Tagged Analysis of $B^0 \to \pi^- \ell^+ \nu_\ell$ at Belle

Jonathan Wolfgang Grumke

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, ...

Date Signature

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

Acknowledgements

I would like to thank Jochen Dingfelder for giving me the opportunity to conduct my studies as a member of his Belle analysis group. After completing my bachelor thesis in his group, it was clear to me that I also want to carry out my master thesis with this beautiful bunch of physicists. I thank Peter Lewis for his supervision and guidance throughout the past year, he was a major reason why I wanted to return to Belle. I appreciate the input by Florian Bernlochner and thank him for refereeing this thesis. Furthermore, I want to thank Svenja Granderath for her countless advice and for always bringing positive vibes to the office. It was always great with her, Stephan, Daniel and Rieka, as long as we were allowed to be together in one office. Not being able to share an office for the majority of the past year due to the Covid pandemic was mentally exhausting, but also caused negative repercussions on the physics side of things. The lack of both face-to-face discussions with a whiteboard at hand, and of unscheduled coffee breaks, which often lead to new ideas and insights, overshadowed the whole period of this study. This was partially countered by the much appreciated introduction of a virtual office by Stephan Duell. I wish him all the best for the final strides of his PhD. Finally, I want to thank my family for their ongoing support over the past years.

Contents

CHAPTER 1

Introduction

The Standard Model of particle physics has proven to be successful in describing interactions of the most fundamental particles in our universe. B-factories, such as KEKB, provide a great opportunity for testing predictions made by the Standard Model and measuring parameters that are needed as input to the Standard Model. The advantage of B-factories over hadron colliders is the precise knowledge of the initial state of a physical process and the resulting clean outcome containing low backgrounds.

A set of parameters whose values are not predicted by the Standard Model are the elements of the Cabbibo-Kobayashi-Maskawa (CKM) matrix which describes mixing of quark flavors via the weak interaction. The matrix element $\left|V_{\mathrm{ub}}\right|$ is of special interest to this study and can be extracted in two different ways: exclusively and inclusively. The inclusive method does not concern itself with the precise identity of the final state, whereas the exclusive method only looks at a specific final state. There exists a persistent tension between the inclusive and exclusive measurement of $\left|V_{\mathrm{ub}}\right|$ of $\sim 3.5\sigma$ [1]. The most accessible channel for an exclusive measurement of $\left|V_{\mathrm{ub}}\right|$ is the decay $B^0 \to \pi^- \ell^+ \nu_\ell$, where $\ell = e, \mu$.

In this thesis $B^0 \to \pi^- \ell^+ \nu_\ell$ is studied with hadronic tagging on data from the Belle experiment. Tagging is a procedure where one B meson is reconstructed from its decay products, allowing to recover information of the other B in an event. In signal events only two B mesons and no other particles are initially created. A similar study on the same dataset was performed previously [2] and a branching fraction of $\mathcal{B}(B^0 \to \pi^- \ell^+ \nu_\ell) = (1.49 \pm 0.09) \times 10^{-4}$ was determined. Since the start of the Belle II experiment in 2019 a new tagging algorithm, the Full Event Interpretation (FEI) [3], has been introduced. The FEI achieves a higher efficiency than the previous algorithm, mostly by including more hadronic decay channels of B mesons and intermediate particles. These additional modes may, however, degrade the resolution of reconstructed properties like the four-momentum of a B meson. Whether the trade-off between increased efficiency and decreased resolution is worthwhile is still an open question and is part of my study. Multiple studies investigating the performance of the FEI in charmed B decays have been performed before, whereas in this study I investigate the performance in a charmless B decay.

The method discussed in this thesis is a tagged exclusive measurement. After reconstructing the signal decay $B^0 \to \pi^- \ell^+ \nu_\ell$ from simulated data, the analysis is split into two versions to study effects of using the new tagging algorithm compared to the previous one. Selections aiming at the suppression of backgrounds, while retaining signal events, are deployed. This is partially achieved by making use of multivariate analysis. The agreement between simulated and experimental data will be inspected at in the sideband of the missing mass squared (m^2_{miss}) distribution.

The following two chapters contain an overview of the theoretical background relevant to this study and an overview of the Belle detector. Chapter 4 provides an overview of two different tagging algorithms, an introduction to Monte Carlo simulations and the course of action taken in order to reconstruct the signal decay $B^0 \to \pi^- \ell^+ \nu_\ell$. In Chapter 5 the suppression of continuum backgrounds with multivariate analysis is explained. Tag and signal selections are described in Chapters 6 and 7, respectively. Chapter 8 presents the fit procedure used to extract the signal from the m^2_{miss} distribution, while paying special attention to the resolution of m^2_{miss}. Chapter 9 contains the conclusion.

CHAPTER 2

Theory

This chapter provides an introduction to theoretical concepts relevant to this analysis. The first section provides an overview of the Standard Model of particle physics, followed by further details on the weak interaction and the production of B mesons at Belle. Furthermore, semileptonic B decays and a multivariate analysis method is explained.

2.1 The Standard Model of Particle Physics

The Standard Model of particle physics is a quantum field theory that describes a set of fundamental particles and their interactions. There are two types of particles in the Standard Model, which are called fermions with half-integer spin and bosons with integer-spin. According to the Standard Model, all matter is made from fermions while bosons are the force carriers that allow matter to interact. A schematic overview of the Standard Model can be seen in Figure 2.1. Fermions are divided into leptons and quarks, which are split into three generations of quark pairs and lepton pairs. This makes six quarks and six leptons in total. One of the two quarks in each generation has an electric charge of $+\frac{2}{3}e$ and one of $-\frac{1}{3}e$, where e is the elementary charge. Quarks also carry a color charge which is a property resulting from the strong interaction explained below. Of the two leptons in each generation one has a charge of $-e$ and the other lepton is its corresponding chargeless neutrino. Stable matter is composed of quarks and leptons of the first generation. For example, protons and neutrons, which make up the nuclei of atoms, are composed of up- and down-quarks. The fermions of the second and third generation are heavier than the ones of the first generation and therefore are unstable. For each of the 12 fermions there exists an antiparticle with the same mass but opposite electric charge and for quarks also color charge.

Interactions between fermions are mediated by bosons. The three fundamental forces described by the Standard Model are the strong, the weak and the electromagnetic force, each having their own corresponding bosons. Gravity is currently not incorporated into the standard model. The strong force is described by Quantum Chromodynamics (QCD) and is an SU(3) symmetry. The mediator of the strong force is the massless, color-charged gluon. Besides interacting with quarks, gluons can also interact with each other. One effect of QCD is that quarks can not be observed as free particles and only occur as bound states, hadrons. When bound quarks are separated from each other their binding energy increases until it becomes large enough to produce another quark pair from the vacuum. This is called confinement [4].

All electrically charged particles interact with the electromagnetic interaction described by Quantum Electrodynamics (QED), which is a U(1) symmetry. The electromagnetic interaction is mediated by the charge- and massless photons.

The weak interaction, which is the weakest of the three fundamental forces, is described by a SU(2) symmetry and mediated by the massive $W^{\pm}$ and Z^{0} bosons. The high masses of $m_W \approx 80\,\text{GeV}$ and $m_Z \approx 90\,\text{GeV}$ are caused by spontaneous electroweak symmetry breaking, caused by the Higgs mechanism. A result of the Higgs mechanism is the spin-1 Higgs boson.

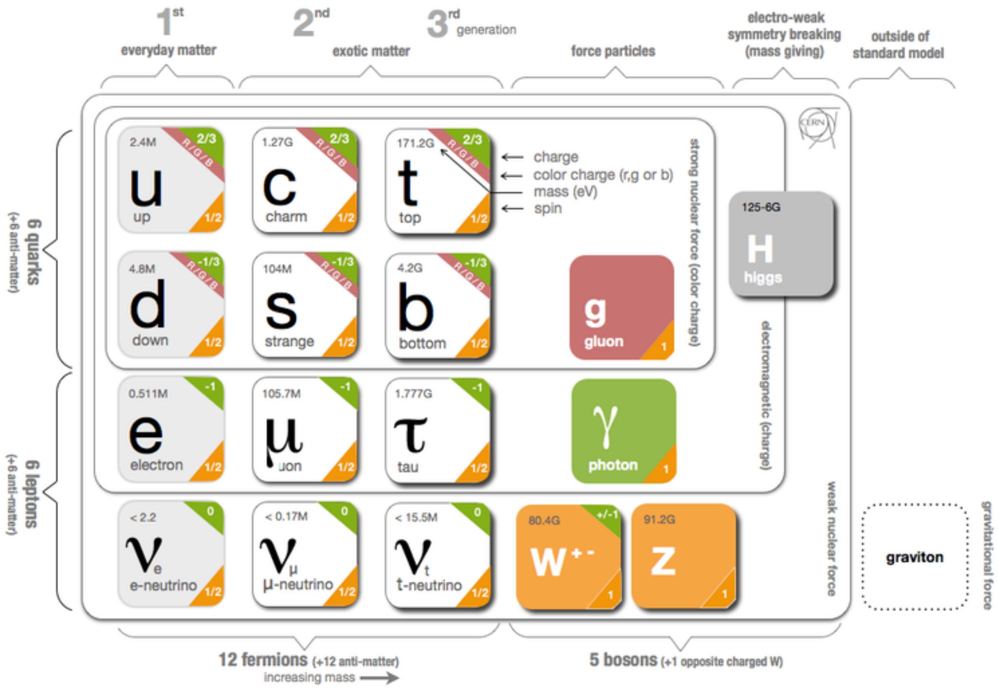


Figure 2.1: Schematic overview of the Standard Model of particle physics [5].

2.2 Weak Interaction

In the Standard Model, the weak interaction is the only interaction that allows quarks to change flavor. This is described by the Cabbibo-Kobayashi-Maskawa (CKM) matrix, which relates the mass eigenstates of quarks to the eigenstates of the weak interaction:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.1}$$

The elements of the CKM matrix and their experimentally determined magnitudes are the following [6]:

$$V_{\text{CKM}} = \begin{pmatrix} V_{\text{ud}} & V_{\text{us}} & V_{\text{ub}} \\ V_{\text{cd}} & V_{\text{cs}} & V_{\text{cb}} \\ V_{\text{td}} & V_{\text{ts}} & V_{\text{tb}} \end{pmatrix} \quad , \quad |V_{\text{CKM}}| = \begin{pmatrix} 0.974 & 0.224 & 0.004 \\ 0.221 & 0.987 & 0.041 \\ 0.008 & 0.039 & 1.013 \end{pmatrix}. \tag{2.2}$$

The probability of a quark of flavor i decaying into a quark of flavor j is proportional to the CKM matrix element $|V_{ij}|$ squared. As can be seen in Equation (2.2), the elements on the diagonal are close to unity. This means, that in case of a flavor change quarks are most likely to transition into the other quark of the same generation. Decays from one generation to a neighboring generation are suppressed, whereas decays between first and third generation involving the elements V_{ub} and V_{td} are doubly suppressed.

2.3 $B\overline{B}$ Production at Belle

At B-factories like Belle, the goal is to produce large numbers of B mesons. At the Belle experiment this was achieved by colliding electrons and positrons at a center-of-mass (CM) energy corresponding to the $\Upsilon(4S)$ resonance with $m_{\Upsilon(4S)} = 10.579\,\text{GeV}$. The $\Upsilon(4S)$ is a bound state of a b quark and its antiparticle $\overline{b}$ and decays via the strong interaction at least 96 % of the time into a pair of B mesons, with $\mathcal{B}(\Upsilon(4S) \to B^+B^-) = (51.4 \pm 0.6)\%$ and $\mathcal{B}(\Upsilon(4S) \to B^0\overline{B}^0) = (48.6 \pm 0.6)\%$ [6]. B mesons have a mass of $m_B = 5.28\,\text{GeV}$ and a lifetime of $\approx 1.6\,\text{ps}$. A Feynman style diagram for the production of the $\Upsilon(4S)$ from electron positron annihilation is shown in Figure 2.2.

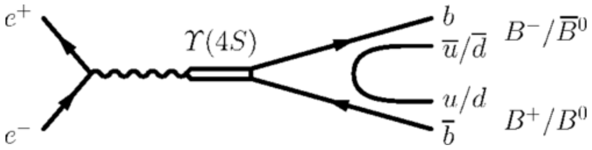


Figure 2.2: Feynman style diagram for the production and decay of the $\Upsilon(4S)$ resonance [7].

Colliding electrons and positrons at a CM energy of the $\Upsilon(4S)$ mass does not always result in the production of an $\Upsilon(4S)$. Among other processes, the non-resonant production of quark pairs $e^+e^- \to q\overline{q}$, $q = u, d, s, c$, and the production of lepton pairs $e^+e^- \to \ell^+\ell^-$ is also abundant. The cross sections of different processes can be found in Table 2.1.

2.4 Semileptonic B Decays

B mesons can only decay via the weak interaction. One category of B decays are semileptonic decays, where a B meson decays into a hadron, a lepton and a neutrino. The process $B^0 \to \pi^-\ell^+\nu_\ell$ investigated in this study is one example for a semileptonic B decay and the corresponding Feynman diagram is shown in Figure 2.3. In this process a $\overline{b}$ quark decays into a $\overline{u}$ quark by emitting a W^+ boson, which decays into a lepton-neutrino pair. The resulting $\overline{u}$ quark together with the remaining d quark forms a π^-. Because this B decay involves the transition of a $\overline{b}$ quark to a $\overline{u}$ quark, the rate of this process is proportional to $|V_{ub}|^2$.

Table 2.1: Cross sections of different physics processes at a CM energy of the $\Upsilon(4S)$ mass [8].

Process	Cross Section in nb
$\Upsilon(4S)$	1.11
$u\bar{u}$	1.61
$d\bar{d}$	0.40
$s\bar{s}$	0.38
$c\bar{c}$	1.30
$e^+e^-(\gamma)$	300
$\mu^+\mu^-(\gamma)$	1.15
$\tau^+\tau^-(\gamma)$	0.92
$\gamma\gamma(\gamma)$	4.99
$e^+e^-e^+e^-(\gamma)$	39.7
$e^+e^-\mu^+\mu^-(\gamma)$	18.9

Besides the weak component this process also involves a strong contribution from quark-gluon interactions and gluon self-interactions, depicted by the curled lines in Figure 2.3. The strong effects are described by a from factor (FF), which is typically parameterized as a function of q^2. q^2 is the four-momentum transfer squared and is defined as

$$q^2 = (p_B - p_\pi)^2 = (p_\ell + p_\nu)^2, \tag{2.3}$$

with p_B, p_π, p_ℓ and p_ν being the four-momenta of the B meson, pion, lepton and neutrino, respectively. The differential rate is then [9]

$$\frac{d\Gamma(B \to \pi\ell\nu)}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{24\pi^3} |p_\pi|^3 |\mathrm{FF}(q^2)|^2, \tag{2.4}$$

where G_F is the Fermi constant and p_π is the pion momentum in the B meson rest frame. By measuring the differential decay rate of $B^0 \to \pi^-\ell^+\nu_\ell$ and applying a form factor model from theoretical predictions, $|V_{ub}|$ can be extracted.

The current world average from the exclusive determination described above is $|V_{ub}| = (3.70 \pm 0.10 \pm 0.12) \times 10^{-3}$, where the first error is experimental, and the second arises from theoretical models. The average obtained from inclusive $B \to X_u\ell\nu$ decays is $|V_{ub}| = (4.25 \pm 0.12^{+0.15}_{-0.14} \pm 0.23) \times 10^{-3}$ [9]. By combining these values and weighting them by their relative uncertainties, one obtains

$$|V_{ub}| = (3.82 \pm 0.24) \times 10^{-3} \quad \text{(average)}. \tag{2.5}$$

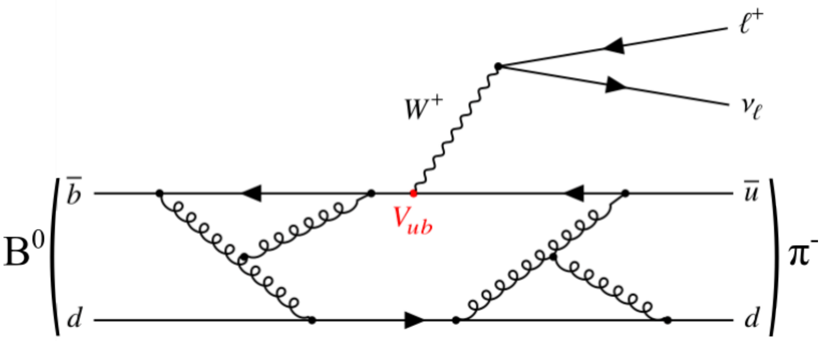


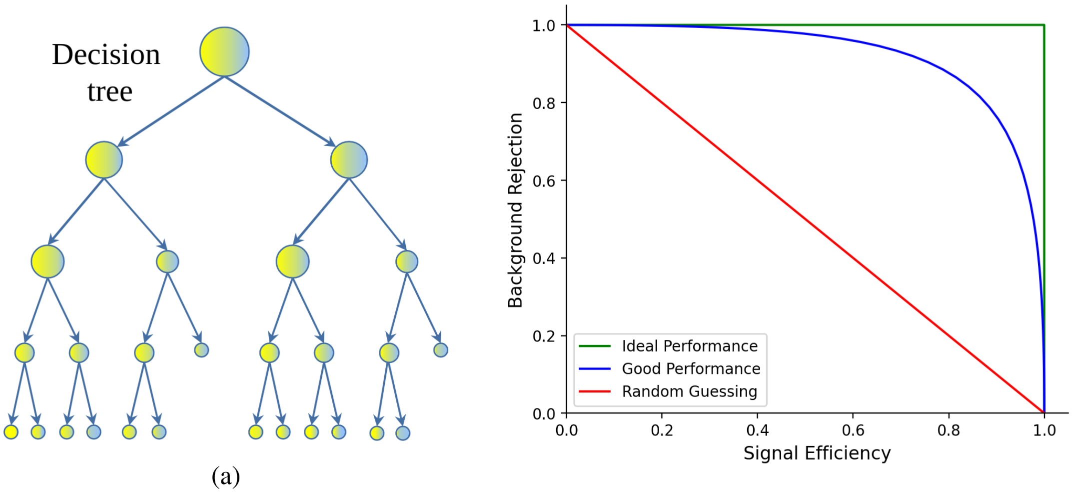
Figure 2.3: Feynman diagram for the decay process $B^0 \to \pi^- \ell^+ \nu_\ell$.

