Investigation of $B \to \mu\nu_\mu$ with Inclusive Tagging at Belle II

Daniel Jacobi

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

August 2021

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, 31.08.2021
Date

Signature

1. Gutachter: Prof. Dr. Jochen Dingfelder
2. Gutachter: Prof. Dr. Florian Bernlochner

Acknowledgments

While writing this thesis, I have been fortunate to be surrounded by many people who have supported me in every way. First of all, I would like to thank my supervisor Jochen Dingfelder, who gave me the opportunity to work in his group and thus also gave me the chance to learn a lot about the work in particle physics. It was his interesting lectures that guided me towards writing my master's thesis in his group. I would also like to thank Florian Bernlochner, who agreed to read this thesis as a second reviewer. It was also thanks to his very informative seminar on the physics of particle detectors that I was encouraged to go into particle physics. Next, I would like to thank Peter Lewis, who has supported me in every way possible. It didn't matter what questions or problems I had, he always had time for me and had the right advice. Without the much appreciated meetings we had, none of this would have been possible. The other person I want to thank for unceasing support is Svenja. She supported me throughout with her help and always had a solution to my problems and ideas that I could pursue. She is another person without whom it would not have been possible to the write this thesis. Furthermore I would like to thank Markus. He has worked on this topic before and was always there to give me advice. He also gave me access to his project and notes, which helped me a lot in understanding and researching the topic. A thank you is also addressed to all the other Belle II Group members who made my days in the office very pleasant. This includes Stephan, Jonathan, Henrik, Christian and Dennis. Each of them has contributed in one way or another to making it possible to write this thesis. I would also like to thank the people who made this time so special in a different way. Luisa has always supported me during this time and set my head straight. She was always the person I could turn to if I had any doubts and always had the right words to get me back on track. Finally, I would like to thank my parents for always believing in me and for always having an open ear for me when things didn't go as planned. I couldn't ask for better parents who support me tirelessly and without doubt. Thank you for everything.

Contents

Introduction

The Standard Model of particle physics provides an understanding of the fundamental building blocks of our known universe. The elementary particles as well as the mediators of the fundamental forces are united in this theory. Since the 1970s, this theory was able to explain almost all experimental results [1]. With the discovery of the Higgs boson at the Large Hadron Collider (LHC) [2], another missing piece in the Standard Model was discovered. Although this theory was proven successful over time, there are still phenomena which can not be described by the Standard Model, leading to physics beyond the Standard Model. The Standard Model seems to be incomplete, since it does not contain gravity as an additional fundamental force and also since no particle candidate has been found to date, which could make up dark matter [3].

An important experiment pursuing the precise measurement of the Standard Model parameters as well as the search for physics beyond the Standard Model is the Belle II experiment. In the SuperKEKB asymmetric-energy collider electrons and positrons are collided at a center of mass energy at the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ resonance decays with at least 96 % [4] into a B meson pair without any additional particles present in the hadronization process, giving the Belle II experiment an alternative description as a B-factory. The asymmetry in the energies is chosen in order to allow time-dependent charge-parity symmetry violation measurements [5]. The decay products of the B mesons are then measured by the Belle II detector acting as a general purpose spectrometer. In chapter 3, a detailed description of the SuperKEKB collider and the Belle II experiment is provided.

B mesons can decay via various decay channels. Leptonic B meson decays, as the $B \rightarrow \mu\nu_{\mu}$ decay, which investigation is presented in this thesis, are really interesting since they serve as probes to physics beyond the Standard Model. The observation of the $B \rightarrow \mu\nu_{\mu}$ decay could lead to first hints for charged Higgs bosons [6] and sterile neutrinos [7]. But in contrast, leptonic B meson decays are difficult to observe experimentally since they are highly suppressed due to various reasons. The most recent study investigated the $B \rightarrow \mu\nu_{\mu}$ decay using the full 711 fb^{-1} data set recorded by the predecessor to Belle II, the Belle experiment, resulted in a one-sided significance of 2.8 standard deviations [7]. Following this analysis closely, a first study for the Belle II experiment will be presented.

Chapter 4 shows the reconstruction process of particles measured by the Belle II experiment and gives a short introduction to the Monte Carlo simulations used in this thesis. This is followed by

chapter 5, where the analysis strategy and the event selections are presented. Furthermore a Boosted Decision Tree was trained in the course of this thesis, which results are also shown in this chapter. A first study of errors in the reconstruction of the B_{tag} is presented in chapter 6. The aim of this study is to understand these errors in more detail and subsequently draw measures in order to improve the chances of discover the $B \rightarrow \mu\nu_{\mu}$ decay. Looking for improvements in the reconstruction is taken a step further by investigating different methods to precisely estimate the four-momentum of the B_{tag}, presented in chapter 7, followed by some additional studies in chapter 8 and the discussion of the obtained results in chapter 9.

Background Information

In this chapter an overview of relevant background information is provided. The Standard Model of Particle Physics is introduced, the physics of the rare leptonic $B \to \ell \nu_\ell$ decay explained, as well as Multivariate Analysis discussed. If not otherwise mentioned, the content of this chapter is based on [2], [6], [8] and [9].

2.1 The Standard Model of Particle Physics

Particle physics aims to explore the elementary building blocks of our universe and to understand the fundamental interactions that act between them. According to our current understanding, the elementary particles, as well as the mediators of the fundamental forces in the form of gauge bosons, can be united in a theory known as the Standard Model of particle physics, which is shown in Figure 2.1.
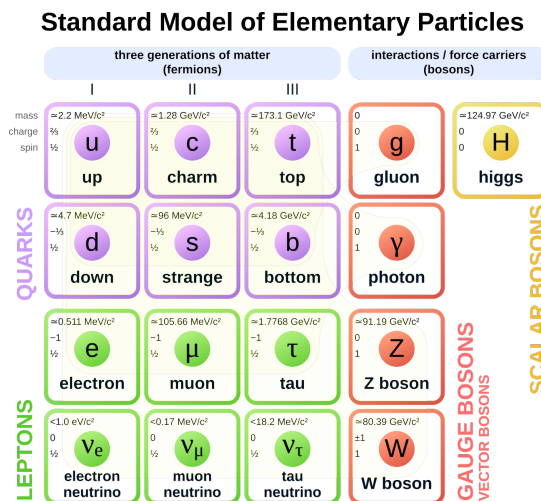

Figure 2.1: The Standard Model of particle physics. The mass, charge and spin of the quarks (purple), leptons (green), gauge bosons (red) and the Higgs boson (yellow) are shown. Figure was taken from [10].

The particles listed in the Standard Model can be divided into bosons or fermions depending on their spin. The fermions carry a half-integer spin and can be further divided into three generations of quarks and leptons. There are six different types of quarks. They all carry a spin of $1/2$ and differ essentially only in mass and electric charge. The up, charm and top quark all carry a charge of $+2/3$, while the down, strange and bottom quarks carry a charge of $-1/3$. As with quarks, there are six different types of leptons with spin $1/2$ that differ in mass and electric charge. The electron, muon, and tau carry a charge of -1, while the electron neutrino, muon neutrino and tau neutrino carry no charge. Each particle has a corresponding antiparticle. In principle, they carry the same quantities except for an inverse electric and color charge.

In the Standard Model, there are also four types of gauge bosons that carry spin 1. They act as mediators of the electromagnetic, strong and weak interaction. Gravity, as the fourth fundamental interaction, is not mentioned in the Standard Model since it is too weak on the particle scale to play a role. The strong interaction is described by the theory of Quantum Chromodynamics (QCD) and occurs exclusively between quarks since they carry color charge. The mediator of the strong interaction is the gluon with spin 1, but without electric charge and mass. The gluon itself carries a color charge, which can lead to self-interaction of gluons. An important result of the strong interaction is the fact that quarks cannot be observed as free particles, but always in bound states called hadrons. The electromagnetic interaction is described by the theory of Quantum Electrodynamics (QED) and only occurs between the quarks and the charged leptons. The interaction between the charged particles is mediated by the chargeless and massless photon. The weak interaction is the only interaction that occurs for all elementary particles. The weak interaction can be further divided into the weak charge-current and neutral-current interaction. The weak charge-current interaction is mediated by the massive and charged W^+ or W^- boson, while the weak neutral-current interaction is mediated by the massive but uncharged Z^0 boson. On a particle scale, the weak interaction is weaker than the strong and electromagnetic interaction.

The Standard Model was completed with the discovery of the Higgs boson at the LHC in 2012 [2]. The massive Higgs boson is the only particle in the Standard Model with spin 0. The Higgs boson is a result of the Higgs mechanism by which all other elementary particles acquire their masses. The particle can be viewed as an excitation of the Higgs field. By interacting with the Higgs field the elementary particles obtain their masses.

2.2 The Leptonic $B \to \ell \nu_\ell$ Decay

B mesons are the lightest bound state of a b quark and a light quark. B meson pairs are produced in the decay of the $\Upsilon(4S)$ resonance created in e^+e^- collisions, as occurring at the SuperKEKB collider, described more in detail in chapter 3. B mesons decay in different modes and in the course of this thesis the leptonic $B \to \mu \nu_\mu$ decay is investigated. In the following, the $B \to \ell \nu_\ell$ decay is described in more detail using the $B^+ \to \ell^+ \nu_\ell$ decay as an example. It should be noted that the given information is also valid for the $B^- \to \ell^- \bar{\nu}_\ell$ case.

The u and $\bar{b}$ quarks forming the charged B^+ meson can eventually annihilate into a virtual W^+ boson, which subsequently decays into a charged lepton and a neutrino, forming the $B^+ \to \ell^+ \nu_\ell$ decay. The probability of the b and u quarks, forming the bound state of the B meson, to annihilate is determined by the decay constant f_B [8]. Due to the nature of this two body decay the lepton momentum in the

rest frame of the decaying B meson is precisely known and given by [11]

$$p_\ell^B = \frac{\sqrt{\left(m_B^2 + m_\ell^2 - m_\nu^2\right)^2 - 4m_B^2 m_\ell^2}}{2m_B} = 2.6386 \, \text{GeV} \tag{2.1}$$

with the mass of the B meson m_B, the mass of the lepton m_ℓ and the mass of the neutrino m_ν, respectively. Although this decay, whose tree-level Feynman diagram is shown in Figure 2.2, is easy to describe and understand theoretically, it is quite a challenge to observe this decay experimentally. The involvement of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element $|V_{ub}|$ in the u and b quark annihilation process as well as helicity suppression makes this decay quite rare.
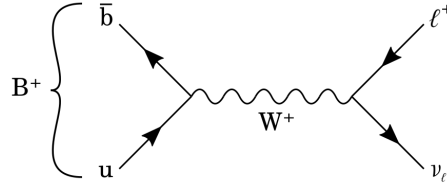
Figure 2.2: $B^+ \rightarrow \ell^+ \nu_\ell$ tree-level Feynman diagram, adapted from [7].

The branching ratio of the $B^+ \rightarrow \ell^+ \nu_\ell$ decay is calculated in the Standard Model by [9]

$$\mathcal{B}(B^+ \rightarrow \ell^+ \nu) = \frac{G_F^2 m_B m_\ell^2}{8\pi} \left(1 - \frac{m_\ell^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B \tag{2.2}$$

with the Fermi constant being G_F, the decay constant of the B meson f_B, the CKM matrix element $|V_{ub}|$ for the weak transition, as well as the the B meson lifetime τ_B.
The numerical values of these parameters for the $B^+ \rightarrow \mu^+ \nu_\mu$ case are listed in Table 2.1 and the calculated values for the branching ratio are shown in Table 2.2. The helicity suppression is expressed via the m_ℓ^2 factor in the branching ratio. The suppression arises from the fact, that the spin-0 B meson decays into two fermions with spin 1/2 each. To conserve the angular momentum, the ℓ^+ and ν_ℓ must have left-handed helicity. Since the ν_ℓ is assumed to be a massless particle, it always has left-handed helicity. The ℓ^+ however is forced into its wrong helicity state in order to conserve the momentum and the probability to find the ℓ^+ in this state is much greater for heavier leptons. The allowed helicity configuration is depicted in Figure 2.3. The available phase space in the $B^+ \rightarrow \ell^+ \nu_\ell$ decay is determined in the branching ratio by the $(1 - m_\ell^2/m_B^2)^2$ term. Due to the phase space argument, the branching fraction is higher for more massive leptons and, in addition to the helicity suppression, this is the main reason this analysis aims for an investigation of $B^+ \rightarrow \mu^+ \nu_\mu$ instead of $B^+ \rightarrow e^+ \nu_e$. Although $B^+ \rightarrow \tau^+ \nu_\tau$ promises a higher branching ratio, it is quite hard to observe this process since the further decay of the tau produces a higher number of missing neutrinos in the final state and thus leads to more missing information in the event.
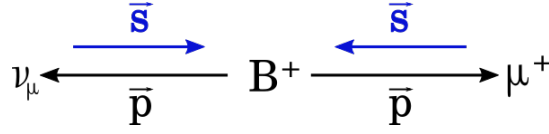
Figure 2.3: The allowed helicity configuration in the $B^+ \rightarrow \mu^+ \nu_\mu$ decay. Figure adapted from [2]. The spin configuration $\vec{s}$ is shown in blue and the momentum configuration $\vec{p}$ in black.

Parameter	Value
$G_F/(\hbar c)^3$	$(1.17 \pm 0.06) \cdot 10^{-5}$ GeV^{-2}
m_B	$5.28 \pm (1.2 \cdot 10^{-4})$ GeV
m_μ	$105.66 \pm (2.4 \cdot 10^{-6})$ MeV
f_B	190.0 ± 1.3 MeV
τ_B	1.638 ± 0.004 ps
$|V_{ub}|_{\text{inc}}$	$(4.25 \pm 0.12^{+0.15}_{-0.14} \pm 0.23) \cdot 10^{-3}$
$|V_{ub}|_{\text{exc}}$	$(3.70 \pm 0.10 \pm 0.12) \cdot 10^{-3}$
$|V_{ub}|_{\text{avg}}$	$(3.82 \pm 0.24) \cdot 10^{-3}$

Table 2.1: Numerical values of the parameters needed to calculate the $B^+ \rightarrow \mu^+ \nu_\mu$ branching ratio defined in Equation 2.2. Values are taken from [4].

$|V_{ub}|$	$\mathcal{B}(B \rightarrow \mu\nu_\mu)$
inclusive	$(5.02 \pm 0.65) \cdot 10^{-7}$
exclusive	$(3.80 \pm 0.36) \cdot 10^{-7}$
average	$(4.05 \pm 0.54) \cdot 10^{-7}$

Table 2.2: Inclusive, exclusive and averaged branching ratio of $B \rightarrow \mu\nu_\mu$ calculated with the values listed in Table 2.1.

2.3 Multivariate Analysis and Decision Trees

In addition to the references given above, references [12] and [13] were consulted for this section. Multivariate analysis is a tool for retrieving optimal differentiating information in a given data set and is widely used in particle physics for background suppression. Various types of multivariate analysis are based on different mathematical techniques. One of these techniques is the binary decision tree, which uses machine learning methods. First the decision tree is trained on simulated data, where signal and background are clearly labeled. In the training process of the decision tree, various discriminating variables are examined. Different cuts on these variables are made and the one resulting in the highest figure of merit, which denotes the signal-to-background ratio, is applied, resulting in the creation of two nodes. If necessary, these nodes are further separated by cuts on other variables. Ideally, there are two nodes at the end of the branches of the decision tree, which clearly distinguish signal from background. A schematic illustration of the procedure is pictured in Figure 2.4.

The cut-flow within the decision tree is then combined into a one-dimensional output classifier C_{out} and applied to a test data sample to evaluate its performance, which furthermore can be improved

by boosting. The boosted decision trees, or BDTs, use the idea of combining a set of weak learners into a strong one. A BDT takes the result of a binary decision tree and when an event is incorrectly classified, it is assigned a new weight. This weight is then taken into account when training the next binary decision tree. This process is repeated for several given trees and the output of each one is then combined into the output classifier. An appropriate choice of the so-called hyperparameters, for example the number of trees to be trained, plays a role in the performance. In Belle II an optimized BDT, FastBDT, is widely used.

When evaluating the performance, the p-value can be considered, which describes the difference between the training and test samples after the output classifier has been applied. In order to prevent over- or undertraining, the p-value should have a value above 0.05 and below 0.95 as a rough guideline. Overtraining occurs when the training picks up on features that are intrinsic to the training sample, such as statistical fluctuations. Undertraining occurs when key features may not be recognized at all. The Receiver Operating Characteristic (ROC) curve provides information about the performance of the BDT. This curve describes the relationship between signal efficiency and background rejection. In the case of an ideal BDT that can completely separate signal from background, the area under this curve (AUC) is 1. In general, a BDT performs well when the AUC is greater than 0.9. In the case of random guessing, this value drops to 0.5. This behavior is depicted in Figure 2.5.
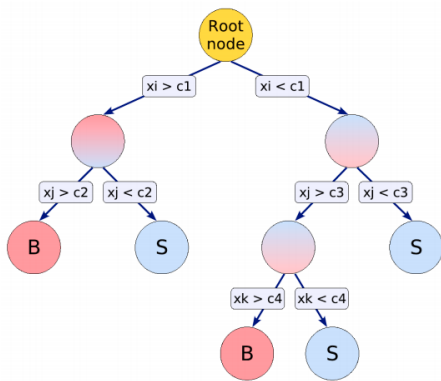
Figure 2.4: Schematic illustration of a binary decision tree, taken from [14]. Each node is depicted by a circle. Different cuts $c_{1,2,3,4}$ on discriminating variables $x_{i,j,k}$ are performed in order to separate signal (S) from background (B).
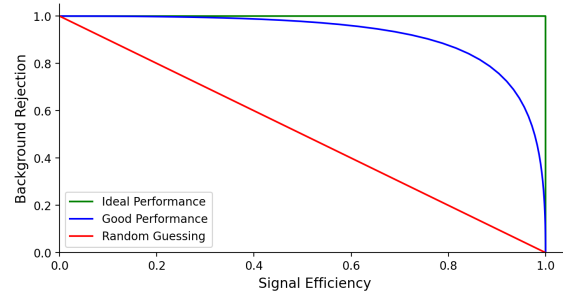
Figure 2.5: ROC curve for different performances of a BDT, produced with information obtained from [15].

Introduction to SuperKEKB and Belle II

In this chapter a general overview of the SuperKEKB collider and the Belle II experiment including all relevant detector components is given. Unless otherwise indicated, the references [16], [5], [17], [18] and [19] were consulted in order to write the content of this chapter.

3.1 The SuperKEKB Collider

SuperKEKB is an asymmetric-energy electron-positron collider located at the High Energy Research Organization (KEK) in Tsukuba in Japan and was launched as the successor to the KEKB collider. Its construction began in 2010 and it was commissioned in 2016. In Phase 1 the collider ran without collisions, the final focus and Belle II detector components. In the following Phase 2 in 2018, the first collisions were recorded while most of the detector components were in operation, except for the inner detectors. During this period, SuperKEKB reached a peak luminosity of $0.5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and recorded data with integrated luminosity of the order of $0.5\,\mathrm{fb}^{-1}$. As of early 2019, Phase 3 is operating using the full-scale collider with a fully installed detector. The design luminosity of the SuperKEKB is $8 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and exceeds that of KEKB by about 40 times [16].
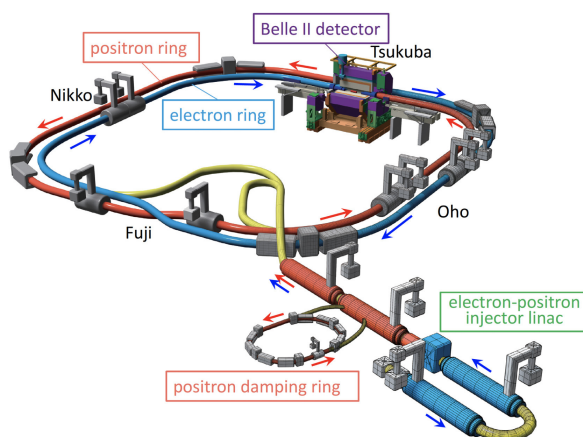
Figure 3.1: Schematic illustration of SuperKEKB, taken from [16]. The positron and electron rings are depicted as well as the linear accelerator and the damping ring.

In the SuperKEKB collider complex, electrons and positrons are accelerated in two different rings with asymmetric energies. This asymmetry is chosen so that the resulting B mesons travel a sufficient distance in the detector to allow precise measurement of various physical quantities like the B meson decay time via the vertex position. Electrons are generated in a pre-injector located at the beginning of the linear accelerator. They are then also used to generate positrons by the irradiation of a tungsten target. The generated positrons are guided through a damping ring, which is put into place to reduce the emittance. In the linear accelerator electrons are accelerated to an energy of 7 GeV and are stored in the high-energy ring (HER) while the positrons are accelerated up to 4 GeV and are stored in the low-energy ring (LER). A schematic layout of the SuperKEKB complex is shown in Figure 3.1.

3.2 Particle Collisions and Background Events

The SuperKEKB collider is operating at the center of mass energy of $\sqrt{s} = 10.58$ GeV. This is exactly the energy required to produce the $\Upsilon(4S)$ resonance. Operating at this resonance has the main advantage of producing a clean sample of $B\bar{B}$ pairs via the decay of the $\Upsilon(4S)$ resonance. This process is shown in Figure 3.2. At least 96% of the time the resonance decays into a $B\bar{B}$ pair [4]. Since the mass of the $\Upsilon(4S)$ is only approximately 21 MeV higher than twice the B meson mass, no additional particles are produced in the hadronization process and furthermore the two B mesons are emitted back-to-back in the center of mass frame of the $\Upsilon(4S)$ with a momentum of approximately 332 MeV [8].
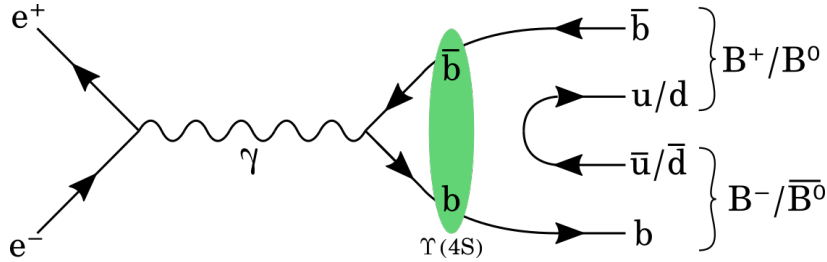
Figure 3.2: Feynman-style diagram of the creation of the $\Upsilon(4S)$ resonance by electron and positron annihilation into a virtual photon. Subsequently, the resonance decays into a $B\bar{B}$ pair via the strong interaction, adapted from [8].

The collision does not always result in the production of the $\Upsilon(4S)$ resonance. The production of non-resonant final states is also possible. These are namely continuum events, where the collision produces lighter quark-antiquark pairs if the center of mass energy is at least twice as high as the mass of the quark or antiquark. At Belle II continuum events consist of u, d, s and c quark-antiquark pairs given the center of mass energy of 10.58 GeV. Additional backgrounds are introduced by the electromagnetic interaction between the colliding e^+ and e^- producing a lepton and antilepton pair. Two-photon processes can also occur [8]. The corresponding cross sections to the named physical processes are shown in Table 3.1.

In addition to the backgrounds due to non-resonant final states, backgrounds originating from the beam also occur. Compared to the KEKB collider, the background due to Touschek and beam-gas scattering is enhanced. Particles in the same bunch can either scatter with each other via the Coulomb interaction (Touschek scattering) or with the gas molecules in the beam pipe via Coulomb interaction

or Bremsstrahlung (beam-gas scattering). Due to the scattering process a particle might leave the bunch and in the worst case leave a trace in the detector. Another background occurs due to synchrotron radiation emitted from the beam. Furthermore, radiative Bhabha scattering can lead to an additional background. The photons emitted can interact with the iron of the magnets, subsequently producing neutrons, leaving signatures in the detector. The last background to be mentioned here is the one caused by two photon processes $ee \rightarrow eeee$ and $ee \rightarrow ee\mu\mu$. A produced electron positron pair can spiral around the field lines of the superconducting solenoid and in the worst case leave multiple hits in the inner detectors.

Physics process	Cross section in nb
$\Upsilon(4S)$	1.11
$u\bar{u}(\gamma)$	1.61
$d\bar{d}(\gamma)$	0.40
$s\bar{s}(\gamma)$	0.38
$c\bar{c}(\gamma)$	1.30
$e^+e^-(\gamma)$	300
$\gamma\gamma(\gamma)$	4.99
$\mu^+\mu^-(\gamma)$	1.148
$\tau^+\tau^-(\gamma)$	0.919
$\nu\bar{\nu}(\gamma)$	$0.25 \cdot 10^{-3}$
$e^+e^-e^+e^-$	39.7
$e^+e^-\mu^+\mu^-$	18.9

Table 3.1: Cross sections corresponding to the $\Upsilon(4S)$ production and various background processes, taken from [5].

3.3 The Belle II Detector

Due to the update of the KEKB collider to the SuperKEKB, a much larger amount of data is expected. As a result, an update of the Belle detector was first presented in 2004. The Belle II detector, whose schematic layout is shown in Figure 3.3, was introduced. Various modifications and changes have been made to maintain performance at higher background rates, which are expected with the updated collider. The Belle II detector consists of several subdetectors, each of which performs different measurements such as tracking, particle identification or energy measurements. The individual detector components are described and explained in the following. In addition, a large superconducting solenoid is installed around all subdetectors except for the K_L^0 and Muon Detector (KLM). It provides a magnetic field of 1.5 T.
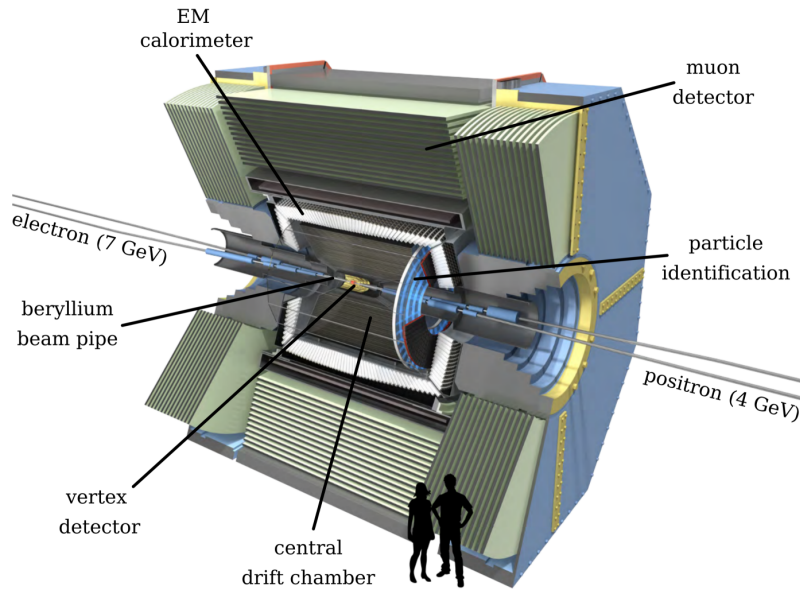
Figure 3.3: Illustration of the layout of the Belle II detector. Shown are the vertex detectors (PXD and SVD), the CDC, the detectors used for particle identification (TOP and ARICH), the EM calorimeter (ECL) as well as the muon detector (KLM). All the detector components are built around the beam pipe. Figure taken from [20] and labeling adapted from [21].

3.3.1 Pixel Detector (PXD)

The Pixed Detector makes up the two inner layers of the vertex detector, which is mainly responsible for measuring the vertex positions of the B meson decays. The PXD is the detector closest to the beam pipe and thus has to deal with very high hit rates due to additional beam-related background and low-momentum-transfer QED processes. The two layers of the detector consist of sensors based on the DEPleted Field Effect Transistor technology. These monolithic DEPFET sensors are made out of silicon. The inner layer is placed at a radius of 14 mm from the beam axis and consists of 8 laddern with 16 modules, while the outer layer is placed at a radius of 22 mm holding 12 ladders with 24 modules. Each ladder consist of 250×1600 DEPFET pixels.

3.3.2 Silicon Vertex Detector (SVD)

The next four layers of the vertex detector are formed by the Silicon Vertex Detector which, in addition to the PXD, is responsible for measuring the decay vertices of the B meson decay and other decay channels. The SVD sensors are double-sided silicon strip detectors located radially between 38 mm and 140 mm from the beam pipe. One sensor consists of a silicon bulk with p-doped strips on the one side and n-doped strips on the other side. A charged particle passing through the detector creates an electron-hole pair. The electron drifts to the p-strips, while the hole drifts to the n-strips. If enough electron-hole pairs are produced, this can be measured.

3.3.3 Central Drift Chamber (CDC)

The Central Drift Chamber provides a number of important properties of the particles produced in the B meson decay. With the CDC it is possible to reconstruct the tracks of charged particles and to measure their momentum. In addition, the energy loss is measured as a particle traverses the gas volume, and this information can be used to identify the particle. Furthermore, the CDC can provide a trigger signal for these charged particles. The main working principle of the CDC is as follows. When a charged particle passes through the gas-filled CDC volume, it ionizes the gas. The electrons produced during ionization drift towards built-in sense-wires due to an applied electric field. Secondary ionization occurs near the wires due to the high electric field, and this avalanche of charge induces a measurable signal in the wires. The current measured in the wires is proportional to the energy loss of the particle, from which the information dE/dx can be extracted and thus particles can be identified. Field-wires are installed in order to shape the electric field. A total of 14336 sense-wires made out of tungsten and 42240 field-wires made out of aluminum are installed in the CDC in an alternate wire configuration. The volume of the CDC is filled with a gas mixture He-C_2H_6, in the relation of 50% helium and 50% ethane.

3.3.4 Time of Propagation Counter (TOP)

In order to identify particles in the barrel region, a Time of Propagation Counter is built into the Belle II detector. In order to separate mainly pions from kaons, the TOP uses the Cherenkov effect. When a charged particle passes through a medium with refractive index n at a speed v greater than the speed of light in that medium $c_n = c_0/n$, photons are emitted in a cone with an characteristic angle of

$$\cos\theta_c = \frac{1}{\beta n} \qquad \text{with} \qquad \beta = \frac{v}{c_0}. \tag{3.1}$$

Measuring the Cherenkov angle enables the determination of β and subsequently the velocity v of the particle. Due to the relativistic momentum relation

$$p = \gamma m \beta c \qquad \text{with} \qquad \gamma = \frac{1}{\sqrt{1 - \beta^2}} \tag{3.2}$$

it is possible to determine the mass if the momentum is known and thus to identify the particle. The TOP at Belle II is build of an array of quartz radiators and photo multipliers (PMTs) at the end surfaces of the quartz bars, distributed over 16 modules around the CDCs outer wall. A particle traverses the quartz crystal and emits Cherenkov photons, which are fully reflected inside the crystal until registered by the PMT. Measurement of the propagation time of the photons in the corresponding crystal allows to obtain 3-dimensional track information of the particle. Furthermore, kaons emit Cherenkov photons at a smaller angle than pions with the same momentum, thus these photons are reflected more often inside the crystal resulting in a larger time of propagation. Using this information, one can distinguish between kaons and pions.

3.3.5 Aerogel RICH Detector (ARICH)

The Aerogel RICH Detector is installed in the forward region of the detector. Next to separating kaons and pions, it also aids in the differentiation between pions, muons and electrons in the low momentum region below 1 GeV. A charged particle traversing the aerogel radiator emits Cherenkov photons in a cone. Inside an expansion volume the Cherenkov photons form a ring which is subsequently projected onto a hybrid avalanche photon detector surface. The radius R of the Cherenkov ring in approximated by

$$R \approx d\sqrt{1 - \frac{1}{\beta^2 n^2}} \tag{3.3}$$

where d is the distance from the radiator to the photo detectors surface. Measuring the radius allows a determination of β and thus the identification of the particle.

3.3.6 Electromagnetic Calorimeter (ECL)

The main purpose of the Electromagnetic Calorimeter is to measure the energy of photons and electrons accurately and subsequently to separate them from pions and other hadrons. Upon entry into the ECL, the charged particle interacts with the calorimeter material and a cascading effect of interactions occurs, subsequently resulting in the development of a shower. The shape, depth and starting point of the shower differs for different kinds of particles. The energy deposited in the ECL is measured and used for further separation. A muon as an example of a minimum ionizing particle shows a constant energy loss in the ECL regardless of its momentum. In total there are 6624 Caesium Iodide crystals doped with Thalium employed in the barrel part and 2112 pure Caesium Iodide crystals located in the endcaps. Photodiodes are attached to the rear surface of each crystal in order to read out the scintillation light.

3.3.7 K_L^0 and μ Detector (KLM)

The K_L^0 and μ Detector is built of alternating layers of iron plates and active detector elements. The iron plates have two different tasks. Since the KLM lies outside the superconducting solenoid, the iron plates serve as a magnetic flux return. Furthermore, the iron plates serve as a calorimeter in which the K_L^0 shower hadronically. In the barrel region there are in total 15 layers of glass-electrode resistive plate chambers acting as the active detector elements and 14 iron plates. In both endcaps scintillator strips are installed in 14 layers as the active detector material, in addition to 14 iron plates. Muons are identified as such if a hit in the KLM is found and it matches with a track extrapolated from the CDC. A particle is identified as a kaon if at least one hit is registered in the ECL and the KLM.

3.4 Trigger System

With the update to the SuperKEKB a much larger background rate is expected. In order to cope with a high number of events containing processes without great scientific relevance, a trigger system is employed to discard these events. A vast majority of unwanted processes are two photon processes, such as Bhabha scattering at low angles and e^+e^- interactions which do not produce the $\Upsilon(4S)$ resonance. The low level trigger (L1) and the high level trigger (HLT) make up the Belle II trigger

system. The decision of the L1 is dependent on the information given by all subdetectors except for the PXD. The sofware-based HLT is running an online reconstruction of the event in real time. If an event is tagged as interesting, the information in combined with the one obtained by the PXD and stored for further processing.

Reconstruction and Monte Carlo Simulation

In this chapter an overview on the reconstruction of charged particles and clusters, as well as on the particle identification at Belle II is provided. Furthermore, the Monte Carlo event generators and additional analysis-specific aspects are introduced. Unless otherwise indicated, the references [5] and [9] were consulted.

4.1 Reconstruction of Charged Particles

The reconstruction of charged particles is performed in two steps, namely track finding and track fitting. For the track finding different recognition algorithms are applied to the hits registered in the VXD and CDC. Different tracking algorithms are applied due to the differences in the corresponding subdetectors. An algorithm that follows the cellular automaton model is applied to the hits in the VXD. If two hits in adjacent layers are compatible with each other, they are grouped into cells. Whether hits are compatible is determined by consulting a so-called sector map obtained through simulations. Subsequently, it is checked whether cells containing the same hit are geometrically adjacent according to the sector map. For track finding in the CDC, a global and a local track finder are employed. For the global track finder, all hits in the CDC are considered and a Hough transformation is applied. The local track finder also employes the cellular automaton model as explained above. Identified tracks in the VXD are extrapolated to the CDC's outer wall and if this extrapolated track is compatible with one found by the CDC, it is kept as a track candidate.

The next step is track fitting. In an ideal case, a charged particle would follow a helix trajectory in a constant magnetic field. Multiple scattering of the particle in the detector and inconsistencies of the magnetic field can divert the particle from the ideal trajectory and must be considered by the track fitting algorithm. Furthermore the energy loss due to scattering results in a helix trajectory with smaller diameter. The algorithm used is a deterministic annealing filter based on Kalman filters. A more detailed description of the filter is omitted here and a more detailed description is given in reference [19] and [22]. Weather the track finding and fitting is performing well is quantified by the tracking efficiency which is defined by the ratio of tracks correctly reconstructed to the number of generated charged particles.

4.2 Reconstruction of Clusters

In addition to the charged particle reconstruction, the clusters in the ECL produced by charged and neutral particles must be reconstructed. The reconstruction of the energy and position of the clusters will provide additional information for the identification of the particle. The algorithm used for reconstruction starts with seed finding. A crystal with an energy of at least 10 MeV and which energy is higher than the ones of the adjacent crystals is picked as the seed for a cluster. Then all the crystals surrounding the seed are added to a connected region if they have an energy higher than 0.5 MeV. The neighbours of the crystals are further added if they have an energy of a least 1.5 MeV. This is repeated until no more crystals are found that meet the energy threshold. If two formed connected regions added the same crystal, they are merged und subsequently the connected regions are split into clusters.

