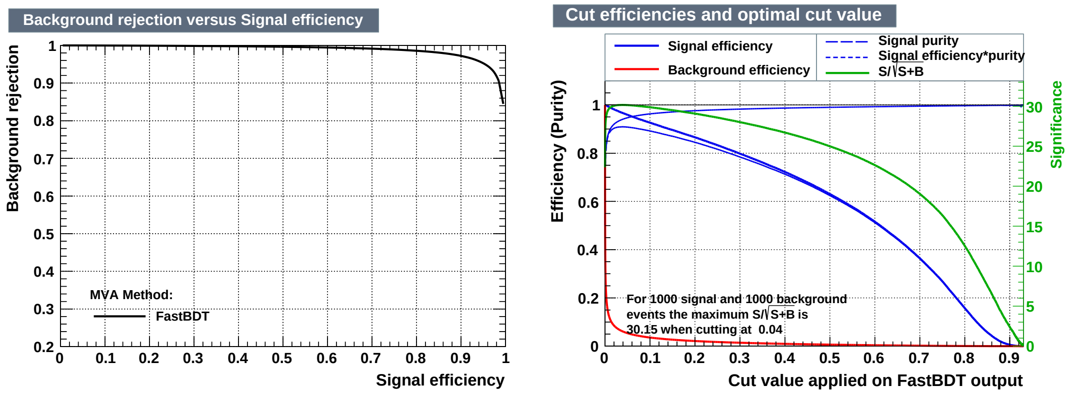
(b) Receiver operator characteristic curve

(c) Cut efficiency evaluation

Figure A.2.: Results of the training evaluation performed by TMVA for the **veto classifier** training described in Section 5.6. The output distributions obtained from an application of the trained classifier on the training and validation sample is shown in (a). An over-training did not occur, as the distribution of the independent validation sample agrees with the one of the training sample. The ROC curve with an area under the curve of 96.3 % is shown in (b) and the validation of different cuts on the output is displayed in (c).



(a) TMVA over-training check plot



(b) Receiver operator characteristic curve

(c) Cut efficiency evaluation

Figure A.3.: Results of the training evaluation performed by TMVA for the **signal selection classifier** training described in Section 5.6. The output distributions obtained from a application of the trained classifier on the training and validation sample is shown in (a). An over-training did not occur, as the distribution of the independent validation sample agrees with the one of the training sample. The ROC curve with an area under the curve of 98.6 % is shown in (b) and the validation of different cuts on the output is displayed in (c).