2.5 Multivariate Analysis

A commonly used tool in high energy particle physics is multivariate analysis (MVA). MVA algorithms are used to describe observed data and to deduce properties of the processes that underlie the generation of data. One task of special interest to this analysis is the classification of data into categories like signal and background. A way to achieve this is with machine learning, which provides an effective way to learn a statistical model from a (training) dataset and use this model to infer information for a new independent dataset [10].

One MVA tool, which is based on machine learning, is the boosted decision tree (BDT). A decision tree uses a series of binary selections on input variables with the aim to maximize the signal to background ratio at nodes, as is illustrated in Figure 2.4 (a). In the end, each data-point (event) is assigned a classifier between 0 and 1, which indicates how signal- or background-like an event is. A single decision tree itself is a weak learner because its predictions are often dominated by statistical fluctuations in the training dataset. This is called overtraining and results in a poor performance of the classifier on new data-points. A BDT uses many trees in series while each tree focuses on the errors of the previous tree by assigning a higher weight to wrongly classified events. The final BDT classifier is the weighted sum of the outputs of all the trained decision trees. How well a BDT can discriminate between signal and background events can be evaluated with the Receiver Operating Characteristic (ROC) curve. This curve shows the background rejection for different signal efficiencies achieved by cuts on the output classifier. A measure for the performance of a BDT is the area under the ROC curve. Ideally one could reject all background events at a signal efficiency of 100 %, which would result in an area under the curve of 1. ROC curves for differently performing BDTs are shown in Figure 2.4 (b).

Different hyperparameters which affect the performance of a BDT can be adjusted. The first parameter is the *number of trees* and determines the number of decision trees that are combined to form the BDT. The *depth of trees* is the number of cut levels in each separate tree, while the *number of cuts* is the number of cuts in each tree. The *shrinkage* controls the learning rate of the procedure.



(a)

(b)

Figure 2.4: (a) Schematic overview of a decision tree [11]. (b) Example ROC curves showing the background rejection vs. the signal efficiency for different BDTs, adapted from Ref. [12].

CHAPTER 3

KEKB and the Belle Experiment

This chapter provides an overview over the Belle experiment and the Belle detector. For more detailed information see Ref. [13].

3.1 KEKB

The Belle experiment ran at the KEKB e^+e^- asymmetric energy collider in Tsukuba, Japan between 1999 and 2010. Most of the data was recorded on the $\Upsilon(4S)$ resonance at 10.58 GeV. This CM energy was achieved by colliding 3.5 GeV positrons with 8 GeV electrons. The $\Upsilon(4S)$ resonance decays almost 100 % of the time into a pair of B mesons.

The full Belle data set of $711\,\text{fb}^{-1}$ at the $\Upsilon(4S)$ resonance corresponds to 772×10^6 pairs of B mesons with about half of them being $B^0\overline{B}^0$ pairs.

3.2 The Belle Detector

The Belle detector was located at the interaction region of the KEKB collider. It was optimized to measure time-dependent CP violation in B meson decays. To do so the detector needed good vertex resolution and good particle identification capabilities for leptons and hadrons over a range of momenta. Most of the sub-detectors were located radially inside a 1.5 Tesla superconducting solenoid to bend the trajectories of charged particles. The Belle detector covered a polar angle of $17^\circ < \theta < 150^\circ$. This asymmetric coverage stems from the asymmetric beam energies which leads to a boost of the CM frame with respect to the laboratory frame. The longitudinal cross section of the detector can be seen in Figure 3.1.

The following sections contain an overview over the different detector components and their purposes.

3.2.1 Silicon Vertex Detector

The silicon vertex detector (SVD) was the sub-detector located closest to the e^+e^- interaction point and had the purpose of measuring B decay vertices. After the first 3 years of operation of the KEKB collider the SVD was upgraded. The first system, called SVD1, consisted of 3 layers of double-sided silicon-strip detectors and had a coverage of $23^\circ < \theta < 140^\circ$. In 2003 the system was upgraded to

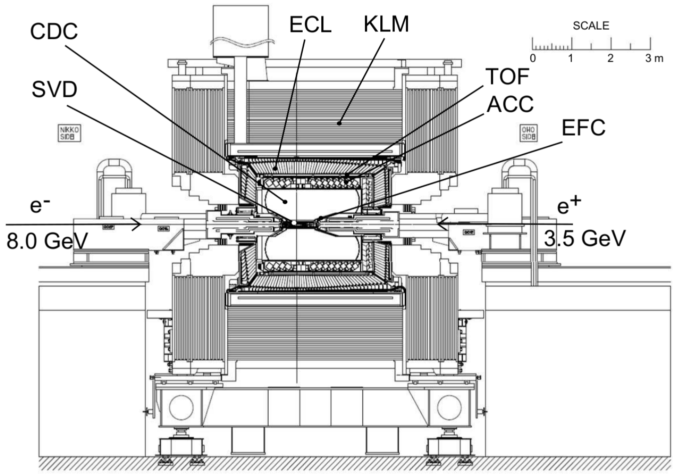


Figure 3.1: Longitudinal cross section of the Belle detector [13].

SVD2. Consisting of 4 layers now covering the full angular acceptance of $17^\circ < \theta < 150^\circ$, SVD2 was also developed to be more radiation hard than SVD1. 85 % of Belle data were taken with SVD2.

3.2.2 Central Drift Chamber

The central drift chamber (CDC) was located radially outside of the SVD. The CDC was a wire chamber consisting of 8400 drift cells filled with a mixture of He and C_2H_6 and wires up to 2 400 mm long. Charged particles that traveled through the CDC ionized gas molecules which then accelerated towards the wires and produced an electrical signal. The hit coordinates were used to reconstruct tracks of charged particles. The magnetic field in the detector bent the trajectories of charged particles which enabled momentum measurements. The CDC also had the purpose of measuring track ionization losses ($\mathrm{d}E/\mathrm{d}x$) used for particle identification.

3.2.3 Particle Identification System

Particle Identification (PID) was accomplished with information from the aerogel Cherenkov counter (ACC) and time-of-flight (TOF) system, combined with $\mathrm{d}E/\mathrm{d}x$ measurements from the CDC. The ACC consisted of modules of aerogel blocks in the barrel part outside the CDC and the forward end-cap of the detector with a total acceptance of $13.6^\circ < \theta < 127.9^\circ$. Depending on their velocity, some traversing charged particles would produce Cherenkov light which was detected by photomultiplier tubes attached to the aerogel blocks.

The TOF system consisted of a barrel of plastic scintillator counters and thin trigger scintillation counters (TSC) covering a polar angle of $33^\circ < \theta < 121^\circ$. Charged particles that crossed through the scintillators produced scintillation light which was detected by photomultiplier tubes. The time difference between a hit in the TSC and the corresponding TOF counter was used to determine the velocity of a charged particle. A time resolution on the order of 100 ps allowed to distinguish between kaons and pions with momenta below 1.2 GeV [14]. Details about the methods and performance of the PID system will follow in Section 3.3.

3.2.4 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECL) consisted of CsI(Tl) crystals located in the barrel region and both end-caps. The crystals had a tower-like shape and pointed towards the interaction point. Attached to the rear-end of each crystal was a photodiode to detect light from electromagnetic showers. The primary purpose of the ECL was to measure the energy of photons, but charged particles could also deposit energy in the ECL. The ratio of the shower energy to the track momentum E/p provided the main parameter for electron/hadron separation. In addition, the ECL was used as a trigger source and as a luminosity monitoring system.

3.2.5 K_L and Muon Detector

The outermost sub-detector was the K_L and muon detector (KLM) and had the purpose to detect high-momentum K_L and muons. The KLM consisted of alternating layers of glass-electrode resistive plate counters as charged particle detectors (RPC) and 4.7 cm-thick iron plates. There were 15 detector layers in the barrel region and 14 layers in each end-cap covering a polar angular range of $20^\circ < \theta < 155^\circ$ in total. Only a few long-living particles, like K_L and muons, could reach the KLM. K_L were identified from clusters in the KLM that did not match with any charged track extrapolated from the CDC. Above 1.5 GeV the muon identification efficiency was more than 90 % at a fake rate of less than 5 %. Muons with a momentum below 500 MeV did not reach the KLM [15].

3.3 Particle Identification

The identification of charged particles stable enough to be detected plays an important role in the physics of the Belle experiment. Good PID performance is needed to reduce backgrounds in form of final states different from that under study by misidentifying particles. PID is also required for separating hadronic final states of B decays used for hadronic tagging and is crucial for flavor tagging of B mesons, which means distinguishing between B^0 and $\overline{B}^0$ candidates.

PID at Belle is based on likelihood ratios. For the identification of hadrons, likelihoods for a candidate α ($\alpha = \pi, K, p$) are calculated based on dE/dx measurements of a charged track by the CDC ($\mathcal{L}_\alpha^{\mathrm{CDC}}$), time of flight measurements from the TOF ($\mathcal{L}_\alpha^{\mathrm{TOF}}$) and the number of photons from the ACC ($\mathcal{L}_\alpha^{\mathrm{ACC}}$). Then the likelihood ratio

$$\mathcal{L}_{\alpha|\beta} = \frac{\mathcal{L}_\alpha^{\mathrm{CDC}}\mathcal{L}_\alpha^{\mathrm{TOF}}\mathcal{L}_\alpha^{\mathrm{ACC}}}{\mathcal{L}_\alpha^{\mathrm{CDC}}\mathcal{L}_\alpha^{\mathrm{TOF}}\mathcal{L}_\alpha^{\mathrm{ACC}} + \mathcal{L}_\beta^{\mathrm{CDC}}\mathcal{L}_\beta^{\mathrm{TOF}}\mathcal{L}_\beta^{\mathrm{ACC}}} \tag{3.1}$$

is calculated as a measure for how likely the candidate α is to actually be a particle of type β. For example, pions can be selected by requiring a high $\mathcal{L}_{\pi|K}$.

For electron identification, information from the ECL, including E/p, the transverse shower shape and matching of the track position and energy cluster, combined with $\mathcal{L}_\alpha^{\mathrm{CDC}}$ and $\mathcal{L}_\alpha^{\mathrm{ACC}}$ is used to form a likelihood ratio which is called `eID`. To identify muons, reconstructed hits in the KLM were compared to the extrapolation of charged tracks in the CDC. The difference ΔR between the measured and expected range of a track, and the statistic χ_r^2, which is constructed from transverse deviations of hits associated with a track were used to form likelihoods for the muon, pion and kaon hypotheses. The likelihood ratio $\mathcal{L}_\mu/(\mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K)$ is used as a discriminating variable, called `muID`.

CHAPTER 4

Reconstruction

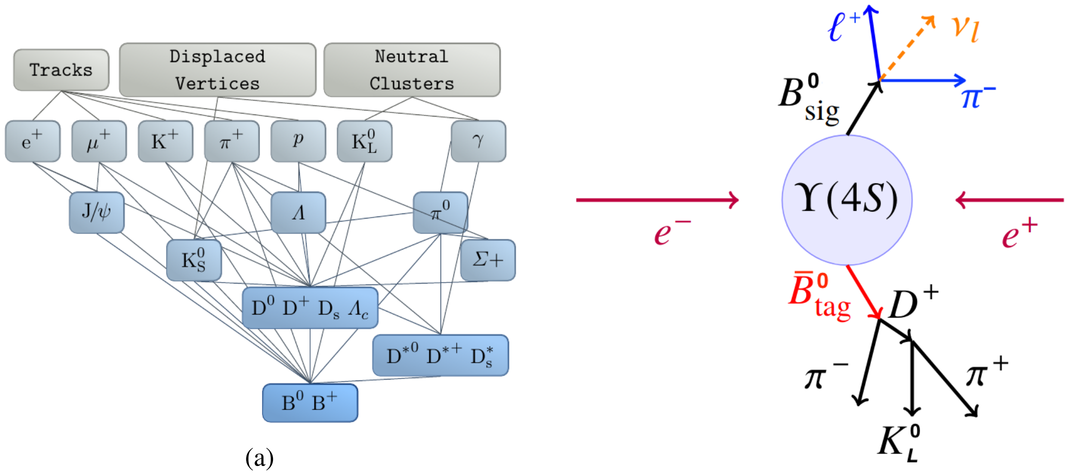
This chapter contains an overview of the steps taken in order to reconstruct the signal decay $B^0 \to \pi^- \ell^+ \nu_\ell$. I start by introducing the procedure of hadronic tagging, followed by an overview of Monte Carlo simulations and different corrections that have to be applied. Then, I explain the selections used to reconstruct the signal decay and introduce the missing mass squared.

4.1 Hadronic Tagging

The main goal of the Belle detector was to detect, reconstruct and identify different decays of B mesons. The detector can only detect stable particles and particles that live long enough to interact with the different sub-detectors. These particles are: $e^\pm, \mu^\pm, \pi^\pm, K^\pm, p, \overline{p}, \gamma$ and K_L^0 and are referred to as final state particles (FSPs). B mesons have a lifetime on the order of 10^{-12} s and therefore decay close to the interaction point inside the beam pipe. In order to study B mesons and their properties, they must first be reconstructed from their final state particles. Hadronic tagging aims at identifying all of the FSPs that originate from the same B meson and summing their four-momenta to measure the four-momentum of the original B meson.

In hadronic tagging the full decay chain of one B meson, called B_{tag}, is reconstructed from the bottom up. Clusters that are not associated with any tracks are used to construct photon candidates, and tracks to construct charged FSP candidates. As the next step, intermediate particles are reconstructed. For example, π^0 candidates are formed from two photon candidates. K^-, π^+ and π^0 candidates are combined to form D^0 candidates. This goes on until finally a B^0 meson is reconstructed by combining for example D^{*-} and π^+ candidates. The four-momentum of each reconstructed particle is the sum of the four-momenta of its decay products. An illustration of this hierarchical reconstruction structure is shown in Figure 4.1 (a).

Hadronic tagging is especially useful for studies of semileptonic B decays. Since these decays involve neutrinos which do not interact with the detector and therefore can not be reconstructed, the signal-side B meson (B_{sig}) and its four-momentum can not be fully reconstructed from its decay products. In the rest-frame of the $\Upsilon(4S)$, i.e. the CM frame, the two B mesons are produced back-to-back and no other particles are created at the initial e^+e^- collision. The precise knowledge of the initial four-momentum of the $\Upsilon(4S)$ therefore allows the recovery of the four-momentum of B_{sig} by reconstructing the four-momentum of B_{tag}. Measuring the four-momenta of the pion and lepton from the signal decay $B^0 \to \pi^- \ell^+ \nu_\ell$ then allows the inference of the four-momentum of the undetected neutrino. An illustration of hadronic tagging for the signal decay can be seen in Figure 4.1 (b).



(a)

(b)

Figure 4.1: Diagrams illustrating the hierarchical reconstruction structure used by the Full Event Interpretation tagging algorithm (a) and the use of tag-side reconstruction to search for $B^0 \to \pi^- \ell^+ \nu_\ell$ (b). The B meson on the tag-side decays hadronically, allowing the recovery of information of the neutrino on the signal-side. Both images are adapted from Ref. [16].

4.1.1 Full Reconstruction

The hierarchical reconstruction procedure described above was first implemented in form of the Full Reconstruction (FR) algorithm [17] for the Belle experiment. The FR used the multivariate analysis package NeuroBayes and reconstructed B candidates in four distinct stages, which are shown in Table 4.1. In total the FR reconstructed 1104 exclusive decay chains resulting in tag-side efficiencies of 0.18 % for B^0 and 0.28 % for $B^\pm$, where the efficiency is defined as the number of correctly reconstructed B mesons divided by the total number of produced $B\overline{B}$ pairs.

4.1.2 Full Event Interpretation

The Full Event Interpretation (FEI) [3] is a new tagging algorithm based on the FR, that was developed for the Belle II experiment. The FEI also follows a hierarchical approach, now with six instead of four stages as shown in Table 4.1. The FEI reconstructs a larger number of explicit B meson and intermediate particle decay channels, resulting in the reconstruction of $O(10000)$ distinct decay chains. Also the candidate selection criteria were improved, in particular a best-candidate selection is performed during the reconstruction of intermediate particles instead of fixed cuts. The FEI achieved a maximum tag-side efficiency of 0.46 % for B^0 and 0.76 % for $B^\pm$ on simulated Belle events. For each reconstructed B_{tag} candidate the FEI calculates a signal probability, which is an estimate for how likely this B_{tag} candidate is to be reconstructed correctly. This FEI signal probability is built from the signal probabilities assigned to all particles in the reconstructed decay chain. Although it is called “probability”, the FEI signal probability is not the actual probability of a B_{tag} candidate being correct.

Table 4.1: Stages of the reconstruction of B candidates used in the FR and FEI.

	Particles	
Stage	**FR**	**FEI**
1	$e^\pm$, $\mu^\pm$, $K^\pm$, $\pi^\pm$, K_S^0, γ, π^0	$e^\pm$, $\mu^\pm$, $K^\pm$, $\pi^\pm$, K_L^0, γ
2	$D^\pm$, $D_s^\pm$, D^0, J/Ψ	J/Ψ, π^0, Λ
3	$D^{*\pm}$, $D_s^{*\pm}$, D^{*0}	K_S^0, Σ^+
4	$B^\pm$, B^0	$D^\pm$, $D_s^\pm$, D^0, Λ_c
5		$D^{*\pm}$, $D_s^{*\pm}$, D^{*0}
6		$B^\pm$, B^0

4.2 Monte Carlo Simulation

In this study I work with Monte Carlo (MC) simulated data. MC simulations are commonly used in high energy particle physics. The reason for this will become apparent in the following lines. In order to study a specific decay process (e.g. of a B meson), one looks for events in recorded data where the same final state particles were reconstructed as expected for the decay of interest. Not all events that match this criterion contain the signal decay. Events where this is not the case are called background events. There can be other processes involving different intermediate particles that result in the same final state. Also the identity hypothesis of a reconstructed FSP can be wrong. In order to reduce backgrounds I need to understand its origin, which can be accomplished by using simulated data.