4.3 Identification of Particles

The information obtained by the subdetectors about particle identification and energy loss $(\mathrm{d}E/\mathrm{d}x)$ is combined in order to ensure an efficient particle identification (PID). The identification of charged particles works as follows. The information obtained by each subdetector able to perform PID is evaluated independently. It is determined what the likelihood is that a reconstructed track or cluster belongs to a certain type of charged particle. The likelihoods from the subdetectors $\mathcal{L}^{\mathrm{det}}$ are then combined. A likelihood ratio can de defined as [23]

$$P(x)_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j} \qquad \text{with} \qquad \mathcal{L}_i = \prod_{\mathrm{det}}^{\{SVD,CDC,...\}} \mathcal{L}_i^{\mathrm{det}} \qquad (4.1)$$

with $i, j \in \{e, \mu, \pi, K, p\}$ or their antiparticles and x being a set of measurements. A selection on the likelihood ratio can be used in order to identify certain types of particles. Neutral particles are identified differently. Photons are identified in the ECL by investigation of the cluster shape. Additionally, there is no reconstructed track matching that cluster. K_L^0 mesons are identified with the information obtained from the ECL and KLM. After gathering all the information it is possible to construct an event which can be investigated in an analysis.

4.4 Monte Carlo Simulation

Monte Carlo (MC) generators are widely used in particle physics. For the physical processes occurring at the Belle II experiment different event generators are used in order to simulate the corresponding processes. The B meson decays are simulated using the EvtGen event generator. The generator strictly follows a decay table, that lists each possible B meson decay with its branching ratio and decay model. The $B \to \mu\nu$ decay is not included in the decay table and therefore has to be simulated individually with EvtGen [24]. This MC sample is hereafter referred to as signal MC. In the signal MC simulation, one B meson always decays into the muon and the neutrino while the other B meson decays generically following the decay table. The signal MC has to be scaled according to the expected branching ratio. The non-resonant final states $e^+e^- \to q\bar{q}$, where $q \in \{u, d, s, c\}$, are generated using the KKMC generator [25][26] and the fragmentation process of these events is simulated by PYTHIA 8.2 [27]. The $e^+e^- \to \tau^+\tau^-(\gamma)$ process is also simulated with the KKMC generator.

This analysis divides the generic B decays ($B \to \mu \nu_\mu$ excluded) into three categories. Decays involving the $b \to c$ transitions are further referred to as the $B \to X_c \ell \nu$ sample and decays involving the $b \to u$ transition as the $B \to X_u \ell \nu$ sample. All remaining $B\bar{B}$ decays, including events where the muon is misidentified or originates from a secondary decay, are combined into an *other* $B\bar{B}$ sample. The generated $B \to \mu \nu$ events are further called signal or $\mu \nu$ sample. The $e^+ e^- \to q\bar{q}$ processes together with the $e^+ e^- \to \tau^+ \tau^- (\gamma)$ processes are merged into the *continuum* sample.

For this analysis we look at $100 \, \text{fb}^{-1}$ of MC, which corresponds to about 54 million $B^+ B^-$ events. In order to match the recorded data (see section 4.4.2) the MC was scaled down to $62.8 \, \text{fb}^{-1}$. We expect only approximately 27 signal events in $62.8 \, \text{fb}^{-1}$ considering the average branching ratio shown in Table 2.2. For the training of the BDT in section 5.2 we use $200 \, \text{fb}^{-1}$ of MC simulated data.

4.4.1 Hybrid Model

The hybrid method was first introduced in [28]. The short explanation of this model given in this section is based on [8], [7] and [29]. For further information please refer to these references.

The hybrid model is introduced to describe the $B \to X_u \ell \nu$ decays using a combination of inclusive and exclusive predictions. The resonant $B \to \pi \ell \nu$, $B \to \rho \ell \nu$ and $B \to \omega \ell \nu$ decays are modeled using the Bourrely-Caprini-Lellouch (BCL) form factor parametrization, while the $B \to \eta \ell \nu$ and $B \to \eta' \ell \nu$ decays are modeled using the LCSR calculation. In addition non-resonant $B \to X_u \ell \nu$ decays with more than one pion in the final state are modeled using the De Fazio and Neubert (DFN) model. In such semileptonic decays, the hadronic and leptonic parts factorise, and the form factor entails the QCD contribution. The resonant and non-resonant predictions are then combined and weights w are assigned to the inclusive contributions. This is done so that the inclusive values can be reproduced by the combination of the triple differential rate of the inclusive $\Delta \mathcal{B}^{\text{incl}}$ and exclusive $\Delta \mathcal{B}^{\text{excl}}$ predictions, expressed with

$$\Delta \mathcal{B}_{ijk}^{\text{incl}} = \Delta \mathcal{B}_{ijk}^{\text{excl}} + w_{ijk} \times \Delta \mathcal{B}_{ijk}^{\text{incl}}. \tag{4.2}$$

The indices i, j and k denote the corresponding bin of the split inclusive distribution in three dimensions

$$
\begin{aligned}
m_X &= [0.00, 1.40, 1.60, 1.80, 2.00, 2.50, 3.00, 3.50] \, \text{GeV} \\
q^2 &= [0.00, 2.50, 5.00, 7.50, 10.0, 12.5, 15.0, 20.0, 25.0] \, \text{GeV}^2 \\
E_l^B &= [0.00, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 3.00] \, \text{GeV}
\end{aligned} \tag{4.3}
$$

where m_x is the hadronic invariant mass, q^2 the momentum transfer and E_l^B the energy of the lepton. Subsequently the number of events in a particular bin i predicted by the hybrid model is given by

$$H_i = R_i + w_i I_i \qquad \text{with} \qquad w_i = \frac{I_i - R_i}{I_i}. \tag{4.4}$$

The number of events given by the inclusive predictions in bin i is denoted as I_i and the number of events resulting from the resonant predictions as R_i. The weight in the corresponding bin is defined as w_i. The $B \to X_u \ell \nu$ sample introduced in section 4.4 is replaced by the distribution obtained with the hybrid model. A higher number of events is expected around the kinematic endpoint of this sample [8]. This must be kept in mind since signal is also expected in this region.

4.4.2 Blinded Data Sample

In order to compare the generated MC to a real dataset, we look at $62.8\,\text{fb}^{-1}$ of data collected by Belle II until summer 2020. The muon momentum in the center of mass frame p_μ^* is blinded in the region of $2.3\,\text{GeV} \leq p_\mu^* \leq 3.0\,\text{GeV}$ to avoid bias. In this region we would expect our signal distribution. In order to improve the agreement between data and MC, leponID and momentum corrections are deployed.

Analysis Strategy and Event Selection

In this chapter the analysis strategy for the investigation of the $B \rightarrow \mu\nu_\mu$ decay is explained. Furthermore the event selections are presented. The BDT training is explained and the results shown. The events were processed using the Belle II software *basf2*.

5.1 Analysis Strategy

In the rest frame of the decaying B meson, the muon momentum is expected to be a monochromatic line at $p_\mu^B = 2.64\,\text{GeV}$ as shown in Equation 2.1. The muon momentum in the B rest frame can be obtained by performing a Lorentz boost on the muon momentum in the center of mass frame p_μ^*. The monochromatic line in p_μ^B is broadened by mainly two reasons. The detector resolution when measuring the muon momentum p_μ^* as well as the resolution on the boost vector have an subsequent impact on the resolution of p_μ^B. In order perform a high-precision boost, the kinematics of the B_sig and the μ_sig must be reconstructed as accurately as possible. The precise reconstruction of the B_sig is a difficult task, since the signal side neutrino is not reconstructed and the kinematic information is lost. Therefore a different approach is used.

We assume that the $\Upsilon(4S)$ decays approximately at rest in the center of mass frame, thus emitting the B_sig and B_tag meson back-to-back. Assuming no specific B_tag decay it is possible to gather more information about the B_tag kinematics than about the B_sig kinematics, since more particles can be reconstructed there. The ultimate goal therefore is to reconstruct B_tag and use

$$\begin{pmatrix} E_{B_\text{tag}} \\ \vec{p}_{B_\text{tag}} \end{pmatrix} = \begin{pmatrix} E_{B_\text{sig}} \\ -\vec{p}_{B_\text{sig}} \end{pmatrix} \tag{5.1}$$

to construct the kinematics of the signal side B meson, and subsequently perform the boost of the muon momentum p_μ^* into the B_sig rest frame. In the ideal case we would recover a δ peak.

To obtain a high precision Lorentz boost vector we thus need to determine the kinematics of the B_tag and the μ_sig very accurately. The uncertainty in the reconstruction of the μ_sig kinematics is dominated by the resolution of the detector. The uncertainty in the reconstruction of the B_tag kinematics on the other hand is dominated by missing particles in the reconstruction as well as other factors, which are studied and discussed in chapter 6. Figure 5.1 shows the p_μ^B distribution for two independent cases

using unscaled signal MC. In the one case the boost is performed using the true $B_{\text{tag}}^{\text{gen}}$ kinematics and reconstructed $\mu_{\text{sig}}^{\text{reco}}$ kinematics, both simulated with MC. The other case shows the p_μ^B distribution using the true $\mu_{\text{sig}}^{\text{gen}}$ kinematics and reconstructed $B_{\text{tag}}^{\text{reco}}$ kinematics. It is clearly visible to see that the uncertainty in the estimation of the B_{tag} kinematics has a much larger impact on the p_μ^B resolution, and therefore the focus of this analysis is on the accurate estimation of B_{tag}, in order to minimize the resolution of p_μ^B. The main driver of the signal significance is therefore the resolution of the Lorentz boost vector.
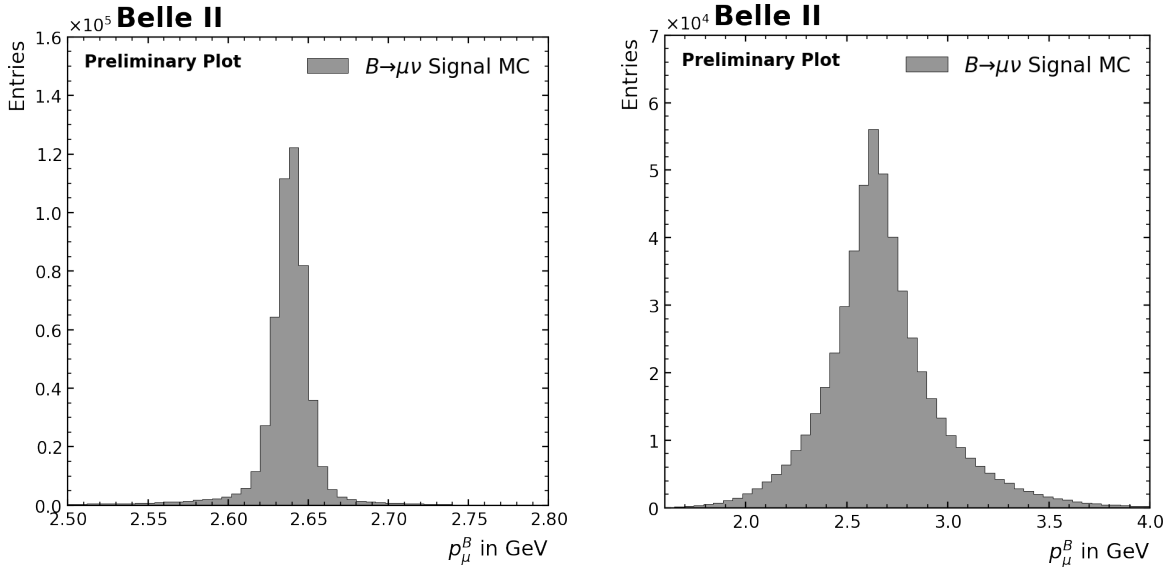
Figure 5.1: Comparison of reconstructed B_{tag} momentum components to the generated ones for signal MC. The left plot shows the distribution using the boost obtained from the true B_{tag} and reconstructed muon kinematics. The right plot shows the distribution using the boost obtained from the true muon kinematics and reconstructed B_{tag} kinematics.

There are mainly three methods one can use to reconstruct B_{tag}. The small overview below is based on [12] and further information can be found there. In hadronic tagging, only purely hadronic B_{tag} decays are considered. On the one hand, the kinematics of the B_{tag} are very well reconstructed using this method, but on the other hand, the branching fraction of purely hadronic decays is small. This tagging method shows a high purity but a low efficiency. But as a result, the B_{tag} is reconstructed quite well leading to a high resolution in p_μ^B. In semileptonic tagging, as the name suggests, only semileptonic decays are used for the B_{tag} reconstruction. The branching fraction for semileptonic decays is considerably higher than for hadronic decays, but due to the neutrino in the final state, kinematic information is lost. Semileptonic tagging thus has a higher efficiency than hadronic tagging but a lower purity.

This analysis is performed using the third method, which is the inclusive tagging approach. In inclusive tagging no explicit decay channel is assumed for the B_{tag}. Every track and cluster, which is not reconstructed as part of the B_{sig} decay, is assumed to originate from the B_{tag} decay and so called the rest of event (ROE). The information of the tracks and clusters are combined into the kinematic

information of the B_{tag}. This method shows a high efficiency since all events containing a signal candidate are considered, but also suffers from a low purity due to missing particles and also extra tracks due to, for example, tracks originating from beam background. In chapter 6 a study on the errors in the reconstruction of the B_{tag} using inclusive tagging is shown. Inclusive tagging is expected to show the worst resolution for the boost vector and therefore for p_μ^B. But due to the low branching ratio of the $B \to \mu\nu_\mu$ decay we require a high efficiency and thus use the inclusive approach.

5.1.1 Event Selection

In this section the event selections are shown. These are selections on certain reconstruction variables, which ideally suppress the background events without loosing signal events. The selections used are following closely the ones chosen for the previous analysis of $B \to \mu\nu_\mu$ at Belle [7] but only after a check to see, whether these also apply to Belle II. An event is kept, if the following conditions are met. We require that at least two tracks are reconstructed in the ROE. Additionally an event is kept if less than three leptons are reconstructed. The number of leptons is determined by taking the sum of reconstructed electrons and muons, that are selected based on a cut on the particle identification and a cut on the impact parameters z_0 and d_0. We reconstruct a track as an electron if the muon ID is below 0.9 while the electron ID is above 0.9. Muons are reconstructed if they show a muon ID above 0.9. The electrons as well as the muons track should be consistent with originating from the interaction point, therefore cuts on the impact parameters $|z_0| < 2\,\text{cm}$ and $d_0 < 0.5\,\text{cm}$ are performed. We furthermore define a normalized beam constrained mass $\hat{m}_{bc}^{ROE}$ and normalized energy difference $\Delta\hat{E}^{ROE}$, following [8], via

$$\hat{m}_{\text{bc}}^{\text{ROE}} = \sqrt{\frac{E^{*2} - |\vec{p}_B|^2}{E^{*2}}} \qquad \text{and} \qquad \Delta\hat{E}^{\text{ROE}} = \frac{E_B - E^*}{E^*} \tag{5.2}$$

where E^* is half of the beam energy, E_B is the energy and $|\vec{p}_B|$ the momentum magnitude of the reconstructed B_{tag} from the ROE. The selections to keep an event on these variables are $\hat{m}_{\text{bc}}^{\text{ROE}} > 0.964$ and $-0.5 < \Delta\hat{E}^{\text{ROE}} < 0.1$. The beam constrained mass $m_{\text{bc}}^{\text{ROE}} = \sqrt{E^{*2} - |\vec{p}_B|^2}$ as well as the energy difference $\Delta E^{\text{ROE}} = E_B - E^*$ can both be used in order to test how well the B_{tag} is reconstructed from the ROE. The beam constrained mass describes the invariant mass, which is given if we replace the energy of the ROE with the half of the beam energy. For a perfectly reconstructed momentum of the B_{tag}, $m_{\text{bc}}^{\text{ROE}}$ should lie at the B mass at $5.28\,\text{GeV}$ and the normalized beam constraint mass therefore at 1. The energy difference shows the difference between the reconstructed energy of B_{tag} and the energy given by the beam. For an ideally reconstructed B_{tag} energy, the energy difference as well as the normalized energy difference should be consistent with zero. Lastly, an event is kept if it has an R_2 value smaller than 0.5. An explanation of this variable is given in equation 5.5. The distributions of all variables for which a selection was made are given in Figure 5.2, the cuts chosen are also shown. The efficiencies after each performed cut is shown in Table 5.1 for signal as well as for the background.

5.1.2 Signal Side and Rest of Event Selection

The following selections are performed on the signal side. The reconstructed charged track has to meet a requirement of a muon ID > 0.9 and has to originate from the interaction point. We thus select
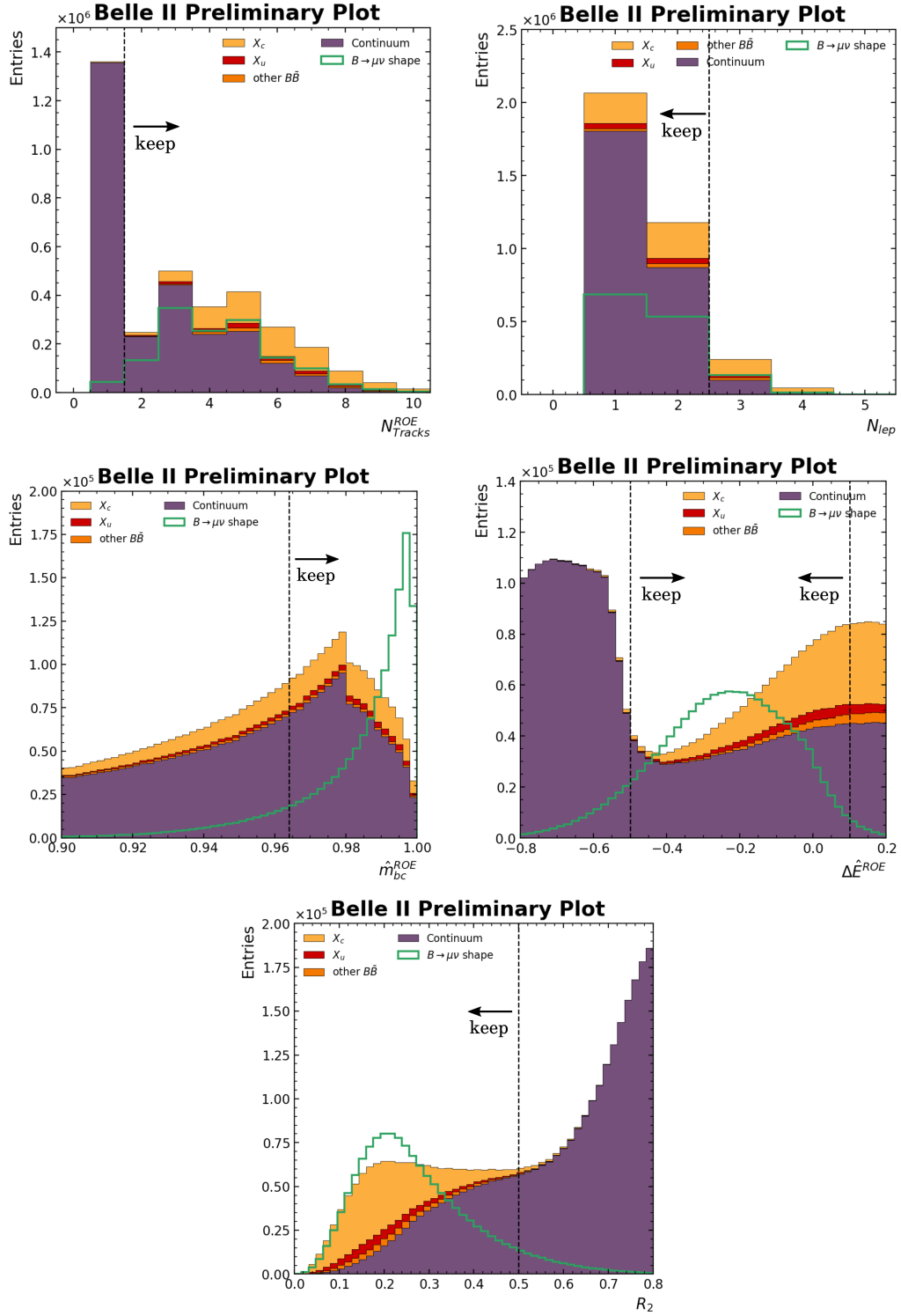
Figure 5.2: Distributions of the variables used in order to perform the event selection. The distributions of the number of tracks in the ROE N_{Tracks}^{ROE}, the number of leptons in the event N_{lep}, the normalized beam constraint mass $\hat{m}_{bc}^{ROE}$, the normalized energy difference $\Delta\hat{E}^{ROE}$ as well as the distribution of R_2 are depicted for the MC samples. The signal shape is overlaid and the chosen cut values are shown with the dotted line.

Selection	$B \to \mu\nu_\mu$	Generic	Continuum
$N_{\text{Tracks}}^{\text{ROE}}$	0.966	0.995	0.494
N_{lep}	0.886	0.760	0.916
$\hat{m}_{\text{bc}}^{\text{ROE}}$	0.881	0.516	0.454
$\Delta\hat{E}^{\text{ROE}}$	0.912	0.745	0.710
R_2	0.949	0.977	0.516

Table 5.1: Efficiencies of the selections for the signal, generic and continuum sample. It should be noted, that the efficiencies were calculated after loose precuts.

on the z coordinate of the point of closest approach $|z_0| < 2\,\text{cm}$ and the distance to the point of closest approach in the $r - \phi$ plane $d_0 < 0.5\,\text{cm}$. The momentum in the center of mass frame of the charged track has to be greater than $1.8\,\text{GeV}$. Although we expect the p_μ^* at a higher value, this loose selection is chosen in order to perform solid comparisons of data to MC. Later in this analysis we will only look at events where the boosted muon momentum p_μ^B is greater than $2.2\,\text{GeV}$.

Selections performed on the ROE are used to remove tracks and clusters that are unlikely to originate from the B_{tag} decay. Therefore a mask is applied to the ROE, selecting only tracks with a transverse momentum $p_t > 0.1\,\text{GeV}$ that are compatible with coming from the interaction point, $|z_0| < 3\,\text{cm}$ and $d_0 < 0.5\,\text{cm}$. Furthermore, tracks are further considered if they have an energy less than $5.5\,\text{GeV}$ and have a polar angle θ which lies inside the acceptance of the CDC, $17° < \theta < 150°$. Only clusters with an angle that is also within the acceptance of the CDC and that have a cluster energy $0.1\,\text{GeV} < E < 5.5\,\text{GeV}$ are retained. The selections on the ROE were adapted from reference [30] since the $B^+ \to K^+ \nu\bar{\nu}$ decay investigated using inclusive tagging is similar to the one investigated in this thesis.

After selecting the tracks and clusters, a cleanup of the ROE is performed, following [7]. Reconstructed photons are kept in the ROE if there is no associated track found and if they have an energy greater than $50\,\text{MeV}$ in the barrel region, greater than $100\,\text{MeV}$ in the forward endcap region or greater than $150\,\text{MeV}$ in the backward endcap region. All charged tracks are further treated with the pion mass hypothesis. A loose impact parameter selection of $|z_0| < 20\,\text{cm}$ and $d_0 < 10\,\text{cm}$ is performed. Pions originating from K_s decays are not further considered. Additionally, tracks tagged as curlers by the corresponding *basf2* module are rejected. Using all the leftover clusters and charged tracks the ROE is then built. The efficiencies after the event, the signal side and the ROE selections are shown in Table 5.2.

5.2 Continuum Suppression

The background in the signal region is mainly due to continuum events. In order to separate signal from continuum, a multivariate analysis is performed employing the training of FastBDTs. A MC generated data sample of $200\,\text{fb}^{-1}$ is used for the training of signal events versus continuum events. We use approximately twice as many signal events as continuum events, since it became evident that this avoids undertraining. 75% of the sample is used for training while 25% are used for testing. Only

events with $p_\mu^* \geq 2.2\,\text{GeV}$ were considered since we primarily want to suppress the continuum in the signal region. The event shape of continuum events is quite different to that of $B\bar{B}$ events, which is depicted and explained in Figure 5.3 and Figure 5.4.
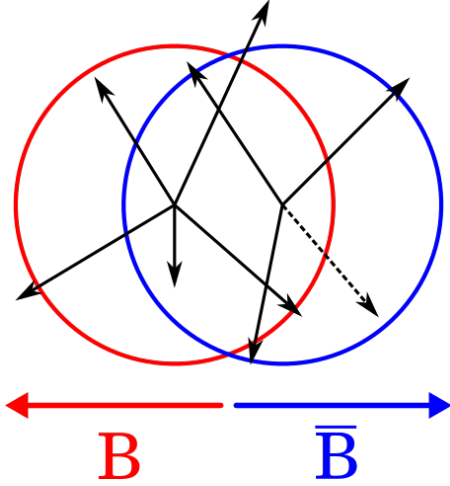
Figure 5.3: Depiction of the spherical shape of $B\bar{B}$ events. The two B mesons are produced approximately at rest leading to an isotropic distribution of its decay products. Figure adapted from [31].
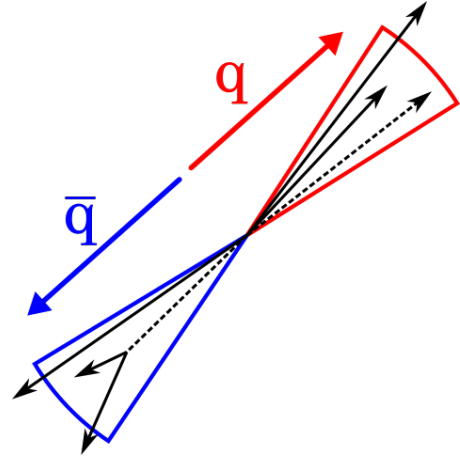
Figure 5.4: Depiction of the jet-like shape of $q\bar{q}$ events. At production, the two quarks each carry away half of the center of mass energy. Jets are formed in the hadronization process. Figure adapted from [31].

The event shape is described by continuum suppression variables explained in the following, which subsequently are used for the training of the BDT. The explanation is based on references [9], [5] and the *basf2* documentation.

Thrust

The thrust scalar T is defined as

$$T = \frac{\sum_{i=1}^{N} |\vec{T} \cdot \vec{p}_i|}{\sum_{i=1}^{N} |\vec{p}_i|} \tag{5.3}$$

for a set of N particles with an individual momentum $\vec{p}_i$ and a thrust axis $\vec{T}$, which is defined as the unit vector along the maximal projection of the particle set. The thrust axis can be calculated for the signal side B meson Thrust_B and for the ROE Thrust_O respectively.

$\cos\theta_T$

$\cos\theta_T$ defines the cosine of the angle between the thrust axis of the B_{sig} candidate and the thrust axis constructed from tracks and clusters in the ROE. Particles resulting from $B\bar{B}$ decays are showing an isotropic distribution resulting in randomly distributed thrust axes and thus display an uniform distribution in $|\cos\theta_T|$. The $|\cos\theta_T|$ distribution of continuum events, instead, peaks at higher values, since these events have a more jet-like structure.

cos θ_B

The angle between the thrust axis of the particles originating from the B meson decay and the beam axis is called θ_B. The distribution of the cosine of this angle is expected to be uniform for $B\bar{B}$ events. For continuum events, one expects instead a $1 + \cos^2 \theta_B$ dependence, due the more jet-like structure.

Cleo Cones

The Cleo Cones essentially describe the sum of all momenta in nine different cones around the thrust axis of the B candidate. The Cleo Cones are calculated in bins of $10°$ around the axis. Since continuum events are less spherical, the momentum flux in these events is expected to be higher in the Cleo Cones with small opening angles. The first Cleo Cone cc_1 is not included in the BDT training since signal and continuum are highly correlated in this variable.

Fox-Wolfram moments

The Fox-Wolfram moments (FWM) describe the phase-space distribution of both momentum and energy flow occurring in a event. The k-th order FWM for a collection of N particle which each carry the momentum $\vec{p}_i$ is defined as

$$H_k = \sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| P_k(\cos \theta_{ij}) \tag{5.4}$$

where P_k is the Legendre polynomial of k-th order and θ_{ij} is the angle between the particles with $\vec{p}_i$ and $\vec{p}_j$. Normalized FWM ratios are often used, defined as

$$R_k = \frac{H_k}{H_0} \tag{5.5}$$

and in the BDT training we use the ratio of the muon on event level $R_2^{\mu, event}$. Furthermore, the Fox-Wolfram moments can be modified with

$$h_k^l = \frac{\sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| P_k(\cos \theta_{ij})}{\sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j|}. \tag{5.6}$$

If particle i and particle j are both taken from the ROE, l is denoted as $l = oo$. If particle i is taken from the signal side and j from the ROE, l takes on $l = so$.

Event variables

In addition to the continuum supression variables, some event variables are also included in the training process. The number of tracks in the ROE N_{Tracks}^{ROE}, the number of leptons in the event N_{lep}, as well as the normalized beam constrained mass $\hat{m}_{bc}^{ROE}$ and normalized energy difference $\Delta\hat{E}^{ROE}$, as defined in Equation 5.2, are also considered for signal and continuum separation.

5.3 Results of the BDT Training

The training is performed using the following set of hyperparameters. We use 200 trees, each with 4 levels and each level containing 8 cuts. A sampling rate of 0.5 and a shrinkage of 0.05 is chosen. A grid search was performed, in order to determine the set of hyperparameters resulting in the highest area under the ROC curve. Unfortunately, the optimal set led to undertraining, so the next best set of parameters not leading to undertraining was kept. The BDT was firstly trained using all the continuum suppression and event variables introduced before and the output was evaluated. The eight most discriminating variables were kept as long as they are not highly correlated to each other. The training was performed again using the reduced set of variables and the results are shown in the following. In Figure 5.6 the distributions of the variables used in the training are shown. A data sample of $200\,\mathrm{fb}^{-1}$ was used and the background contribution represents the continuum. The BDT shows a good separation power which is expressed in the area under the ROC curve of 0.95. The ROC curve is shown in Figure 5.7. The overtraining plot shown in Figure 5.5, as well as the normalized classifier output for signal and continuum, shows good separation. The BDT training obtained can then be applied to the MC samples with $62.8\,\mathrm{fb}^{-1}$, resulting in a C_{out} distribution shown in Figure 5.8. The plots showing the correlations as well as the distributions of all other continuum suppression variables are shown in the appendix A.1.
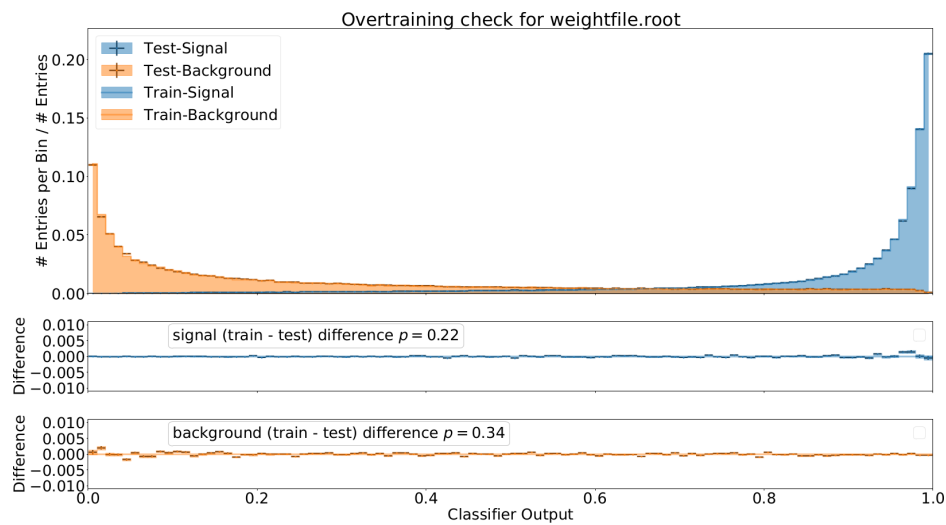
Figure 5.5: Overtraining plot obtained in the BDT training with p-values in the acceptable region.

(a) $R_2^{\mu,event}$

(b) h_{22}^{so}

(c) $\cos\theta_T$

(d) $\cos\theta_B$

(e) $\hat{m}_{bc}^{ROE}$

(f) h_{20}^{so}

(g) $\Delta\hat{E}^{ROE}$

(h) h_{24}^{so}

Figure 5.6: Normalized distributions of the most discriminating variables used in the BDT training process. The background represents the continuum. The distributions for the variables $R_2^{\mu,event}$, h_{22}^{so}, $\cos\theta_T$, $\cos\theta_B$, $\hat{m}_{bc}^{ROE}$, h_{20}^{so}, $\Delta\hat{E}^{ROE}$ and h_{24}^{so} are shown.
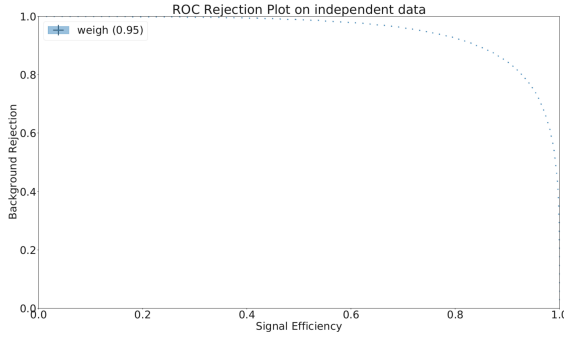
Figure 5.7: Receiver Operating Characteristic curve obtained in the BDT training. The area under the curve is shown.
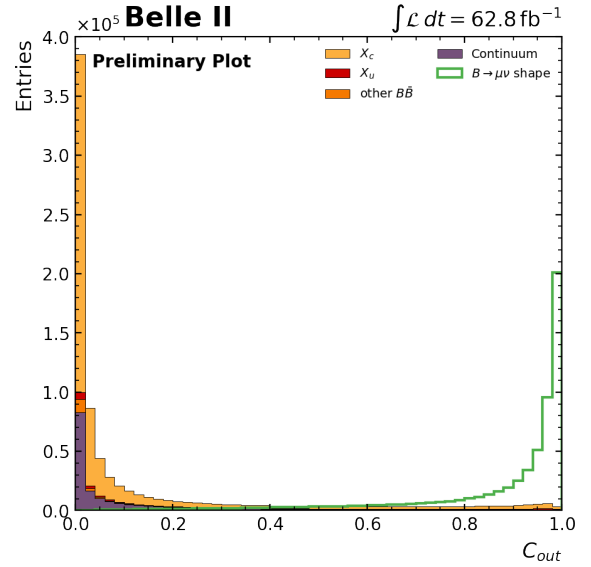
Figure 5.8: Distribution of the output classifier C_{out} for the scaled MC samples.

5.3.1 Data-MC Agreement of the BDT Input Variables

In Figure 5.9 the comparison of the blinded data set to the generated MC data of the eight most discriminating variables is shown. The data-MC agreement for all the other parameters is shown in appendix A.3. This is done after a cut on the output classifier $C_{out} > 0.98$ has been performed. This cut is chosen based on information gained in section 7.1.2. The efficiencies of the cut on the output classifier are shown in Table 5.2. We have removed more than 99.99 % of background while retaining approximately 19 % of signal. The data-MC agreement is good for the variables in most regions and no significant differences are observed. The normalized residuals do not show large discrepancies.

Sample	Event, signal and ROE selection	C_{out} cut
$B \to \mu\nu_\mu$	56.937 %	18.968 %
Generic	0.506 %	(3.658×10^{-3}) %
Continuum	0.055 %	(0.112×10^{-3}) %

Table 5.2: Efficiencies of the selection on the signal, continuum and generic MC as described in section 5.1.1 and 5.1.2 and the efficiencies after a cut on the BDT output classifier $C_{out} > 0.98$.
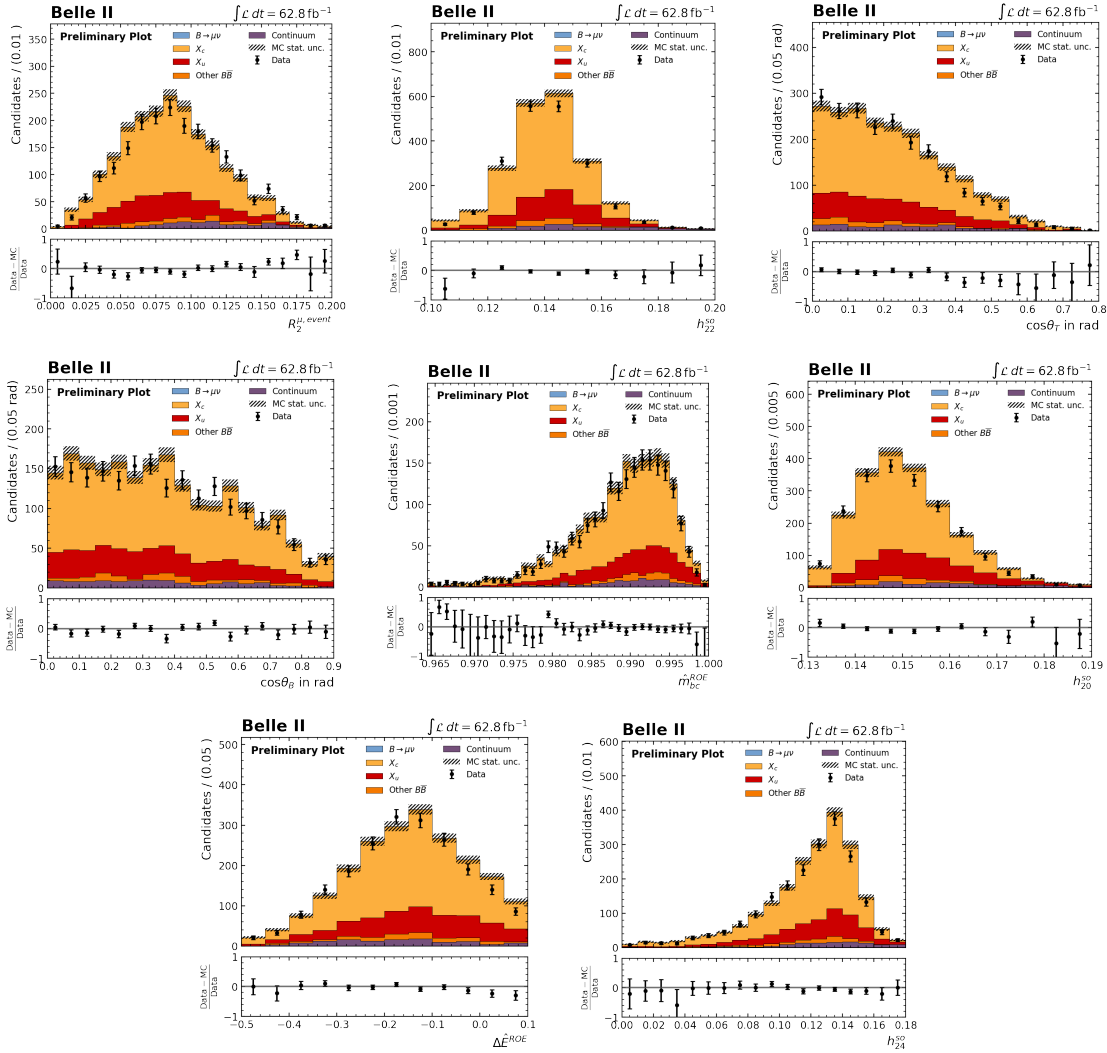
Figure 5.9: Comparison of the blinded data set to generated MC samples for the variables used in the BDT training process.

Reconstruction Errors of B_tag

The resolution of the boost vector constructed from the B_tag kinematics has a high impact on the resolution of the muon momentum in the B meson rest frame, as stated in section 5.1. Different errors in the reconstruction of the B_tag are presented in this chapter. As the reconstruction errors of p_z in Figure 6.1 shows, which is given as an example, the generated B_tag kinematics differ highly from the reconstructed ones, resulting in a broad muon momentum. For this part of the analysis, approximately 90000 generated signal events were investigated. The main aim of this study is to identify all the errors in the reconstruction, in order to explain the resolution of p_z and to draw conclusions from this.
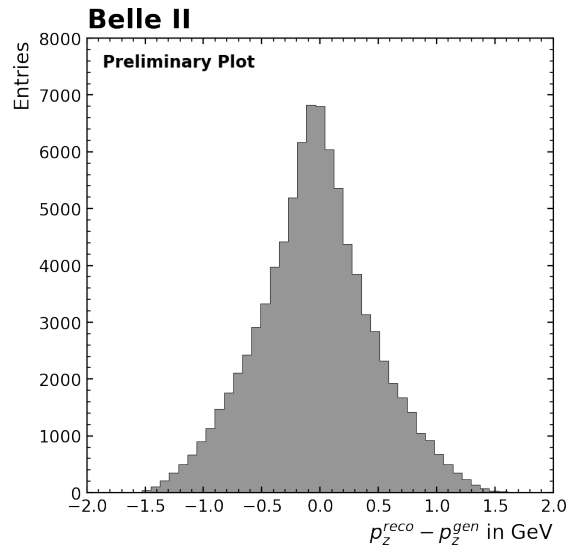
Figure 6.1: Error of the p_z component of the B_tag reconstructed from the ROE.

The differences between the reconstructed and generated B_tag are further referred to as errors denoted as δ. There are mainly two different types of errors in the reconstruction. The first type are errors where tracks and clusters from the B meson decay are generated, but not reconstructed. Thus they are errors due to missing four-momentum. The other type describes tracks and clusters which are reconstructed and thus appear in the ROE, but were not generated as decay products. Namely these are errors due to extra four-momentum. Furthermore, the errors are divided into different subcategories

which are explained in the following. These subcategories are fully exclusive, so if a reconstructed or generated particle falls in one subcategory it will not appear in another one. The errors are shown for the p_z component of the B_{tag}, since the p_z component shows a worse resolution compared to the other components and a calibration of this component proves to be useful at a later point in section 7.1.2. The errors in the other components are shown in appendix B.

6.1 Missing Information Errors

In this section the errors due to missing information in the reconstruction of the B_{tag} are shown. The errors are divided into the following subcategories.

Detector Acceptance Error

Particles from the B meson decay that lie outside the acceptance of the Belle II detector are lost and cannot be reconstructed. The angles of the generated particles in the B_{tag} decay are compared with the acceptance of every sub-detector. If the particle lies outside the acceptance of every single sub-detector, the momentum of every generated particle within the event that lies outside the acceptance is added up, to estimate how much four-momentum is lost due to the full detector acceptance. The distribution of the error on p_z is shown in Figure 6.2.

Missing Neutrinos Error

Neutrinos cannot be observed by the Belle II detector and are missing in the reconstructed, but not in the generated, B meson decay. The error on p_z due to missing four-momentum due to unreconstructed neutrinos is shown in Figure 6.2.

Other Missing Particles Error

The majority of events contains at least one missing particle in addition to the ones described above. A particle is generated as a decay product of a B meson, but has no matching reconstructed particle. These particles are mainly generated K_L, $\pi^\pm$ and γ. The γ in the most cases originate from π^0 decays. In the reconstruction process the γ is not assigned to the π^0 and thus is missing. This error can be attributed to the non-ideal detector efficiency. The error distribution is shown in Figure 6.2.

Error due to K_L and KLM clusters

On the reconstruction side, neutral particles which are created from a KLM cluster are not included in this study. Since the KLM is only used for particle identification the energy of the traversing particle is not fully measured. So even if a K_L is detected, its four-momentum cannot be used. To compare the reconstructed B_{tag} to the generated one, we subsequently have to introduce this error. This is different to the previously discussed error, since the K_L here do have a matching reconstructed particle. We expect an error due to reconstructed K_L with an assigned KLM cluster $\delta p_{z,K_L}^{\text{KLM cluster}}$, as well as an error due to reconstructed K_L without a KLM cluster $\delta p_{z,K_L}^{\text{no KLM cluster}}$. Additionally, an error due to any generated particle matched to a reconstructed one created from a KLM cluster $\delta p_z^{\text{KLM cluster}}$ is expected. All three error distributions are shown in Figure 6.3.
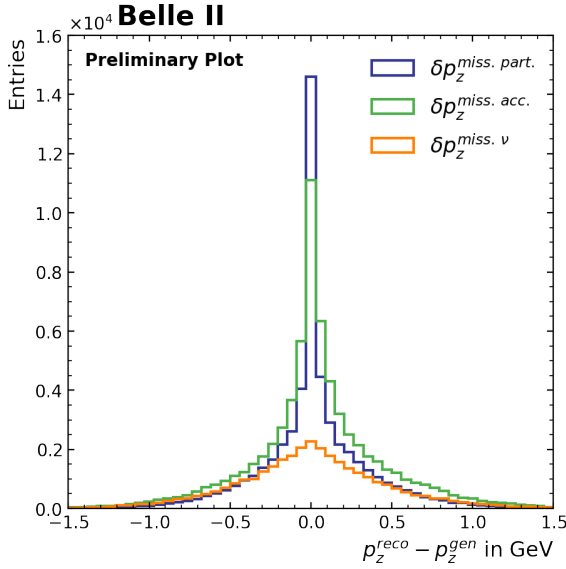
Figure 6.2: Distribution of the errors due to the acceptance, missing neutrinos and missing particles for the p_z component of the B_{tag}.
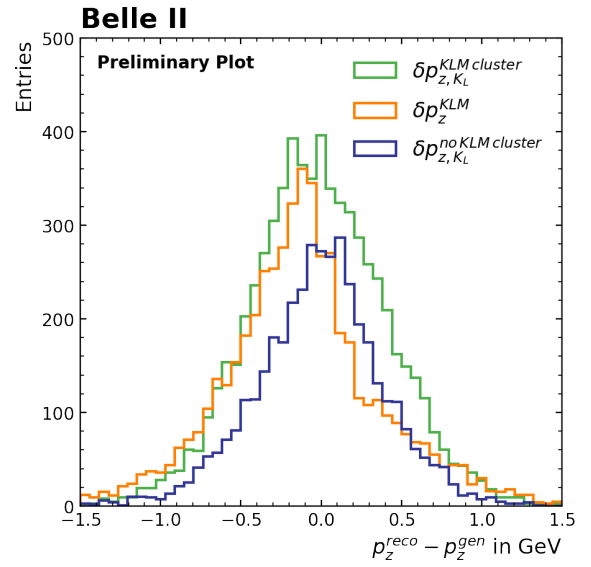
Figure 6.3: Distribution of the errors due to not including KLM information for the p_z component of the B_{tag}.

Final State Radiation Error

Photons which originate from final state radiation can also be an error source. The radiative photons are generated by PHOTOS and are not always reconstructed. The distribution of the error due to unreconstructed final state radiation is shown in Figure 6.4.

Wrongly Generated Decays Error

During this study we observed decays that were wrongly generated. Mainly the decay $a_1(1260)^+ \to f_0(500)\pi^+$, where $a_1(1260)^+$ is the decay product of a $B^\pm \to a_1(1260)^+ X$ decay, does not follow the energy and momentum conservation. This problem was reported but not resolved up to this point. We still include this as an error, which is depicted in Figure 6.4.

Unexplained Matching Error

The error described here is observed for particles which do not fit into one of the categories listed above, but that are still responsible for differences between the reconstructed and generated B_{tag}. It is not fully understood where the error arises from. Mainly we observe a reconstructed particle which is matched to a generated final state particle from the B decay, but the matched reconstructed particle is not included in the reconstructed ROE. The error is not dominant, but it is still present and shown in Figure 6.4.
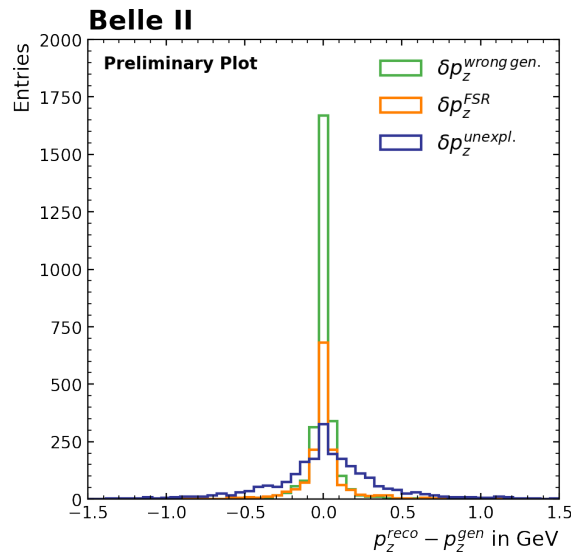
Figure 6.4: Distribution of the errors due to final state radiation, wrongly generated decays and unexplained matching for the p_z component of the B_{tag}.

6.2 Extra Information Errors

The errors described in this section are due to extra particles which are reconstructed but are not generated as final state particles of the B meson decay.

Non-Primary Particles Errors

Non-primary particles, such as particles from beam background and material interaction, are reconstructed, but they do not originate from the B meson decay. Thus they add extra four-momentum to the ROE, leading to the error shown in Figure 6.5.

Misidentified Particles Errors

This error arises from reconstructed particles with a wrongly assigned mass hypothesis. Since the calculation of the energy of the reconstructed particle is based on a given mass hypothesis of the particle, an error arises if the mass hypothesis is wrong. This error is shown in the distribution depicted in Figure 6.5. The misidentification of a particle should result in an error in the reconstructed energy, but not in the reconstructed three-momentum. It is not fully understood why we observe an error in the p_z component.

Curling Particles Error

If a particle has a low transverse momentum it is possible for it to curl back into the detector due to the applied magnetic field. In the reconstruction process this is shown by two reconstructed particles that match to one generated one. If the reconstructed particle is tagged by the corresponding *basf2* module, the error is calculated and it is shown in Figure 6.6.

Clone Error

Double-matched particles can occur in two cases. Either two reconstructed particles are matched to one generated one or two generated particles can match to one reconstructed one. Neither one comes from curling particles, but in both cases this leads to an error in the reconstruction. The error due to clones is shown in Figure 6.6.
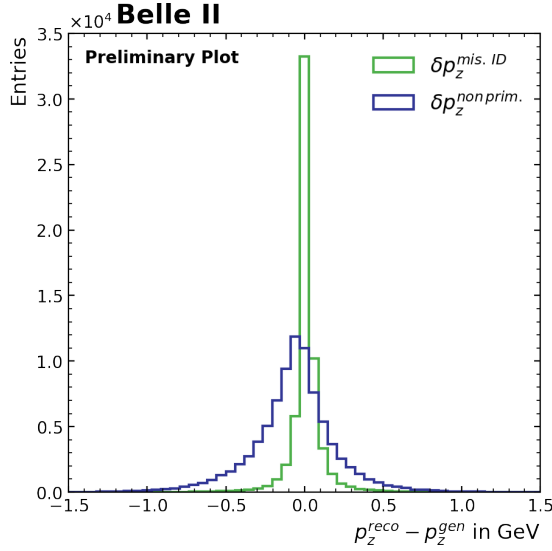
Figure 6.5: Distribution of the errors due to non primary and misidentified particles for the p_z component of the B_{tag}.
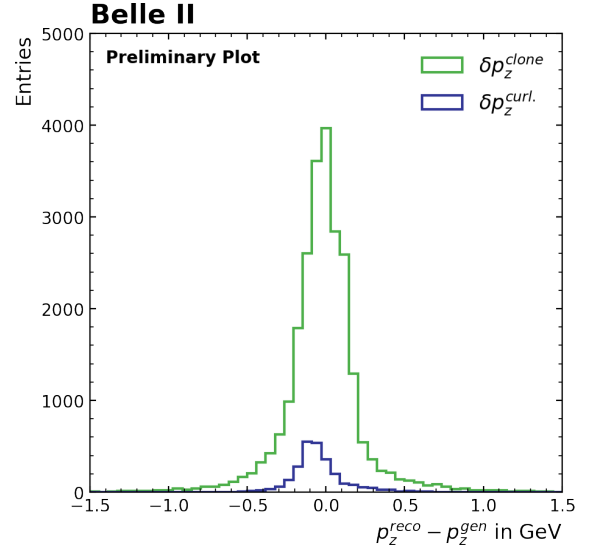
Figure 6.6: Distribution of the errors due to curlers and clones for the p_z component of the B_{tag}.

6.3 Results and Measures

Now that all the errors in the reconstruction of the B_{tag} are identified, they should explain the resolution shown in Figure 6.1. In order to test this, we subtract all the errors due to extra four-momentum from the resolution $\Delta p_z = p_z^{\text{reco}} - p_z^{\text{gen}}$. Furthermore, we add all the errors due to missing four-momentum. If all errors have been accounted for, the result of $\Delta p_z - \Sigma \delta p_z$ should equal zero. Figure 6.7 shows the distribution of this value for the p_z component. The distributions for the other kinematic variables are shown in appendix B.

It can clearly be seen, that the result is not consistent with zero. There could be specific event typologies where particles do not fall into one of the error categories explained above. Nevertheless, the events are distributed close to zero with a standard deviation of $\sigma \approx 0.164\,\text{GeV}$ which indicates that the errors that have the biggest impact on the resolution have been identified. The impact of the different errors on the reconstruction is shown in Table 6.1.

The main errors in the reconstruction of the B_{tag} kinematics are due to missing information due to particles which are not reconstructed, as well as due to non-primary particles. The issue is that very little done to avoid these errors. Particles which are lost due to the detector acceptance cannot be recovered, just as lost four-momentum due to neutrinos and other missing particles cannot.
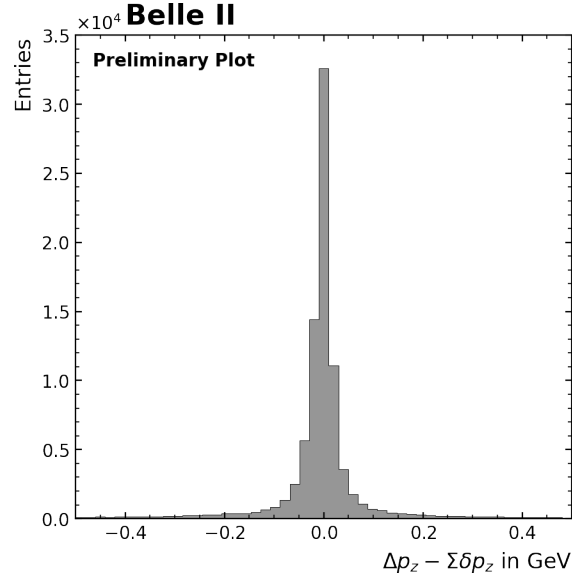
Figure 6.7: Verification that most errors in the p_z component of the ROE have been identified. If all errors are correctly identified, $\Delta p_z - \Sigma \delta p_z$ equals zero.

Error	Impact on p_z resolution
Detector acceptance	30.92 %
Non primary	18.49 %
Missing neutrinos	18.35 %
Missing particles	16.11 %
Clone	4.16 %
KLM clusters	3.75 %
K_L^0 with KLM cluster	3.27 %
Misidentified	2.07 %
K_L^0 without KLM cluster	1.68 %
Unexplained matching	0.83 %
Curling particles	0.22 %
Final state radiation	0.11 %
Wrongly generated	0.05 %

Table 6.1: Impact of the different errors on the reconstruction of the p_z component of B_{tag}.

Misidentification of particles can be somewhat reduced by setting more stringent cuts on the PID. However this would also result in particles that are not reconstructed if they do not survive the cut, and thus this results in additional reconstruction errors due to missing particles.

Non-primary particles can be removed by the application of an ROE mask. The ROE mask, adapted from [30], as well as the ROE cleanup is described in section 5.1.2. The effect on the errors were studied and therefore two categories are introduced. Category I shows the errors without applying the mask and ROE cleanup, while Category II shows the errors after the mask has been applied and the cleanup has been performed. As mentioned earlier, most of the errors simply cannot be avoided, so the mask and cleanup have no effect on them.

In the following three individual errors are shown and described, where an effect was observed. The distribution of the error due to curling particles is shown in Figure 6.8. In the process of the ROE cleanup particles tagged as curlers by the *basf2* module are not further considered. The efficiency of the tagger does not seem to be very high, but still the standard deviation of the error distribution is reduced by approximately 11.8%. The distribution of the error due to non-primary particles is shown in Figure 6.9. The applied mask removes a small number of non-primary particles leading to a reduction of 13.7% of the standard deviation. Looking at the distribution of the error due to final state radiation in Figure 6.10 shows that the cleanup and mask application has the opposite of the desired effect. Photons which lead to a small error are discarded and thus only events with higher errors remain. This results in an increase of 38.4% of the standard deviation of the error distribution. Since this only occurs in a small number of events compared to the other errors, this has a negligible effect.
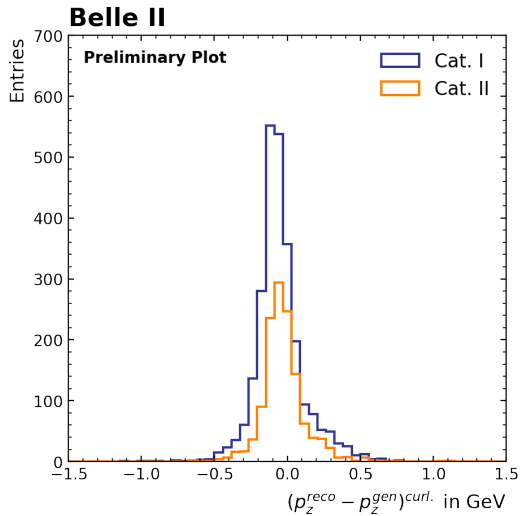
Figure 6.8: Distribution of the error in the p_z component due to curling particles, shown for category I and category II.
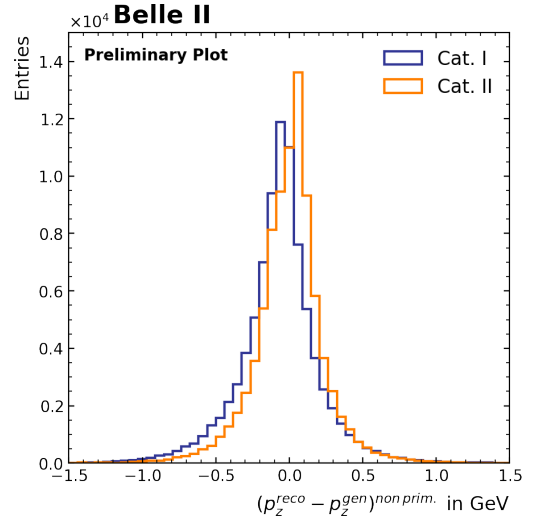
Figure 6.9: Distribution of the error in the p_z component due to non-primary particles, shown for category I and category II.

The overall effect of the applied mask and ROE cleanup is shown in the resolutions in Figure 6.11 for the two categories. A decrease of 4.2% of the standard deviation of the resolution is observed after applying the mask and cleanup of the ROE. Nevertheless, the main errors in the reconstruction of the B_{tag} cannot be avoided, therefore other methods are used in order to precisely estimate the B_{tag} kinematics, which are described in the next chapter.
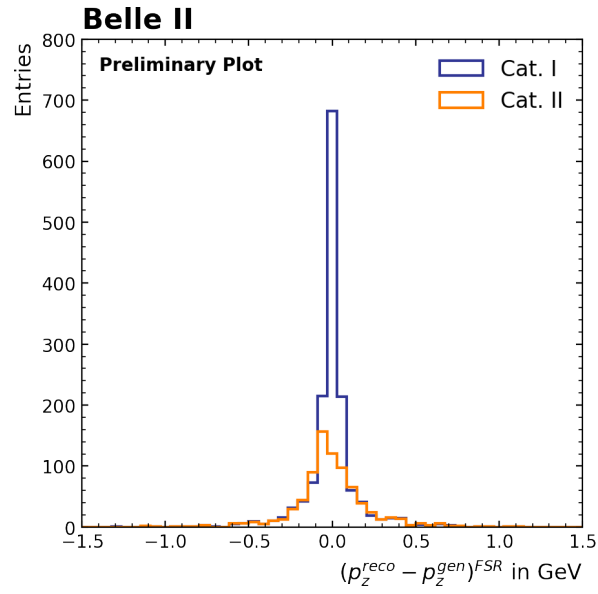
Figure 6.10: Distribution of the error in the p_z component due to final state radiation, shown for category I and category II.
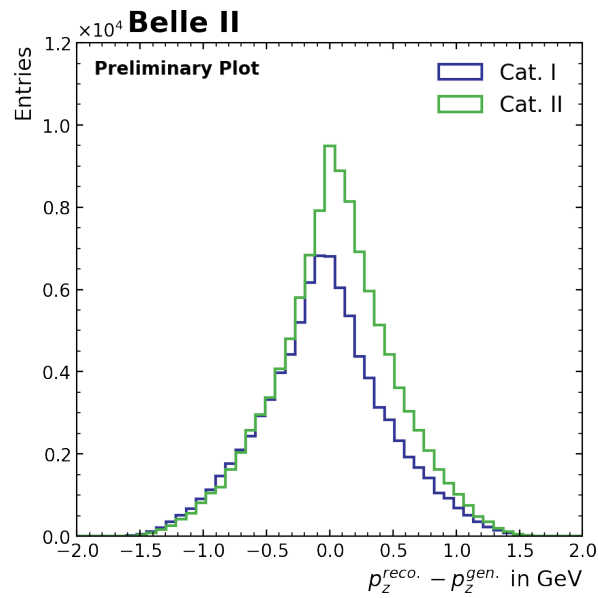
Figure 6.11: Overall resolution of the p_z component, shown for category I and category II.

Precise Estimation of the B_{tag} Four-Momentum

In this chapter the main results of this thesis are presented. As shown in the previous chapter 6 the reconstruction of the B_{tag} kinematics is highly deteriorated by missing information which is unavoidable. Therefore, the main task is to develop methods to improve the estimation of the B_{tag} four-momentum. This was done before in the Belle analysis of $B \to \mu\nu_\mu$ by Markus Prim et al. [7] and the methodology is explained in section 7.1. This method is further denoted as the *Belle method*. In order to perform a high precision boost of the muon momentum into the rest frame of the signal side B meson, the momentum components p_x, p_y and p_z, as well as the energy E of the B_{tag} have to be determined accurately. Subsequently Equation 5.1 can be used to construct the B_{sig} kinematics. In section 7.2 newly developed methods are presented and compared to the *Belle method*.

Several things need to be considered when comparing two different methods. The first and highest priority is the minimization of the p_μ^B resolution. In addition though, it is crucial to investigate the influence of the methods on the background. If a method leads to high sculpting of the background, it is not further considered. The significance from a binned maximum likelihood fit to an Asimov data set is extracted in order to compare two methods. The Asimov data set is constructed by taking the expected values from the MC generated data samples. We therefore basically fit the MC data to its own expectation, giving us the possibility to estimate the statistical significance of the different MC samples and allowing us to predict which method will lead to a higher significance. The fit to the Asimov data set is a good measure of the impact of different methods, since it is sensitive to both the signal to background ratio and background sculpting.

7.1 Belle Analysis Method

The general strategy of the *Belle method* is the constraining of the four-momentum. In addition, a calibration of the p_z component of B_{tag} is performed.

7.1.1 Constrained B_{tag} Four-Momentum

The first step of the *Belle method* is constraining the momentum components. Due to the nature of the two-body decay of the $\Upsilon(4S)$, which is approximately at rest upon its decay, into a $B\bar{B}$ pair, the theoretically expected mean momentum magnitude of B_{tag} and B_{sig} in the center of mass frame,

further denoted with *, is known as

$$|\vec{p}_{\text{sig}}^{\,*}| = |\vec{p}_{\text{tag}}^{\,*}| = \frac{\sqrt{m_{\Upsilon(4S)}^4 - 4m_{\Upsilon(4S)}^2 m_B^2}}{2m_{\Upsilon(4S)}} = 0.332 \,\text{GeV} \tag{7.1}$$

with $m_{\Upsilon(4S)} = 10.579 \,\text{GeV}$ and $m_B = 5.279 \,\text{GeV}$. Using the expected momentum magnitude, we can also calculate the expected mean energy of B_{tag} and B_{sig} at the center of mass to be

$$E_{\text{sig}}^* = E_{\text{tag}}^* = \sqrt{|\vec{p}_{\text{tag}}^{\,*}|^2 + m_B^2} = 5.289 \,\text{GeV}. \tag{7.2}$$

This energy is always used to perform the boost, since with no here-investigated method a better E_{sig}^* resolution is achieved. Assuming that the B_{tag} always has the fixed mean energy, this allows us to fix the momentum magnitude as well. Thus we are only sensitive to the direction of the B_{tag} momentum. In the following, the use of indices for the B_{tag} is omitted. If we refer to the momentum components, this will always refer to the B_{tag} ones. The B_{sig} kinematics are subsequently designated as such. The momentum components are constrained by

$$p_{i,\text{constr}}^* = p_i^* \cdot \frac{0.332 \,\text{GeV}}{|\vec{p}_{\text{tag}}^{\,*}|} \qquad , i \in x, y, z \tag{7.3}$$

in order to rescale the three-momentum to that its magnitude is equal to the expectation. This preserves the polar angle θ as well as the azimuth angle ϕ. The energy is subsequently set to

$$E_{\text{constr}}^* = \sqrt{|\vec{p}_{\text{constr}}^{\,*}|^2 + m_B^2}. \tag{7.4}$$

This leads to a significant improvement in the resolution of the B_{tag} momentum components shown in Figure 7.3, compared to the resolution of the raw reconstruction of the B_{tag} shown in Figure 7.1. Comparing the constrained B_{tag} momentum components to the generated ones in Figure 7.4, differences in the shapes of the distributions are visible, but the outer boundaries of the distributions are showing better agreement confirming the constraining leads to a better result, compared to the distributions before constraining shown in Figure 7.2. The p_z component shows a bias towards negative values, which is probably a consequence of the fact, that many particles are lost in the forward direction of the detector due to the boost [8].

7.1.2 Calibration and Optimization of p_z

In the *Belle method* an additional calibration of the p_z component of B_{tag} is employed. We now reproduce this very well-designed calibration for Belle II. The constrained p_z component of the B_{tag} is grouped in bins in a range of $-0.332 \,\text{GeV} \leq p_{z,\text{constr}}^* \leq 0.332 \,\text{GeV}$ and then mapped against the mean of the generated p_z-component in the corresponding bin. The goal here is that we extract the
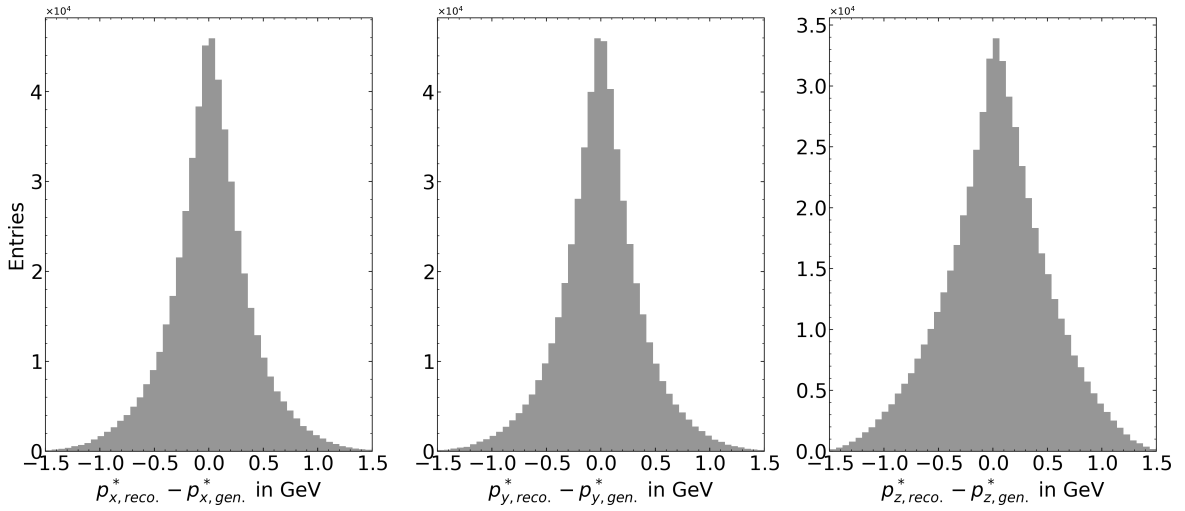
Figure 7.1: Resolution of the reconstructed B_{tag} momentum components for signal MC.
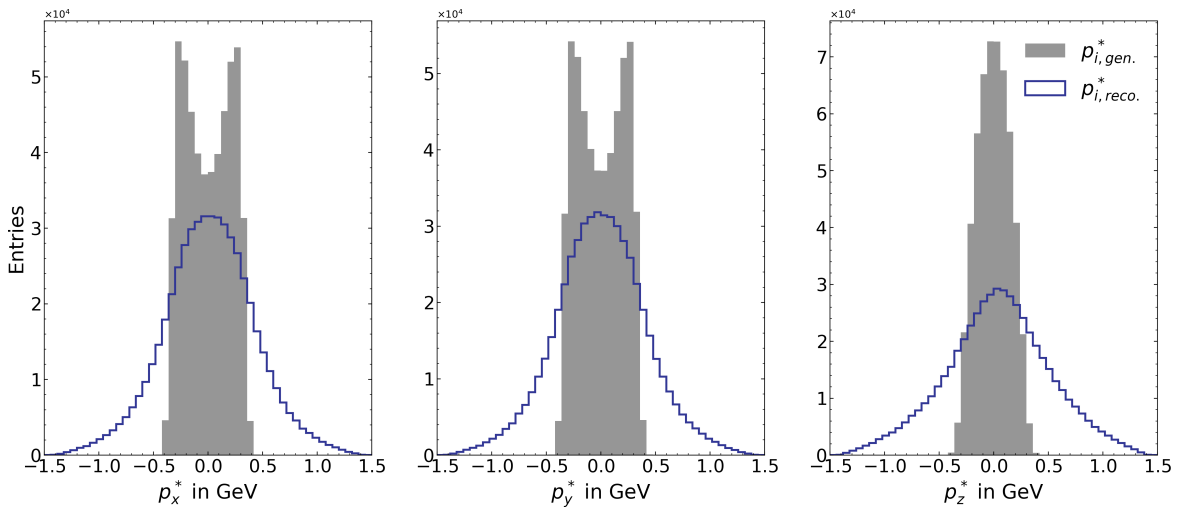
Figure 7.2: Comparison of the reconstructed B_{tag} momentum components to the generated ones for signal MC.
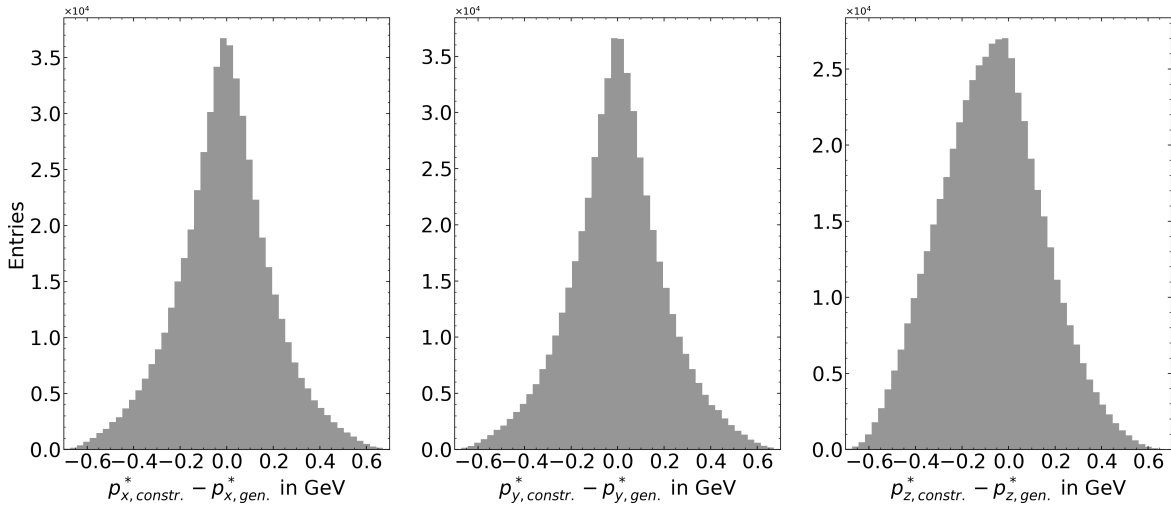
Figure 7.3: Resolution of the constrained B_{tag} momentum components for signal MC after constraining the B_{tag} kinematics.
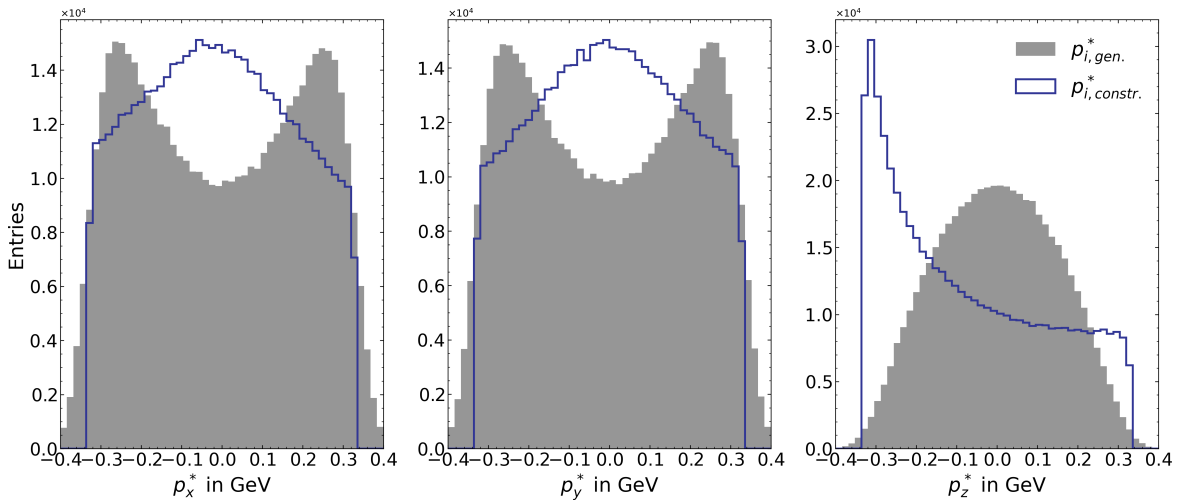
Figure 7.4: Comparison of constrained B_{tag} momentum components to the generated ones for signal MC after constraining the B_{tag} kinematics.

functional dependence via a smoothed spline fit, and thus later apply the function to $p^*_{z,\text{constr}}$ via

$$
\begin{aligned}
p^*_{z,\text{calib}} &= f(p^*_{z,\text{constr}}) \\
p^*_{T,\text{calib}} &= \sqrt{|\vec{p}^{\,*}_{\text{constr}}|^2 - (p^*_{z,\text{calib}})^2} \\
p^*_{i,\text{calib}} &= \frac{p^*_{T,\text{calib}}}{p^*_{T,\text{constr}}} \cdot p^*_{i,\text{constr}} \qquad , i \in x, y
\end{aligned}
\tag{7.5}
$$

in order to match it to the generated value and get a better estimate for the kinematics of B_{tag}. The reconstructed mean $p^*_{z,\text{constr}}$ versus the mean of the generated $p^*_{z,\text{gen}}$ in 26 bins and the smoothed spline fit are shown in Figure 7.5. A grid search was performed to determine the optimal number of bins. This was done by calculating the root mean square (RMS) of the difference between $p^*_{z,\text{calib}}$ and $p^*_{z,\text{gen}}$ for a different number of bins between 4 and 30, which is shown in Figure 7.6. For a perfectly reconstructed event, the RMS equals zero. The RMS does not change significantly above a value of 20 bins, but to ensure that the RMS does not get significantly smaller with a higher number of bins, we also considered much higher and random bin values as shown in Figure C.1 in the appendix. We decide to use 26 bins where a value of $RMS_{min} = 0.1447\,\text{GeV}$ was determined.
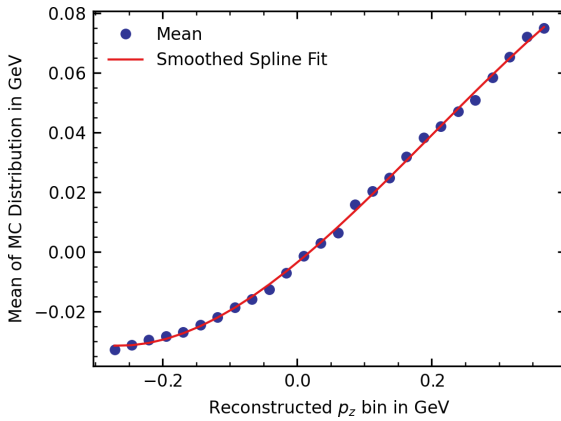
Figure 7.5: Binned constrained p_z versus generated p_z and smoothed spline fit.
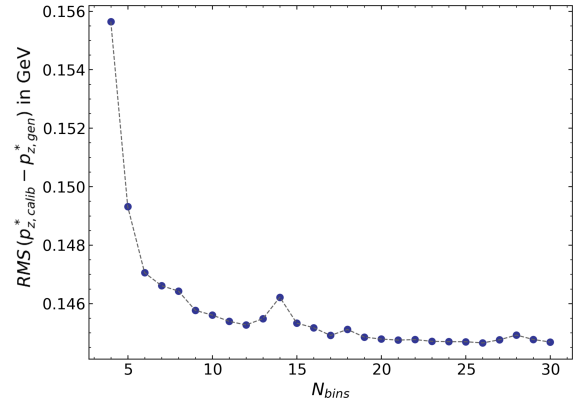
Figure 7.6: RMS of the difference between the calibrated and generated p_z component for different number of bins from 4 to 30.