The MC generator `EvtGen` [18] simulates e^+e^- collisions and the decays of resulting $B\overline{B}$ pairs, as well as background processes like $e^+e^- \to \ell^+\ell^-, q\overline{q}$. EvtGen returns the four-momentum vectors of the resulting FSPs, which are then passed to the detector simulation `GEANT3` [19]. The interactions of FSPs with the detector are simulated, such as the bending of charged tracks due to the magnetic field, scattering or further decays. In a next step the response of the detector is simulated: energy depositions from FSPs and their locations. These data are mixed with simulated beam-induced backgrounds to make them more realistic and are then passed to reconstruction algorithms. This way the simulated data look like recorded data, with one difference being that for simulated data the *truth*-information of the reconstructed FSPs including information on their ancestors and true particle ID is available.

Belle MC is produced separately for different physics processes in so-called streams. One stream of a certain MC type contains the number of events of this process expected in the full Belle data set of 711 fb^{-1}. The amount of MC data and the contents of the different MC types are shown in Table 4.2. For every MC type I use more than one stream in order to increase statistics.

Simulated MC data do not represent recorded data perfectly. Different efficiency corrections have to be applied to MC in order to improve the agreement between data and MC. These corrections are explained below. In principle every correction results in a weight for MC events. The total weight of a MC event is then the product of all weights that were assigned to this event by the different corrections.

Table 4.2: Different types of Belle MC and the number of streams used in this analysis.

MC type	**physics process**	**number of streams used**
charged	$e^+e^- \to \Upsilon(4S) \to B^+B^-$ both B mesons decay generically	2
mixed	$e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ both B mesons decay generically	2
uds	$e^+e^- \to q\overline{q}, q = u, d, s$	2
charm	$e^+e^- \to c\overline{c}$	2
charged $X_u\ell^+\nu_\ell$	$e^+e^- \to \Upsilon(4S) \to B^+B^-$ $B^+ \to X_u\ell^+\nu_\ell$ $B^- \to$ generic	10
mixed $X_u\ell^+\nu_\ell$	$e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$ $B^0 \to X_u\ell^+\nu_\ell$ $\overline{B}^0 \to$ generic	10

4.2.1 Hybrid Model

The model described in the following lines was originally proposed by Ref. [20] and my implementation follows Ref. [21]. Charmless semileptonic decays are produced as a mixture of specific exclusive modes and non-resonant contributions. Due to the discrepancy between the inclusive and exclusive measurements of $|V_{\mathrm{ub}}|$ the non-resonant part typically gets overestimated. The exclusive modes are $B \to \pi\ell^+\nu_\ell$, $B \to \rho\ell^+\nu_\ell$, $B \to \omega\ell^+\nu_\ell$ and $B \to \eta^{(\prime)}\ell^+\nu_\ell$. The non-resonant $B \to X_u\ell^+\nu_\ell$ decays are produced with at least two pions in the final state using the De Fazio-Neubert (DFN) model [22]. The triple differential rate $\Delta\mathcal{B}$ in this model is a function of the four-momentum-transfer squared (q^2), the lepton energy in the B rest-frame (E_ℓ^B) and the hadronic invariant mass squared (m_X^2). The inclusive and exclusive predictions for $B \to X_u\ell^+\nu_\ell$ are combined using the so-called "hybrid" approach. Both predictions get combined such that the partial branching fractions in the triple differential rate of the inclusive prediction $\Delta\mathcal{B}_{ijk}^{\mathrm{incl}}$ and the combined exclusive predictions $\Delta\mathcal{B}_{ijk}^{\mathrm{excl}}$ reproduce the inclusive values. To do so, weights w_{ijk} are assigned to inclusive contributions, such that

$$\Delta\mathcal{B}_{ijk}^{\mathrm{incl}} = \Delta\mathcal{B}_{ijk}^{\mathrm{excl}} + w_{ijk} \times \Delta\mathcal{B}_{ijk}^{\mathrm{incl}} \tag{4.1}$$

where i, j, k denote the bin in the three dimensions of q^2, E_ℓ^B and m_X:

$$q^2 = [0, 2.5, 5, 7.5, 10, 12.5, 15, 20, 25]\ \mathrm{GeV}^2$$
$$E_\ell^B = [0, 0.5, 1, 1.25, 1.5, 1.75, 2, 2.25, 3]\ \mathrm{GeV}$$
$$m_X = [0, 1.4, 1.6, 1.8, 2, 2.5, 3, 3.5]\ \mathrm{GeV}$$

The effect of the hybrid model re-weighting on the distributions of the three variables can be seen in Figure 4.2 for neutral B mesons. On the left side the initial generator level distributions of q^2, E_ℓ^B and m_X for resonant and non-resonant contributions are shown. In m_X the two peaks of the resonant part correspond to $\pi^\pm$ at 139.6 MeV and $\rho^\pm$ at 775.1 MeV. On the right side the same distributions after applying hybrid weights to the non-resonant contribution are shown. For E_ℓ^B and m_X we can see that the hybrid model, which is the sum of resonant and non-resonant contributions, has the same shape and height as the non-resonant part before the re-weighting. The largest difference can be seen in m_X, where the non-resonant part only gets scaled down for masses below 1.5 GeV. The same figures for charged B mesons can be found in the Appendix in Figure A.1.

4.2.2 Lepton ID Corrections

Although MC simulations are generated in a way to look as similar to recorded data as possible, particle identification performs slightly different on MC and data. This results in different efficiencies for MC and data and is accounted for with correction factors. These correction factors are calculated in bins of lepton momentum $|\vec{p}_\ell|$ and polar angle θ in the laboratory frame. The correction factors used in this analysis are provided by the Belle collaboration and were obtained by comparing data and MC efficiencies for two-photon processes $e^+e^- \longrightarrow e^+e^-\ell^+\ell^-$ ($\ell = e, \mu$) [23].

4.2.3 Branching Fraction Corrections

The branching ratios of different B decays that are used in the production of MC are stored in a *decay file*. For some decays the current world average of the corresponding branching ratio has changed since production of the MC. This is accounted for by defining a weight w_{BF}, which is calculated as

$$w_{\mathrm{BF}} = \frac{\mathcal{B}^{\mathrm{PDG}}}{\mathcal{B}^{\mathrm{DEC}}} \tag{4.2}$$

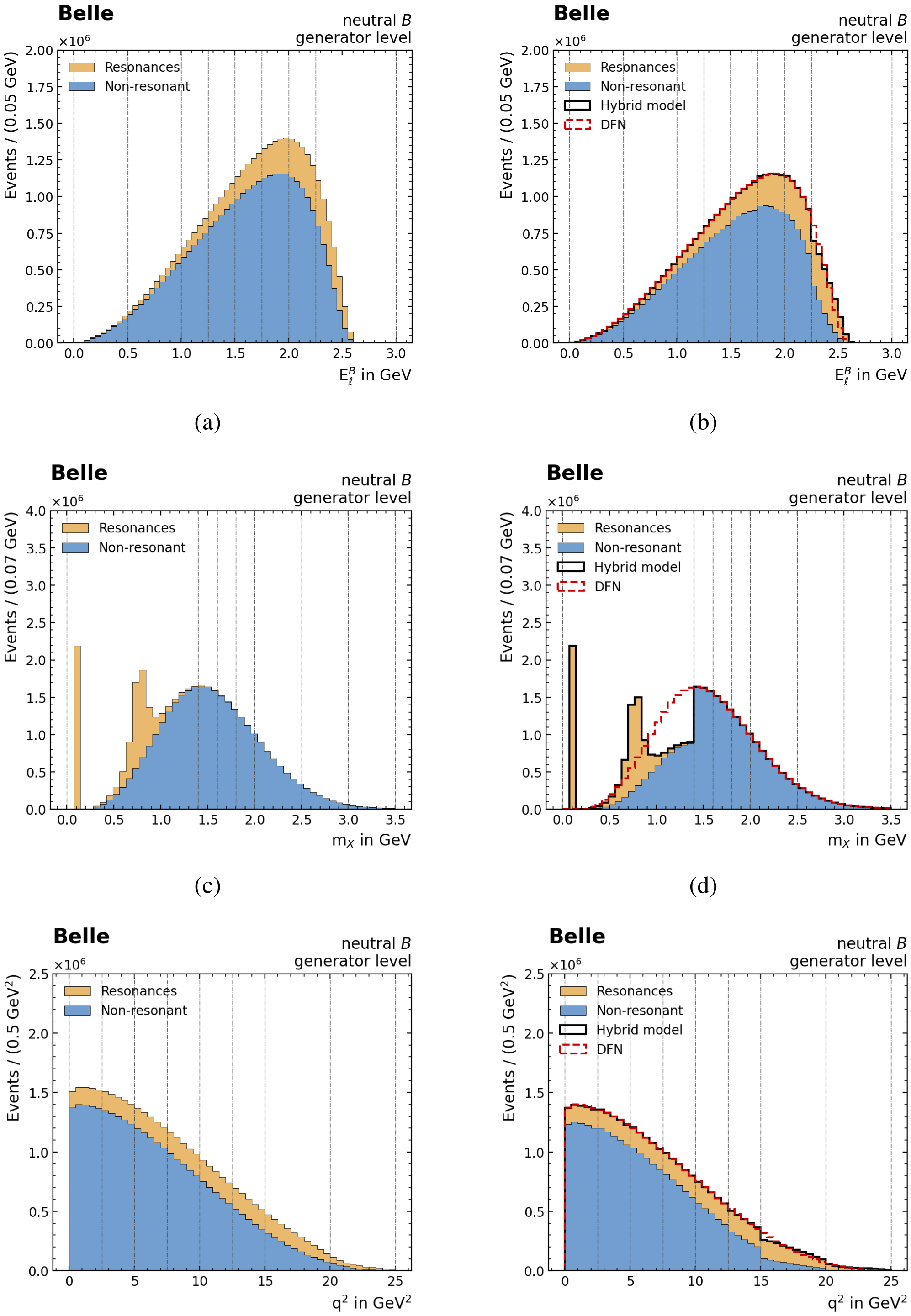
where $\mathcal{B}^{\mathrm{PDG}}$ is the current world average of the branching ratio from [6] and $\mathcal{B}^{\mathrm{DEC}}$ is the branching ratio from the decay file. In my analysis I correct the branching fractions of the following B decays:

$$\begin{aligned} B^0 &\longrightarrow D^-\ell^+\nu_\ell \\ B^0 &\longrightarrow D^{*-}\ell^+\nu_\ell \\ B^+ &\longrightarrow \overline{D}^0\ell^+\nu_\ell \\ B^+ &\longrightarrow \overline{D}^{*0}\ell^+\nu_\ell \end{aligned}$$

The exclusive charmless semileptonic decays $B \longrightarrow X_u\ell^+\nu_\ell$ were produced with the correct branching ratios and therefore did not have to be corrected.

4.2.4 FEI Efficiency Correction

The tagging efficiency is different on MC and data because of mismodelled contributions in the MC. There can be contributions in data that have not yet been seen or understood and are therefore not correctly simulated. Also the response of the detector may not be accurately simulated. Correction



(a) (b)

(c) (d)

(e) (f)

Figure 4.2: Generator level $B \rightarrow X_u \ell^+ \nu_\ell$ distributions of E_ℓ^B, m_X and q^2 before (left) and after (right) re-weighting for neutral B mesons. The black line on the right shows the hybrid model, composed of resonant and non-resonant contributions.

factors have to be determined to compensate for different tagging efficiencies on data and MC. In general, these correction factors depend on the precise selections on B_{tag} candidates, but also on the B decay that is reconstructed on the signal-side. Therefore a FEI efficiency calibration specific to the selections used in this analysis is necessary for absolute measurements. FEI correction factors were calculated for different charmed B decays on the signal-side accompanied by different B_{tag} decay chains reconstructed on the tag-side [24]. In doing so, average calibration factors of

$$\begin{aligned} c_{\text{charged}} &= 0.810 \\ c_{\text{neutral}} &= 0.853 \end{aligned} \tag{4.3}$$

were obtained for charged and neutral B_{tag} for Belle MC. In this analysis only neutral B_{tag} candidates are of relevance, therefore every MC event is assigned an additional weight of 0.853.

4.3 Event Reconstruction

During the reconstruction of the signal decay I specify a set of preselections for the whole event and both the tag- and signal-side. These preselections have to be met by an event in order to pass the reconstruction. The preselections are loose cuts on certain characteristic variables that remove a large portion of backgrounds to reduce the amount of data that is left after the reconstruction, while retaining a high signal efficiency. In the following two sections I present my preselections for the tag- and the signal-side. In addition, I require each event to not contain any additional charged tracks that are not associated with either the tag- or the signal-side B candidate. This is called the *completeness constraint*.

4.3.1 Tag-Side Reconstruction

The preselections for the tag-side can be seen in Table 4.3.

Table 4.3: Preselections for B_{tag} candidates.

Variable	Selection
FEI signal probality	$> 10^{-10}$
M_{bc}	$> 5.26\,\text{GeV}$
$\lvert\Delta E\rvert$	$< 0.2\,\text{GeV}$

Here M_{bc} is the beam-energy constrained mass and is defined as

$$M_{\text{bc}} = \sqrt{E_{\text{beam}}^{*2} - \left(\vec{p}^{\,*}_{B_{\text{tag}}}\right)^2} \tag{4.4}$$

where E^*_{beam} is the beam energy in the CM frame, which is equal to half of the CM energy, and $\vec{p}^{\,*}_{B_{\text{tag}}}$ is the reconstructed momentum of B_{tag}, also in the CM frame. For a perfectly reconstructed B_{tag} its CM energy is equal to E_{beam}. Therefore, M_{bc} is similar to the invariant mass of the B_{tag} meson.

ΔE is the difference between the beam energy and the reconstructed energy of B_{tag}, both in the CM frame:

$$\Delta E = E^*_{\text{beam}} - E^*_{B_{\text{tag}}} \tag{4.5}$$

In the CM frame both B mesons have the same energy and since the $\Upsilon(4S)$ decays into two B mesons and no other particles their energy equals the beam energy E^*_{beam}. Therefore, if the energies of all FSPs were measured perfectly and B_{tag} was correctly reconstructed ΔE is 0. ΔE is sensitive to the reconstructed energy while M_{bc} is sensitive to the reconstructed momentum of B_{tag}.

There can be multiple B_{tag} candidates in each event, therefore, I perform a best candidate selection by only accepting the B_{tag} candidate with the highest FEI signal probability.

4.3.2 Signal-Side Reconstruction

The preselections for the signal-side can be seen in Table 4.4. The PID variables eID, muID and $\mathcal{L}_{\pi|K}$ were introduced in Section 3.3. The lepton and pion candidate are required to originate from the region near the interaction point of the electron and positron beams. In a cylindrical system with origin at the interaction point, axis of symmetry z along the electron beam direction and radial coordinate r, r_{PCA} and z_{PCA} are the coordinates of the point of closest approach (PCA) of a reconstructed charged particle to the z axis. $\vec{p}_{\text{lab}}$ is the measured lepton momentum in the laboratory frame. A vertex fit is performed for signal-side B candidates using `TreeFitter` [25]. Only B_{sig} candidates that survive the vertex fit are considered.

Table 4.4: Preselections for B_{sig} candidates.

Variable	$e^\pm$	$\mu^\pm$	$\pi^\pm$
eID	> 0.5	-	< 0.98
muID	-	> 0.8	< 0.98
$\mathcal{L}_{\pi\|K}$	-	-	> 0.5
r_{PCA} in cm	< 2	< 2	< 2
$\lvert z_{\text{PCA}}\rvert$ in cm	< 4	< 4	< 4
$\lvert\vec{p}_{\text{lab}}\rvert$ in GeV	> 0.3	> 0.6	-
$-3 < \cos\theta_{\text{BY}} < 3$			

In the electron case, for all photons that are not associated with the tag-side and are found within a 3° cone around the electron direction, the photon four-momentum is added to that of the electron and the photon is excluded from further analysis.

By combining the four-momentum vectors of the pion and the lepton candidates on the signal-side we can construct a pseudoparticle Y with $p_Y = p_\pi + p_\ell$. The cosine of the angle between the direction of Y and B_{sig} in the CM frame is then

$$\cos\theta_{\text{BY}} = \frac{2E_{B_{\text{sig}}}E_Y - m_B^2 - m_Y^2}{2|\vec{p}_{B_{\text{sig}}}||\vec{p}_Y|} \tag{4.6}$$

where m_Y is the invariant mass of the pseudoparticle Y. For signal decays this cosine can only lie in the interval $\left|\cos\theta_{\mathrm{BY}}\right| < 1$, although a small fraction of signal events is reconstructed outside this interval because of finite detector resolution. Background events are not restricted to this range and populate a much wider region. I use the loose selection $\left|\cos\theta_{\mathrm{BY}}\right| < 3$ to retain all correctly reconstructed signal events and reject a majority of background events while providing enough background events to fix background shapes during the signal extraction.