We again consider the resolution of the B_{tag} momentum components and the direct comparison of the calibrated momentum components to the generated ones. Looking at the resolution of the $p^*_{x,\text{calib}}$ and $p^*_{y,\text{calib}}$ components in Figure 7.7, we see that the calibration leads to a slightly smaller width of the distributions compared to the constrained distributions in Figure 7.3. This is mainly explained by the improvement in the shapes shown in Figure 7.8. The distributions of $p^*_{x,\text{calib}}$ and $p^*_{y,\text{calib}}$ show a similar shape to the distributions of the generated components with more entries at the outer momentum boundaries and fewer events around zero. Looking at the resolution of the $p^*_{z,\text{calib}}$ component, we observe a better resolution compared to the $p^*_{z,\text{constr}}$ component. However, this occurs only because the calibration leads to a $p^*_{z,\text{calib}}$ distribution around zero, as can be seen in Figure 7.8.

To further improve the estimation of the B_{tag} kinematics, a correction factor is determined to scale the calibrated momentum with a fixed value in order to match to the generated momentum. A grid search is performed for different values of the correction factor ξ. The correction factor that leads to a
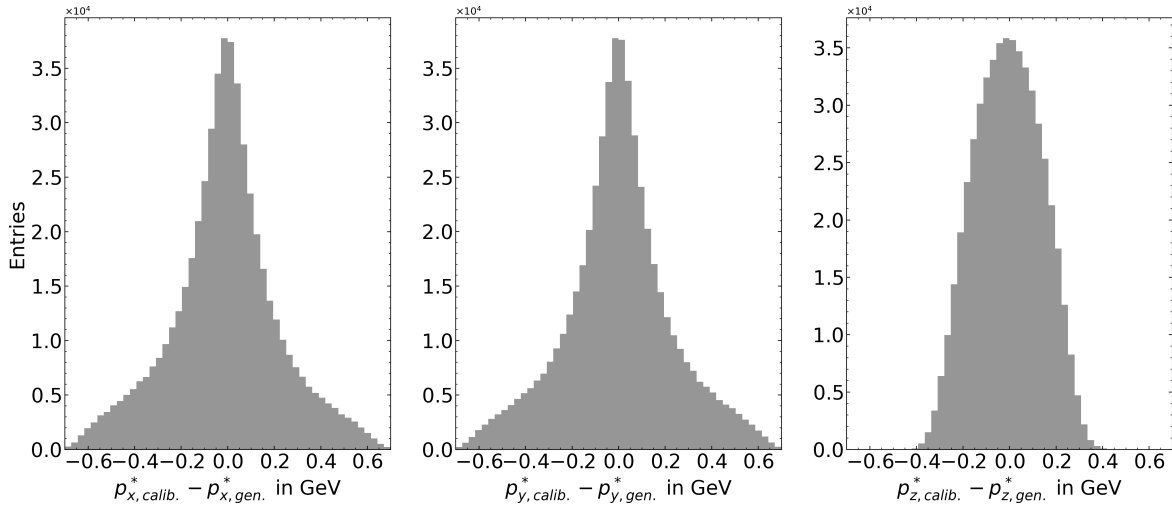
Figure 7.7: Resolution of the calibrated B_{tag} momentum components for signal MC after calibrating the p_z component.
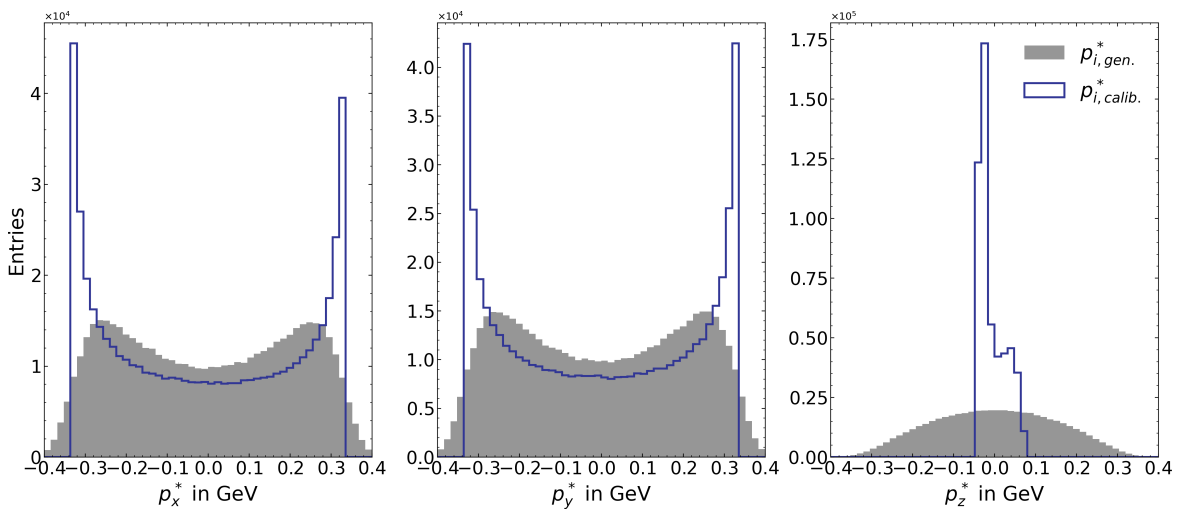
Figure 7.8: Comparison of calibrated B_{tag} momentum components to the generated ones for signal MC after calibrating the p_z component.

minimization of

$$D = \xi \vec{p}^{\,*}_{\text{calib}} - \vec{p}^{\,*}_{\text{gen}} \tag{7.6}$$

is maintained. For ideally reconstructed events, D would be zero. The value at which D is minimal is calculated to occur at $\xi = 0.607$ and is then used to construct the optimized momentum components of the B_{tag} via

$$p^*_{i,\text{opt}} \quad = \quad \xi \cdot p^*_{i,\text{calib}} \qquad , i \in x, y, z. \tag{7.7}$$

In the optimization process the polar angle θ changes while the azimuth angle ϕ is kept constant as depicted in Figure 7.9 and Figure 7.10.
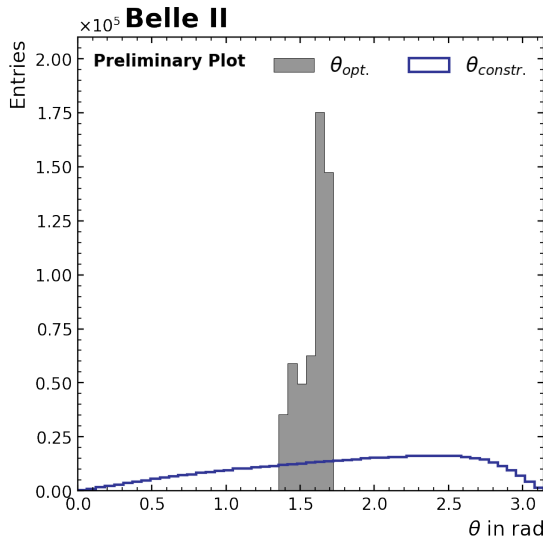
Figure 7.9: Distribution of the polar angle θ using the constrained and optimized momentum components. The calibration and optimization process changes the polar angle.
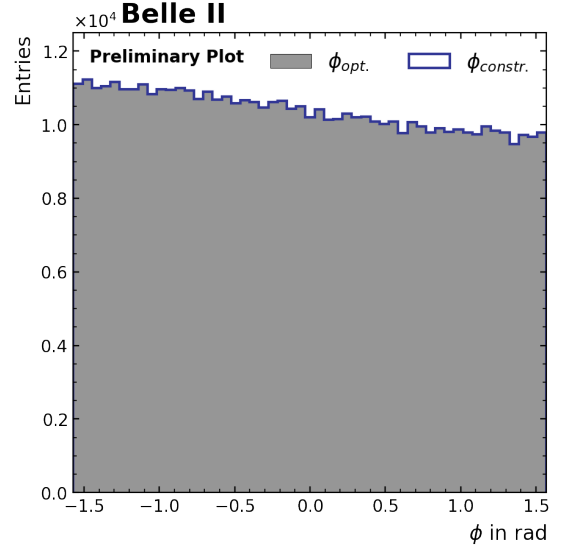
Figure 7.10: Distribution of the azimuth angle ϕ using the constrained and optimized momentum components. The calibration and optimization process changes the polar angle.

The resolution and comparison between the optimized and generated momentum components are shown in Figure 7.11 and Figure 7.12. The resolution of the $p^*_{z,\text{opt}}$ component does not change significantly compared to the calibrated component, as it is still mainly distributed around zero. The shapes of the $p^*_{x,\text{opt}}$ and $p^*_{y,\text{opt}}$ components change due to scaling with the correction factor. The outer limits of the distribution shift closer to zero, leading to a change in resolution.

Using the optimized kinematics, the momentum components of the B_{sig} can be constructed with

$$p^*_{B,\text{sig}} = \begin{pmatrix} \sqrt{\left|\vec{p}^{\,*}_{\text{constr}}\right|^2 + m_B^2} \\ -\vec{p}_{\text{opt}} \end{pmatrix} \tag{7.8}$$

and the boost of the muon momentum into the B_{sig} rest frame can be performed. The result is shown in Figure 7.13. Optimization of the B_{tag} kinematics leads to a more peaked structure. We then apply the same calibration and optimization to all generated MC samples as well as to the real data sample.
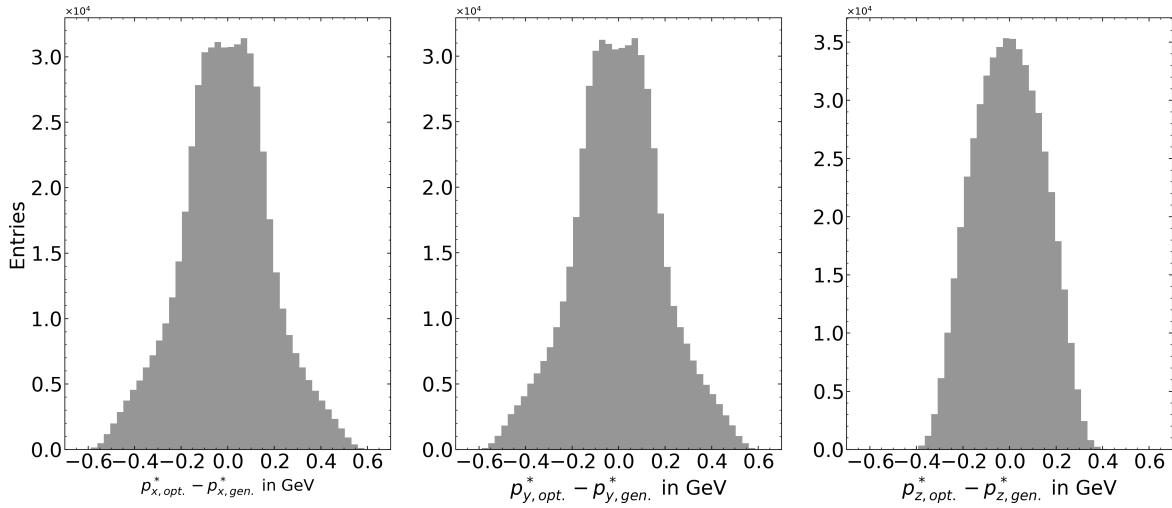
Figure 7.11: Resolution of the optimized B_{tag} momentum components for signal MC after optimizing the p_z component.
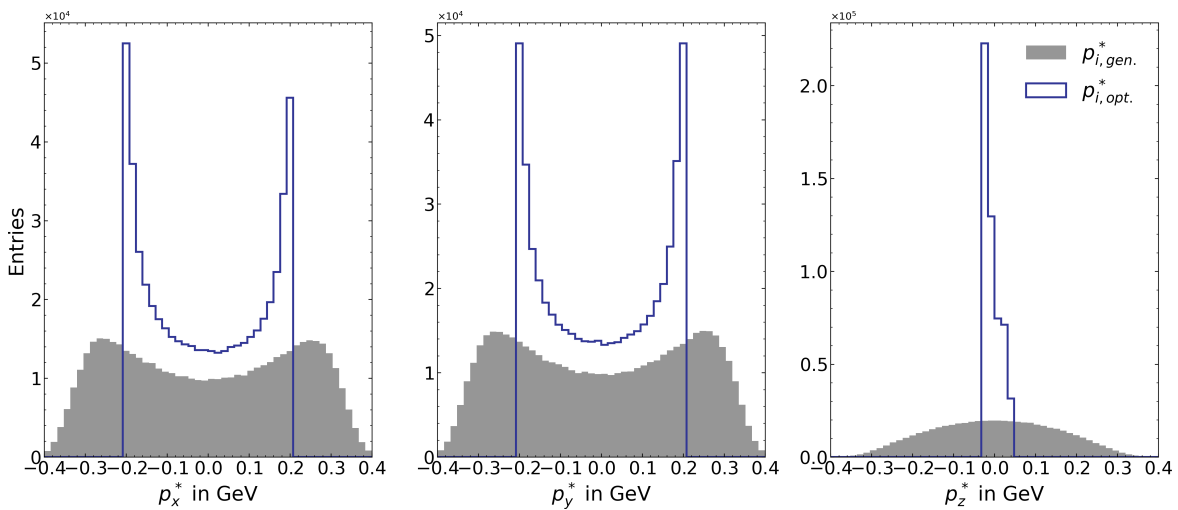
Figure 7.12: Comparison of optimized B_{tag} momentum components to the generated ones for signal MC after optimizing the p_z component.
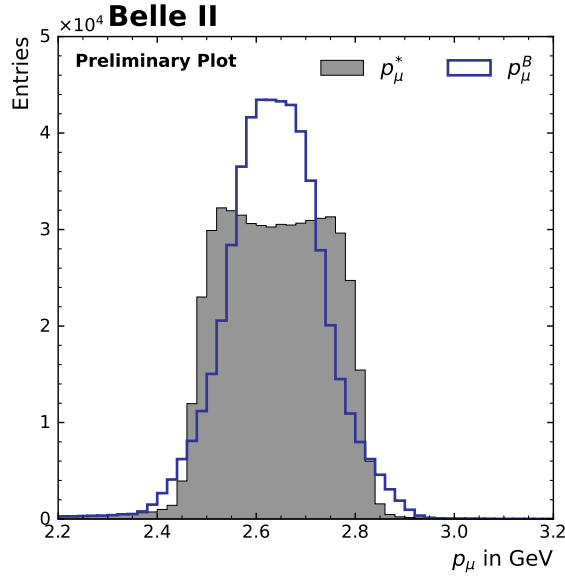
Figure 7.13: p_μ distributions in the center of mass and B rest frame of the signal MC sample, after optimizing the p_z component.

The next step is to suppress the background, which consists of the generated X_u, X_c, and continuum samples, as well as the other $B\bar{B}$ events. Since we are really only interested in background suppression in the signal region, we calculate the figure of merit

$$\text{FOM} = \frac{N_{\text{sig}}}{\sqrt{N_{\text{sig}} + N_{\text{bg}}}} \tag{7.9}$$

for different cuts on the output classifier C_{out}, that was obtained as described in section 5.2, in a 2σ momentum region of the signal distribution. N_{sig} is the number of signal events and N_{bg} is the number of background events. The FOM for different cuts of the output classifier C_{out} are shown in Figure 7.14. We obtain the highest FOM for $C_{\text{out}} > 0.98$. After applying this cut to all MC samples, we obtain the p_μ^B distribution shown in Figure 7.15. The data-MC agreement in the sidebands are shown in Figure 7.16 and Figure 7.17 and looks reasonable.

Finally, we perform a fit to an Asimov data set to determine the statistical significance of signal in the range $2.2\,\text{GeV} \leq p_\mu^B \leq 3.0\,\text{GeV}$. We choose fine binning to better compare the significance with the ones obtained using other methods for estimating the B_{tag} kinematics explained in the following section 7.2. The fit to the Asimov data set using 40 bins is shown in Figure 7.18, and in addition the correlation of the MC samples is shown in Figure 7.19. The expected $B \to \mu\nu_\mu$ contribution is barely visible in the Asimov Fit, but has no high correlations with the other MC samples. The statistical significance of the signal is determined to

$$\sigma_{\text{stat}} = \frac{N_{\text{fit}}^{\text{sig}}}{\sigma_{\text{fit}}^{\text{sig}}} = \frac{5.01790}{14.41718} = 0.34805. \tag{7.10}$$
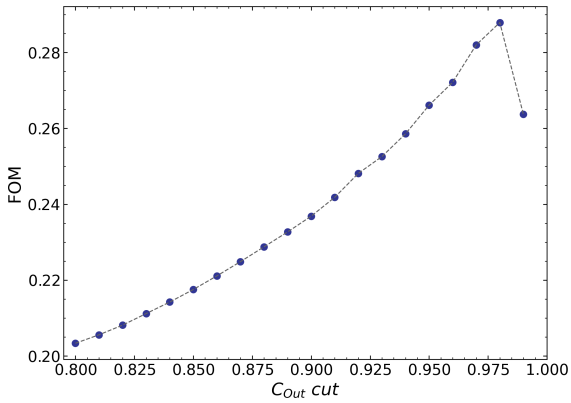
Figure 7.14: FOM obtained from different cuts on the output classifier C_{out}, after optimizing the p_z component.
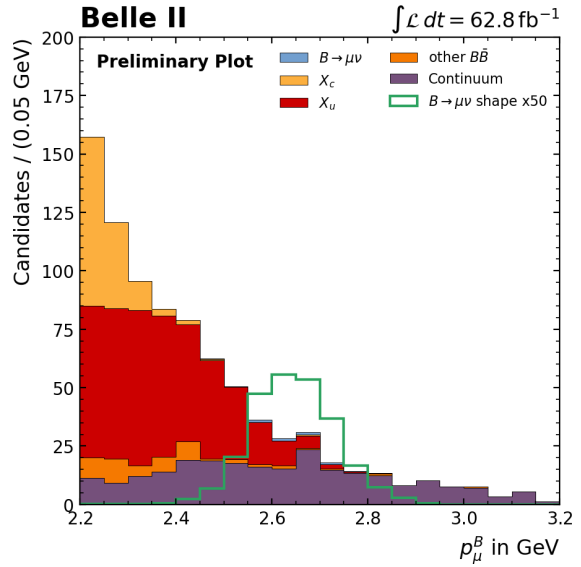
Figure 7.15: p_μ distribution of all MC samples after the optimization and cut on the output classifier and after optimizing the p_z component.
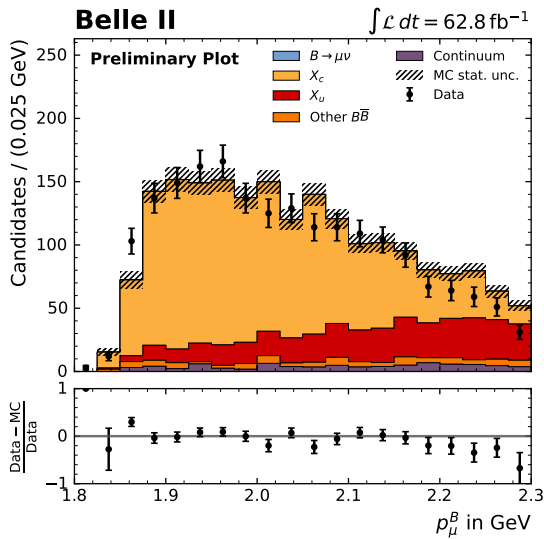
Figure 7.16: Data-MC agreement in the low momentum sideband, after optimizing the p_z component.
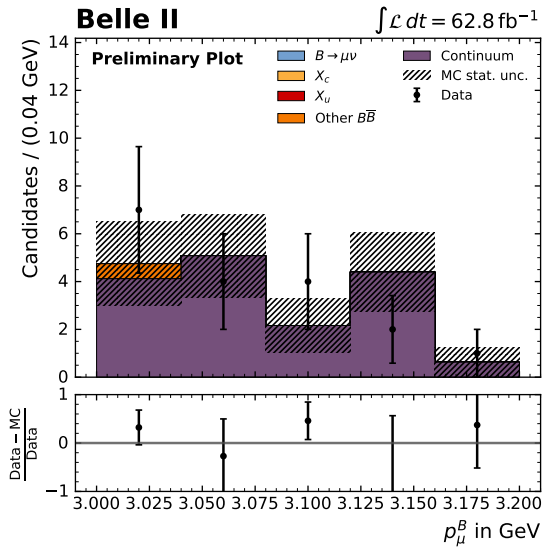
Figure 7.17: Data-MC agreement in the high momentum sideband, after optimizing the p_z component.
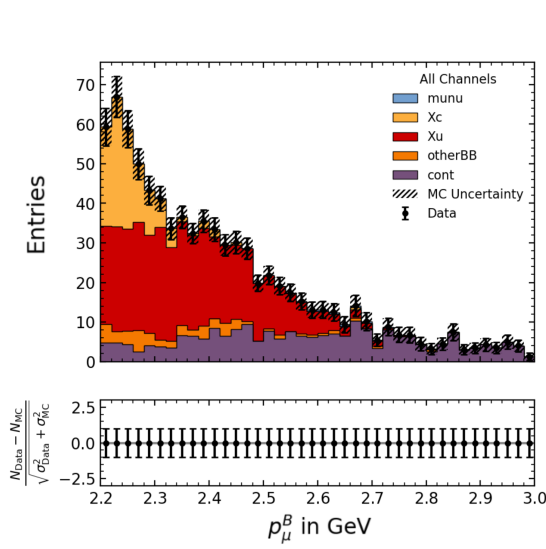
Figure 7.18: Fit to the Asimov data set, after optimizing the p_z component.
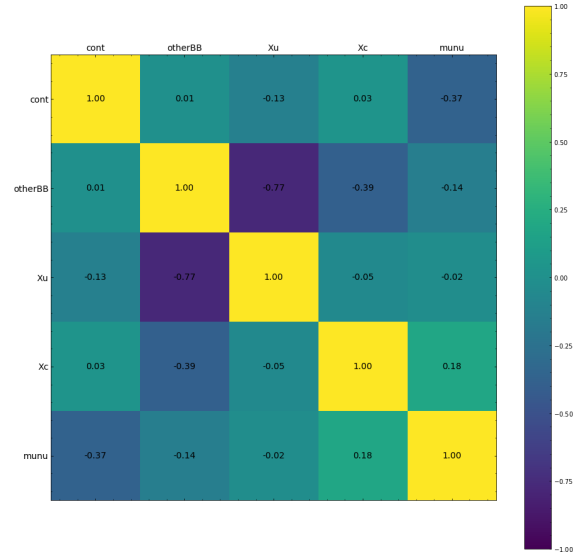
Figure 7.19: Correlations between the MC samples, after optimizing the p_z component.

7.2 Improving the Estimation of the B_{tag} Four-Momentum

In this section, different methods for improving the estimation of the kinematics of B_{tag} are presented and investigated. The different sections in which the methods are presented have more or less the same structure. First, an explanation of each method is given. We again consider the resolution of the B_{tag} momentum components and compare the shapes. Then, the optimal cut on the output classifier C_{out} in the 2σ region of the signal distribution is determined, as shown in the previous section. Then, the fit to the Asimov data set is performed for each of the methods and the significance is compared to that determined using the *Belle method*. We perform the fit in the same p_μ^B region and use the same number of bins as before. The data-MC agreement is also shown for each method in the appendix. Other methods were investigated but showed greatly reduced significance or were tested without scientific justification, so they were not considered further. A brief account of these methods is nevertheless included in section 7.2.5.

7.2.1 $p_{x,y,z}$ Calibration and Optimization

As seen in Figure 7.11 the resolution of the optimized p_x and p_y component is broader than the resolution of the p_z component. So the main idea of the method investigated in this section is to calibrate the p_x and p_y component in the same way the p_z component was calibrated in the previous section. The same calibration steps are performed independently for each momentum component. We bin the respective momentum component and map the values against the mean of the generated momentum component. Thus we get one functional dependency for each momentum component individually. The constrained mean $p_{i,\text{constr}}^*$, $i \in x, y, z$ versus the mean of the generated component and the smoothed spline fits are shown in Figure 7.20. Each extracted function is then applied to the
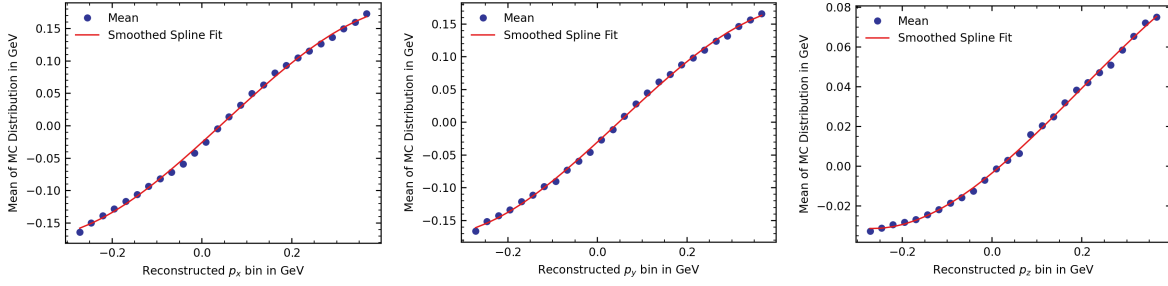
Figure 7.20: Binned constrained momentum component versus generated one and smoothed spline fit. The left plot shows p_x, the middle one p_y and the right one p_z.

momentum component

$$
\begin{aligned}
p_{x,\text{calib}}^* &= g(p_{x,\text{constr}}^*) \\
p_{y,\text{calib}}^* &= h(p_{y,\text{constr}}^*) \\
p_{z,\text{calib}}^* &= f(p_{z,\text{constr}}^*)
\end{aligned}
$$

and following the same procedure as in section 7.1.2 a correction factor leading to a minimization of Equation 7.6 is calculated to be $\xi = 1.249$. The optimized momentum components are then calculated according to Equation 7.7.

The resolution and comparison of the optimized momenta are shown in Figure 7.21 and Figure 7.22.
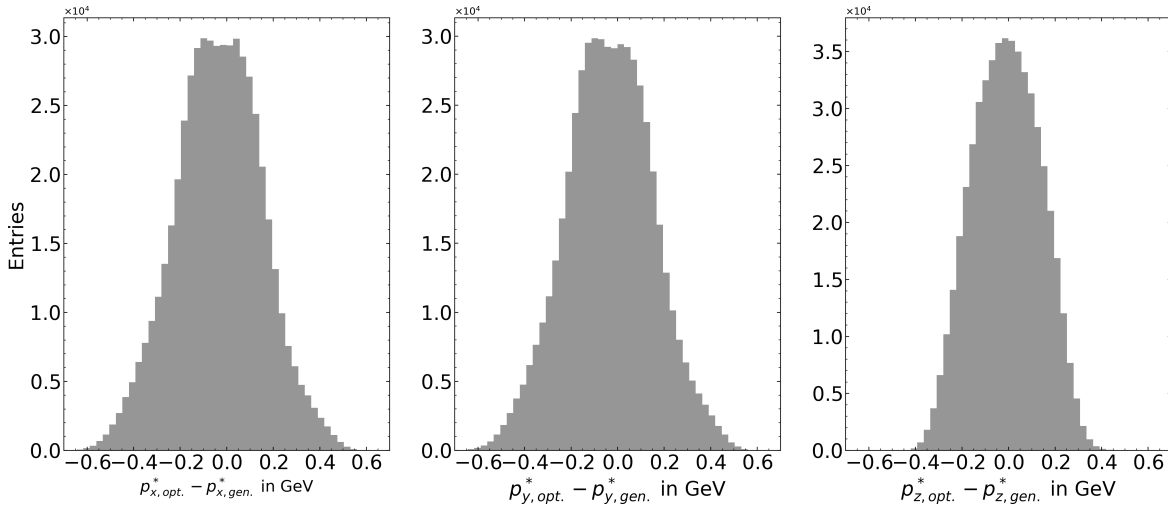
Figure 7.21: Resolution of the optimized B_{tag} momentum components for signal MC after optimizing the $p_{x,y,z}$ component.

Again, we use the optimized components to construct the kinematics of the B_{sig} using Equation 7.8 and boost the muon momentum into the B_{sig} rest frame as depicted for the signal MC in Figure 7.23 and for all scaled MC samples after a optimal cut on $C_{\text{out}} > 0.98$ in Figure 7.24. The grid search to determine the optimal cut on the output classifier is shown in the appendix in Figure C.4. The data-MC agreement in the sidebands is also shown in the appendix in Figure C.2 and Figure C.3.
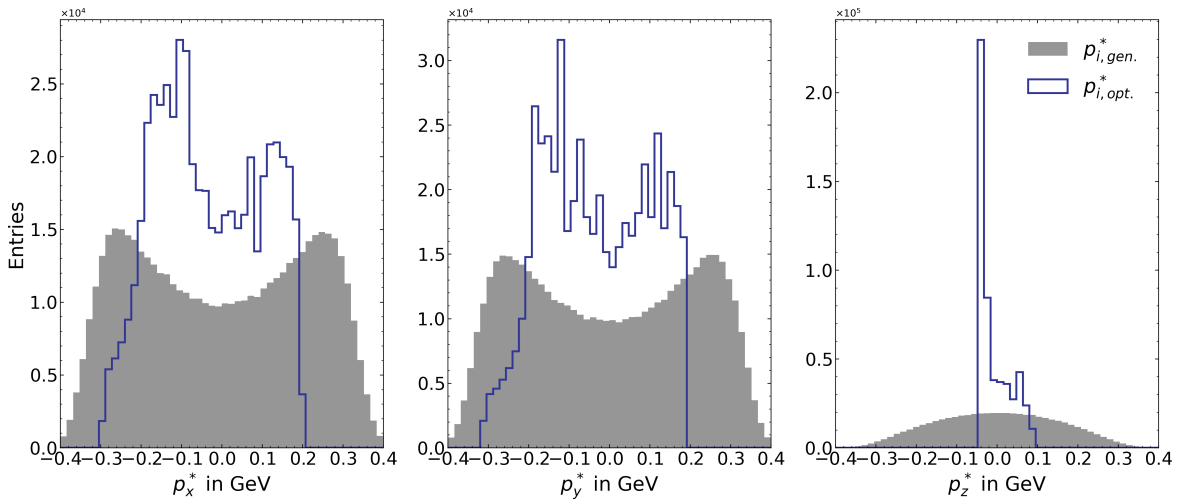
Figure 7.22: Comparison of optimized B_{tag} momentum components to the generated ones for signal MC after optimizing the $p_{x,y,z}$ component.
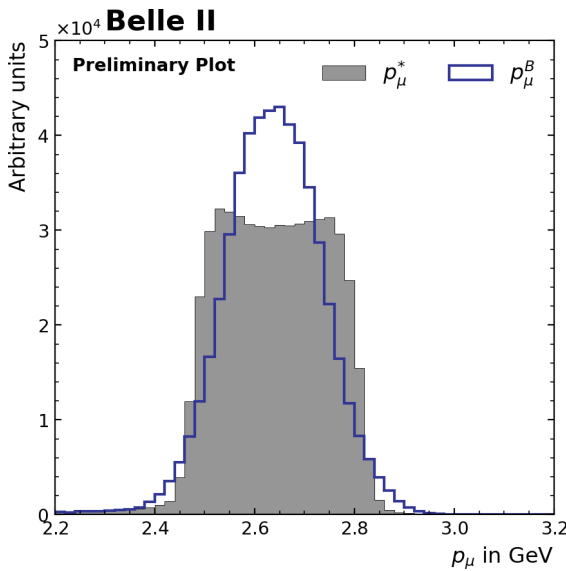
Figure 7.23: p_μ distributions in the center of mass and B rest frame of the signal MC sample, after optimizing the $p_{x,y,z}$ component.
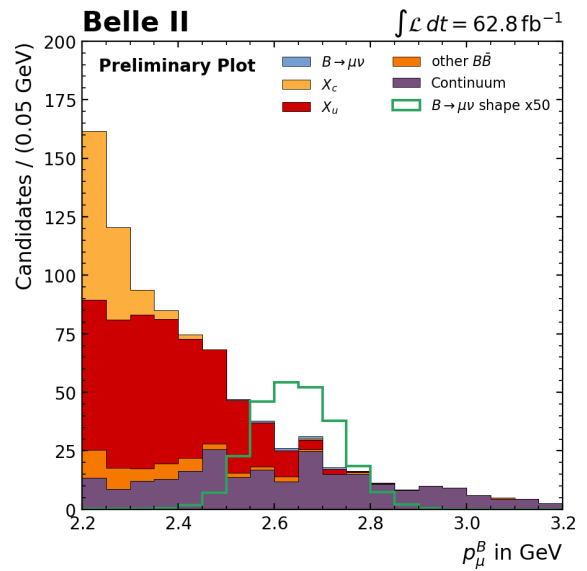
Figure 7.24: p_μ distribution of all MC samples after optimization and cut on the output classifier, after optimizing the $p_{x,y,z}$ component.

Again, a fit to an Asimov data set shown in Figure 7.25 is performed in order to determine the statistical significance of the signal to be

$$\sigma_{\text{stat}} = \frac{5.01801}{15.08727} = 0.33260. \tag{7.11}$$
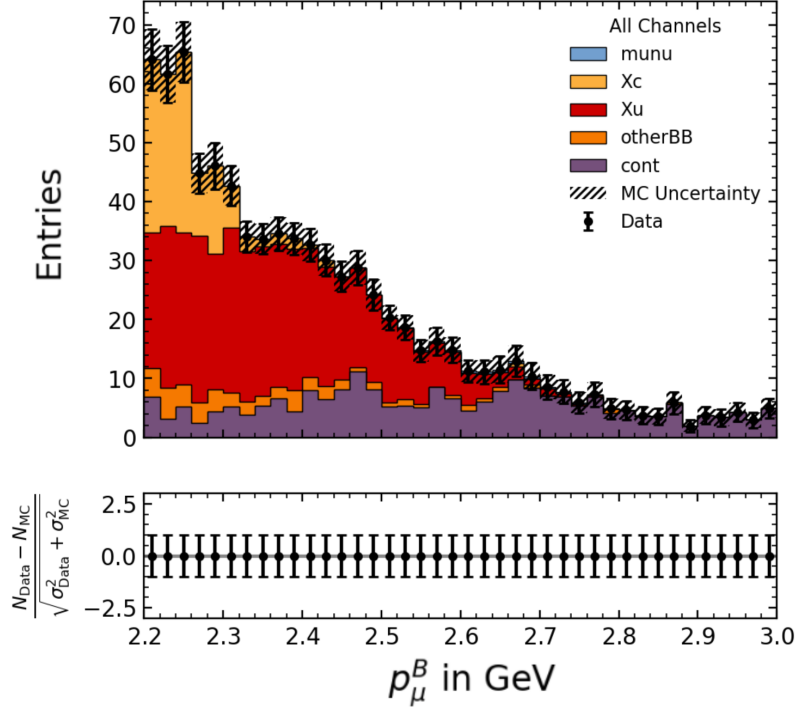
Figure 7.25: Fit to the Asimov data set, after optimizing the $p_{x,y,z}$ component.

The plot showing correlations between the MC samples is shown in the appendix in Figure C.5. In order to get an estimate whether the new method is superior to the *Belle method* we compare the statistical significances to each other. The $p_{x,y,z}$ calibration and optimization method has a significance 4.44% lower due to a slightly larger error. Thus this method does not yield the intended result. The main reason for this is that due to the additional calibration of the p_x and p_y the azimuth angle is changed in the optimization process in addition to the polar angle, which was not the case using the *Belle method*. This results in a slightly worse resolution in ϕ which is the expected reason for the lower significance. The azimuth angle resolution is shown in Figure 7.26 and Figure 7.27 for the *Belle Method* and the $p_{x,y,z}$ optimization, respectively.

7.2.2 Calibration and Optimization in Bins of N_{lep}

As shown in the previous section 7.2.1, calibrating and optimizing only the p_z component of the B_{tag} leads to better results than optimizing the p_x, p_y and p_z components. Therefore, we use only the p_z calibration and optimization for further considerations. The general idea is to perform individual calibrations of events, depending on certain properties. As shown in chapter 6, one of the main
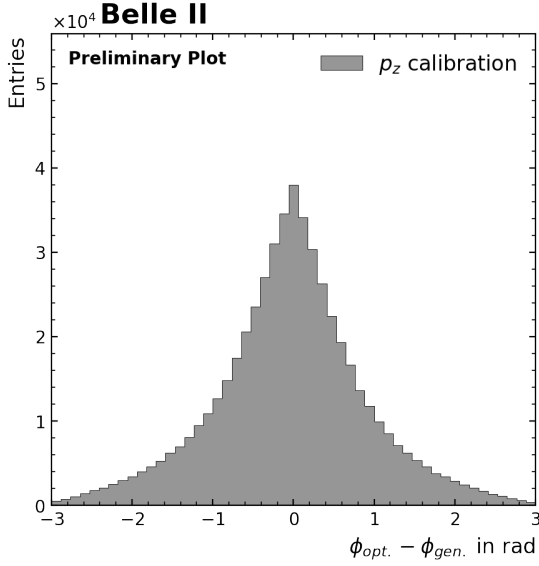
Figure 7.26: Resolution of the azimuth angle ϕ after optimizing the p_z component.
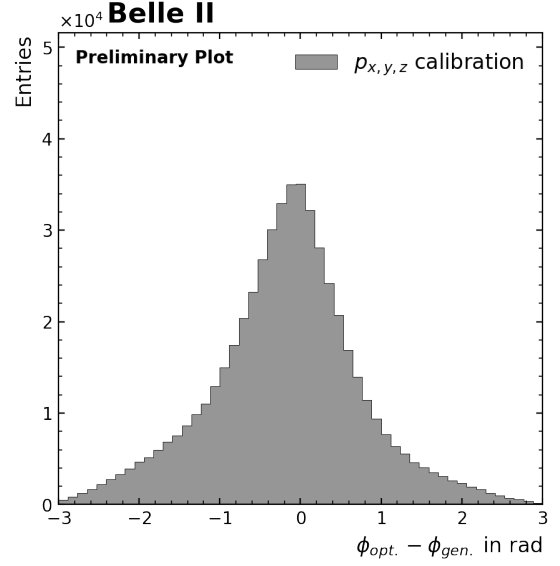
Figure 7.27: Resolution of the azimuth angle ϕ after optimizing the $p_{x,y,z}$ component.

uncertainties in the estimation of the B_{tag} kinematics are neutrinos on the B_{tag} side which are not reconstructed. For events in which the B_{tag} decays without a neutrino in the final state, no reconstruction error due to missing neutrinos is expected. Considering only these events, one would expect better agreement between the optimized B_{tag} kinematics and the generated ones. However, since far too many events are lost if we consider only the hadronic decays on the B_{tag} side, we want to perform a different calibration and optimization depending on the number of neutrinos N_ν on the B_{tag} side. Since we cannot know how many neutrinos are present in the event, we need a variable that we can reconstruct that gives us a sense of the value of N_ν. We test if a relation between the charged leptons and neutrinos in an event exists by looking at a 2D histogram of the generated number of charged leptons versus the generated number of neutrinos present in an event. This is shown in Figure 7.28 and a relation is clearly visible. Usually there is no neutrino in the event, if there is no charged lepton. If there is one charged lepton, in most cases there is one neutrino as well. Additionally, there are exceptions to the relation, but only in a few cases. However, the number of charged leptons depicted in Figure 7.28 is the number of generated ones and we need to look at the relation for leptons which actually can be reconstructed. Since we only look at the reconstruction of electrons and muons, we plot $N_{\text{lep}}^{\text{reco}}$ against $N_{\text{gen}}^{\nu_{e,\mu}}$, shown in Figure 7.29. A linear trend is still visible, but there are more exceptions to the relation, especially in the case $N_{\text{lep}}^{\text{reco}} = 1$ and $N_{\nu_{e,\mu}}^{gen.} = 0$. This is mainly due to pions which are misidentified as electrons and muons and they are therefore included in the number of reconstructed leptons.

We still use the this relation and perform two different calibrations depending on the number of reconstructed leptons: $N_{\text{lep}}^{\text{reco}} = 0$ and $N_{\text{lep}}^{\text{reco}} = 1$. This method is justified, if we look at the reconstruction errors due to missing neutrinos for these two cases. As visible in Figure 7.30 the error due to missing neutrinos reduces for the $N_{\text{lep}}^{\text{reco}} = 0$ cases and thus we expect a better result in the calibration for these events. On the contrary, we expect a slightly worse result from the calibration for the $N_{\text{lep}}^{\text{reco}} = 1$ cases, since the average is no longer taken over all events. We thus perform two independent calibrations and
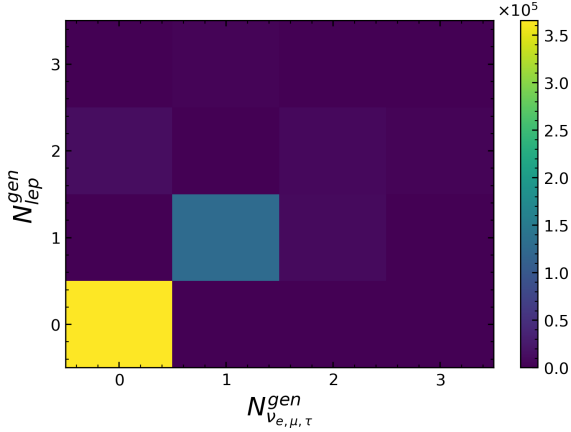
Figure 7.28: Number of generated leptons versus the number of generated neutrinos in an event.
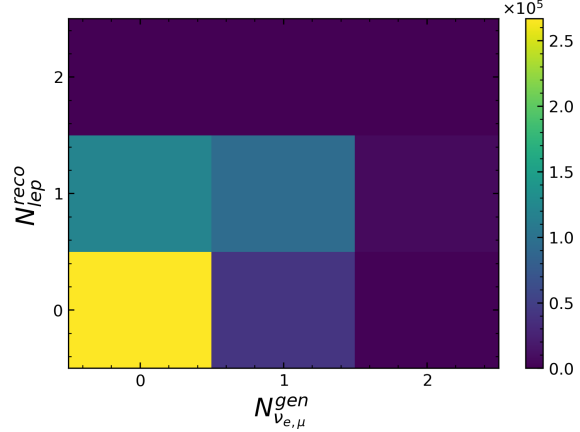
Figure 7.29: Number of reconstructed leptons versus the number of generated neutrinos (ν_e and ν_μ) in an event.

optimizations for these two cases. The mean of $p^*_{z,\text{constr}}$ versus the mean of $p^*_{z,\text{gen}}$ and the smoothed spline fits are shown in Figure 7.31 and Figure 7.32.

We additionally determine two independent correction factors for the two cases to be

$$\xi(N^{\text{reco}}_{\text{lep}} = 0) = 0.641 \qquad \text{and} \qquad \xi(N^{\text{reco}}_{\text{lep}} = 1) = 0.550 \tag{7.12}$$

and calculate the optimized momentum components using Equation 7.7. The resolutions and comparison of the optimized momentum components to the generated components are shown in Figure 7.33 and 7.34, but they do not show obvious improvements.

We again construct the B_{sig} after recombining the two cases using Equation 7.8 and boost the muon momentum into the rest frame of the B_{sig}. The distribution for the signal sample is shown in Figure 7.35. We again search for the optimal cut on the output classifier C_{out} resulting in the highest FOM in the 2σ region, also shown in the appendix in Figure C.8. The cut is performed on each MC sample which results in the distribution shown in Figure 7.36. We also look at the data-MC agreement depicted in the appendix in Figure C.6 and Figure C.7.

The fit to the Asimov data set shown in Figure 7.37 yields a statistical significance of

$$\sigma_{\text{stat}} = \frac{5.01798}{14.60016} = 0.34369. \tag{7.13}$$

This thus is about 1.25% lower than the one obtained by the *Belle method*. This is still not an improvement, but a better result that the individual calibration of $p^*_{x,\text{constr}}$, $p^*_{y,\text{constr}}$ and $p^*_{z,\text{constr}}$ yielded. The plot showing the correlations of the MC samples after the fit are shown in Figure C.9 in the appendix. The reason for the slightly lower significance could be due to the following. Due to a calibration in different bins we loose statistics in each bin resulting in a more inaccurate mapping of the binned constrained p_z component to the generated one, resulting in a slightly worse optimization of the components. This is not compensated by the calibration in different bins, which is mainly due to the not optimal relation between the number of reconstructed leptons and generated neutrinos, which could be improved upon in the future.
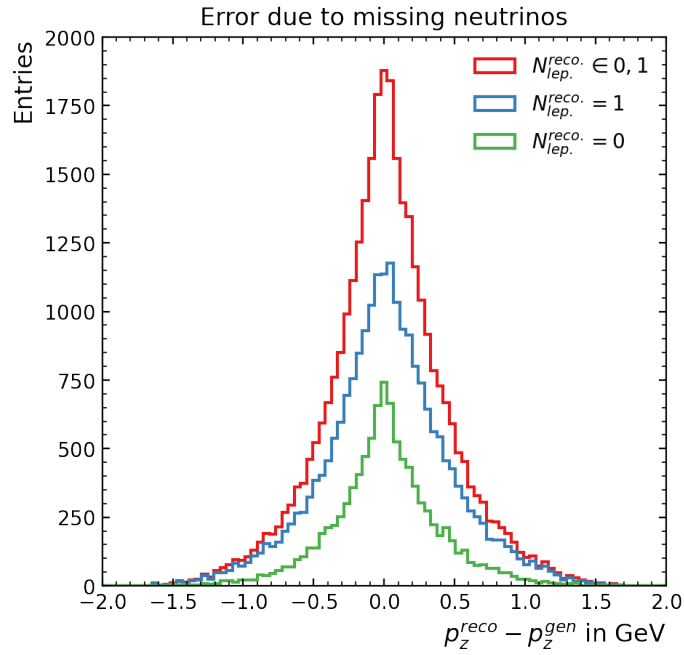
Figure 7.30: Error in the reconstruction of the p_z component of the B_{tag} due to missing neutrinos for different numbers of $N_{\text{lep}}^{\text{reco}}$.
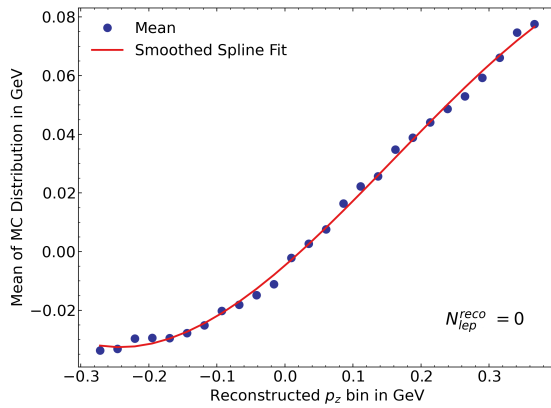
Figure 7.31: Binned constrained p_z momentum component versus generated one and smoothed spline fit for the case $N_{\text{lep}}^{\text{reco}} = 0$.
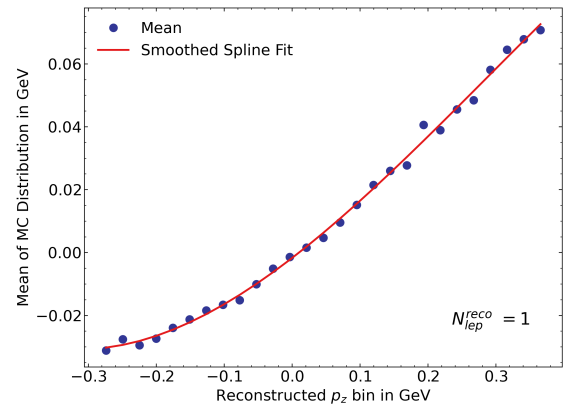
Figure 7.32: Binned constrained p_z momentum component versus generated one and smoothed spline fit for the case $N_{\text{lep}}^{\text{reco}} = 1$.
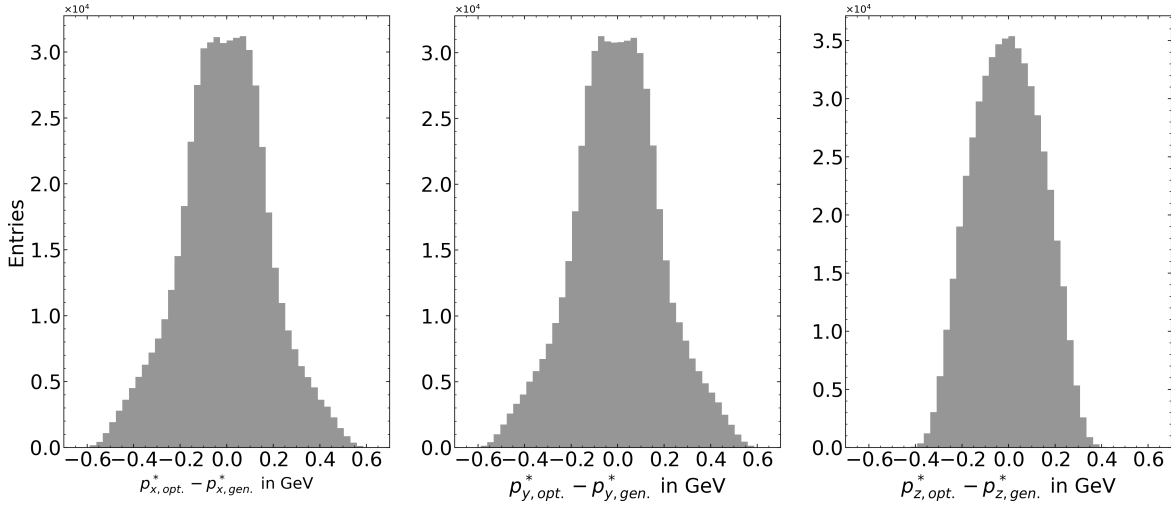
Figure 7.33: Resolution of the optimized B_{tag} momentum components for signal MC after optimizing in bins of $N_{\mathrm{lep}}^{\mathrm{reco}}$.
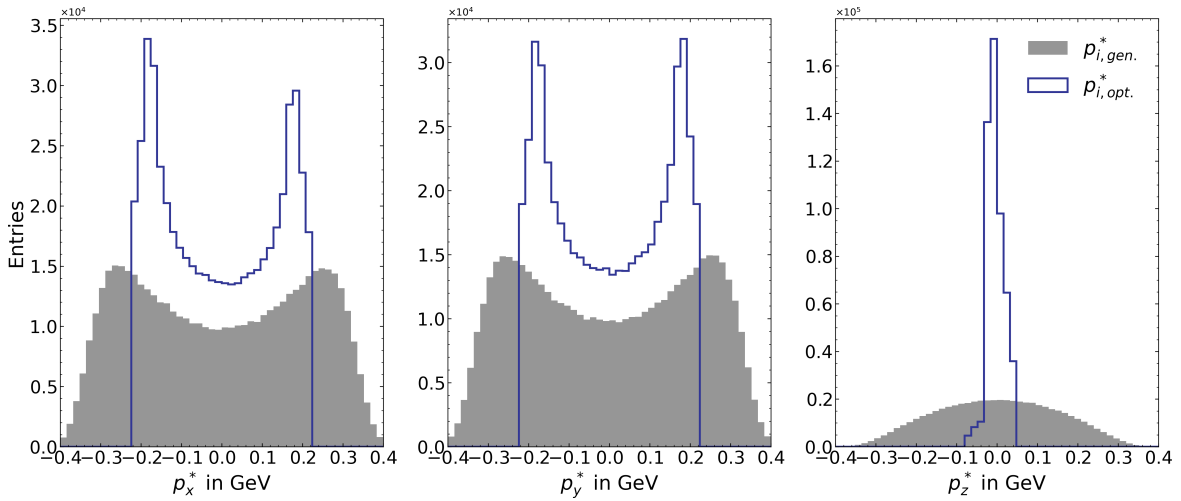
Figure 7.34: Comparison of the optimized B_{tag} momentum components to the generated ones for signal MC after optimizing in bins of $N_{\mathrm{lep}}^{\mathrm{reco}}$.
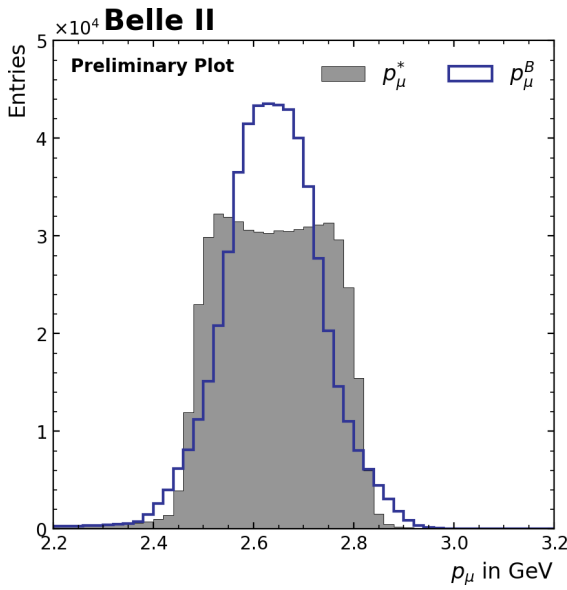
Figure 7.35: p_μ distributions in the center of mass and B rest frame of the signal MC sample, after optimizing in bins of $N_{\text{lep}}^{\text{reco}}$.
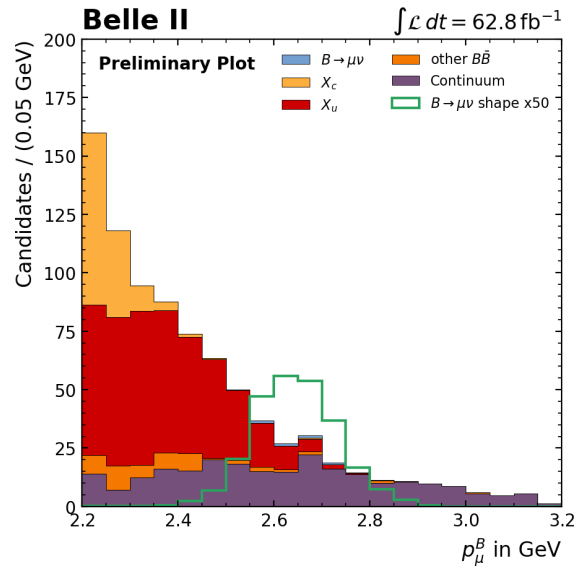
Figure 7.36: p_μ distribution of all MC samples after optimization and cut on the output classifier, after optimizing in bins of $N_{\text{lep}}^{\text{reco}}$.
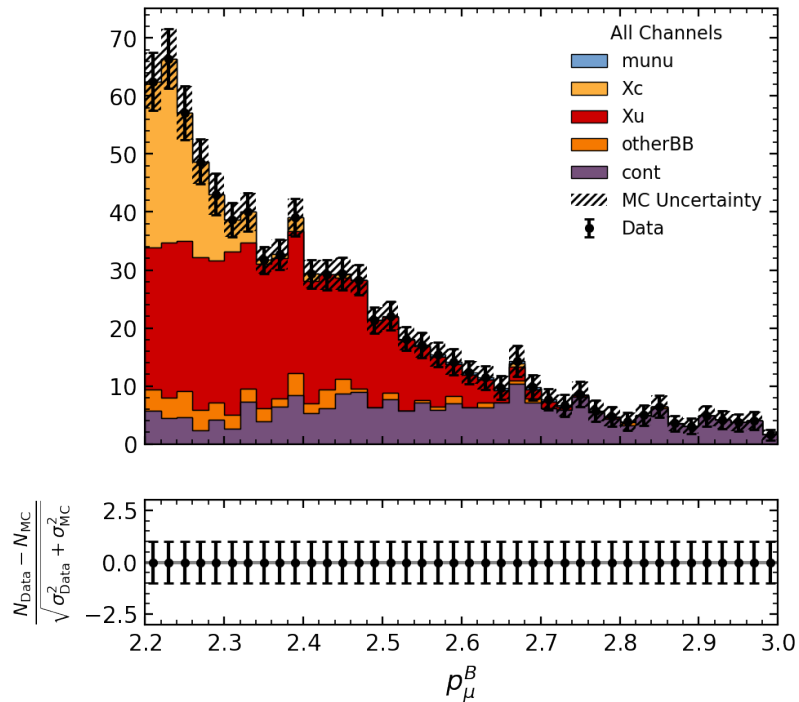
Figure 7.37: Fit to the Asimov data set, after optimizing in bins of $N_{\text{lep}}^{\text{reco}}$

7.2.3 Calibration and Optimization in Bins of θ_{ROE}

The next method is the calibration and optimization in bins of the ROE polar angle. The main idea for this method is similar to the calibration and optimization in bins of N_{lep} method. Particles lost due to the detector acceptance have the largest contribution to the deterioration of the B_{tag} kinematics, next to missing particles. The angle of the ROE is correlated to this error. If θ_{ROE} is small, more particles are lost in the forward direction. If the angle is high, more particles are lost in the backward direction. For an angle θ_{ROE} of $\sim 90^\circ$ the error due to the acceptance of the detector is minimal. As visible in Figure 7.38 it is not feasible to remove events with a small or high θ_{ROE}, since signal events are lost without gaining additional background suppression. Therefore, we define three equally sized regions of θ_{ROE} where we calibrate and optimize the constrained p_z component independently. The three regions of θ_{ROE} are defined as

$$\theta_{ROE} \leq 60^\circ \quad , \quad 60^\circ < \theta_{ROE} \leq 120^\circ \quad \text{and} \quad \theta_{ROE} > 120^\circ \tag{7.14}$$

and are also shown in Figure 7.38. The p_z resolution errors in these bins are depicted in Figure 7.39. The standard deviation is minimal for the $60^\circ < \theta_{\text{ROE}} \leq 120^\circ$ distribution, therefore, a better result for the optimization is expected. In contrast the standard deviation for the $\theta_{\text{ROE}} > 120^\circ$ distribution is higher than the one of the unbinned distribution, thus a worse result for the optimization is expected. We still perform the calibration, optimization and fit to the Asimov data set to investigate the performance of this method. The calibration and optimization process is similar to the one used in the *Belle method* described in section 7.1. We get three different smooth spline fits depending on the angle of the ROE, depicted in Figure 7.40, Figure 7.41 and Figure 7.42, but the fits do not show a good agreement for higher angles. Additionally we calculate the three correction factors to be

$$\begin{aligned} \xi(\theta_{ROE} \leq 60^\circ) &= 0.573 \ , \\ \xi(60^\circ < \theta_{ROE} \leq 120^\circ) &= 0.662 \text{ and} \\ \xi(\theta_{ROE} > 120^\circ) &= 0.530 \end{aligned} \tag{7.15}$$

and perform the optimization. The resolution of the optimized B_{tag} kinematics as well as the shape comparison are depicted in Figure 7.43 and Figure 7.44 but they again do not show significant improvement.

We apply the binned calibration and optimization to all the MC samples and perform the cut on the output classifier C_{out} based on the highest FOM optimization. The muon momentum is boosted into the signal side B rest frame, which kinematics are constructed using to Equation 7.8. The distribution is shown in Figure 7.46 next to the p_μ distributions in the center of mass and B rest frame for the signal sample in Figure 7.45 The data-MC agreement is shown in the appendix in Figure C.10 and Figure C.11. We again perform the fit to the Asimov data set, shown in Figure 7.47 and determine the statistical significance of the signal to be

$$\sigma_{\text{stat}} = \frac{5.01807}{14.44837} = 0.34731. \tag{7.16}$$

Compared to the significance calculated in the *Belle method*, the calibration and optimization in bins of θ_{ROE} yields a significance 0.21% lower. Thus this method is not an improvement, but achieves comparable results.
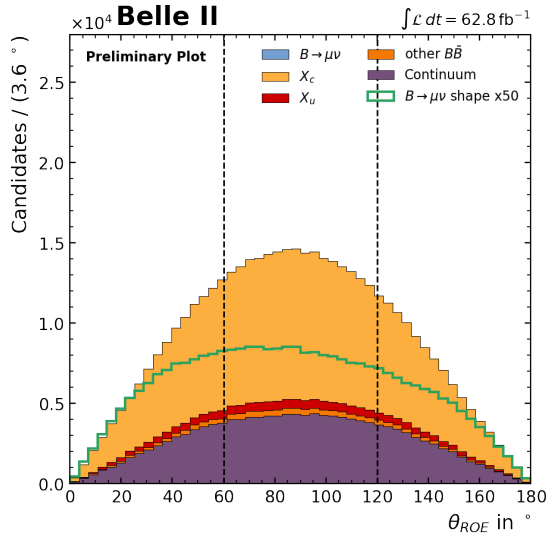
Figure 7.38: θ_{ROE} distributions of the scaled MC samples. The division into the three bins is also shown.
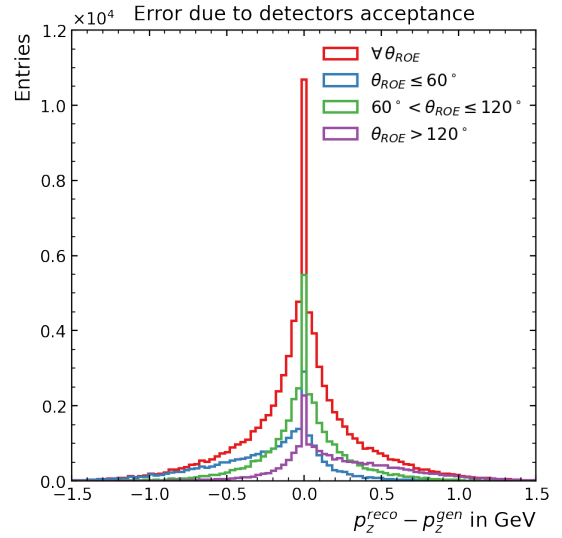
Figure 7.39: Error in the reconstruction of the p_z component of the B_{tag} due to missing particles due to the detector acceptance for different regions of θ_{ROE}.
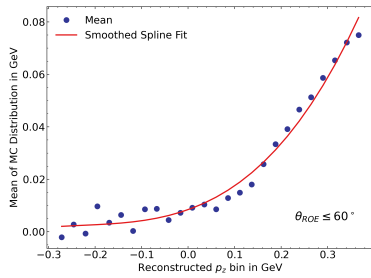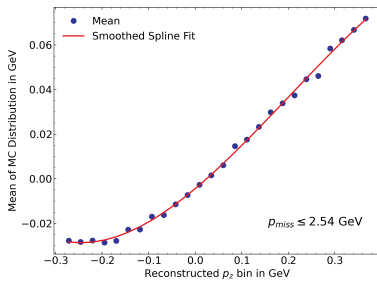

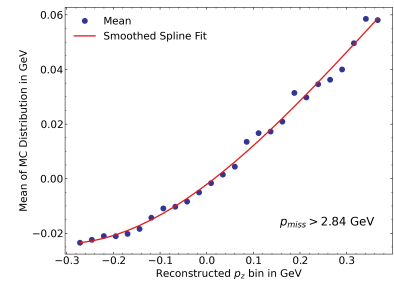Figure 7.40: Binned constrained momentum component versus generated one and smoothed spline fit for the region $\theta_{\text{ROE}} \leq 60°$.
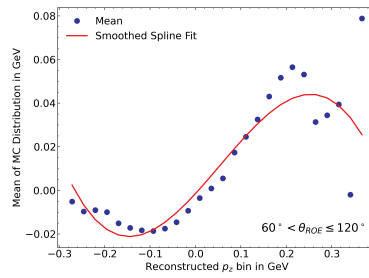
Figure 7.41: Binned constrained momentum component versus generated one and smoothed spline fit for the region $60° < \theta_{\text{ROE}} \leq 120°$.
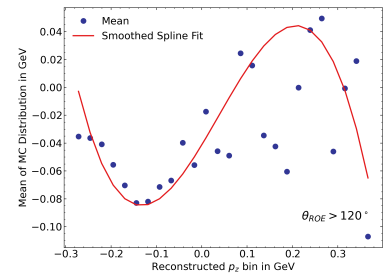
Figure 7.42: Binned constrained momentum component versus generated one and smoothed spline fit for the region $\theta_{\text{ROE}} > 120°$.
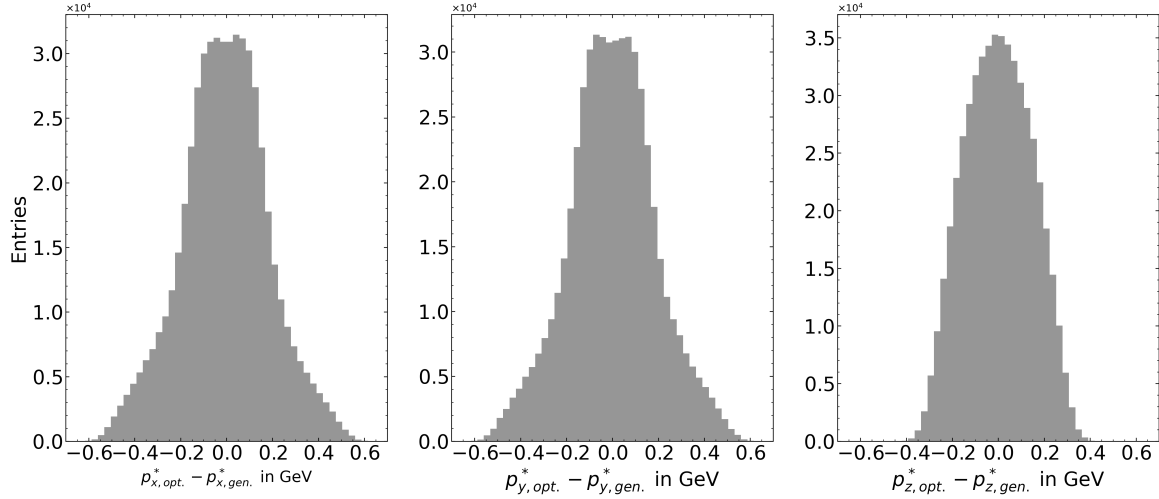
Figure 7.43: Resolution of the optimized B_{tag} momentum components for signal MC after optimizing in bins of θ_{ROE}.
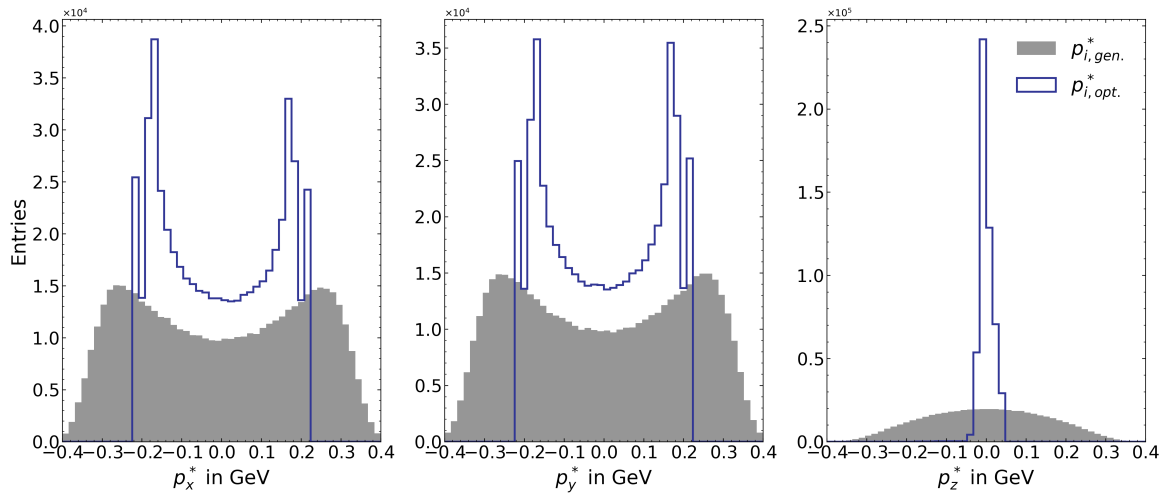
Figure 7.44: Comparison of optimized B_{tag} momentum components to the generated ones for signal MC after optimizing in bins of θ_{ROE}.
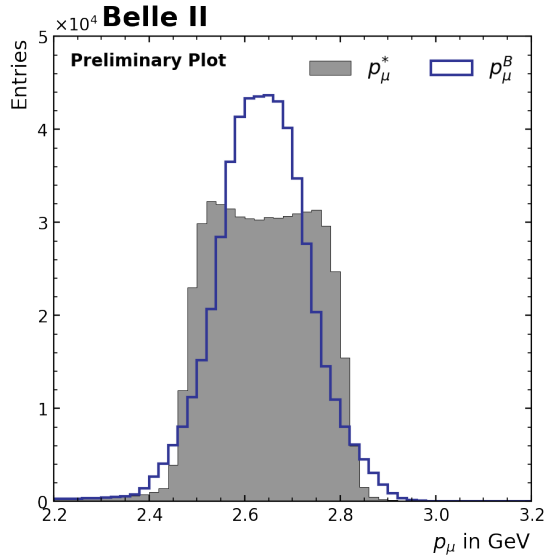
Figure 7.45: p_μ distributions in the center of mass and B rest frame of the signal MC sample, after optimizing in bins of θ_{ROE}.
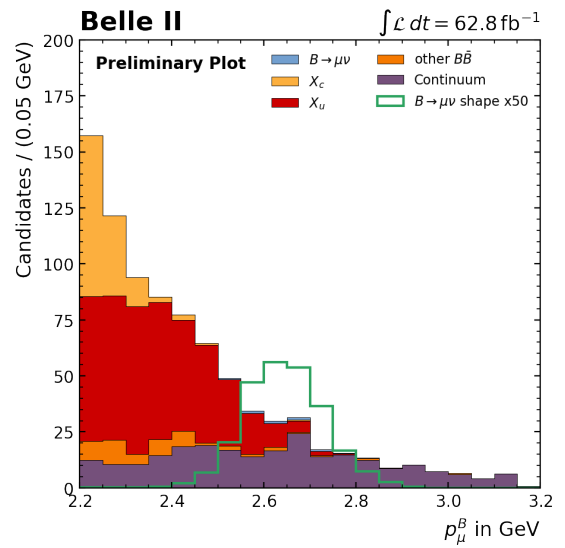
Figure 7.46: p_μ distribution of all MC samples after optimization and cut on the output classifier, after optimizing in bins of θ_{ROE}.
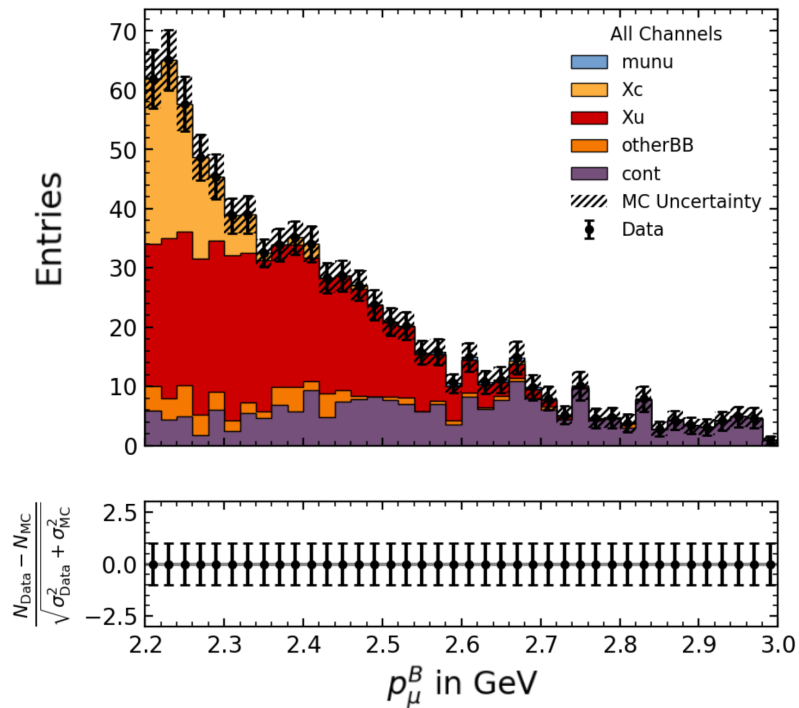
Figure 7.47: Fit to the Asimov data set, after optimizing in bins of θ_{ROE}.