4.4 Reconstruction Efficiency

The signal will be extracted in bins of reconstructed q^2 in order to measure $|V_{\mathrm{ub}}|$. From the reconstructed B_{tag} energy $E^*_{B_{\mathrm{tag}}}$ and momentum $\vec{p}^{\,*}_{B_{\mathrm{tag}}}$, and the reconstructed energy E^*_π and momentum $\vec{p}^{\,*}_\pi$ of the signal-side pion, all in the CM frame, q^2 is calculated as

$$q^2 = \left(m_{\Upsilon(4S)} - E^*_{B_{\mathrm{tag}}} - E^*_\pi\right)^2 - \left|-\vec{p}^{\,*}_{B_{\mathrm{tag}}} - \vec{p}^{\,*}_\pi\right|^2 . \tag{4.7}$$

In the previous analysis by A. Sibidanov et al. [2] the reconstructed q^2 spectrum was divided into 13 bins between $q^2 = 0\,\mathrm{GeV}^2$ and $q^2 = 26\,\mathrm{GeV}^2$, each bin having a width of 2 GeV, as shown in Table 4.5. We can define a reconstruction efficiency in a reconstructed q^2 bin i as

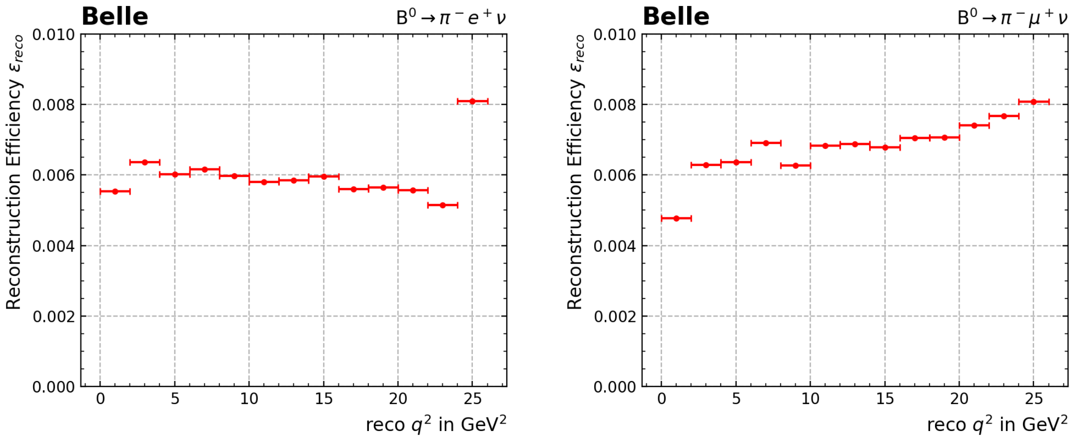
$$\epsilon^i_{\mathrm{reco}} = \frac{N^i_{\mathrm{reco}}}{N^i_{\mathrm{gen}}}, \tag{4.8}$$

where N^i_{reco} is the number of correctly reconstructed signal events in the reconstructed q^2 bin i, and N^i_{gen} is the number of originally generated signal events in the true q^2 bin i, which has the same bin borders as the reconstructed q^2 bin i. The reconstruction efficiencies for the different q^2 bins can be seen in Table 4.5 and Figure 4.3. The reconstruction efficiencies in the electron channel are around 0.6 % for all q^2 bins and between 0.6 % and 0.7 % in the muon channel.

The distributions of the reconstructed q^2 and true q^2 for correctly reconstructed signal events are shown in Figure 4.4 (a) and (b). We can see that the reconstructed q^2 slightly deviates from the true q^2 at the lower and upper ends of the spectrum in both the electron and muon channel. In Figure 4.4 (c) and (d) the reconstructed q^2 is plotted versus the true q^2. When the reconstructed q^2 is not equal to the true q^2, which generally is the case, an event can fall into different q^2 bins of reconstructed and true q^2. For example, an event with a true q^2 within the second bin could have a reconstructed q^2 that falls into the third bin. This is called bin migration.

Table 4.5: Reconstruction efficiencies in the different q^2 bins for $B^0 \to \pi^- e^+ \nu_e$ and $B^0 \to \pi^- \mu^+ \nu_\mu$.

	Reconstruction Efficiency in %												
$B^0 \to \pi^- e^+ \nu_e$	0.55	0.64	0.60	0.62	0.60	0.58	0.58	0.60	0.56	0.57	0.56	0.51	0.81
$B^0 \to \pi^- \mu^+ \nu_\mu$	0.48	0.63	0.64	0.69	0.63	0.68	0.69	0.68	0.70	0.71	0.74	0.77	0.81
Δq^2 in GeV^2	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24	24-26



(a)

(b)

Figure 4.3: Reconstruction efficiencies in bins of q^2 for (a) electron and (b) muon channel.

4.5 Missing Mass Squared

The four-momentum of the signal-side B meson in the $\Upsilon(4S)$ rest-frame is derived only from the reconstructed direction of the tag-side B, by requiring the invariant mass to be the B meson mass [2] and the energy to be half of the CM collision energy:

$$p_{B_{\text{sig}}} = (E_{B_{\text{sig}}}, \vec{p}_{B_{\text{sig}}}) = \left(\frac{m_{\Upsilon(4S)}}{2}, -\frac{\vec{p}_{B_{\text{tag}}}}{|\vec{p}_{B_{\text{tag}}}|} \sqrt{\left(\frac{m_{\Upsilon(4S)}}{2}\right)^2 - m_B^2} \right) \tag{4.9}$$

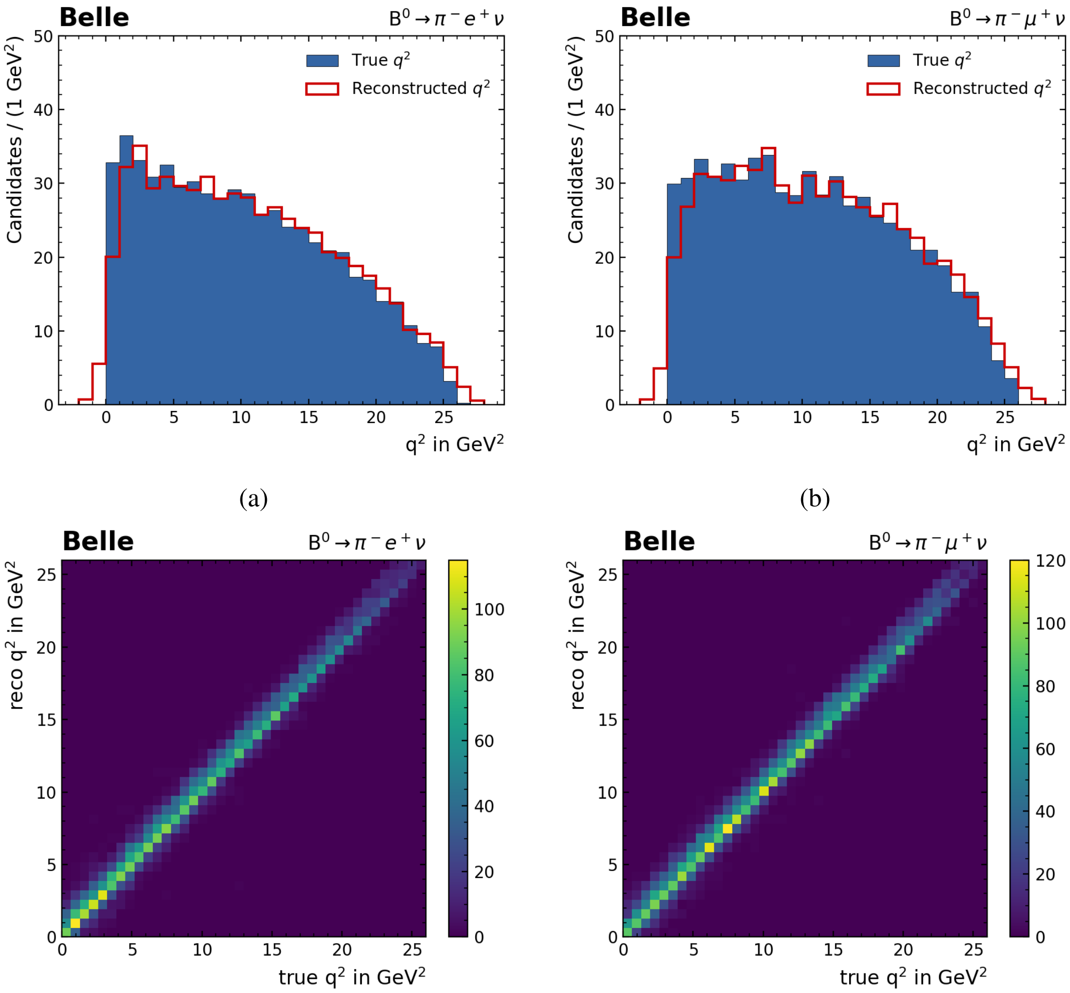
From the reconstructed four-momentum of the signal-side B candidate the missing four-momentum is defined as:

$$p_{\text{miss}} = p_{B_{\text{sig}}} - p_\pi - p_\ell \tag{4.10}$$

As illustrated in Figure 4.1 (b), for a correctly reconstructed tag-side B and a correctly reconstructed semileptonic decay on the signal-side, p_{miss} corresponds to the four-momentum of a single neutrino. The missing mass squared $m^2_{\text{miss}} = p^2_{\text{miss}}$ is in this case consistent with 0. Eventually the signal will be extracted from a fit to m^2_{miss}.

The distribution of m^2_{miss} after reconstruction in the electron and muon channel can be seen in Figure 4.5. We can see that for correctly reconstructed signal events m^2_{miss} peaks at zero as expected. For backgrounds originating from $B\overline{B}$ events m^2_{miss} peaks at larger values because in this case there are particles that were not reconstructed in the event. These missing particles contribute to larger values of m^2_{miss}.

From this point onward I define the signal region as the region within $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$. This region contains 93.4 % of the correctly reconstructed signal events in the electron channel and 95.7 % of the correctly reconstructed signal events in the muon channel. The aim of my further selections is to significantly reduce the amount of background in this signal region while retaining a large amount of the signal.



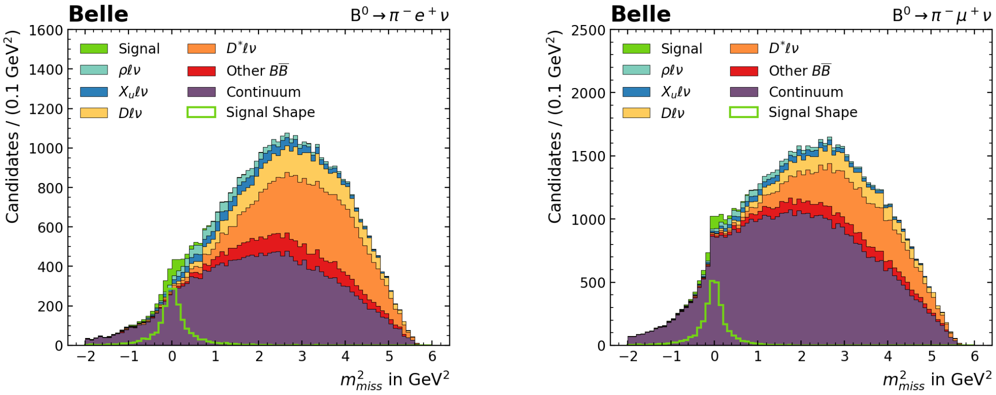
(a) (b)

(c) (d)

Figure 4.4: (a) and (b) show the distributions of reconstructed and true q^2 for signal events in the electron and muon channel. (c) and (d) show the reconstructed versus true q^2 for signal events for both channels.

4.6 Background Classification

In the distribution shown in Figure 4.5 “Signal” denotes correctly reconstructed signal events. This means the decay $B^0 \longrightarrow \pi^- \ell^+ \nu_\ell$ was simulated and both the lepton and pion were correctly identified and reconstructed to be daughters of the signal-side B. If in an event B_{sig} was reconstructed correctly this does not imply that B_{tag} was also reconstructed correctly. Since m^2_{miss} relies on the reconstructed direction of B_{tag}, signal events can have values not equal to 0 if the B_{tag} was reconstructed incorrectly. The categorization of other events will be explained below.



(a)

(b)

Figure 4.5: Distributions of m^2_{miss} for $B^0 \to \pi^- e^+ \nu_e$ and $B^0 \to \pi^- \mu^+ \nu_\mu$ after the reconstruction.

4.6.1 $B\bar{B}$ Background

When classifying $B\bar{B}$ background events, the first step is to check whether the signal lepton candidate is a true lepton. If so, I check if the mother of the lepton is a B meson. Leptons that directly originate from a B decay are called primary leptons, and secondary leptons if they do not. For primary leptons I then check the identity of the first sister particle and assign the event to one of the categories $B \to \rho \ell \nu$, $B \to X_u \ell \nu$, $B \to D \ell \nu$ or $B \to D^* \ell \nu$, accordingly. $B\bar{B}$ events with fake leptons, secondary leptons or events where the first lepton sister is none of the above are labeled as “Other $B\bar{B}$”.

4.6.2 Continuum Background

As mentioned before, e^+e^- collisions at the $\Upsilon(4S)$ resonance only result in pairs of B mesons a fraction of the time. More often pairs of lighter quarks (u, d, s, c) are produced. These events are labeled “Continuum”.

4.7 Two Analysis Versions

So far I have reconstructed the signal decay $B^0 \to \pi^- \ell^+ \nu_\ell$ by reconstructing a B_{tag} candidate and choosing a lepton and pion candidate from the rest of the event with loose selection criteria to reconstruct a B_{sig} candidate. From this point onward I will split the analysis into two versions. One version follows closely the selections applied by Sibidanov et al. in Ref. [2] who used the FR for reconstructing B_{tag} candidates. This version is called *FR version*. In the second version I use my own set of optimized selections. This is called the *FEI version*. The purpose of the FR version is to see how the FEI performs in comparison to the FR when using the same selections. The purpose of the FEI version is to study if further optimizations of these selections are possible when using the FEI. First I will revisit some of the preselections explained in Section 4.3.

4.8 PID Selection

As the first step I revisit the selections on PID variables, which were explained in Section 3.3. During the reconstruction I deliberately choose loose selections. I start with my optimization in the FEI version.

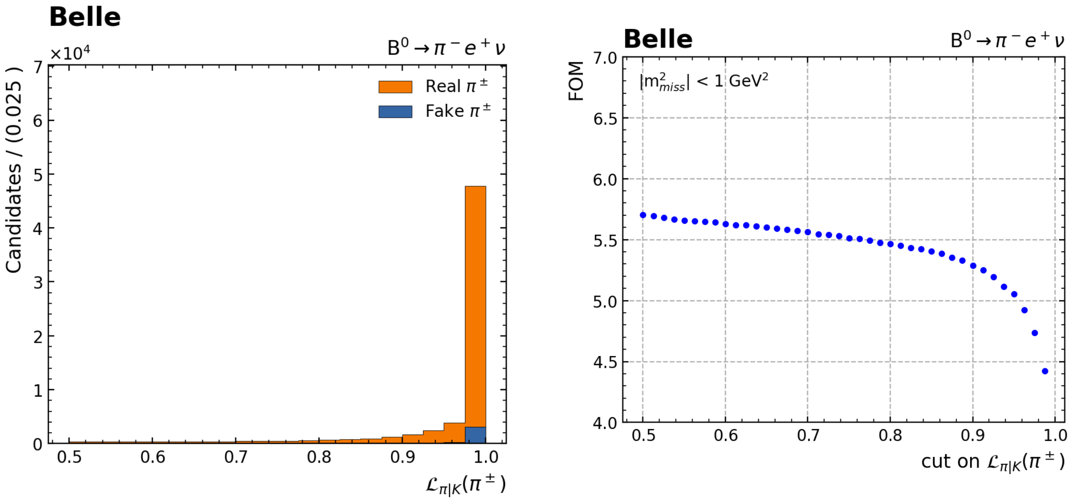
4.8.1 Pion Selection

The most important variable for identifying charged pions is the likelihood ratio $\mathcal{L}_{\pi|K}$, which was introduced in Section 3.3. The distribution of this variable for the pion in $B^0 \rightarrow \pi^- e^+ \nu_e$ after the reconstruction is shown in Figure 4.6 (a). 8.1 % of pion candidates in the electron channel and 6.6 % of pion candidates in the muon channel are other particles misidentified as pions. $\mathcal{L}_{\pi|K}$ peaks at 1 for correctly identified pions.

In the FR version no further cut is placed on $\mathcal{L}_{\pi|K}$ of pion candidates. To find the optimal cut value for the FEI version I calculate the *Figure of Merit* (FOM), which is defined as

$$\text{FOM} = \frac{S}{\sqrt{S+B}}, \tag{4.11}$$

where S is the number of signal events and B the number of background events left after applying a certain selection. I calculate the FOM for events in the previously defined signal region at $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$. The FOM for different cut values is shown in Figure 4.6 (b) for the electron channel. We can see that the FOM is larger for loose cuts. Therefore, I choose to not place a further selection on $\mathcal{L}_{\pi|K}$ for pion candidates. The same plots for the muon channel can be found in the appendix in Figure A.2.



(a)

(b)

Figure 4.6: (a) shows the distribution of $\mathcal{L}_{\pi|K}$ for reconstructed pions in $B^0 \rightarrow \pi^- e^+ \nu_e$ split into real and fake pions. (b) shows the FOM for different cuts on $\mathcal{L}_{\pi|K}$. The FOM is calculated for events in the m^2_{miss} signal region.

4.8.2 Lepton Selection

For the electron and muon channel the distributions of eID and muID of the corresponding lepton candidate are shown in Figure 4.7 (a) and (b). In the electron channel 19.4 % of the electron candidates are fake electrons, which means they actually are other particles misidentified as electrons. For correctly identified electrons the eID peaks at 1, as expected. The red vertical lines in the plot correspond to the working points for which lepton efficiency corrections, as explained in Section 4.2.2, have been obtained.

Sibidanov et al. only keep electron candidates with eID > 0.6, reducing the fake rate to 17.0 % at a signal efficiency of 99.7 % in the FR version. The signal efficiency is defined as

$$\epsilon_{\text{sig}} = \frac{N_{\text{sig}}^{\text{after}}}{N_{\text{sig}}^{\text{before}}}, \tag{4.12}$$

where $N_{\text{sig}}^{\text{after}}$ is the number of signal events that is left after applying a certain selection and $N_{\text{sig}}^{\text{before}}$ is the number of signal events before this selection.

To find the optimal cut for the FEI version I calculate the FOM for different cut values, which is shown in Figure 4.7 (c). We can see that the FOM increases when tightening the eID selection. I therefore choose to cut at eID > 0.9, which yields the highest FOM among the possible cuts indicated by the red lines. This cut reduces the fake rate from 19.4 % to 10.4 % at a signal efficiency of 97.2 %.

In the muon channel the amount of fake leptons is much higher. After the reconstruction the fake rate is 57.3 %; much higher than in the electron channel. One reason for this is that pions can easily get misidentified as muons, due to their similar mass. Also at this stage there is a lot of continuum background. Charged particles that reach the KLM detector are usually identified as muons. Continuum events often result in high-momentum hadrons which in some cases also reach the KLM and get identified as muon.

To find the optimal cut value for the FEI analysis version, I proceed in the same way as for the electron channel. The FOM for different cut values on the muon muID can be seen in Figure 4.7 (d). Similar to the electron channel, the FOM suggests to cut as tightly as possible on the muID. The highest cut value for which muID efficiency corrections have been calculated is 0.97. In order to retain a large number of signal events I choose to select events with muID > 0.9. This reduces the muon fake rate to 52.0 % at a signal efficiency of 98.6 %.