7.2.4 Neutrino Estimation Method

To explain this method a simple decay tree example of a signal event is depicted in Figure 7.48. The B_{sig} decays into the muon and the neutrino while the B_{tag} decays generically, here for example into a pion, a lepton and a neutrino. During the reconstruction of the event, only the pion, the lepton and the signal muon can be reconstructed, while the two neutrinos can not. The kinematics of the two neutrinos are summarized in the missing momentum of the event, which can be calculated during the reconstruction during momentum and energy conservation. The main idea now is to obtain an estimate for the kinematics of the neutrino on the signal side, since then theoretically the B_{tag} four-vector in the center of mass frame can be constructed as

$$p_{B_{\text{tag}}}^{\text{est}} = p_{\text{ROE}}^{\text{reco}} + p_{\text{miss}}^{\text{reco}} - p_{\nu_{\text{sig}}}^{\text{est}}. \tag{7.17}$$

The main challenge of this method is to obtain an accurate estimate of the neutrino kinematics. This is achieved as follows.
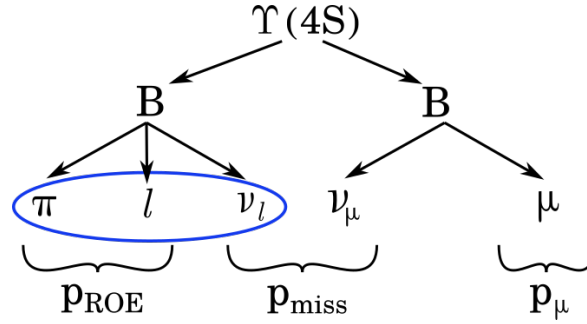
Figure 7.48: Schematic illustration of a simple decay tree of a $\Upsilon(4S)$ decay involving a signal event in order to explain the neutrino estimation method.

Since the $B \to \mu\nu_\mu$ decay is a two-body decay, the muon and the neutrino are emitted back to back in the rest frame of the B_{sig} to conserve the four-momentum. Their three-momentum components are of equal magnitude but with an opposite sign. Additionally, we assume the neutrino to be massless and thus can use the magnitude of the neutrino's three-momentum of

$$|p_\nu^B| = |p_\mu^B| = \frac{m_B^2 - m_\mu^2}{2m_B} = 2.6386\,\text{GeV}. \tag{7.18}$$

We use the muon momentum in the B_{sig} rest frame obtained using the calibration and optimization of the p_z component of the B_{tag} shown in section 7.1.2 to construct the kinematics of the neutrino and introduce a constraint via

$$p_{\nu_{i,\text{sig}}}^{B,\text{est}} = -\frac{|p_\nu^B|}{|p_\mu^B|} \cdot p_{i,\mu}^B \qquad , i \in x, y, z \tag{7.19}$$

where $|p_\nu^B|$ is the theoretically expected magnitude of the neutrino momentum. $|p_\mu^B|$ is the magnitude of the boosted muon momentum and $p_{i,\mu}^B$ the respective momentum component of the boosted muon momentum determined by the calibration and optimization of the p_z component in section 7.1.2. We then boost the hypothetical neutrino back into the center of mass frame and can calculate the four-vector

of the B_{tag} with Equation 7.17. From this information we can estimate the B_{sig} four-momentum in the center of mass frame via

$$p^*_{B,\text{sig}} = \begin{pmatrix} \sqrt{|\vec{p}^*_{\text{constr}}|^2 + m^2_B} \\ -\left(p^{\text{reco}}_{i,\text{ROE}} + p^{\text{reco}}_{i,\text{miss}} - p^{\text{est}}_{\nu_{i,\text{sig}}}\right), \; i \in x, y, z \end{pmatrix}. \tag{7.20}$$

Constraining has the effect, that the three-momentum of the neutrino maintains the direction but is scaled in length in order to keep the expected magnitude. Figure 7.49 and Figure 7.50 shows the resolution of the estimated neutrino kinematics and the comparison to the generated momentum components, respectively. Additionally, in Figure 7.52 and Figure 7.51 the resolution and comparison for the newly estimated B_{tag} kinematics using Equation 7.17 is depicted.
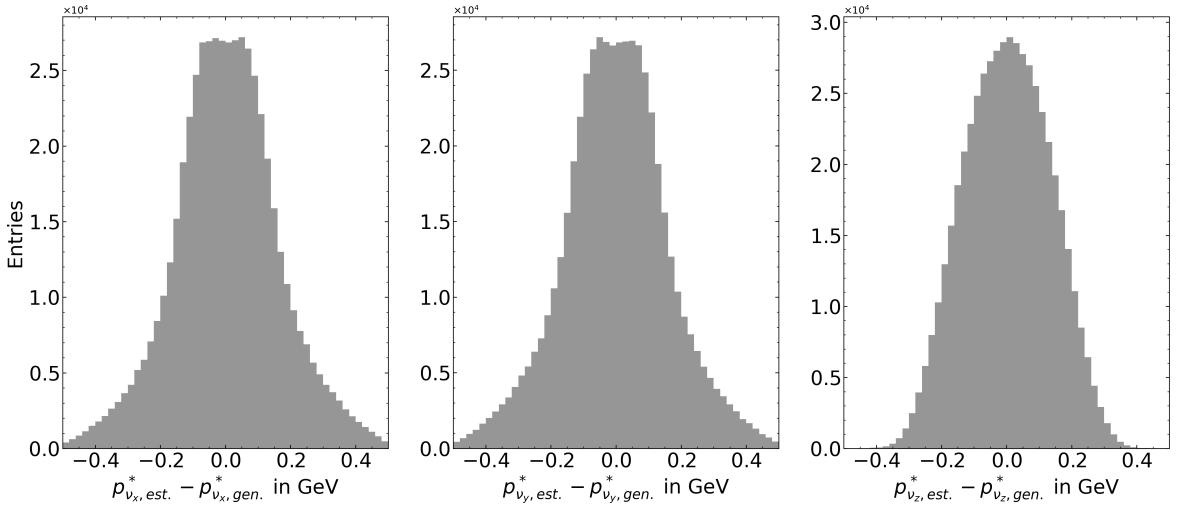
Figure 7.49: Comparison of estimated neutrino momentum components to the generated ones for signal MC after estimating the neutrino kinematics.

Using this method the kinematics of the neutrino can be estimated with good precision which subsequently has a large impact on the estimation of the B_{tag}. This results in a high improvement in the resolution of the muon momentum in the B_{sig} rest frame, as seen in Figure 7.53. As stated before, it is highly important what effect the corresponding method has on the background. Figure 7.54 shows the p^B_μ distribution of all scaled MC samples after an optimized cut on $C_{\text{out}} > 0.98$, that resulted from a grid search is shown in Figure C.16 in the appendix. It is evident, that the used method only works as intended for signal events. For all the other contributions the main effect of the method is a scaling of the momentum towards the expected momentum of 2.64 GeV. Using this method brings a sharper peak for signal events, but shifts background events into this signal region.

We perform the fit to the Asimov data set shown in Figure 7.55 in order to estimate the statistical significance. We get

$$\sigma_{\text{stat}} = \frac{5.02362}{14.17781} = 0.35433. \tag{7.21}$$

and thus get a significance 1.8% higher than the one we achieved using the *Belle method*. The plots
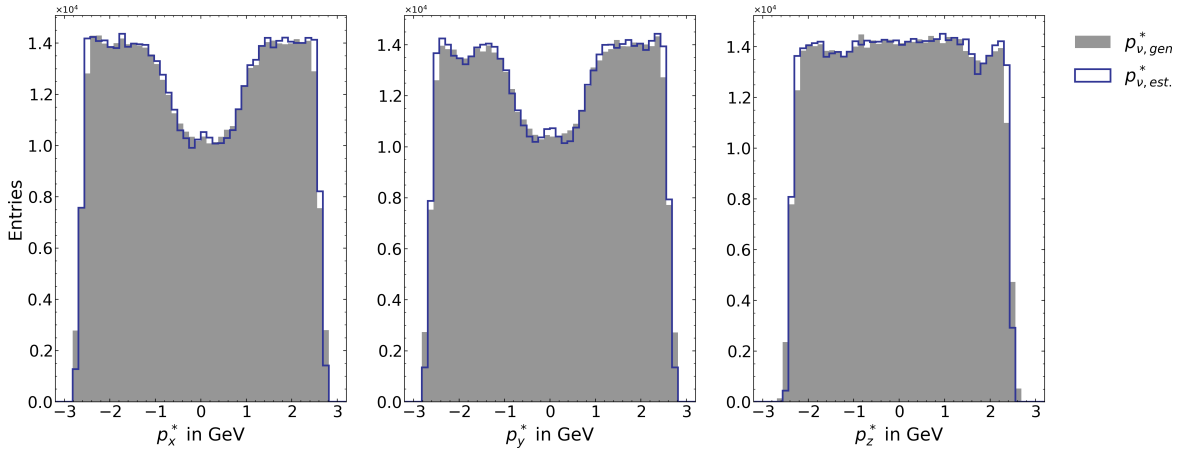
Figure 7.50: Resolution of the estimated neutrino momentum components for signal MC after estimating the neutrino kinematics.
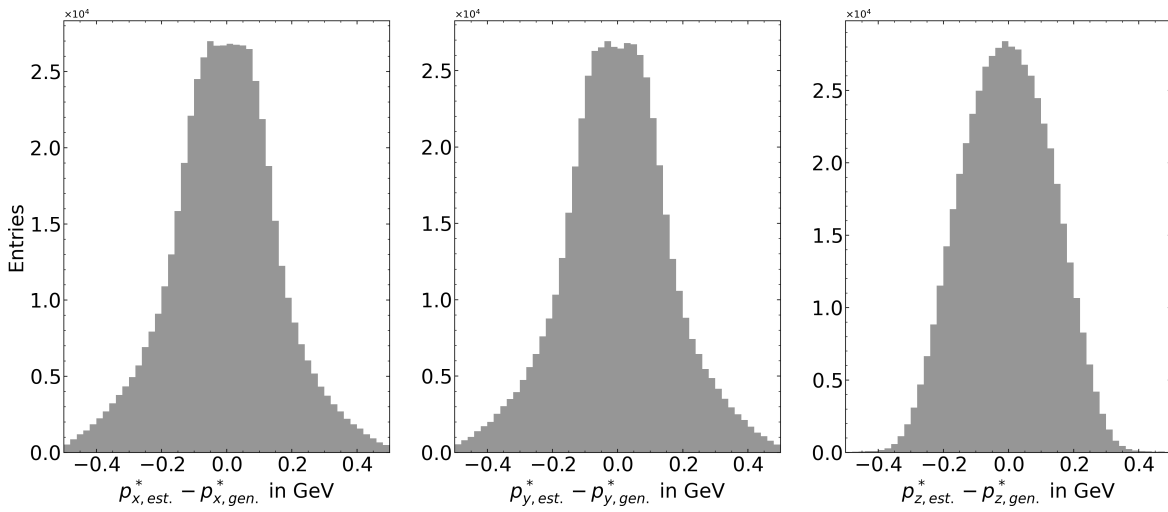
Figure 7.51: Resolution of estimated B_{tag} momentum components for signal MC after estimating the neutrino kinematics.
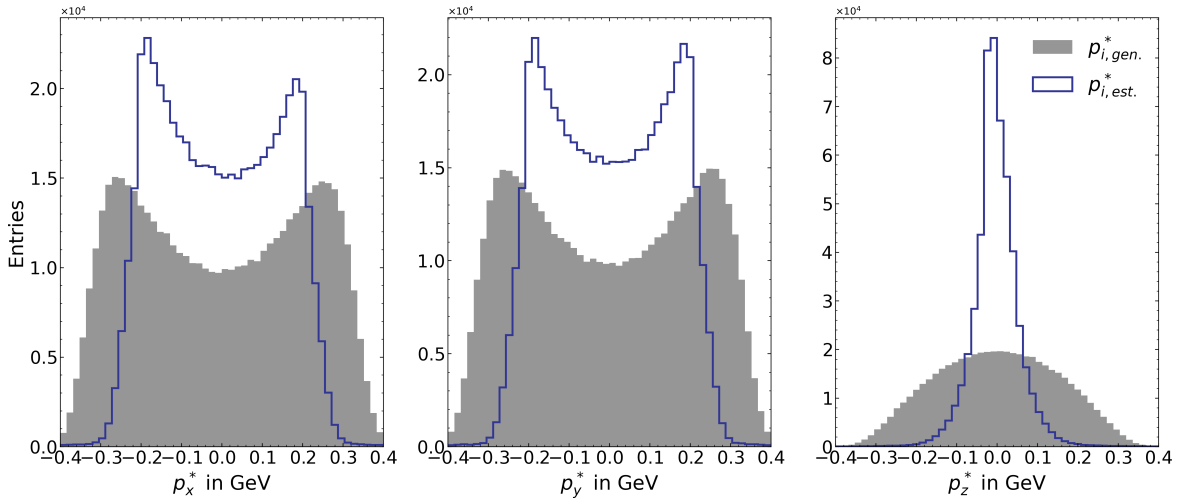
Figure 7.52: Comparison of estimated B_tag momentum components to the generated ones for signal MC after estimating the neutrino kinematics.
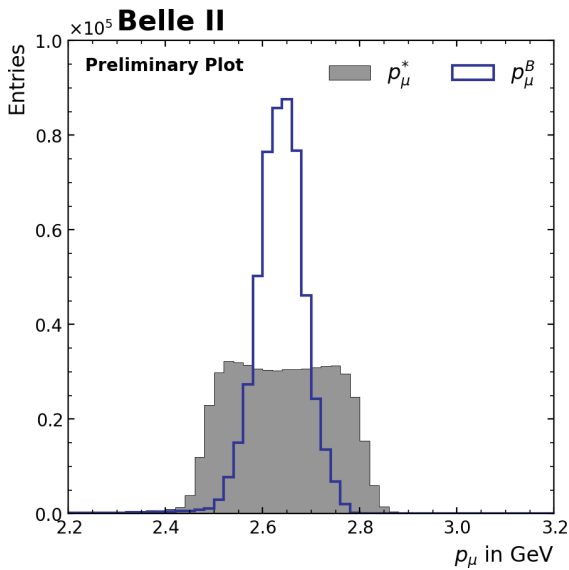
Figure 7.53: p_μ distributions in the center of mass and B rest frame of the signal MC sample after estimating the neutrino kinematics.
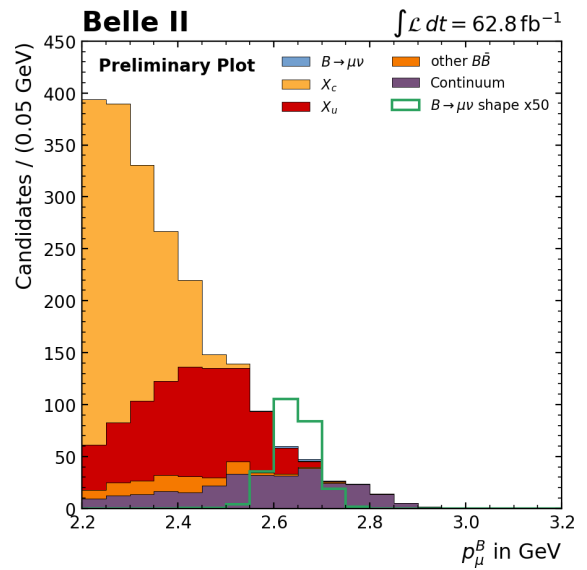
Figure 7.54: p_μ distribution of all MC samples after optimization and cut on the output classifier after estimating the neutrino kinematics.

showing the correlations resulting from the fit are shown in Figure C.17 in the appendix and the data-MC agreement is shown in Figure C.14 and Figure C.15.
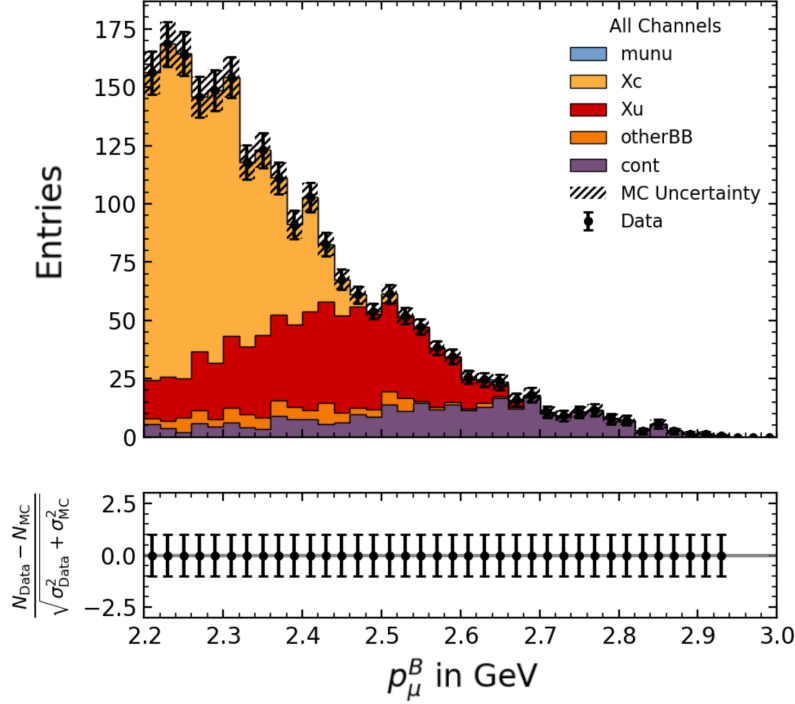
Figure 7.55: Fit to the Asimov data set after estimating the neutrino kinematics.

In addition, we can use this method to revisit the calculation of some variables since we estimated new kinematics for the B_{tag}. We can calculate a new normalized beam constrain mass $\hat{m}_{\text{bc}}^{\text{est}}$ as well as a new normalized energy difference $\Delta\hat{E}^{\text{est}}$ via

$$\hat{m}_{\text{bc}}^{\text{est}} = \sqrt{\frac{E^{*2} - |\vec{p}_{B_{\text{tag}}}^{\text{est}}|^2}{E^{*2}}} \qquad \text{and} \qquad \Delta\hat{E}^{\text{est}} = \frac{E_{B_{\text{tag}}}^{\text{est}} - E^{*2}}{E^{*2}}. \tag{7.22}$$

Their distributions are shown in Figure 7.56 and Figure 7.57. A grid search was performed in order to investigate if a cut on these newly calculated values could lead to a higher FOM. Optimal cuts were found, but as a result of cutting on these values, all events were discarded except for the ones in the signal region, leading to sculpting of the background and a decreased significance.

7.2.5 Additional Methods

The methods presented in this section are methods which were investigated without specific scientific justification, but are sill mentioned and shortly explained here.
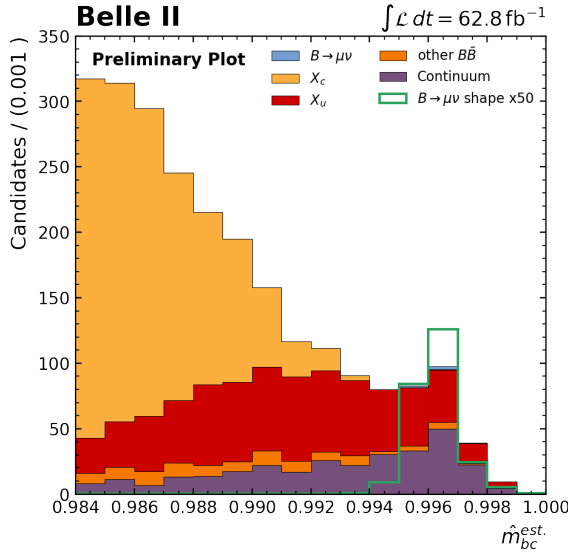
Figure 7.56: $\hat{m}_{\mathrm{bc}}^{\mathrm{est}}$ distribution for the scaled MC samples using the estimated B_{tag} kinematics.
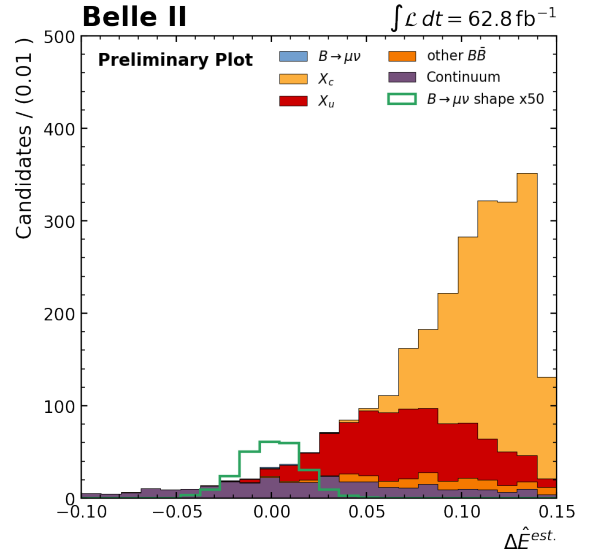
Figure 7.57: $\Delta\hat{E}^{\mathrm{est}}$ distribution for the scaled MC samples using the estimated B_{tag} kinematics.

Diamond Frame

The *Diamond Frame* method was firstly introduced in the "Measurement of $B \to D$ Form Factors in the Semileptonic Decay $B^0 \to D^*\ell\nu$" at BABAR by M. S. Gill [32] and was further optimized by Simon Calò (*S. Calò, The Diamond Frame, Maastricht University, Bachelor Thesis, May 2020*). An explanation of the theory is omitted at this point and in order to obtain more detailed explanations, reference is made to the sources given. The *Diamond Frame* can basically be determined using three different modes. The Y particle in all modes is constructed to be the signal side muon. In mode 1, the B meson is constructed from the reconstructed muon on the signal side. In mode 2, the B meson is constructed from the information gained by the reconstruction of the rest of event using $p_{B_{\mathrm{ROE}}} = -p_{B_{\mathrm{sig}}}$. For mode 3 the information from mode 1 and mode 2 is combined to construct the four-vector of the B_{sig}. Figure 7.58, Figure 7.59 and Figure 7.60 show the p_μ distribution of all MC samples and the scaled signal shape for the three given modes. Figure 7.61, Figure 7.62 and Figure 7.63 show the fit to the Asimov data set.

The significance for the modes are calculated to be

$$\sigma_{\mathrm{stat}} = \frac{5.02099}{28.85974} = 0.17398 \qquad \text{, for mode 1,}$$

$$\sigma_{\mathrm{stat}} = \frac{5.02099}{17.35651} = 0.28929 \qquad \text{, for mode 2,}$$

$$\sigma_{\mathrm{stat}} = \frac{5.02099}{23.78792} = 0.21107 \qquad \text{, for mode 3} \tag{7.23}$$

and are lower than the significance obtained using the *Belle method*. The main problem with the *Diamond Frame* method is the following. The Diamond Frame method was initially intended for semileptonic $B \to X\ell\nu$ decays. The construction of the frame heavily depends on the value of $\cos\theta_{BY}$,
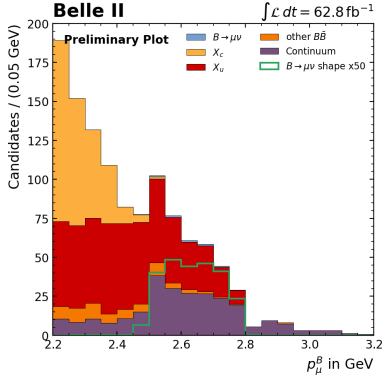
Figure 7.58: p_μ distribution of all MC samples after performing the optimal cut on the output classifier for the *Diamond Frame* method in mode 1.
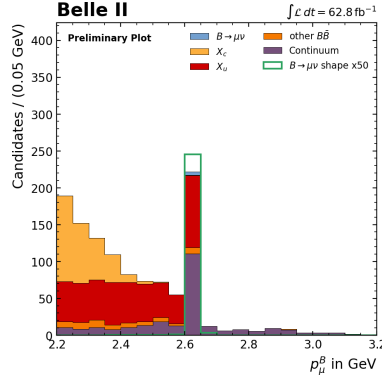
Figure 7.59: p_μ distribution of all MC samples after performing the optimal cut on the output classifier for the *Diamond Frame* method in mode 2.
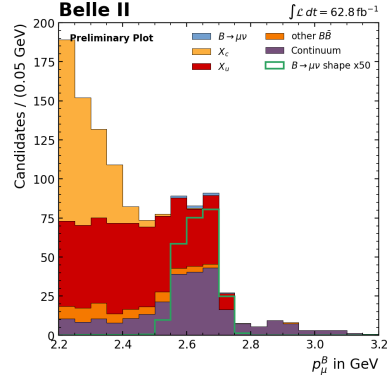
Figure 7.60: p_μ distribution of all MC samples after performing the optimal cut on the output classifier for the *Diamond Frame* method in mode 3.
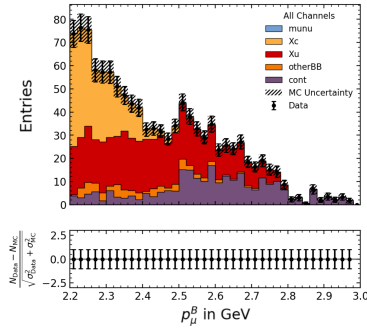
Figure 7.61: Fit to the Asimov data set for the *Diamond Frame* method in mode 1.
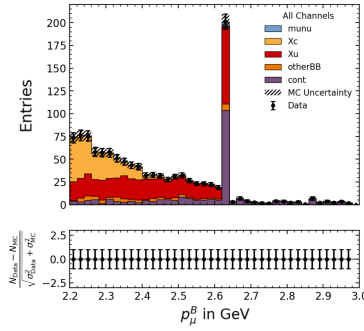
Figure 7.62: Fit to the Asimov data set for the *Diamond Frame* method in mode 2.
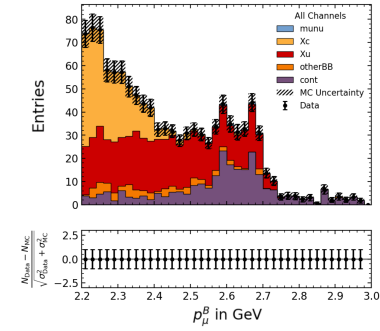
Figure 7.63: Fit to the Asimov data set for the *Diamond Frame* method in mode 3.

which is the cosine of the angle between the B meson and the Y candidate, where Y is the combination of X and ℓ. In the course of the frame construction the value of $\cos\theta_{BY}$ is restricted to have a value between 1 and -1. For the $B \to \mu\nu_\mu$ in mode 1 the B meson kinematics are constructed using only the muon as the daughter particle. Thus the kinematics of the B meson and the Y particle are equal, the θ_{BY} equals zero and the Diamond Frame is not constructed properly. For mode 2 the main issue is the direct correlation of $\cos\theta_{BY}$ to the muon momentum. This forces the signal as well as the background events, which initially have a value of $\cos\theta_{BY}$ between -1 and 1, to have the theoretically expected p_μ^B value resulting in a high sculpting of the background. The problem with mode 3 is then a combination of the issues from method 1 and 2.

p_T Calibration and Optimization

This method is similar to the *Belle method*, but instead of calibrating the p_z component of the B_{tag} the p_T component is used for the calibration. The transverse momentum p_T plays a similar role as the p_z

component, in the sense that it is sensitive to acceptance losses. We therefore map the reconstructed and constrained p_T component in bins in a range of $0.0\,\text{GeV} \leq p^*_{T,\text{constr}} \leq 0.332\,\text{GeV}$ to the mean of the generated component and perform the smooth spline fit. This is shown in Figure 7.64. We apply the function and calculate the other components via

$$
\begin{aligned}
p^*_{T,\text{calib}} &= f(p^*_{T,\text{constr}}) \\
p^*_{z,\text{calib}} &= \sqrt{|\vec{p}^*_{\text{constr}}|^2 - (p^*_{T,\text{calib}})^2} \\
p^*_{i,\text{calib}} &= \frac{p^*_{T,\text{calib}}}{p^*_{T,\text{constr}}} \cdot p^*_{i,\text{constr}} \qquad , i \in x, y.
\end{aligned}
\tag{7.24}
$$

Furthermore, we determine the correction factor to be $\xi = 0.501$ in order to optimize the components and construct the B_{sig} kinematics using Equation 7.8. The mapped components and smooth spline fit are shown in Figure 7.64. The p^B_μ distribution after an optimal cut on the output classifier is shown in Figure 7.65. The significance obtained from the fit to the Asimov data set, shown in Figure 7.66, is determined to be

$$
\sigma_{\text{stat}} = \frac{4.95751}{14.68997} = 0.33748.
\tag{7.25}
$$

The significance is slightly higher than the one determined using the $p_{x,y,z}$ calibration and optimization but it is lower than the one obtained using the *Belle method*. Thus it is not further considered.
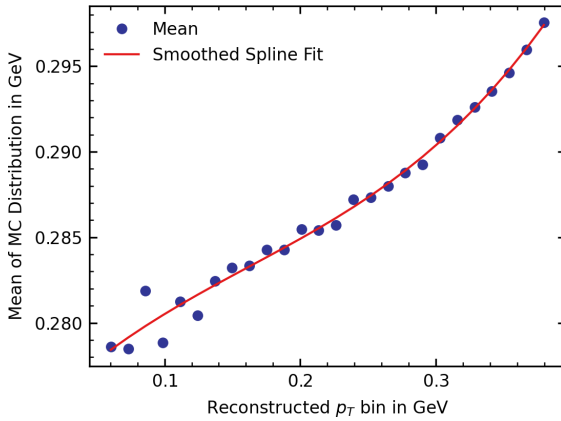
Figure 7.64: Binned constrained p_T versus generated p_T and smoothed spline fit.
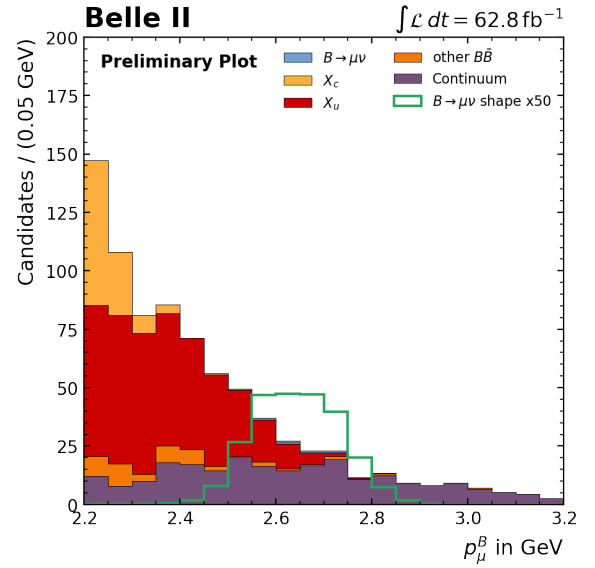
Figure 7.65: p_μ distribution of all MC samples after optimization and cut on the output classifier after calibration and optimization of p_T.

Calibration and Optimization in Bins of p_{miss}

This method was investigated as a combination of the calibration in bins of N_{lep} and θ_{ROE}. Instead of looking at individual errors in the reconstruction of the B_{tag} kinematics, we will look at the overall
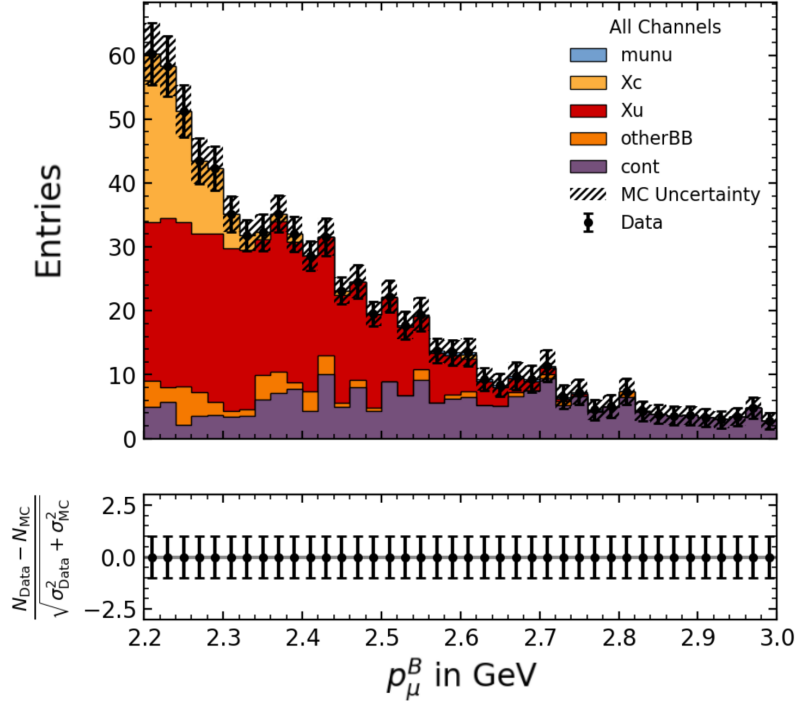
Figure 7.66: Fit to the Asimov data set after calibration and optimization of p_T.

resolution, which is directly correlated with the missing momentum in the event p_{miss}. For small values of p_{miss} the B_{tag} is reconstructed better than for events with a high missing momentum. We therefore propose a calibration in bins of p_{miss}. The missing momentum distribution of the signal MC sample is divided into three independent regions:

$$p_{\text{miss}} \leq 2.54 \,\text{GeV} \quad , \quad 2.54 \,\text{GeV} < p_{\text{miss}} \leq 2.84 \,\text{GeV} \quad \text{and} \quad p_{\text{miss}} > 2.84 \,\text{GeV}. \tag{7.26}$$

They contain the same number of events, which is shown in Figure 7.67. The effect of this division on the resolution of the p_z component is depicted in Figure 7.68. We subsequently perform three individual smoothed spline fits, shown in Figure 7.69, Figure 7.70 and Figure 7.71, and calculate the correction factors to be

$$
\begin{aligned}
\xi(p_{\text{miss}} \leq 2.54 \,\text{GeV}) &= 0.584 \, , \\
\xi(2.54 \,\text{GeV} < p_{\text{miss}} \leq 2.84 \,\text{GeV}) &= 0.696 \text{ and} \\
\xi(p_{\text{miss}} > 2.84 \,\text{GeV}) &= 0.521.
\end{aligned}
\tag{7.27}
$$

The optimization and boost into the B_{sig} rest frame is performed similar to that described in section 7.2.3. The fit to the Asimov data set is shown in Figure 7.72 and yields a significance of

$$\sigma_{\text{stat}} = \frac{5.01838}{14.41533} = 0.34813. \tag{7.28}$$
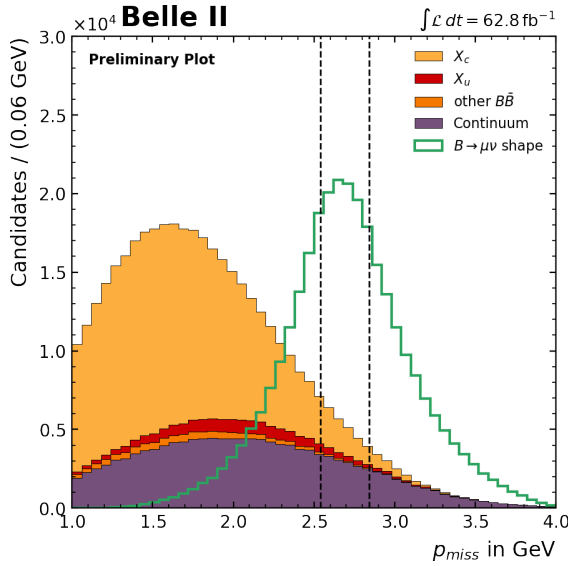
Figure 7.67: p_{miss} distributions of the scaled MC samples. The division into the three bins is also shown.
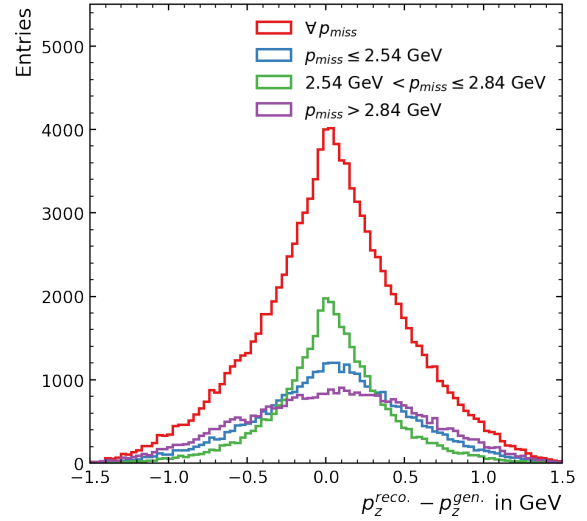
Figure 7.68: Effect on the resolution of the p_z component of B_{tag} introducing different regions of p_{miss}.