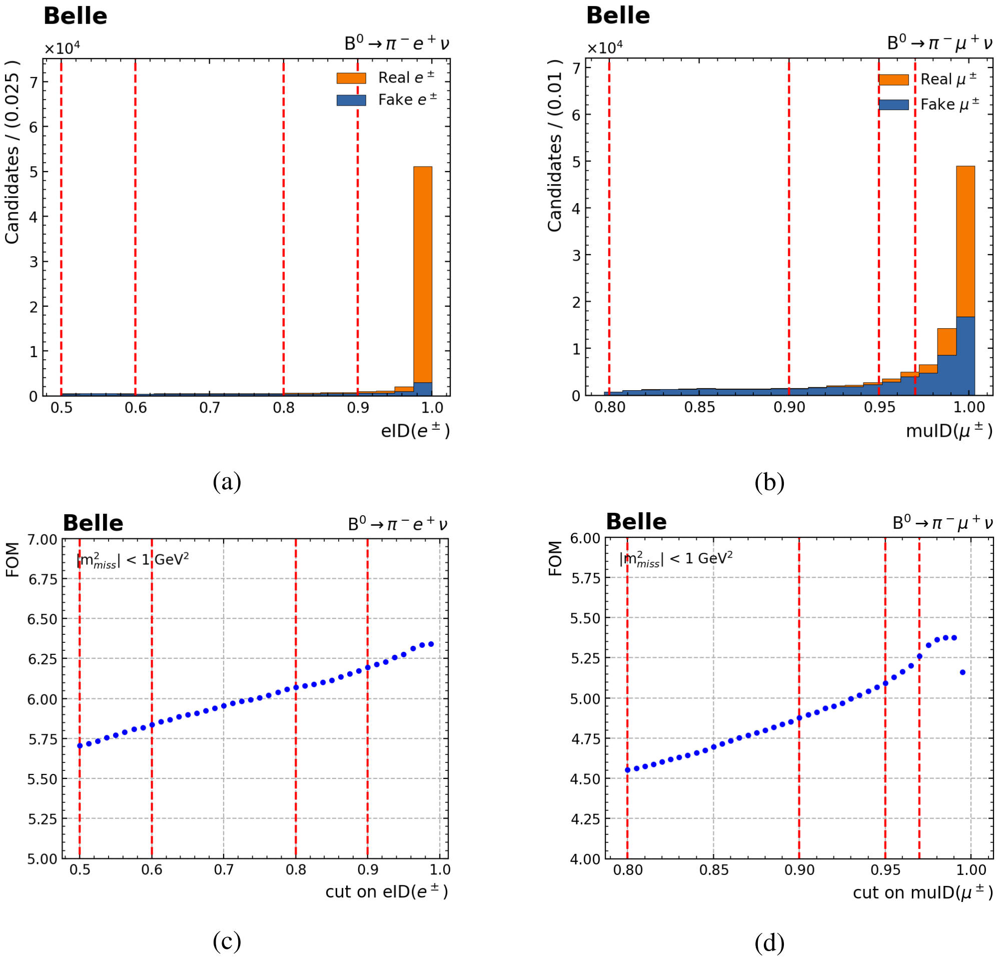
In the FR analysis version I use the same selection for muon candidates, since this was also done by Sibidanov et al.

4.8.3 Best Candidate Selection

After applying the PID selections described above there is still the possibility of having multiple signal candidates for one event. To prevent this from happening in the FR version, only the B_{sig} candidate where the lepton has the highest lepton ID is selected.

In the FEI version I impose an upper limit of 0.9 on the muID of electron and charged pion candidates, and an upper limit of 0.9 on the eID of pion candidates. This results in orthogonal selection criteria and together with the completeness constraint results in there being at maximum one signal candidate per event. An illustration of this can be seen in Figure 4.8.

In both versions a very small number of candidates are rejected by using these mechanisms. The



(a) (b)

(c) (d)

Figure 4.7: (a) and (b) show distributions of electron eID for $B^0 \to \pi^- e^+ \nu_e$ and muon muID for $B^0 \to \pi^- \mu^+ \nu_\mu$ split into real and fake leptons. The vertical red lines correspond to the working points for which lepton ID efficiency corrections have been obtained. (c) and (d) show the FOM for different cuts on the corresponding PID variable. The FOM is calculated for events in the m^2_{miss} signal region.

signal efficiency is in both cases on the order of 1. The final set of PID selections for both analysis versions can be found in Table 4.6.

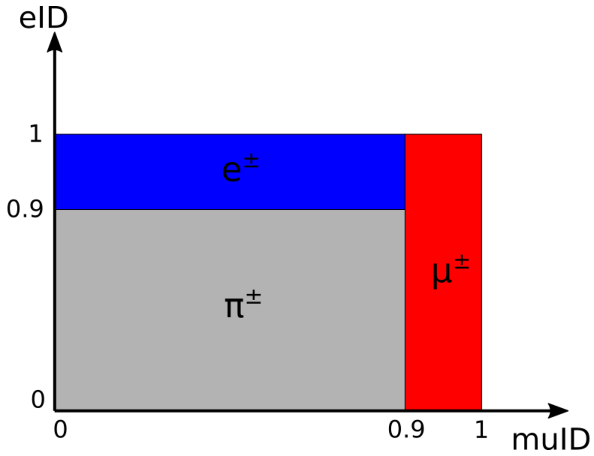


Figure 4.8: PID selection criteria for the FEI analysis version.

Table 4.6: Final PID selections for both FEI and FR analysis versions.

	FEI version			FR version		
	$e^{\pm}$	$\mu^{\pm}$	$\pi^{\pm}$	$e^{\pm}$	$\mu^{\pm}$	$\pi^{\pm}$
eID	> 0.9	-	< 0.9	> 0.6	-	< 0.98
muID	< 0.9	> 0.9	< 0.9	-	> 0.9	< 0.98
$\mathcal{L}_{\pi\|K}$	-	-	> 0.5	-	-	> 0.5

CHAPTER 5

Continuum Suppression

The event rate from e^+e^- collisions at the $\Upsilon(4S)$ resonance is dominated by non-$B\overline{B}$ events. In the m^2_{miss} distributions after the reconstruction shown in Figure 4.5 we saw that a large portion of backgrounds are continuum events in both the electron and muon channel. At this stage continuum events make up about 1/2 of all background events in the electron channel and 2/3 in the muon channel. This fraction is even larger in the m^2_{miss} signal region. For events with $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ 73 % of the background events are continuum events in the electron channel and 88 % in the muon channel. This illustrates how important the suppression of continuum events explained in the following sections is.

5.1 Continuum Characteristics

Continuum events can be distinguished from $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ events by exploiting differences in the angular distributions of the particles that are produced. Because the $\Upsilon(4S)$ mass is just above the $B\overline{B}$ production threshold, in $B\overline{B}$ events both B mesons are produced almost at rest in the $\Upsilon(4S)$ frame. This results in an isotropic distribution of B decay products in the $\Upsilon(4S)$ rest frame. In continuum events, the quarks are produced with a large initial momentum and result in back-to-back jets of light hadrons due to fragmentation. This is illustrated in Figure 5.1. In order to distinguish between $B\overline{B}$

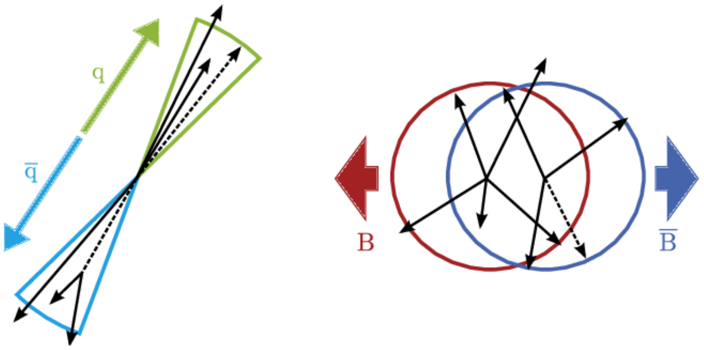


Figure 5.1: Illustration of the jet-like structure of continuum events (left) and the more spherical particle distribution of $B\overline{B}$ events (right) [26].

and continuum events I make use of so-called event shape variables which describe the phase-space distribution of particles detected in an event.

5.2 Event Shape Variables

In the following sections I describe the event shape variables I use to distinguish between continuum and $B\overline{B}$ events. The definitions and explanations are adapted from *The Physics of the B factories* [14].

5.2.1 Thrust

The thrust axis $\vec{T}$ is defined as the unit vector along which the sum of the projected momenta of all particles in an event is maximal:

$$\vec{T} = \max\left(\sum_{i=1}^{N} \vec{T} \cdot \vec{p}_i\right), \tag{5.1}$$

where $\vec{p}_i$ are the momenta of the detected particles.

One variable of interest related to this is $\cos\theta_{\mathrm{T,B}}$, also called cosTBz, where $\theta_{\mathrm{T,B}}$ is the angle between the thrust axis and the z-axis, which corresponds to the beam axis. For $B\overline{B}$ events $|\cos\theta_{\mathrm{T,B}}|$ should be uniformly distributed, while for continuum events the thrust of the particle momenta tend to follow a $1 + \cos^2\theta_{\mathrm{T,B}}$ distribution.

Another interesting quantity is the magnitude of the thrust itself, which is defined as

$$T = \frac{\sum_{i=1}^{N} |\vec{T} \cdot \vec{p}_i|}{\sum_{i=1}^{N} |\vec{p}_i|}. \tag{5.2}$$

5.2.2 Fox-Wolfram Moments

Another parameterization of the phase-space distribution of energy and momentum flow in an event are Fox-Wolfram moments [27], where the k-th order moment is defined as

$$H_k = \sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| P_k(\cos\theta_{ij}). \tag{5.3}$$

Here $\vec{p}_i$ are the momenta of the particles in an event and $\cos\theta_{ij}$ is the angle between $\vec{p}_i$ and $\vec{p}_j$. P_k is the k-th order Legendre polynomial. Often the normalized ratio $R_k = H_k/H_0$ is used. For jet-like events, R_k is close to 0 for odd k, and close to 1 for even k. In this analysis R_2 provides a good discrimination power between $B\overline{B}$ and continuum events, as you will see later on.

Another normalization of Fox-Wolfram moments is defined as

$$h_l^k = \frac{\sum_{m,n} |\vec{p}_m||\vec{p}_n| P_l(\cos\theta_{mn})}{\sum_{m,n} |\vec{p}_m||\vec{p}_n|}, \tag{5.4}$$

where $k = so$ means that particle m is from the signal-side and particle n from the rest of the event. $k = oo$ means that both particles m and n are from the rest of the event.

5.2.3 CLEO Cones

The CLEO collaboration [28] introduced CLEO cones, which describe the momentum flow around the thrust axis. The magnitude of the particle momenta falling into different cones around the thrust axis is measured. The cones are binned in steps of 10 degrees resulting in 9 cones. In $B\overline{B}$ events the flight directions of decay products are not correlated. However, for jet-like continuum events they are correlated.

5.3 FR Version

A. Sibidanov et al. [2] trained a neural net with 18 modified Fox-Wolfram moments and the FR NeuroBayes output variable o_{tag}, which serves a similar purpose as the FEI signal probability, to suppress continuum events. The neural net provided an output variable $o_{\text{tag}}^{\text{cs}}$. The selection criterion on $o_{\text{tag}}^{\text{cs}}$ was placed such that only 0.25 % of events surviving this cut were continuum events.

In order to replicate this I train a BDT to distinguish between signal and continuum events. Both the training and the testing sample contain signal and continuum MC events in a ratio of 50:50, respectively. The training sample has about twice the size of the testing sample. The continuum events I use for training and testing were drawn from the region $|m_{\text{miss}}^2| < 1\,\text{GeV}^2$. I found that this yields a higher background rejection in the signal region.

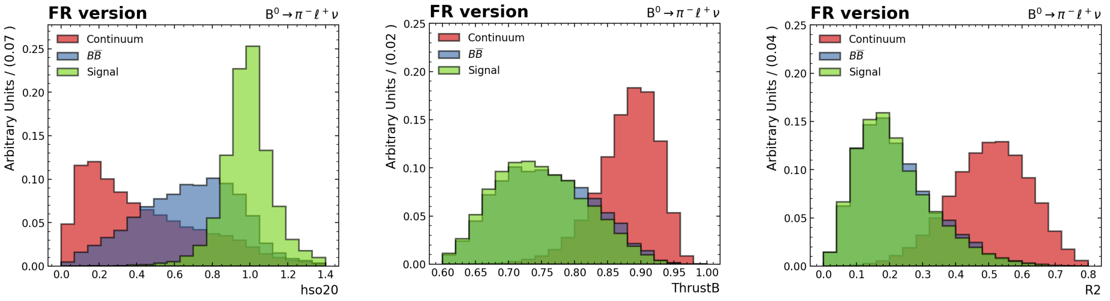
First, I train the BDT with a larger number of event shape variables and the FEI signal probability to find the variables that provide the most discrimination power between signal and continuum events. I select the 8 highest ranking event shape variables together with the FEI signal probability for the final training. The BDT I use is of the type FastBDT and I use the following hyperparameters:

- Number of Trees: 200
- Depth of Trees: 4
- Number of Cuts: 8
- Shrinkage: 0.05
- Sampling Rate: 0.5

The list of variables in decreasing order of importance are:

1. h_{20}^{so}
2. $|\vec{T}|$
3. R_2
4. FEI signal probability
5. E_t
6. $cc1$
7. h_{00}^{so}
8. $\cos\theta_{\text{T,B}}$
9. $cc2$

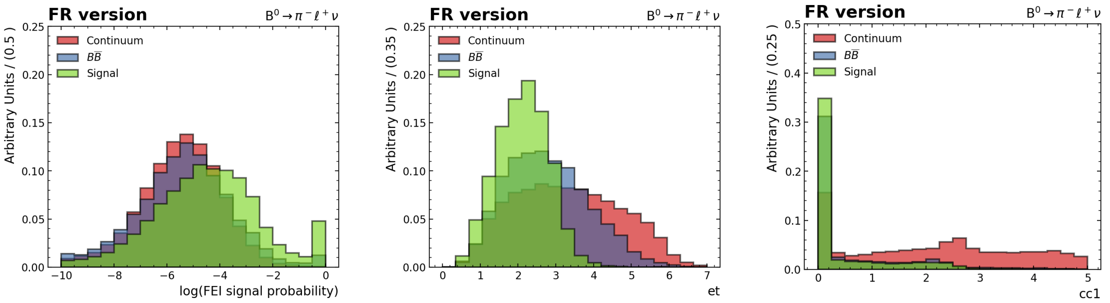
E_t is the transverse energy of an event. The shapes of the distributions for signal, continuum and $B\overline{B}$ events in these variables can be seen in Figure 5.2. We see a good separation between signal and continuum events for all variables. The performance of the BDT on the training and testing sample is shown in Figure 5.3, together with an overtraining test. The Kolmogorov–Smirnov test resulting in p-values of 0.29 for signal and 0.53 for continuum events suggests that no overtraining happened and that both the training and test samples are consistent with being drawn from the same underlying



(a) h_{20}^{so}

(b) $|\vec{T}|$

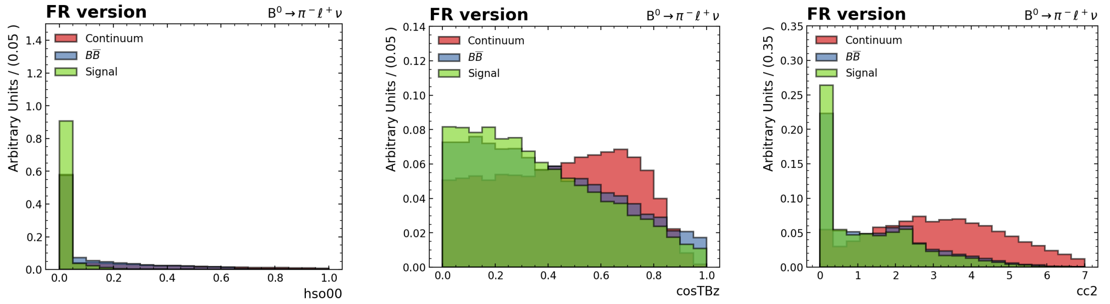
(c) R_2



(d) FEI signal probability

(e) E_t

(f) $cc1$



(g) h_{00}^{so}

(h) $\cos\theta_{\mathrm{T,B}}$

(i) $cc2$

Figure 5.2: Distributions used to train the continuum suppression BDT for $B^0 \to \pi^- \ell^+ \nu_\ell$ in the FR version.

distributions. The receiver operating characteristic (ROC) curve for the trained BDT is shown in Figure 5.4. The area under the curve is 0.99 for the testing sample. As explained in Section 2.5 an area under the ROC curve of 1 means that all background can be rejected while retaining 100 % of the signal. This shows that the BDT can distinguish between signal and continuum well.

Now I want to find the optimal selection on the BDT output classifier. As mentioned above, the continuum suppression neural net used by A. Sibidanov et al. was able to reduce the fraction of continuum events to 0.25 %. Reducing the amount of continuum to a similar level while retaining a large portion of signal events is not possible with the BDT. Instead, to find the optimal cut on the BDT output classifier I calculate the FOM for different cut values. This is shown in Figure 5.5. The maximum FOM is achieved by selecting events with a BDT output classifier > 0.94 in the electron channel and > 0.96 in the muon channel.

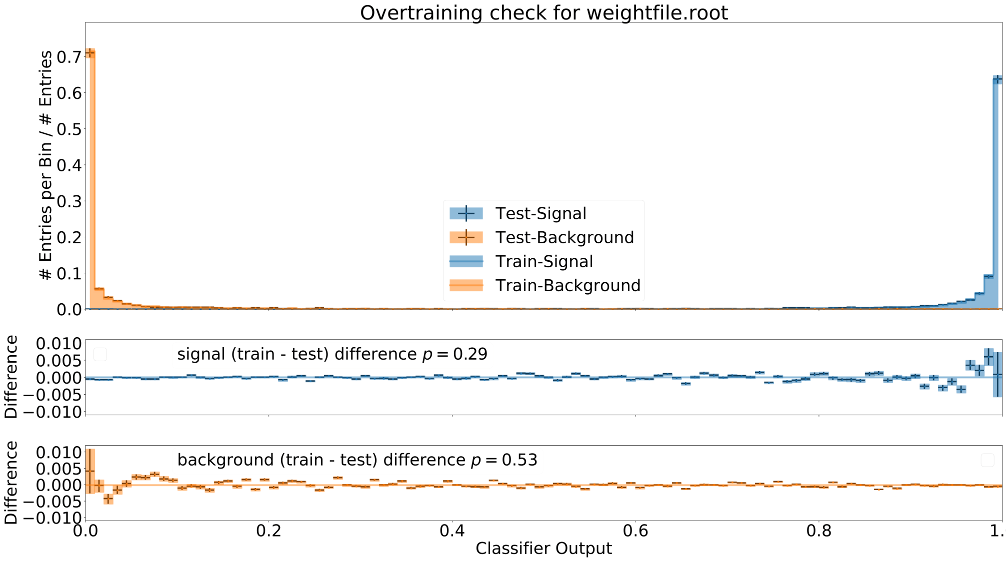


Figure 5.3: BDT classifier output for continuum and signal events split into test and training samples in the FR version. The differences between train and test samples are shown in order to test for overtraining.