Figure 7.69: Binned constrained momentum component versus generated one and smoothed spline fit for the $p_{\text{miss}} \leq 2.54\,\text{GeV}$ region.
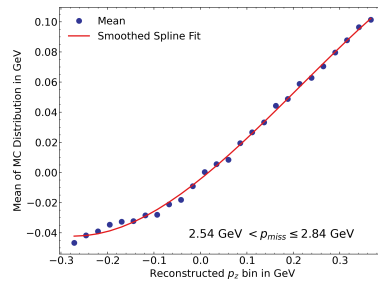
Figure 7.70: Binned constrained momentum component versus generated one and smoothed spline fit for the $2.54\,\text{GeV} < p_{\text{miss}} \leq 2.84\,\text{GeV}$ region.

Figure 7.71: Binned constrained momentum component versus generated one and smoothed spline fit for the $p_{\text{miss}} > 2.84\,\text{GeV}$ region.

The obtained significance is just barely higher than the one obtained with the *Belle method*, thus the calibration and optimization in bins of p_{miss} has to be equated with *Belle method*.
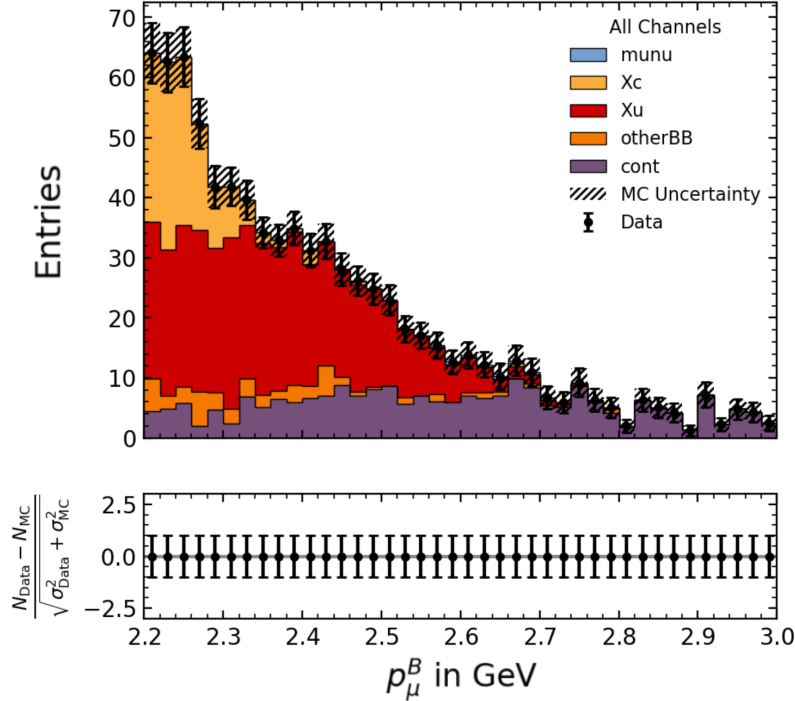
Figure 7.72: Fit to the Asimov data set after optimizing in bins of p_{miss}.

Iterative Neutrino Estimation Method

The proposed method is based on the neutrino estimation method described in section 7.2.4. In the course of exploring this method, we were able to estimate the B_{tag} kinematics using Equation 7.17. The idea of iterating the neutrino estimation is based on a calibration and optimization of the p_z component as done for the *Belle method*, however instead of using the constrained $p_{z,\text{constr}}^{*}$, we use the newly estimated $p_{z,\text{est}}^{*}$ component. We group the estimated component in bins and map it to the generated component, which is shown in Figure 7.73.

The calibration and optimization is then performed the same way as done in the *Belle method* using the newly extracted functional behavior and a calculated correction factor of $\xi = 1.116$.

We subsequently look at the results in the following ways. We perform a fit to the Asimov data set for the muon momentum boosted into the B_{sig} rest frame using the information obtained from the new calibration and optimization. Secondly, we use the gained information to estimate the hypothetical kinematics of the neutrino a second time to improve the resolution and again use Equation 7.17 to construct the signal side B meson kinematics. The fit to the Asimov data for the two cases are shown in Figure 7.74 and Figure 7.75. The main problem of using this method is that due to the iteration of the neutrino method the signal as well as the background MC samples are skewed pushed towards the expected muon momentum, as previously described in section 7.2.4. This is also represented by the

calculated significances:

$$\sigma_{\text{stat}} = \frac{4.96113}{14.53300} = 0.34137 \quad \text{after a second calibration and optimization, and}$$

$$\sigma_{\text{stat}} = \frac{4.96113}{14.92460} = 0.33241 \quad \text{after a second estimation of the neutrino kinematics. (7.29)}$$
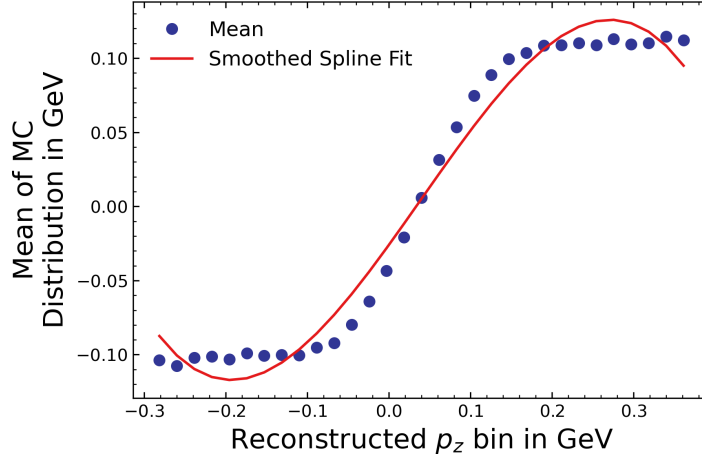
Figure 7.73: Binned newly estimated momentum component versus generated one and smoothed spline fit.
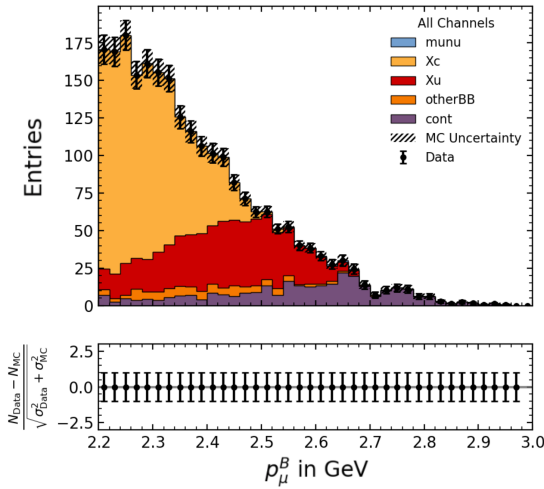
Figure 7.74: Fit to the Asimov data set after a second calibration and optimization.
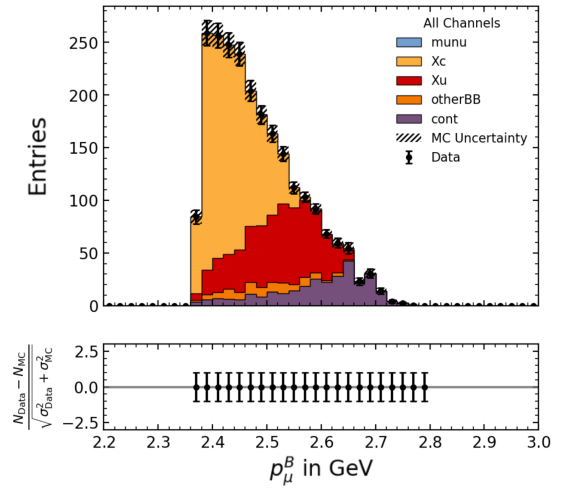
Figure 7.75: Fit to the Asimov data set after a second estimation of the neutrino kinematics.

7.2.6 Conclusion on the Estimation Methods

In the course of this thesis, many different methods were developed and investigated to improve the estimation of the B_{tag} kinematics. Only one method, the method of the neutrino estimation, does show

a small improvement. But this improvement should be treated with caution. Since the significance of the signal is quite small, it can be influenced by the choice of the binning and the range of the fit to the Asimov data set. A fine binning was chosen in order to be sensitive to differences in the shapes of the signal distribution for different methods, but still allow the bins to contain enough statistics. As a result though, the signal contribution is barely visible in the Asimov fits. Since the neutrino estimation method yields the highest significance, it is used for further investigation from here on.

Toy Fit and Significance Estimation

In this chapter a toy fit study is investigated to validate the performance of the fit to the Asimov data set. Furthermore, a significance estimation is presented.

8.1 Toy Fit

The next step is to probe the robustness of the signal extraction fit with realistic statistical fluctuations. We therefore create toy distributions by varying the counts in the individual bins within the expected Poissonian error for an integrated luminosity of $62.8\,\text{fb}^{-1}$. This is done for every MC sample individually. The aim of this study is to investigate if a bias is introduced into the fit result and if the error on the expected yield is estimated correctly. We fit the original MC sample distributions to 1000 generated toy distributions and calculate the difference between the original fit yield N_{fit} and the yield obtained by the fit to the toy distributions N_{toy}. We furthermore normalize the difference by the uncertainty on the toy yield σ_{toy}, which is then denoted as the pull

$$\text{Pull} = \frac{N_{\text{fit}} - N_{\text{toy}}}{\sigma_{\text{toy}}}. \tag{8.1}$$

The normalized pull distribution is expected to be a Gaussian distribution with a mean value of $\mu = 0$ and a standard deviation of $\sigma = 1$. A deviation from the expected mean is a sign for a bias in the parameter extraction during the fit, while a deviation from the expected standard deviation is a sign for an under- or overestimation of the uncertainty [8]. We therefore perform a Gauss fit on the obtained pull distribution and determine the mean and standard deviation to be

$$\mu = 0.053 \pm 0.0208 \qquad \text{and} \qquad \sigma = 0.925 \pm 0.0209. \tag{8.2}$$

The pull distribution as well as the Gaussian fit is shown in Figure 8.1. The mean as well as the standard deviation show good agreement to the expected value but small discrepancies occur. Thus we need to be careful in the future if a fit on real data is performed, since a bias might be introduced.
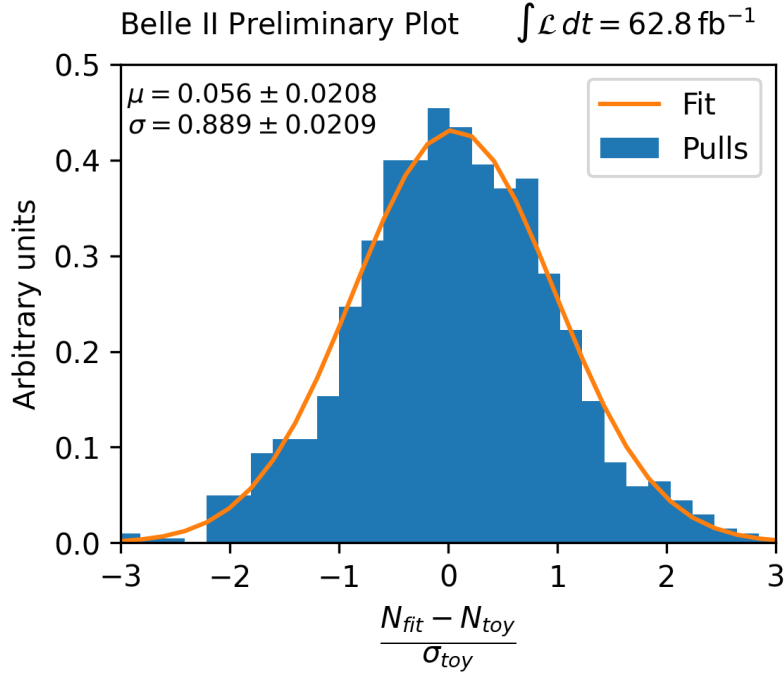
Figure 8.1: Plot of the pull distribution obtained by fitting to 1000 generated toy distributions. The calculated mean shows good agreement, as does the standard deviation, with the expected values.

8.2 Significance Estimation

Next, we perform the fit to the Asimov data set for different integrated luminosities. In total we have $200\,\mathrm{fb}^{-1}$ MC generated data in two batches we use for this study. We calculate the significance for integrated luminosities of 62.8, 100, 150 and $200\,\mathrm{fb}^{-1}$. We expect the significance to scale with the square root of the luminosity and therefore fit the expected functional behavior to the measurements. This is shown in Figure 8.2 and the fit shows good agreement with the measurement. Using the parameters extracted from the fit, we calculate the integrated luminosity needed in order to achieve a significance of 3σ needed for observation to be at least $4.53\,\mathrm{ab}^{-1}$. The 5σ needed for discovery of the $B \rightarrow \mu\nu_\mu$ decay can be achieved with an integrated luminosity of at least $12.57\,\mathrm{ab}^{-1}$. However, it is possible that further improvements to the discussed methods reduce these luminosity targets. It is important to note that only the statistical and not the systematic uncertainties are taken into account in this calculation. Belle II aims for a recording of a dataset with a size of $50\,\mathrm{ab}^{-1}$ over the running time [33]. If that target is met, the $B \rightarrow \mu\nu_\mu$ decay can be discovered in the future.
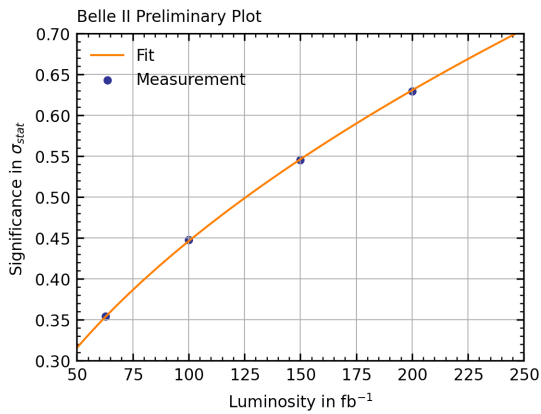
Figure 8.2: Measured significance values as a function of toy luminosity used in the Asimov fit. A fit to the measurement is shown.
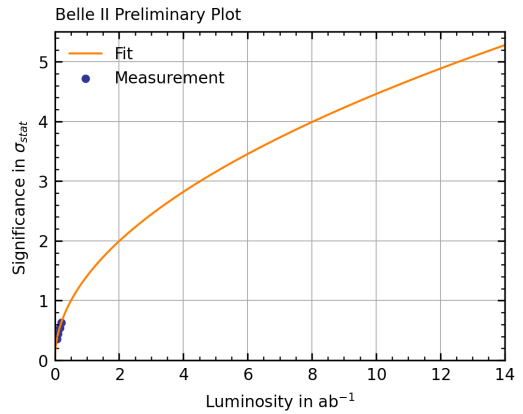
Figure 8.3: Expected significance for higher values of integrated luminosity estimated with the fit to the measurements.

Conclusion and Outlook

9.1 Conclusion

In this thesis the rare $B \rightarrow \mu\nu_\mu$ decay was investigated using the inclusive tagging approach. It is challenging to observe this specific decay channel in the Belle II experiment. The reconstruction of the tag side B meson kinematics is deteriorated by multiple factors, which were investigated by studying the different reconstruction errors. Measures were employed to improve the reconstruction, but many reconstruction errors cannot be avoided and are due to the physical properties of the Belle II detector. A different approach was taken to improve the estimation of the B_{tag} kinematics. Starting with applying the method of constraining, calibrating and optimizing the p_z component of B_{tag} which was previously used at Belle, novel methods were developed. In order to compare two different methods, a fit to an Asimov data set was performed for each method and the statistical significances were compared. A calibration of the p_x, p_y and p_z component, as well as a calibration in bins of N_{lep} and θ_{ROE} has not led to the desired increase in significance. The method of the neutrino estimation was investigated and showed a small increase in the significance compared to the method used by the Belle analysis. Additional methods were also presented. The statistical significances of all investigated methods are summarized in Table 9.1. The neutrino estimation method was used for further considerations. The fit stability was validated by fitting the MC samples to 1000 generated toy distributions and observing the obtained pull distributions. A significance estimation was employed in order to assess the integrated luminosity needed to observe and discover the $B \rightarrow \mu\nu_\mu$ decay at Belle II.

9.2 Outlook

It is possible that there are further methods not yet investigated which could result in a higher significance. As a conclusion from the here presented study, it seems that the method used by Belle does quite well, but we showed that the neutrino estimation method seems to be able to improve on it a bit. It is possible, that an extension of this method can provide even more improvement. In order to increase the chances to discover the $B \rightarrow \mu\nu_\mu$ decay, the ROE mask has to be optimized in order to reduce the error in the reconstruction due to non-primary particles. But more importantly, the background suppression has to be improved upon. The main focus in the analysis was put on new methods to estimate the B_{tag} kinematics, but for following studies a different approach for background

suppression should be employed, investigating other multivariate analysis methods for example. Doing that successfully could lead to the discovery of the $B \to \mu\nu_\mu$ decay at Belle II in the next years.

Method	Significance
Calibration and Optimization of p_z	1.00000
$p_{x,y,z}$ Calibration and Optimization	0.95561
Calibration and Optimization in Bins of N_{lep}	0.98747
Calibration and Optimization in Bins of θ_{ROE}	0.99787
Neutrino Estimation Method	1.01804
Diamond Frame mode 1	0.49987
Diamond Frame mode 2	0.83117
Diamond Frame mode 3	0.60644
p_T Calibration and Optimization	0.96963
Calibration and Optimization in Bins of p_{miss}	1.00023
Iterative Neutrino Estimation Method after a second calibration and optimization	0.98081
Iterative Neutrino Estimation Method after a second estimation of the neutrino kinematics	0.95506

Table 9.1: Values of the statistical significance obtained by the employment of different methods in order to improve the estimation of the B_{tag} kinematics. The values are scaled so that the significance of the *Belle Method* is 1.0, for better comparison.

Boosted Decision Tree Training

A.1 Correlations
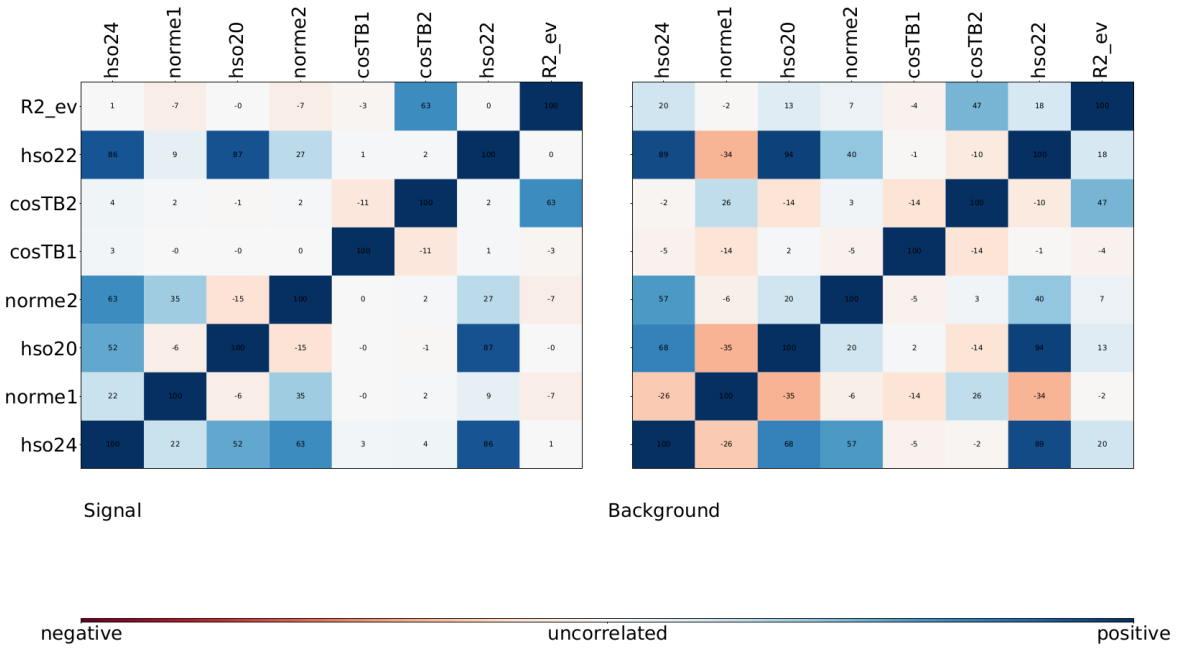
Figure A.1: Correlations between the variables used in the BDT training process.

A.2 Continuum Suppression Variables
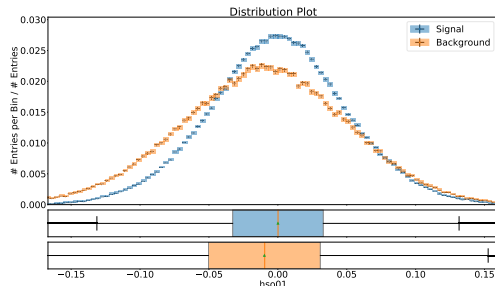
(a) h_1^{so}
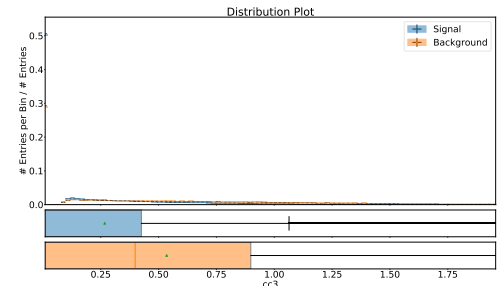
(b) cc_3
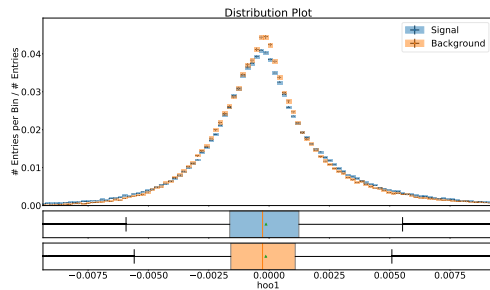
(c) h_1^{oo}
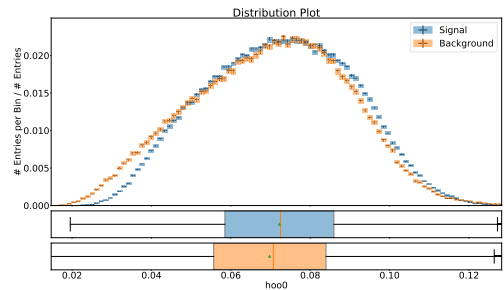
(d) h_0^{oo}

(e) et

(f) cc_6

(g) $Thrust_O$

(h) h_4^{so}
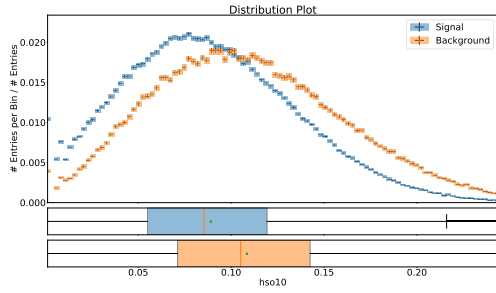
(i) h_{10}^{so}
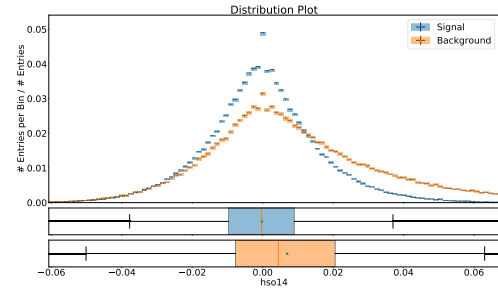
(i) h_{14}^{so}

(k) cc_8

(l) mm_2

(m) cc_5

(n) h_0^{so}

(o) h_4^{oo}

(p) h_3^{so}
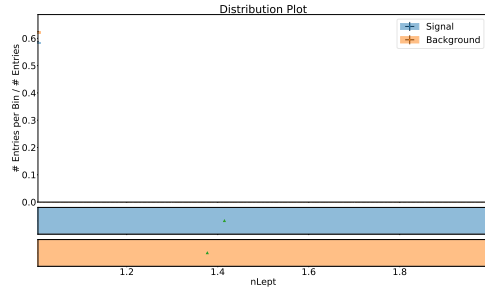
(q) N_{lep}
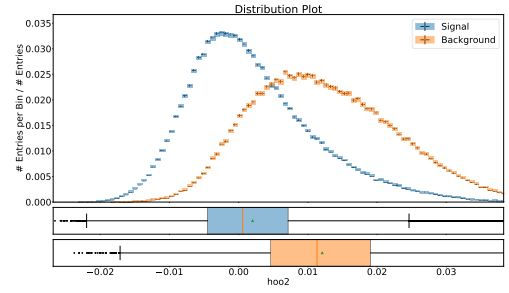
(r) h_2^{oo}
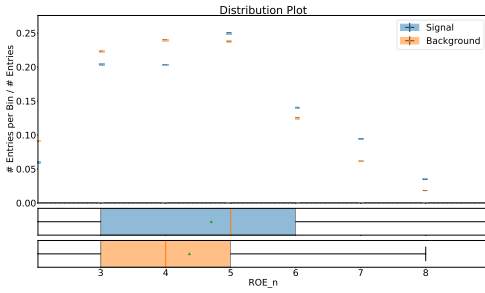
(s) N_{Tracks}^{ROE}
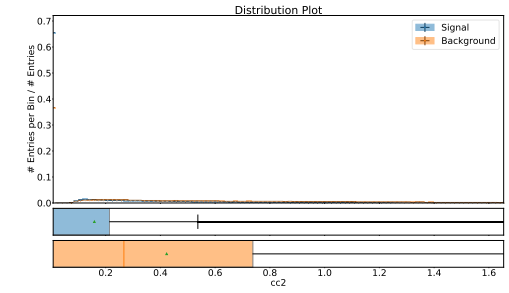
(t) cc_2
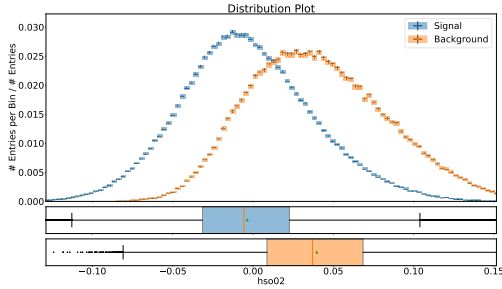
(u) h_2^{so}

(v) h_{12}^{so}

(w) R_2

(x) h_3^{oo}

(y) cc_4

(z) cc_7

(aa) cc_9
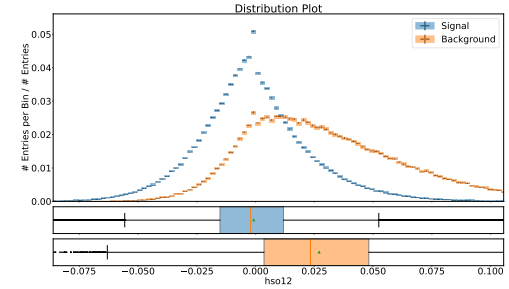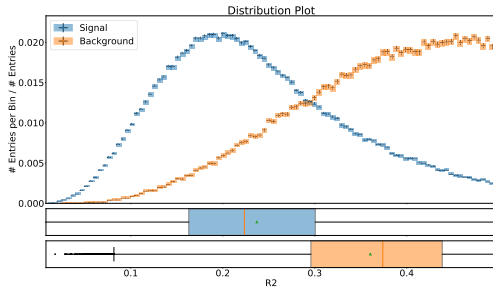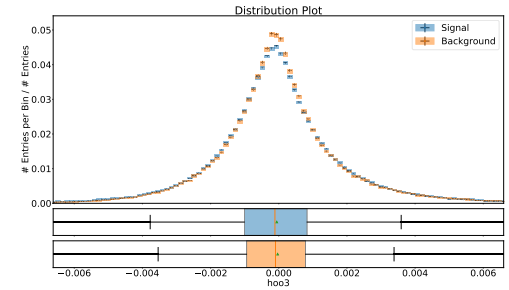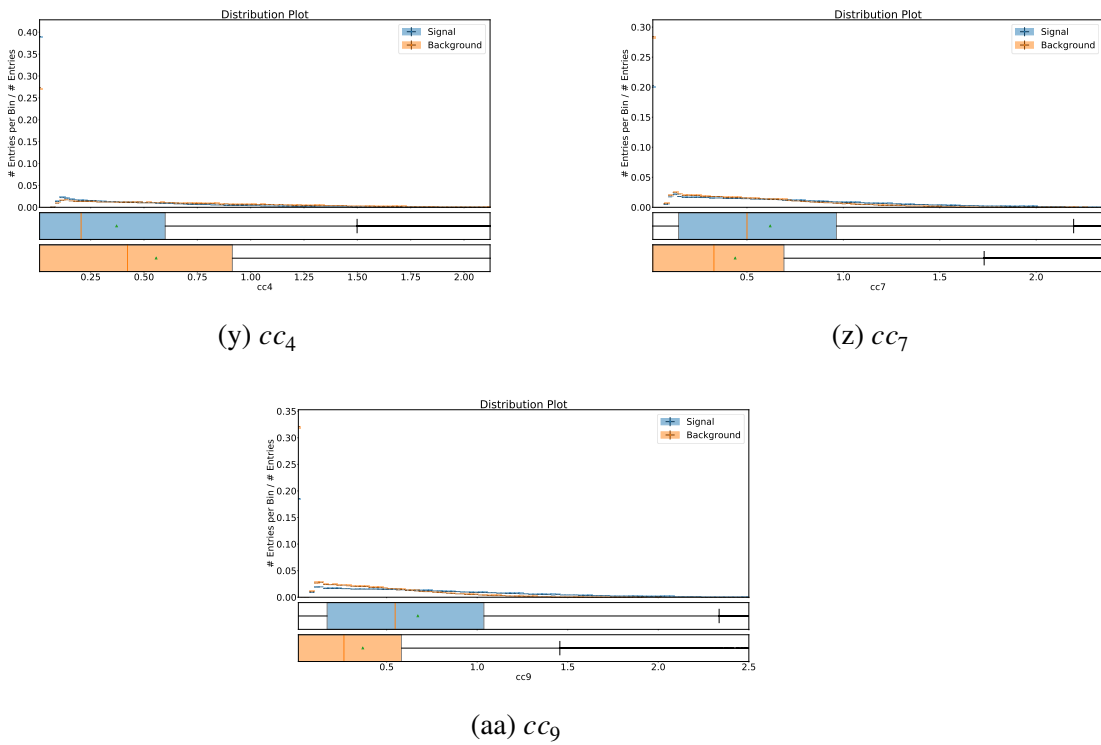
Figure A.2: Normalized distributions of all variables used in the BDT training process. The background represents the continuum.
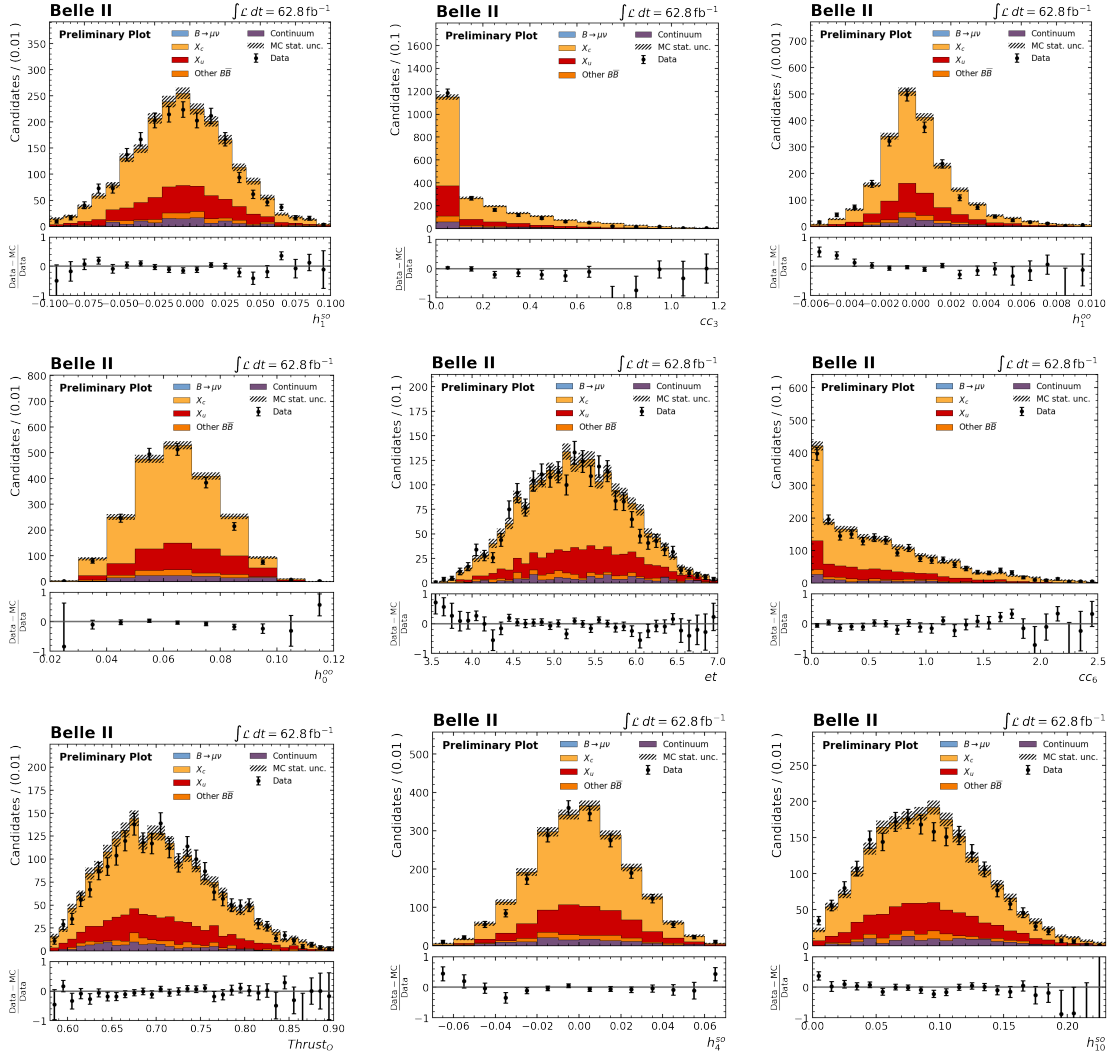
A.3 Data-MC Agreement
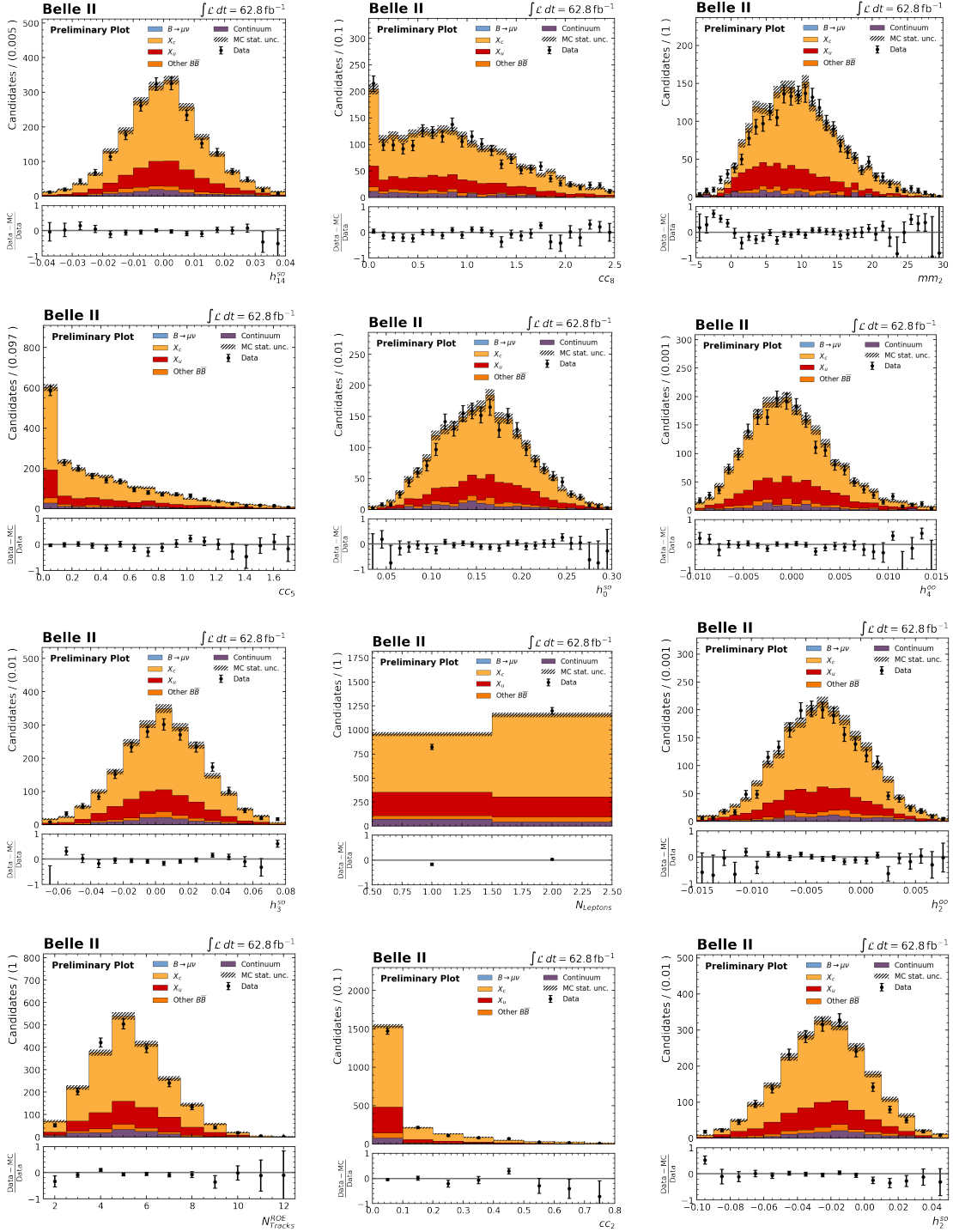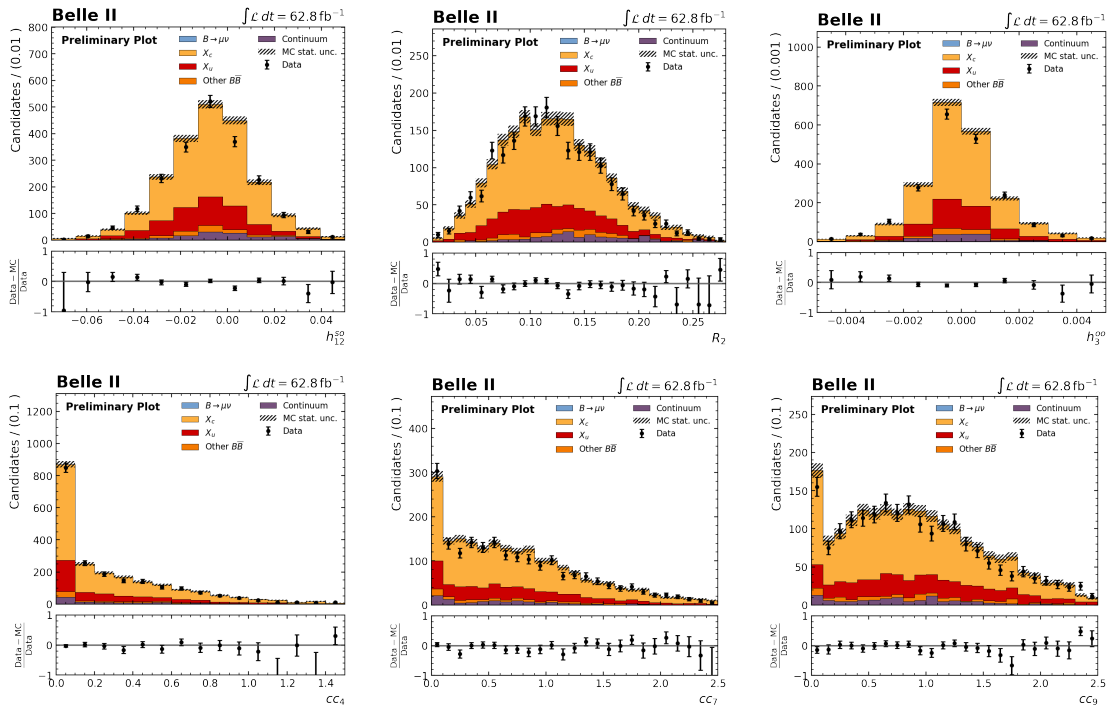
Figure A.3: Comparison of blinded data to generated MC samples for the unused continuum suppression variables