In the electron channel this selection yields a background rejection of 87.3 % in the m^2_{miss} signal region at a signal efficiency of 85.6 %. The background rejection is defined as

$$r_{\text{bkg}} = \frac{B_{\text{before}} - B_{\text{after}}}{B_{\text{before}}}, \tag{5.5}$$

where B_{before} and B_{after} is the number of background events before and after applying a selection, respectively. After the selection the overall fraction of continuum events in the electron channel has been reduced to 2.6 %.

In the muon channel the selection results in a background rejection of 93.9 % in the m^2_{miss} signal region while retaining 82.1 % of the signal events. The selection decreases the fraction of continuum events to 3.3 % of all events.

In Figure 5.6 I show the m^2_{miss} distributions for the electron and muon channel after applying the selection on the continuum suppression BDT classifier. We can see that that the amount of continuum backgrounds has drastically decreased. Now we can clearly see the signal peak at $m^2_{\text{miss}} = 0\,\text{GeV}^2$.

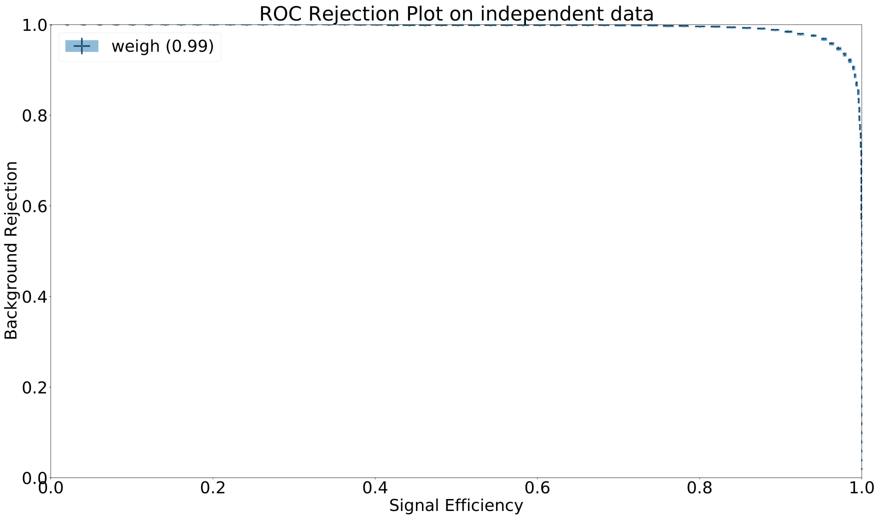
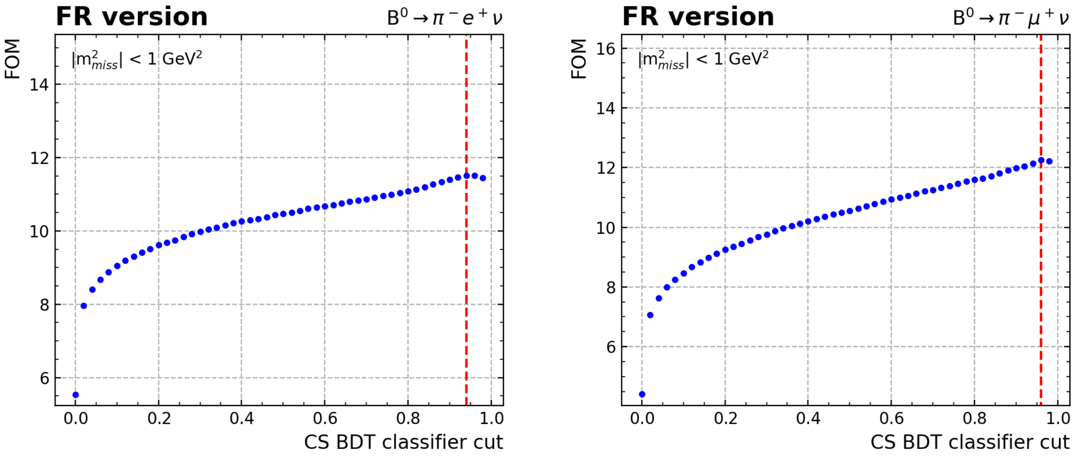


Figure 5.4: Background rejection versus signal efficiency on test sample in the FR version.



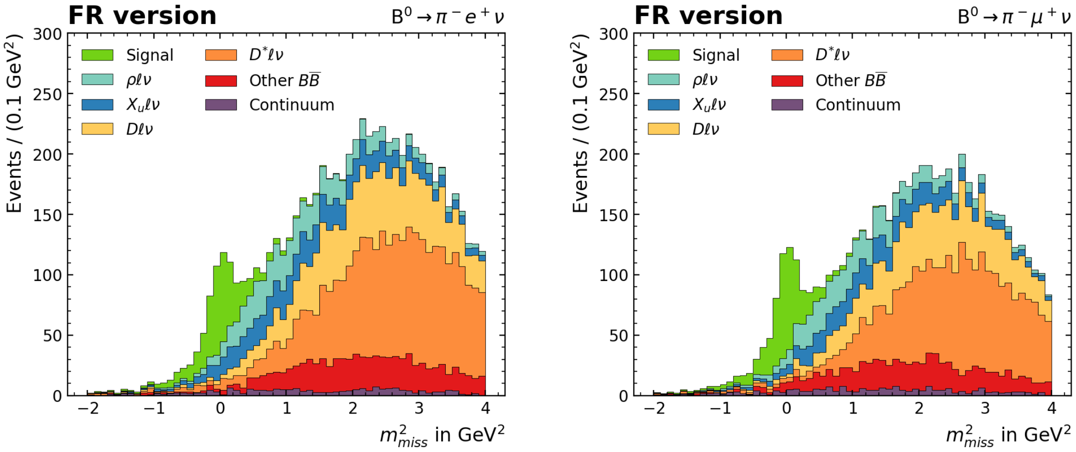
(a)

(b)

Figure 5.5: FOM in the m^2_{miss} signal region for different cuts on the BDT classifier in (a) the electron and (b) muon channel of the FR version.

5.4 FEI Version

For the FEI version I proceed similarly to what I did in the FR version. Again I train a BDT with a training and a testing sample containing signal and continuum MC events in a ratio of 50:50, with the continuum events being drawn from the region $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$. This time I only use event shape variables and do not include the FEI signal probability in the BDT, since the purpose of the FEI is solely to find B_{tag} candidates in an event. As we saw above, the FEI signal probability assigned to



(a)

(b)

Figure 5.6: Distribution of m^2_{miss} for (a) the electron channel and (b) the muon channel after applying the selection on the continuum suppression BDT classifier in the FR version.

a B_{tag} candidate has some discriminating power between signal and continuum events. However, I only want to use variables describing the whole event for continuum suppression and treat variables concerning the quality of a B_{tag} candidate separately in the upcoming chapter.

The 8 event shape variables with the most discriminating power between signal and continuum events are:

1. h^{so}_{20}
2. $|\vec{T}|$
3. R_2
4. E_t
5. $cc1$
6. $\cos\theta_{\text{T,B}}$
7. h^{so}_{00}
8. $cc2$

The distributions of these variables for signal, continuum and $B\bar{B}$ events can be found in the appendix in Figure A.3 and look similar to the ones shown above in Figure 5.2. For the BDT in the FEI version I use the same hyperparameters as in the FR version of the analysis. The performance of the BDT on the training and testing sample can be seen in Figure 5.7 and the ROC curve for the testing sample in Figure 5.8. Again we see a good separation of signal and continuum events with an area of 0.99 under the ROC curve and no signs of overtraining.

The FOM for different cuts on the BDT output classifier for both the electron and muon channel can be seen in Figure 5.9. The FOM suggests to place a tight cut on the classifier again. Despite this, I choose to cut relatively loosely, only keeping events with a BDT output classifier larger than 0.5. This also rejects the majority of continuum events, as shown in Figure 5.7, but retains a larger number of signal events. The signal efficiency of this selection is 96.7 % in the electron channel and 96.7 % in the muon channel, while the background rejection in the m^2_{miss} signal region is 73.8 % in the electron channel and 85.6 % in the muon channel. At a later stage I will revisit the selection on the BDT output classifier with a method other than calculating the FOM.

The distributions of m^2_{miss} for both the electron and muon channel after applying the BDT classifier selection are shown in Figure 5.10.

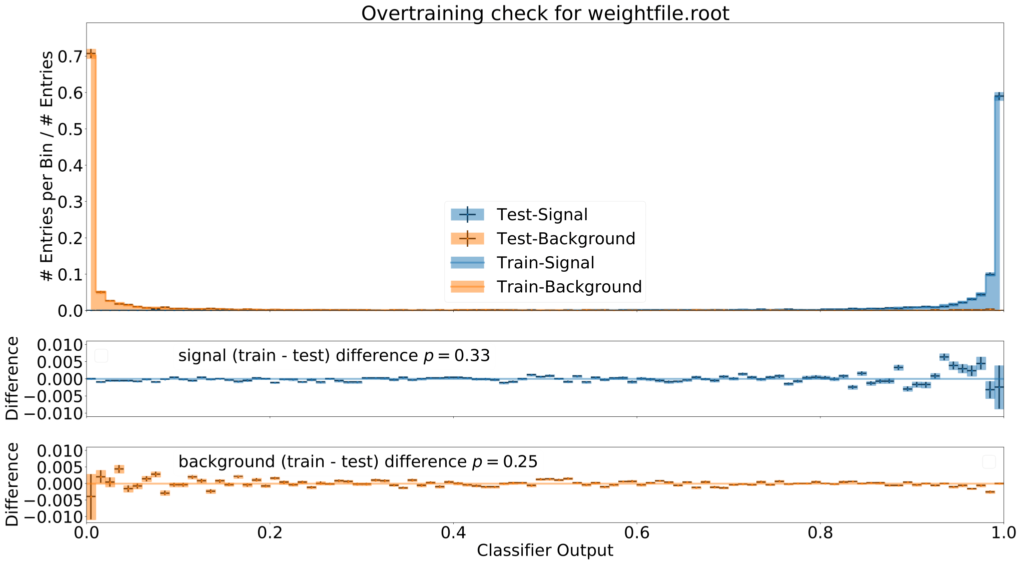


Figure 5.7: BDT classifier output for continuum and signal events split into test and training samples in the FEI version. The differences between train and test samples are shown in order to test for overtraining.

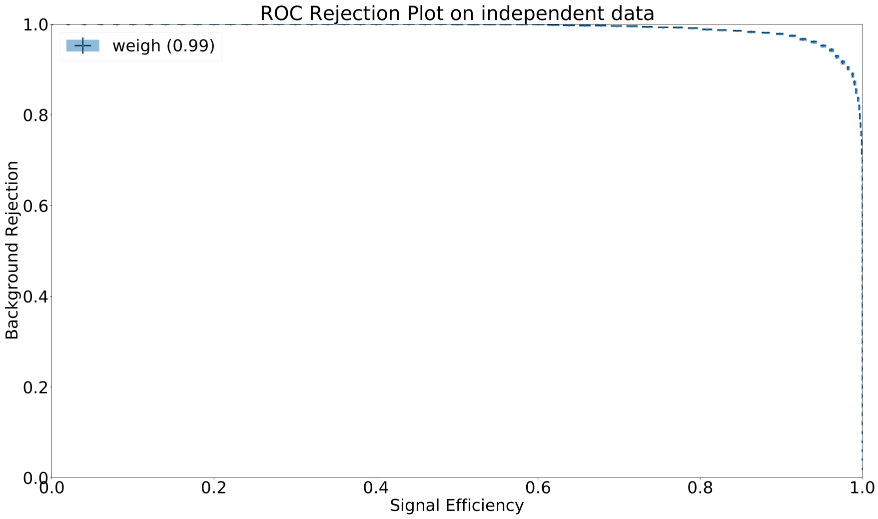
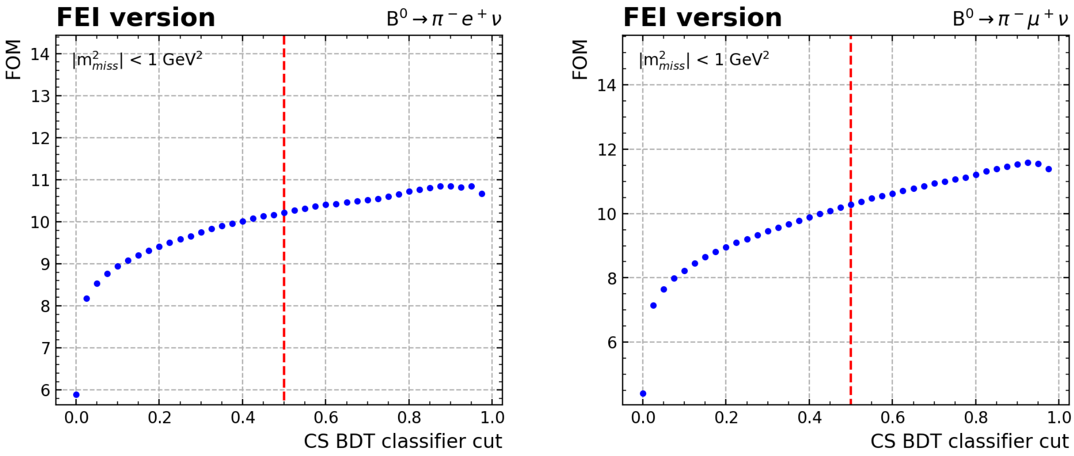


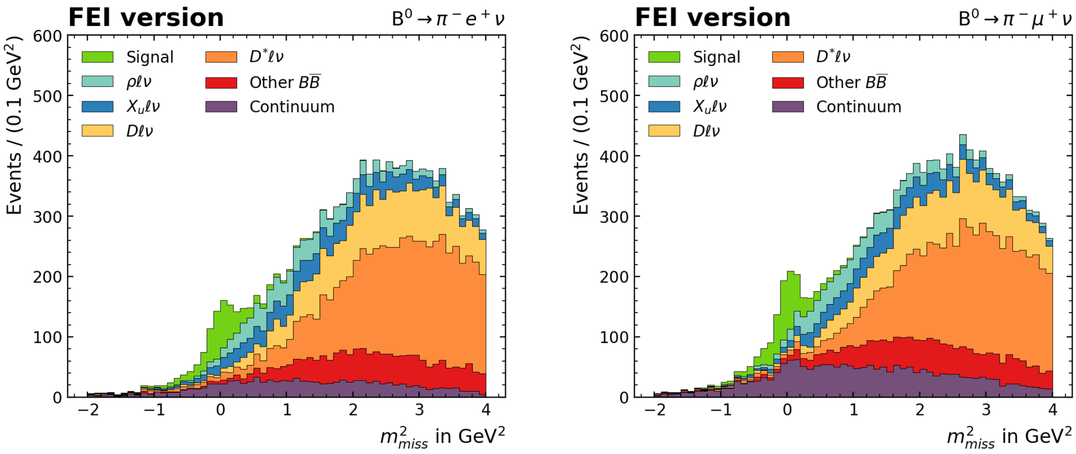
Figure 5.8: Background rejection versus signal efficiency on test sample in the FEI version.



(a)

(b)

Figure 5.9: FOM in the m^2_{miss} signal region for different cuts on the BDT classifier in (a) the electron and (b) muon channel of the FEI version.



(a)

(b)

Figure 5.10: Distribution of m^2_{miss} for (a) the electron channel and (b) the muon channel after applying the selection on the continuum suppression BDT classifier in the FEI version.

CHAPTER 6

Tag Selection

In this chapter I introduce the tag selection which serves the purpose of increasing the fraction of events with correctly reconstructed B_{tag} candidates. A high purity of events with correctly reconstructed B_{tag} candidates reduces the probability of accidentally mixing particles from the signal- and tag-side, which improves the resolution of m^2_{miss}. Also, incorrectly reconstructed B_{tag} candidates have the wrong four-momentum which results in an incorrect value of m^2_{miss} on the signal-side. The tag selection involves selections on the FEI signal probability assigned to each B_{tag} candidate by the FEI, the beam-energy constrained mass M_{bc} and the energy difference ΔE, which were introduced in Section 4.3.1.

6.1 FEI Signal Probability

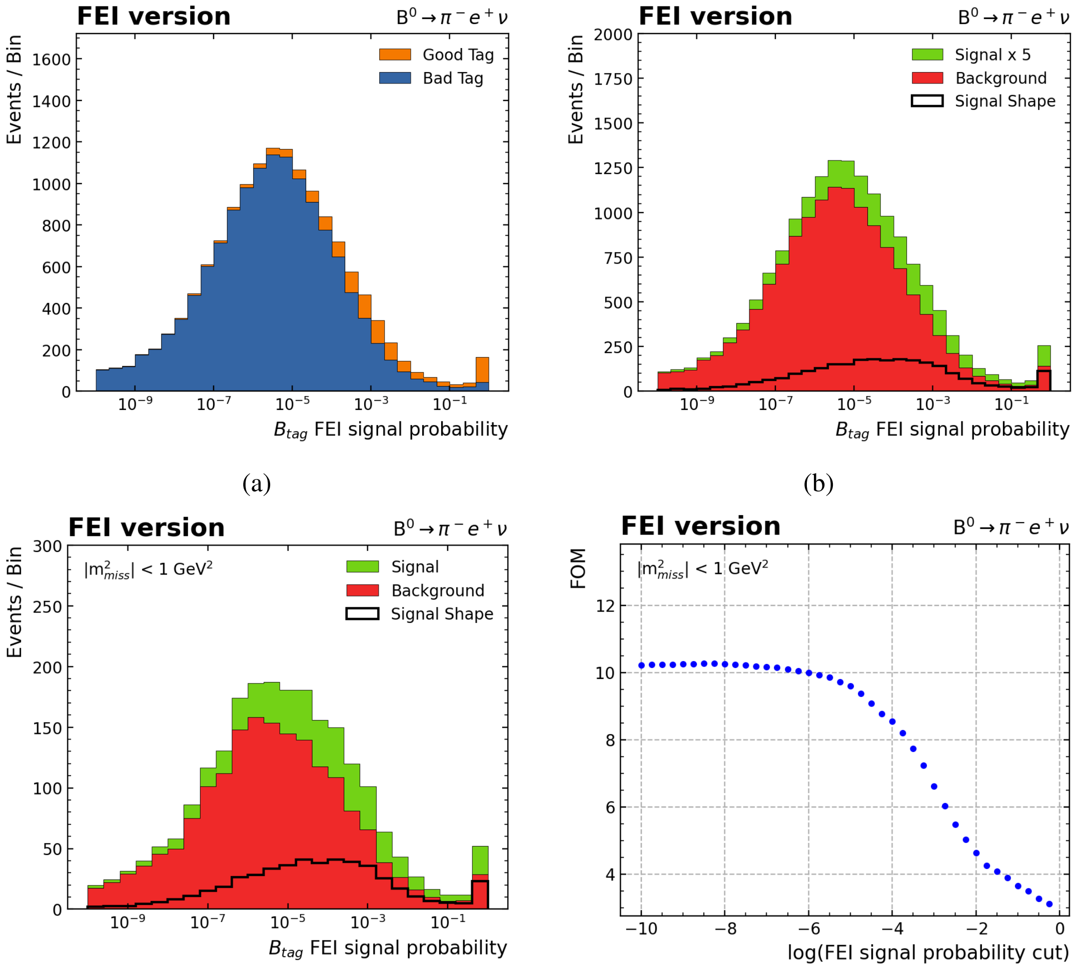
The FEI signal probability takes values between 0 and 1 and tends to have larger values in cases where the B_{tag} candidate is a correctly reconstructed B meson. In the FR analysis version the FEI signal probability is included in the continuum suppression BDT and thus will not be further investigated.

For the FEI version the distribution of the FEI signal probability after applying the continuum suppression is shown in Figure 6.1 for the electron channel. In (a) the distribution is split into *good tags* and *bad tags*. I define good tags as reconstructions where all final state particle candidates assigned to the B_{tag} candidate actually stem from the same B meson. This definition allows for misreconstructed intermediate particles and missing particles on the tag-side, which has a negative impact on the m^2_{miss} resolution. However, this definition delivers a good estimate for whether the the FEI succeeded in properly reconstructing a B_{tag} or not. We can see that for good tags there is a small peak at 1 and most good tags having values greater than 10^{-4}.

For bad tags the distribution has a broad peak around 10^{-5} with tails towards 1 and the lower end of the spectrum. This suggests that placing a cut on the FEI signal probability at around 10^{-4} would be optimal for removing a larger portion of events of bad tags and retaining most events with good tags.

In Figure 6.1 (b) the same distribution is split into signal and background events. The signal component is scaled up by a factor of 5 for better visibility. For signal events we also see a small peak at 1 but also a broad bump around 10^{-4} with signal events extending all the way to the lower end of the spectrum. With the tag selection we not only want to increase the purity of good tags but also retain a large number of signal events. Rejecting all events where the B_{tag} candidate has a FEI signal probability below 10^{-4} would drastically increase the fraction of good tags but also reject more than half of the signal events.

In order to find the optimal cut value, I again calculate the FOM in the signal region of $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ for different cuts. The distribution of the FEI signal probability for signal and background in the signal region and the FOM for different cut values can be seen in Figure 6.1 (c) and (d) for the electron channel. The same distributions for the muon channel can be seen in the appendix in Figure A.4. The FOM increases for loose cuts and has a plateau for cut values below 10^{-6}. Since there is no obvious optimal cut, and the FOM suggests that a loose selection is beneficial for the signal to background ratio, I choose to place no cut on the FEI signal probability at this point.



(a) (b)

(c) (d)

Figure 6.1: Distribution of FEI signal probability for $B^0 \to \pi^- e^+ \nu_e$ after applying the continuum suppression. (a) shows the distribution split into good tags and bad tags. (b) shows the same distribution split into signal and background decays. The signal component is scaled up by a factor of 5. (c) Distribution of FEI signal probability for $B^0 \to \pi^- e^+ \nu_e$ in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region split into signal and background events. (d) FOM in the signal region for different cuts on FEI signal probability.

6.2 Beam-Energy Constrained Mass M_{bc}

As explained in Section 4.3.1 the beam-energy constrained mass M_{bc} should be close to the B meson mass of 5.28 GeV for a correctly reconstructed B_{tag} candidate. Sibidanov et al. chose to reject events where M_{bc} of the B_{tag} candidate is below 5.27 GeV. In my FR analysis version this selection increases the tag purity from around 13 % to 20 % in the electron and muon channel, respectively, at signal efficiencies of 86 % in both channels.

The distribution of M_{bc} after the previous selections in the FEI analysis version is shown in Figure 6.2 for the electron channel. (a) shows the distribution split in good and bad tags. As expected, events with a correctly reconstructed B_{tag} candidate are accumulated around $M_{\mathrm{bc}} = 5.28$ GeV, while for events with badly reconstructed B_{tag} candidates the distribution is much broader. (b) shows the same distribution; this time split into signal and background events. We can see that signal events also show a peaking structure at 5.28 GeV but with a long tail towards lower values.

To find the optimal selection I calculate the FOM in the signal region for different cut values. The distribution of M_{bc} for events with $|m^2_{\mathrm{miss}}| < 1\,\mathrm{GeV}^2$ is shown in Figure 6.2 (c). The FOM for different cut values is shown in (d). We see that the FOM strongly decreases for high cut values, but also slightly decreases for very loose cuts. The cut value of 5.27 GeV appears to be optimal and yields a high background rejection while retaining a large portion of signal events. The same diagrams for the muon channel can be found in the appendix in Figure A.5

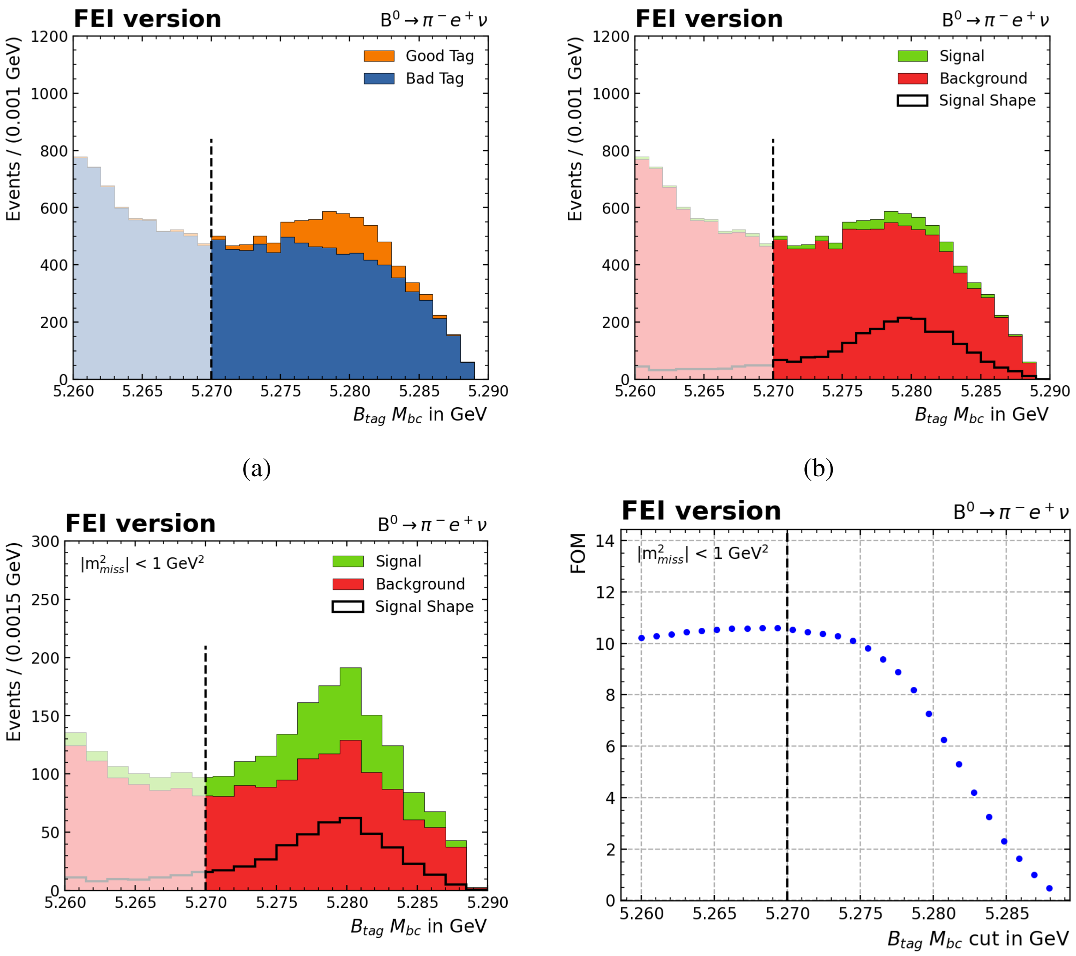
Rejecting all events where the B_{tag} candidate has an M_{bc} value below 5.27 GeV increases the tag purity in the electron channel from 7.6 % to 12.5 % at a signal efficiency of 84.4 %. In the muon channel the same selection increases the tag purity from 7.4 % to 12.2 % at a signal efficiency of 84.6 %. This selection results in the rejection of 37.6 % and 41.3 % of background events in the signal region in the electron and muon channel, respectively.

6.3 Energy Difference ΔE

The energy difference ΔE between the reconstructed energy of a B_{tag} candidate and the beam energy should be close to 0 for a correctly reconstructed B meson. Sibidanov et al. did not place any further selection on ΔE, hence I leave this variable untouched in the FR analysis version.

For the FEI analysis version the distribution of ΔE in the electron channel after applying the M_{bc} selection is shown for the electron channel in Figure 6.3 (a) and (b) for good vs. bad tags and signal vs. background, respectively. In most events with a correctly reconstructed B_{tag} candidate ΔE is close to 0, as expected. For signal events there is also a peak at 0 with a tail towards low values of ΔE. A similar behavior can be observed in the signal region of $|m^2_{\mathrm{miss}}| < 1\,\mathrm{GeV}^2$ in Figure 6.3 (c). Setting a lower limit on ΔE does not appear to be useful since many signal events inhabit the low ΔE range. During the reconstruction I require B_{tag} candidates to fulfill $|\Delta E| < 0.2$ GeV. Loosening this selection during reconstruction would increase the signal efficiency. At the same time, the signal events gained by a looser selection are likely to have poorly reconstructed B_{tag} candidates and thus a poor m^2_{miss} resolution. Due to this reason, I decide to keep the preselection.

Setting an upper limit, however, can reject at least a few background events without loosing many signal events. The FOM for different upper limits on ΔE can be seen in Figure 6.3 (d). For cut values above 0.1 GeV the FOM stagnates. Therefore, I reject events with $\Delta E > 0.1$ GeV increasing the tag purity to almost 14 % in both the electron and muon channel at signal efficiencies of 95 % in both

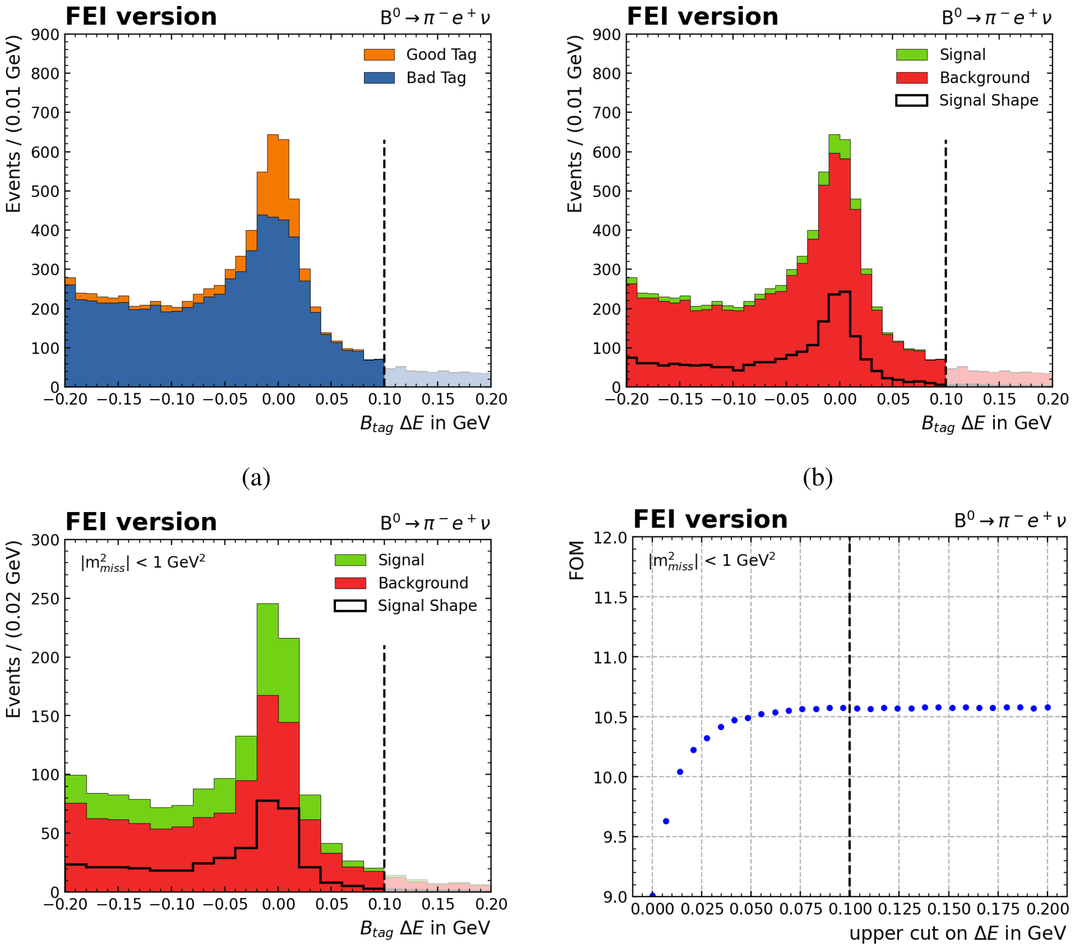


(a) (b)

(c) (d)

Figure 6.2: (a) Distribution of B_{tag} M_{bc} for $B^0 \to \pi^- e^+ \nu_e$ split into good and bad tags. (b) The same distribution split into signal and background events. (c) M_{bc} distribution in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region split into signal and background events. (d) FOM in the signal region for different cuts on M_{bc}. The dashed black line indicates the selected cut.

channels. The same plots for the muon channel can be seen in the appendix in Figure A.6.



(a) (b)

(c) (d)

Figure 6.3: (a) Distribution of B_{tag} ΔE for $B^0 \to \pi^- e^+ \nu_e$ split into good and bad tags. (b) The same distribution split into signal and background events. (c) ΔE distribution in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region split into signal and background events. (d) FOM in the signal region for different cuts on ΔE. The dashed black line indicates the selected cut.

CHAPTER 7

Signal Selection

The selections described in the two previous chapters are aimed at reducing the amount of continuum backgrounds and increasing the purity of correctly reconstructed tag-sides. The selections described in this chapter have the purpose to distinguish between signal and background events specifically for the signal decay $B^0 \to \pi^- \ell^+ \nu_\ell$.

7.1 Point of Closest Approach

For signal events the signal-side lepton and charged pion originate from the same vertex and therefore should have matching coordinates at the start of the track in the vicinity of the interaction point. Sibidanov et al. require for signal-side final state candidates to comply with $|z^{\ell}_{\mathrm{PCA}} - z^{\pi}_{\mathrm{PCA}}| < 1\,\mathrm{mm}$.

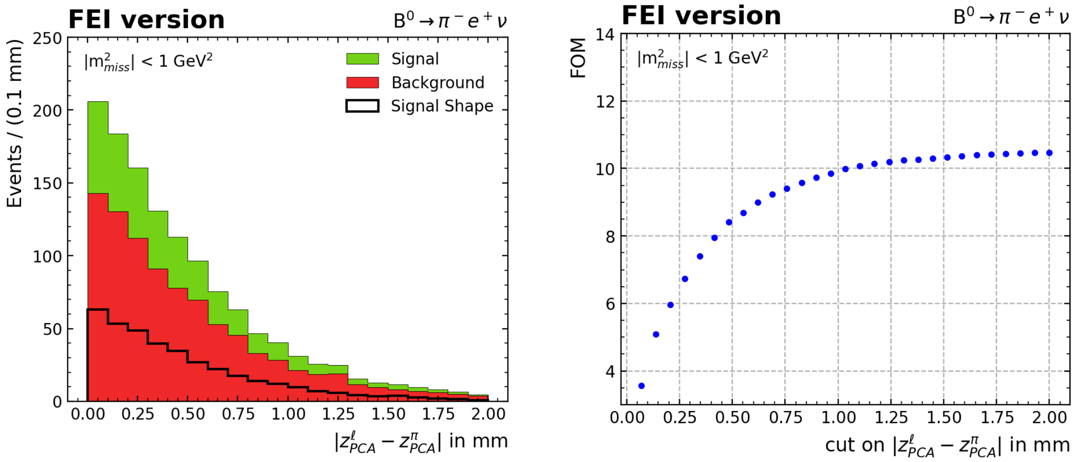
In the FR version of the analysis this selection rejects 21 % and 17 % of the remaining background events in the m^2_{miss} signal region in the electron and muon channel, respectively. In both channels this comes with a signal efficiency of 84 %.