Reconstruction Errors of B_{tag}
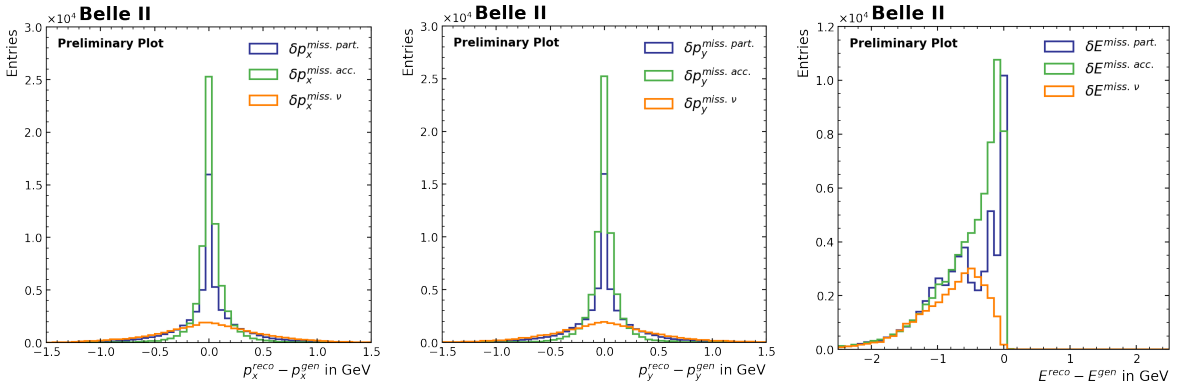
Figure B.1: Distribution of the errors due to the acceptance, missing neutrinos and missing particles for the p_x, p_y and E component of the ROE
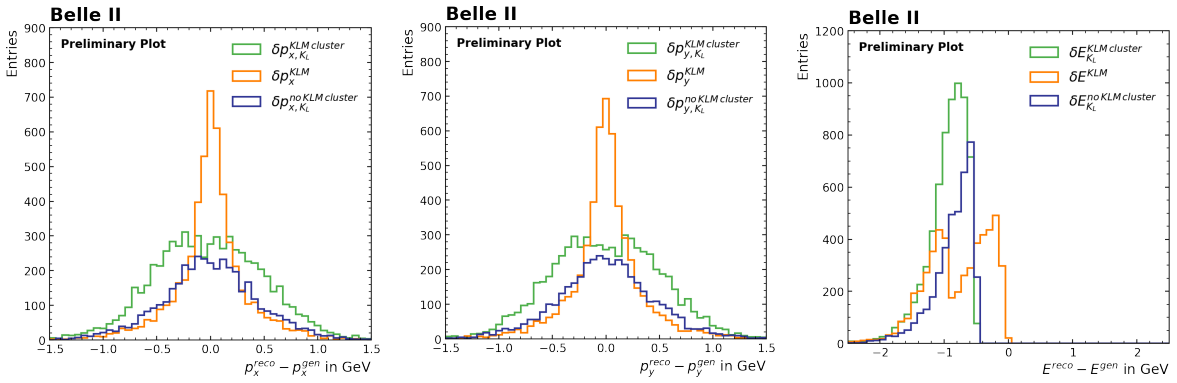
Figure B.2: Distribution of the errors due to the K_L and the KLM cluster for the p_x, p_y and E component of the ROE
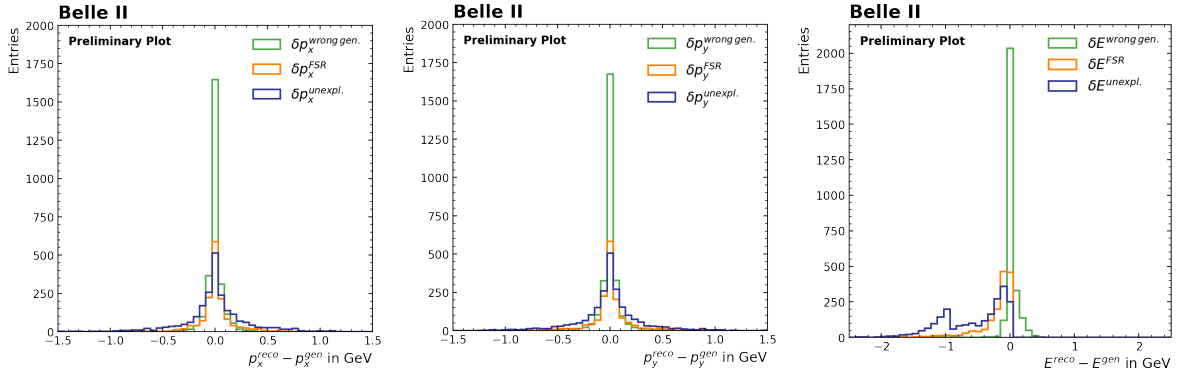
Figure B.3: Distribution of the errors due to final state radiation, wrongly generated decays and unexplained matching for the p_x, p_y and E component of the ROE
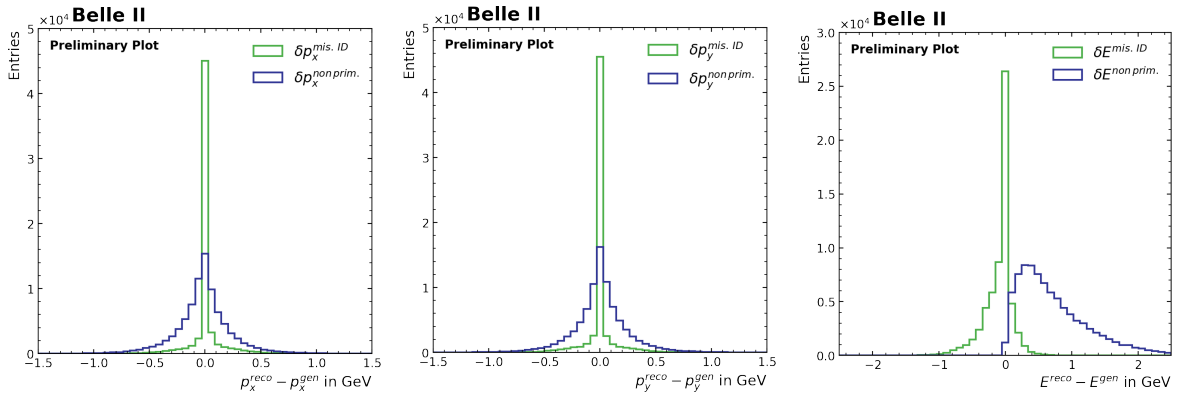
Figure B.4: Distribution of the errors due to non primary and misidentified particles for the p_x, p_y and E component of the ROE
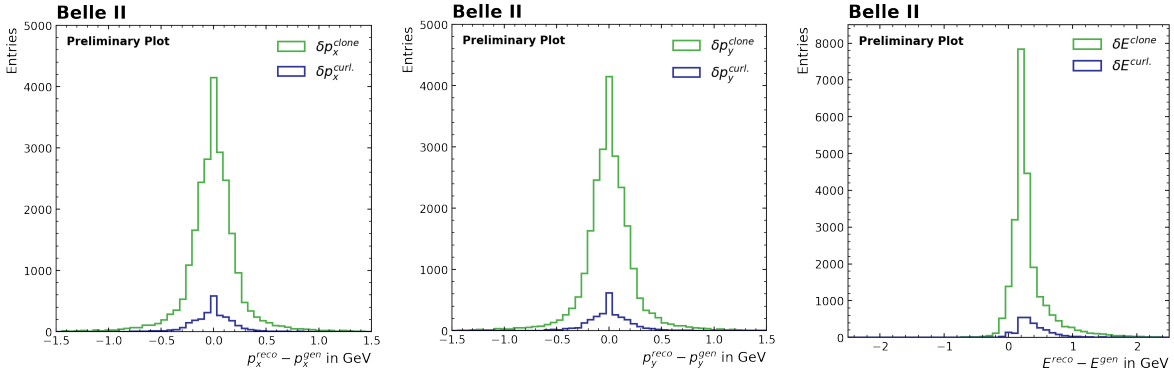
Figure B.5: Distribution of the errors due to curlers and clones for the p_x, p_y and E component of the ROE
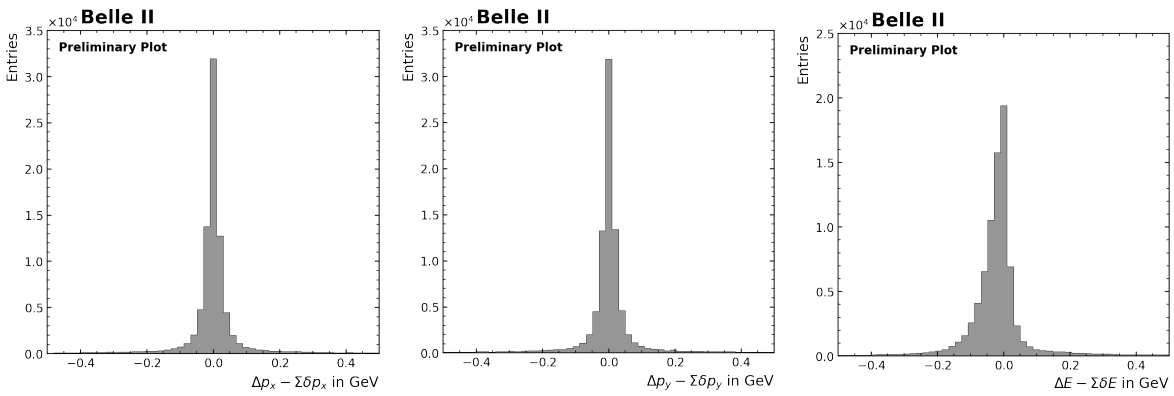
Figure B.6: Verification that most errors have been identified for the p_x, p_y and E component of the ROE

Precise Estimation of the B_{tag} Four-Momentum

C.1 Belle Analysis Method
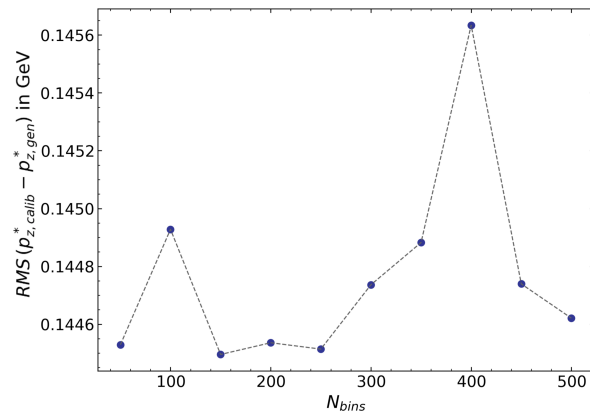
Figure C.1: *RMS* for different number of bins from 50 to 500 in steps of 50

C.2 $p_{x,y,z}$ Calibration and Optimization
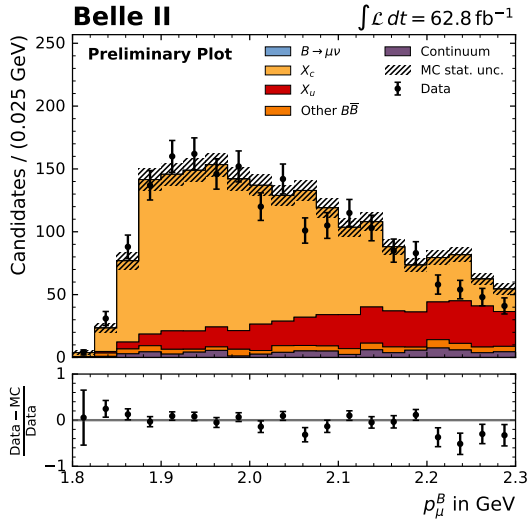
Figure C.2: Data-MC agreement for the left sideband after optimizing the $p_{x,y,z}$ component
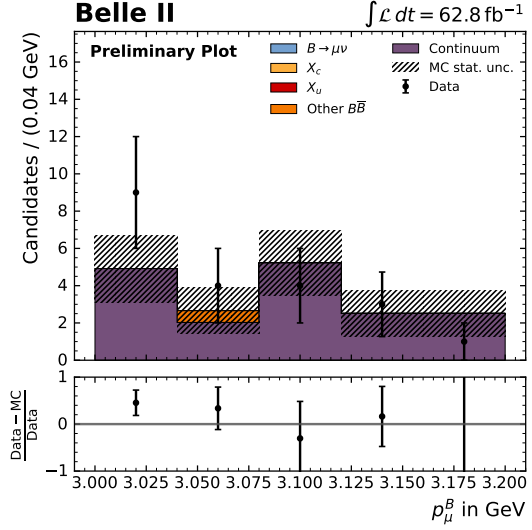
Figure C.3: Data-MC agreement for the right sideband after optimizing the $p_{x,y,z}$ component
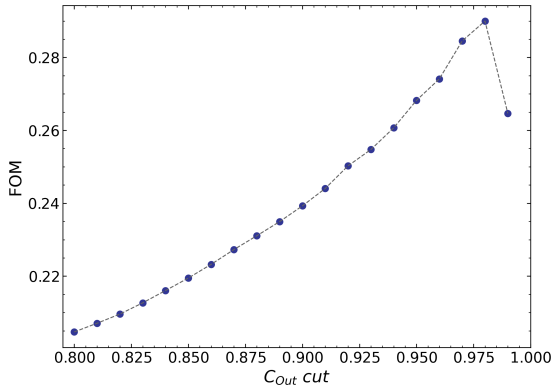
Figure C.4: FOM againt different cuts on the output classifier C_{out} after optimizing the $p_{x,y,z}$ component

Figure C.5: Correlations between the MC samples after optimizing the $p_{x,y,z}$ component

C.3 Calibration and Optimization in Bins of N_{lep}
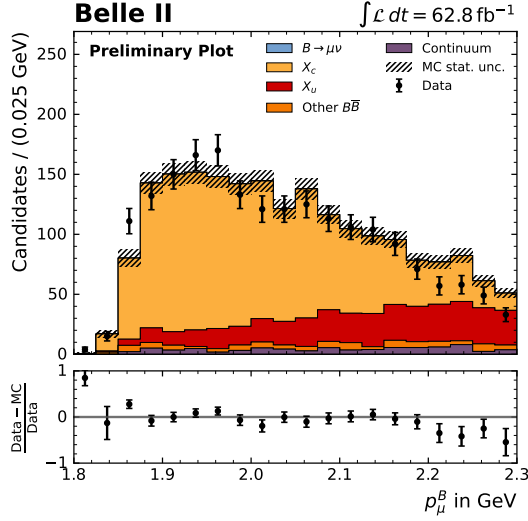
Figure C.6: Data-MC agreement for the left sideband after optimizing in bins of $N_{lep.}^{reco.}$.
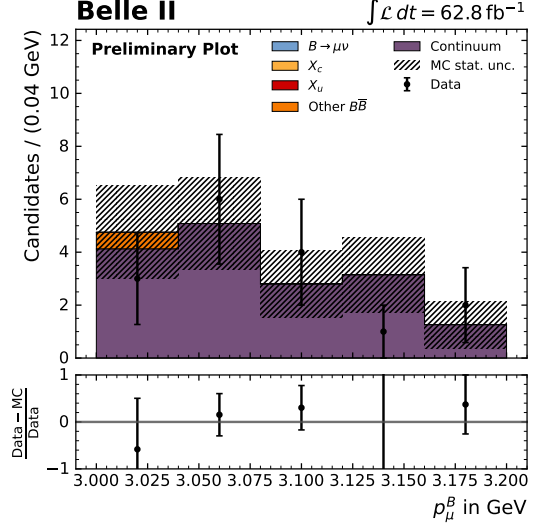
Figure C.7: Data-MC agreement for the right sideband after optimizing in bins of $N_{lep.}^{reco.}$.
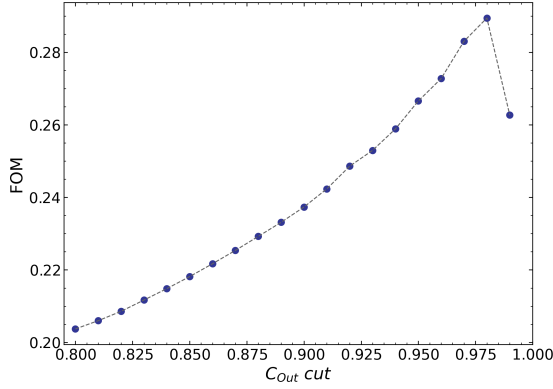
Figure C.8: FOM againt different cuts on the output classifier C_{out} after optimizing in bins of $N_{lep.}^{reco.}$.

Figure C.9: Correlations between the MC samples after optimizing in bins of $N_{lep.}^{reco.}$.

C.4 Calibration and Optimization in Bins of θ_{ROE}
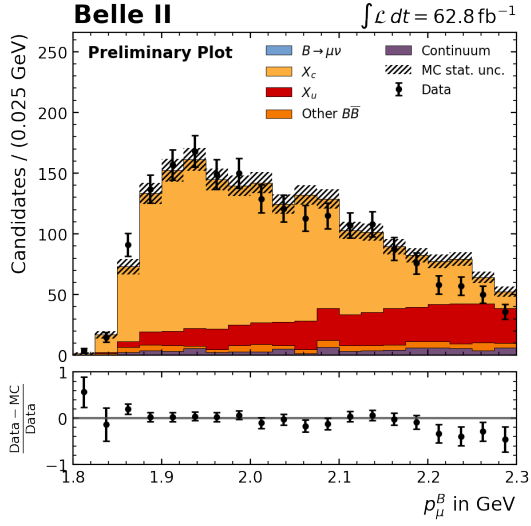
Figure C.10: Data-MC agreement for the left sideband after optimizing in bins of θ_{ROE}
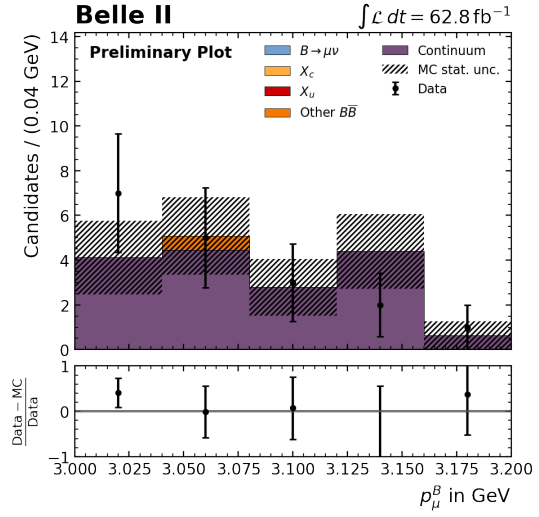
Figure C.11: Data-MC agreement for the right sideband after optimizing in bins of θ_{ROE}
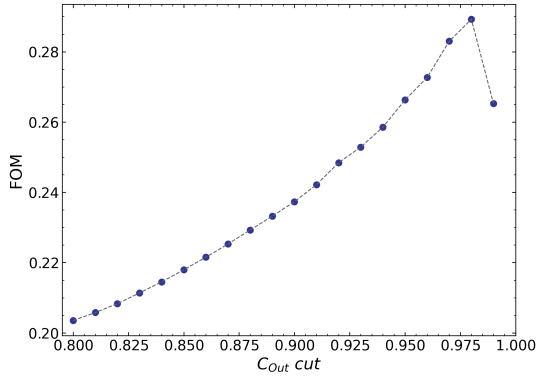
Figure C.12: FOM againt different cuts on the output classifier C_{out} after optimizing in bins of $N_{lep.}^{reco.}$

Figure C.13: Correlations between the MC samples after optimizing in bins of $N_{lep.}^{reco.}$

C.5 Neutrino Estimation Method
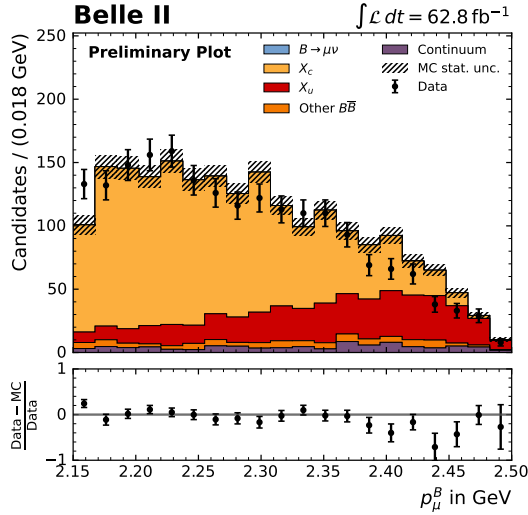
Figure C.14: Data-MC agreement for the left sideband after estimating the neutrino kinematics
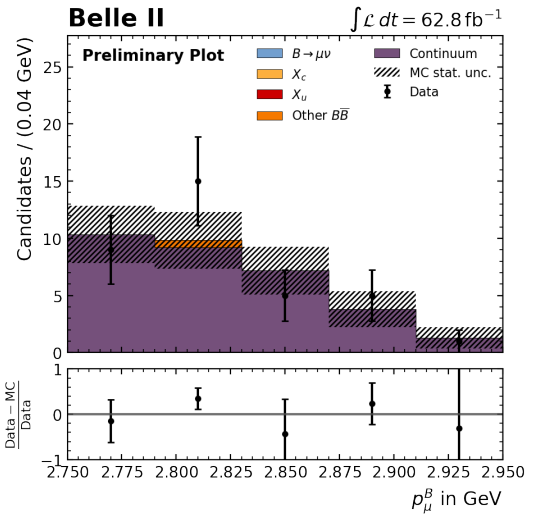
Figure C.15: Data-MC agreement for the right sideband after estimating the neutrino kinematics
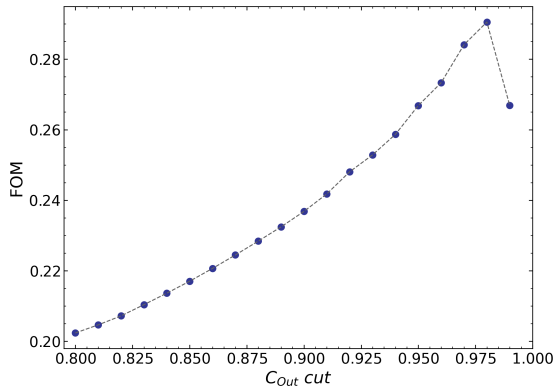
Figure C.16: FOM against different cuts on the output classifier C_{out} after estimating the neutrino kinematics
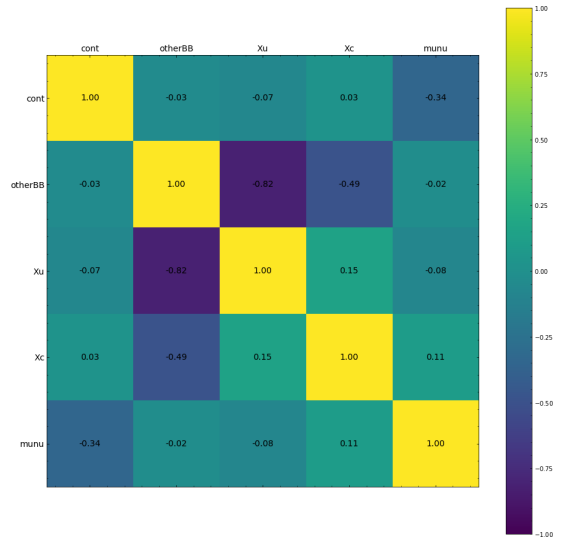
Figure C.17: Correlations between the MC samples after estimating the neutrino kinematics

References

[1] Cern Accelerating science, *The Standard Model*, [Online; accessed 30.08.2021], URL: https://home.cern/science/physics/standard-model (cit. on p. 1).

[2] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013 (cit. on pp. 1, 3, 4, 6).

[3] T. Virdee, *Beyond the standard model of particle physics*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **374** (2016) 20150259 (cit. on p. 1).

[4] P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01 (cit. on pp. 1, 6, 9).

[5] E. Kou et al., *The Belle II Physics Book*, Progress of Theoretical and Experimental Physics **2019** (2019), ISSN: 2050-3911, URL: http://dx.doi.org/10.1093/ptep/ptz106 (cit. on pp. 1, 8, 10, 15, 24).

[6] J. Dingfelder and T. Mannel, *Leptonic and semileptonic decays of B mesons*, Rev. Mod. Phys. **88** (3 2016) 035008, URL: https://link.aps.org/doi/10.1103/RevModPhys.88.035008 (cit. on pp. 1, 3).

[7] M. T. Prim et al., *Search for $B^+ \to \mu^+ \nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging*, Physical Review D **101** (2020), ISSN: 2470-0029, URL: http://dx.doi.org/10.1103/PhysRevD.101.032007 (cit. on pp. 1, 5, 17, 21, 23, 38).

[8] M. T. Prim, *Search for $B \to \mu\nu_\mu$ at Belle*, PhD thesis: Karlsruhe Institute of Technology (KIT), 2019 (cit. on pp. 3, 4, 9, 17, 21, 39, 74).

[9] A. J. Bevan et al., *The Physics of the B Factories*, The European Physical Journal C **74** (2014), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-014-3026-9 (cit. on pp. 3, 5, 15, 24).

[10] Wikimedia Commons, *Standard model of elementary particles: the 12 fundamental fermions and 5 fundamental bosons.* [Online; accessed 13.07.2021], 2019, URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (cit. on p. 3).

[11] Hok-Chuen Cheng, *Two-body decay kinematics*, [Online; accessed 03.08.2021], 2015, URL: https://indico.cern.ch/event/391122/contributions/928962/attachments/782786/1073126/twoBodyDecay.pdf (cit. on p. 5).

[12] T. Keck, *Machine learning algorithms for the Belle II experiment and their validation on Belle data*, PhD thesis: Karlsruher Institut für Technologie (KIT), 2017, 240 pp. (cit. on pp. 6, 20).

[13] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, 2016, arXiv: 1609.06119 [cs.LG] (cit. on p. 6).

[14] Christian Böser, Simon Fink, Steffen Röcker,
Introduction to Boosted Decision Trees, A multivariate approach to classification problems,
[Online; accessed 04.08.2021], URL: https://indico.scc.kit.edu/event/48/
contributions/3410/attachments/1690/2312/BDT_KSETA_Freudenstadt.pdf
(cit. on p. 7).

[15] Syed Sadat Nazrul, *Receiver Operating Characteristic Curves Demystified (in Python)*,
[Online; accessed 04.08.2021], URL: https://towardsdatascience.com/receiver-
operating-characteristic-curves-demystified-in-python-bd531a4364d0
(cit. on p. 7).

[16] K. Akai, K. Furukawa and H. Koiso, *SuperKEKB collider*,
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,
Detectors and Associated Equipment **907** (2018) 188, Advances in Instrumentation and
Experimental Methods (Special Issue in Honour of Kai Siegbahn), ISSN: 0168-9002,
URL: https://www.sciencedirect.com/science/article/pii/S0168900218309616
(cit. on p. 8).

[17] T. Abe et al., *Belle II Technical Design Report*, 2010,
arXiv: 1011.0352 [physics.ins-det] (cit. on p. 8).

[18] H. Kolanoski and N. Wermes, *Teilchendetektoren: Grundlagen und Anwendungen*,
Springer, 2016, ISBN: 978-3-662-45349-0, 978-3-662-45350-6 (cit. on p. 8).

[19] N. Braun, *Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II
Experiment*, PhD thesis: Karlsruhe Institute of Technology (KIT), 2018,
URL: https://doi.org/10.5445/IR/1000089317 (cit. on pp. 8, 15).

[20] Belle II Collaboration, *Belle II Detector 3D model*, [Online; accessed 17.08.2021],
URL: https://www.belle2.org/archives/ (cit. on p. 11).

[21] Bernlochner, Florian et al., *Online Data Reduction for the Belle II Experiment using DATCON*,
EPJ Web Conf. **150** (2017) 00014,
URL: https://doi.org/10.1051/epjconf/201715000014 (cit. on p. 11).

[22] R. Frühwirth and A. Strandlie,
Track fitting with ambiguities and noise: A study of elastic tracking and nonlinear filters,
Computer Physics Communications **120** (1999) 197, ISSN: 0010-4655,
URL: https://www.sciencedirect.com/science/article/pii/S0010465599002313
(cit. on p. 15).

[23] Milesi, Marco, Tan, Justin and Urquijo, Phillip, *Lepton identification in Belle II using
observables from the electromagnetic calorimeter and precision trackers*,
EPJ Web Conf. **245** (2020) 06023,
URL: https://doi.org/10.1051/epjconf/202024506023 (cit. on p. 16).

[24] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. A **462** (2001) 152, ed. by S. Erhan, P. Schlein and Y. Rozen
(cit. on p. 16).

[25] S. Jadach, B. Ward and Z. Was,
The precision Monte Carlo event generator for two-fermion final states in collisions,
Computer Physics Communications **130** (2000) 260, ISSN: 0010-4655,
URL: http://dx.doi.org/10.1016/S0010-4655(00)00048-5 (cit. on p. 16).

[26] S. Jadach, B. F. L. Ward and Z. Was,
Coherent exclusive exponentiation for precision Monte Carlo calculations,
Phys. Rev. D **63** (11 2001) 113009,
URL: https://link.aps.org/doi/10.1103/PhysRevD.63.113009 (cit. on p. 16).

[27] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*,
Computer Physics Communications **191** (2015) 159, ISSN: 0010-4655,
URL: http://dx.doi.org/10.1016/j.cpc.2015.01.024 (cit. on p. 16).

[28] C. Ramirez, J. F. Donoghue and G. Burdman, *Semileptonic $b \to u$ decay*,
Phys. Rev. D **41** (5 1990) 1496,
URL: https://link.aps.org/doi/10.1103/PhysRevD.41.1496 (cit. on p. 17).

[29] L. Cao et al., *Measurements of partial branching fractions of inclusive $B \to X_u \ell^+ \nu_\ell$ decays with hadronic tagging*, Phys. Rev. D **104** (1 2021) 012008,
URL: https://link.aps.org/doi/10.1103/PhysRevD.104.012008 (cit. on p. 17).

[30] B. I. Collaboration et al.,
Search for $B^+ \to K^+ \nu \bar{\nu}$ decays using an inclusive tagging method at Belle II, 2021,
arXiv: 2104.12624 [hep-ex] (cit. on pp. 23, 36).

[31] M. Röhrken, *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment*, PhD thesis, 2012 (cit. on p. 24).

[32] M. S. Gill and /SLAC,
Measurement of B -> D Form Factors in the Semileptonic Decay B -> D* l nu at BaBar,
(2006), URL: https://www.osti.gov/biblio/875810 (cit. on p. 66).

[33] S. H. Robertson, *The Belle II Experiment*,
Journal of Physics: Conference Series **1271** (2019) 012011,
URL: https://doi.org/10.1088/1742-6596/1271/1/012011 (cit. on p. 75).

List of Figures

List of Tables