For the FEI analysis version the distribution of $|z^{\ell}_{\mathrm{PCA}} - z^{\pi}_{\mathrm{PCA}}|$ can be seen in Figure 7.1 (a) for signal and background events in the electron channel. Both signal and background events show similar shapes, peaking at 0 and decreasing asymptotically for larger values. The FOM calculated for different selection cuts in the m^2_{miss} signal region can be seen in Figure 7.1 (b). The FOM increases for looser cuts on $|z^{\ell}_{\mathrm{PCA}} - z^{\pi}_{\mathrm{PCA}}|$ and stagnates for values above 1 mm. Since only few signal and background events possess values of $|z^{\ell}_{\mathrm{PCA}} - z^{\pi}_{\mathrm{PCA}}|$ larger than 1 mm, I choose to not place a restriction on this variable.

7.2 Missing Energy

In correctly reconstructed signal events every particle in the whole event is reconstructed except for a single neutrino. This neutrino carries some energy and thus there should be some missing energy:

$$E_{\mathrm{miss}} = m_{\Upsilon(4S)} - \sum_{i=1}^{N} E_i \tag{7.1}$$



(a) (b)

Figure 7.1: (a) $|z^{\ell}_{\text{PCA}} - z^{\pi}_{\text{PCA}}|$ distribution in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region for signal and background events. (b) FOM in the signal region for different cuts. Both are shown for the electron channel of the FEI version.

E_i is the reconstructed energy of charged final state particles and photons. Background events where the signal-side B meson decays hadronically but a pion is misidentified as a lepton have a missing energy close to 0, if all particles from the signal- and tag-side are detected. In order to avoid such background events Sibidanov et al. requires $E_{\text{miss}} > 300\,\text{MeV}$. In the FR version of this analysis this selection reduces the number of background events in the signal region by 8 % in both the electron on muon channel while retaining 95 % of signal events in each channel.

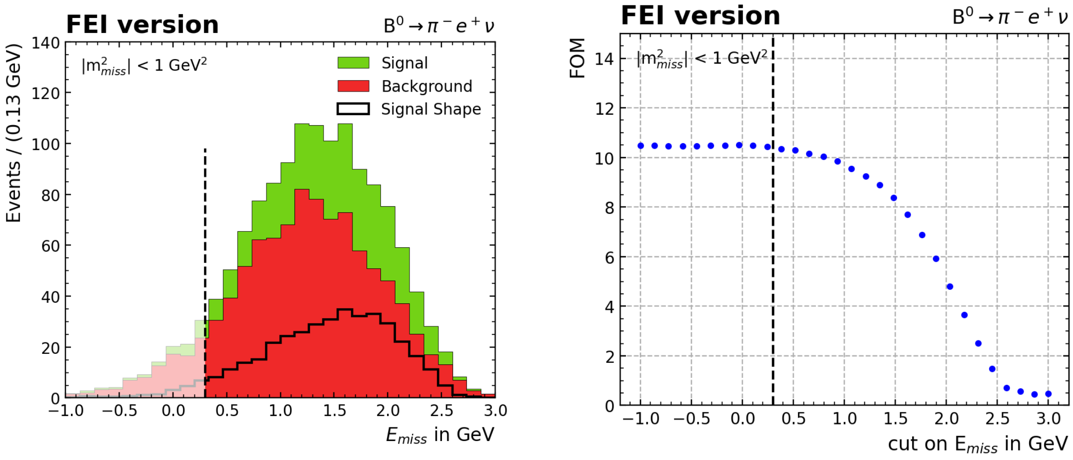
For the FEI version of this analysis the distribution of E_{miss} in the signal region $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ can be seen in Figure 7.2 (a). Signal events tend to have larger values of E_{miss} with most signal events inhabiting the range between 0.5 and 2.5 GeV. For background events the distribution extends to values below 0. In cases where E_{miss} is less than 0, more energy was detected than was provided by the e^+e^- collision. This could be due to beam backgrounds or misreconstructed particles.

The FOM for different lower cut values is shown in Figure 7.2 (b). I choose to also reject events with $E_{\text{miss}} < 300\,\text{MeV}$ which results in the rejection of 9.9 % and 12.7 % of background events in the signal region in the electron and muon channel, respectively. In the electron channel this comes at a signal efficiency of 94.8 % and in the muon channel of 94.5 %.

7.3 Extra Energy in ECL

The extra energy in the ECL (E_{ECL}) is energy that originates from clusters in the ECL not assigned to any reconstructed final state particle. The distribution of E_{ECL} is interesting because for correctly reconstructed signal events there are no unreconstructed particles that deposit energy in the ECL, except for unreconstructed final state radiation (FSR) photons and photons from beam backgrounds. Therefore the distribution should peak at 0. For background events E_{ECL} takes higher values due to unreconstructed γ, π^0, or K^0_L which shower hadronically in the ECL.

Sibidanov et al. require the extra energy to satisfy $E_{\text{ECL}} < 1\,\text{GeV}$. In the FR analysis version this results in background rejections of 16 % and 14 % in the m^2_{miss} signal region in the electron and muon



(a)

(b)

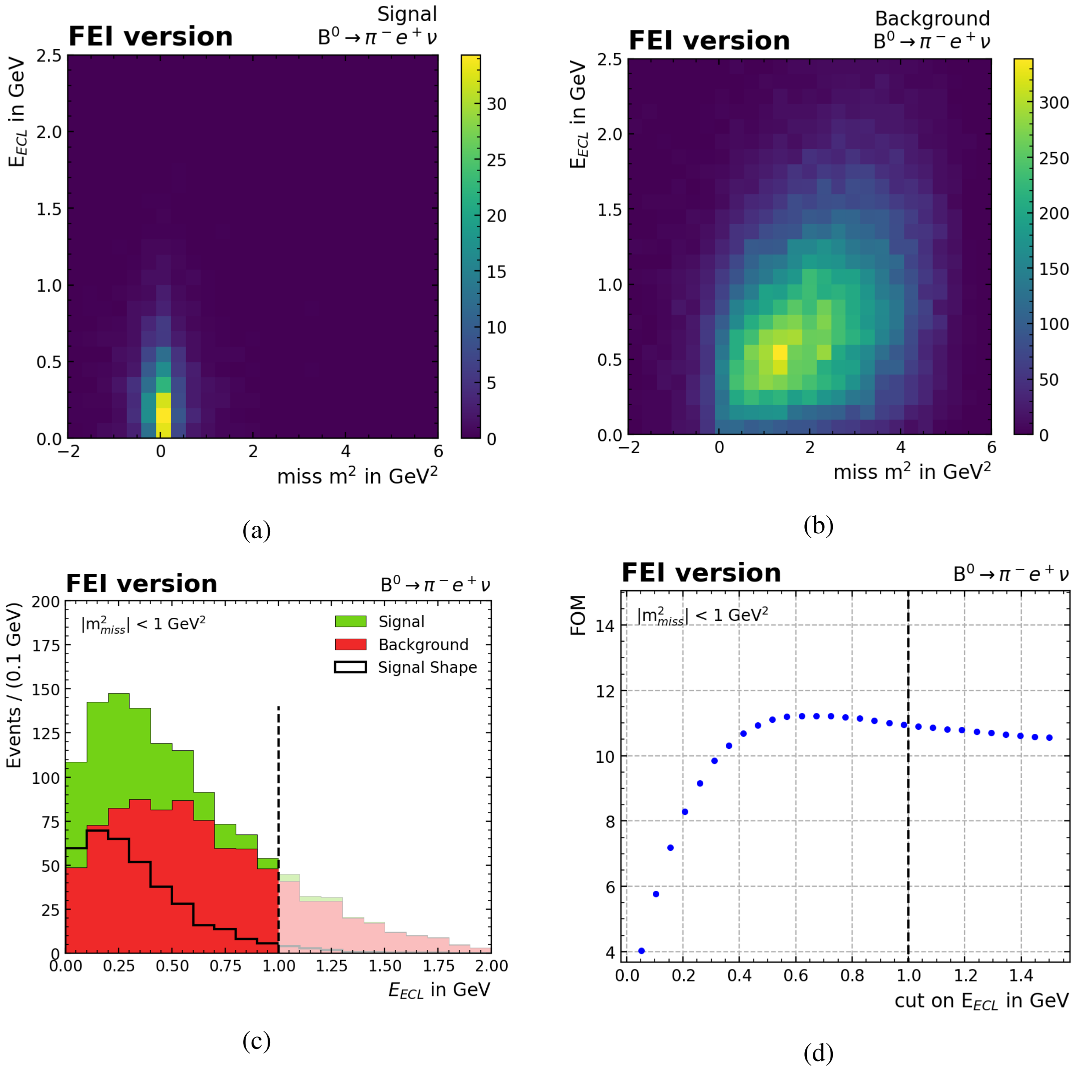
Figure 7.2: (a) Distribution of E_{miss} for signal and background events in the m^2_{miss} signal region for the electron channel of the FEI version. (b) FOM for different lower cuts on E_{miss}.

channel.

The distribution of E_{ECL} versus m^2_{miss} in the FEI version for signal and background events can be seen in Figure 7.3 (a) and (b). As expected, signal events are confined to low values of E_{ECL} at m^2_{miss} close to 0. For background events both variables have a much broader distribution. These plots show that setting an upper limit on E_{ECL} rejects background events over a wide range of m^2_{miss}, including the signal region, while retaining a large portion of background at higher m^2_{miss} needed to fix background shapes in the fit procedure.

The E_{ECL} distribution for signal and background events in the m^2_{miss} signal region is shown in Figure 7.3 (c). The fact that there is no sharp peak at 0 for signal events is mostly due to beam induced backgrounds and unreconstructed particles from the tag-side. The FOM in the signal region, as shown in Figure 7.3 (d), is highest for an upper cut of around 0.6 GeV on E_{ECL}. For higher cut values the FOM has a plateau and does not change much. In order to retain almost all signal events and still reject some background events I choose to require $E_{\text{ECL}} < 1$ GeV. This is also beneficial in a way that this selection is not very sensitive to shape differences in E_{ECL} between MC and data. The shape of E_{ECL} in MC depends strongly on the modeling of beam backgrounds, which is difficult and not very reliable.

The selection of $E_{\text{ECL}} < 1$ GeV in the FEI version reduces the number of background events in the signal region by 21.2 % and 18.1 % at signal efficiencies of 96.3 % in the electron and muon channel, respectively.



(a)

(b)

(c)

(d)

Figure 7.3: E_{ECL} vs. m^2_{miss} for (a) signal events and (b) background events. (c) Distribution of E_{ECL} for signal and background events in the m^2_{miss} signal region for the electron channel of the FEI version. (d) FOM for different lower cuts on E_{ECL}.

CHAPTER 8

Signal Extraction

This chapter will describe the fit procedure for the signal extraction from the m^2_{miss} distribution. An optimization for the selections on the continuum suppression BDT classifier and FEI signal probability in the FEI version is introduced. Results from the FR and FEI versions are compared to the results obtained by Sibidanov et al., followed by an investigation of the m^2_{miss} resolution achieved by the FEI.

8.1 Fit Setup

The number of signal events passing all selection criteria will be extracted through a fit to the m^2_{miss} distribution. A fit to the full Belle data set of 711 fb^{-1} is the ultimate goal. This is beyond the scope of this thesis and will be done at a later time after an official review. The fit procedure is a binned maximum likelihood technique [29], during which a likelihood function for the agreement between the distribution of data and the distributions of the MC components is built. This technique takes into account the finite MC statistics in the template histograms that form the components of the fit [2]. Each MC template has a yield, which is the number of events that this template is composed of. During the fit, components can either be fixed, meaning that the yield of this component will not be changed, or the yield will float within a limited range. The binned maximum likelihood fit returns a yield for each MC component [30].

The components that the MC samples are split into and define the fit templates are the following:

- $B^0 \to \pi^- \ell^+ \nu_\ell$ signal
- $B^0 \to \rho^- \ell^+ \nu_\ell$ cross-feed
- other $B^0 \to X_u \ell^+ \nu_\ell$ cross-feed
- other $B\overline{B}$ backgrounds
- $q\overline{q}$ continuum

The same templates were used by Sibidanov et al., who fixed the continuum component to the MC prediction, and the amount of $B^0 \to \rho^- \ell^+ \nu_\ell$ cross-feed to the value obtained by a separate study of this decay. In the fit I fix both the continuum and $B^0 \to \rho^- \ell^+ \nu_\ell$ components to their MC prediction.

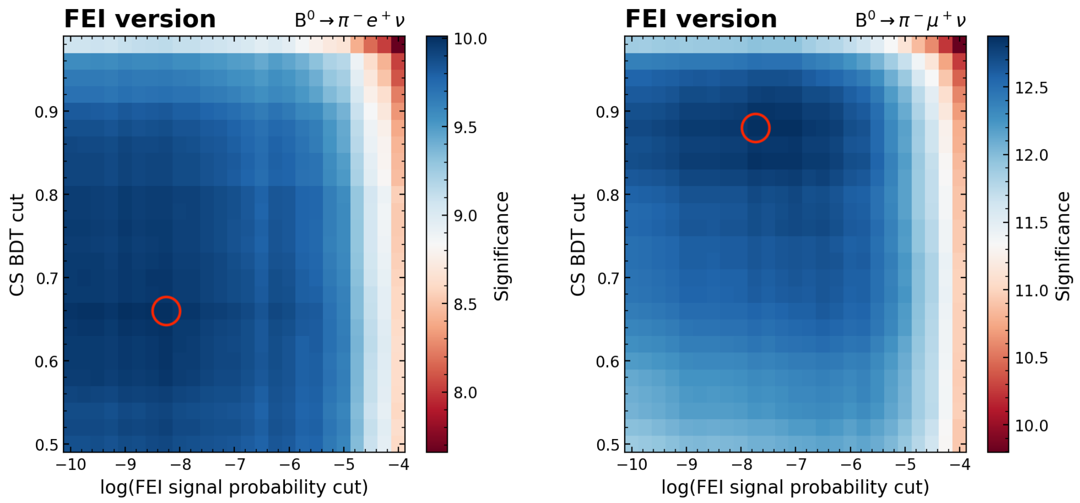
Since I do not fit to real data, I perform an Asimov fit. This means fitting the templates to themselves. The yields of the different fit components, which are estimated during the fit procedure, are equal to the yields of the MC templates for an Asimov fit. The fit parameter uncertainties from an Asimov fit correspond to those coming from statistical fluctuations.

8.2 Significance Optimization

A measure for how well the signal is separated from background is the significance, which is defined as

$$S = \frac{N_{\text{sig}}}{\Delta N_{\text{sig}}}. \tag{8.1}$$

N_{sig} is the signal yield and ΔN_{sig} is the uncertainty on the signal yield, both obtained from the Asimov fit. The uncertainty on the signal yield is proportional to the statistical uncertainty, which is the square root of the yield. For the FEI version of my analysis I use the significance to find the optimal selection on the continuum suppression BDT output classifier and the FEI signal probability of the B_{tag} candidate. To do so I perform Asimov fits for different combinations of selections on these two variables and see which combination yields the highest signal significance. The results of this can be seen in Figure 8.1. The combination of selection cuts yielding the highest signal significance are marked by the red circles and can be found in Table 8.1. The seemingly optimal selection on the FEI signal probability is in both the electron and muon channel quite loose, resulting in a high signal efficiency. Also the selection on the continuum suppression BDT classifier is looser than the selections chosen in the FR analysis version, which were optimized based on the FOM. The optimal BDT cut in the muon channel is higher than in the electron channel. This makes sense because the



(a) (b)

Figure 8.1: Significance from Asimov fits for different combinations of selections on the continuum suppression BDT classifier and FEI signal probability for the (a) electron and (b) muon channel of the FEI analysis version. The red circle in each plot marks the combination of cuts that results in the highest significance.

muon channel is subjected to a larger amount of continuum background, as could be seen in Figure 4.5. The selections shown in Table 8.1 result in background rejections of 15.2 % and 36.7 % in the m^2_{miss} signal region while retaining 94.2 % and 88.7 % of signal events in the electron and muon channel, respectively.

Table 8.1: Selections on the continuum suppression BDT classifier and FEI signal probability in the FEI version from a 2D significance optimization.

	$B^0 \to \pi^- e^+ \nu_e$	$B^0 \to \pi^- \mu^+ \nu_\mu$
BDT classifier	> 0.66	> 0.88
log(FEI signal probability)	> 8.25	> 7.75

8.3 Signal Efficiencies

After applying the selections described in the previous section, I combine the events from the electron and the muon channel. The total signal efficiencies for $B^0 \to \pi^- \ell^+ \nu_\ell$ after all selections are depicted in Figure 8.4 for the FR and FEI analysis versions, as well as the ones obtained by Sibidanov et al.. The corresponding values can be found in Table A.1. We can see that the efficiencies obtained by Sibidanov et al. are around 2×10^{-3} for all 13 q^2 bins. In the FR version, where I apply the same selections as Sibidanov et al., these values are around 3×10^{-3} and therefore about 50 % higher. This increase in efficiency presumably stems from the higher efficiency achieved by the FEI in comparison to the FR algorithm. In the FEI version, where I try to further optimize the selections in regards of efficiency and significance, the efficiencies in all q^2 bins are around 3.5×10^{-3}. This further increase with respect to the FR version probably comes from looser selections in this version. In both the FR and FEI version the efficiencies in the lowest q^2 bins are slightly lower than for the rest. This probably comes from bin migration, which was explained in Section 4.4. The same effect could also cause the higher efficiency in the last q^2 bin for the FEI version. Also, this highest q^2 bin contains the lowest number of events and therefore the uncertainty on this efficiency should be the largest.

The distributions of m^2_{miss} after the final selections can be seen in Figure 8.3. For the FR and FEI versions I uncover experimental data in the sideband of $m^2_{\text{miss}} > 1\,\text{GeV}^2$. We can see that there is good agreement between data and MC, which means that the corrections introduced in Section 4.2 appear to work. The same distributions split into bins of q^2 can be seen in the appendix in Figure A.10 and Figure A.10. The signal yields over the whole q^2 range, which are proportional to the total signal efficiency, can be found in Table 8.2.

8.4 Asimov Fit Results

By performing Asimov fits to the m^2_{miss} distribution, as described above, I can get an estimate for the signal significance. The significance obtained this way for the individual q^2 bins can be seen in Figure 8.4. The values are shown in the appendix in Table A.1. The efficiencies labeled with “Sibidanov et al.” were obtained from fits to data. In general we can see that the significances in all

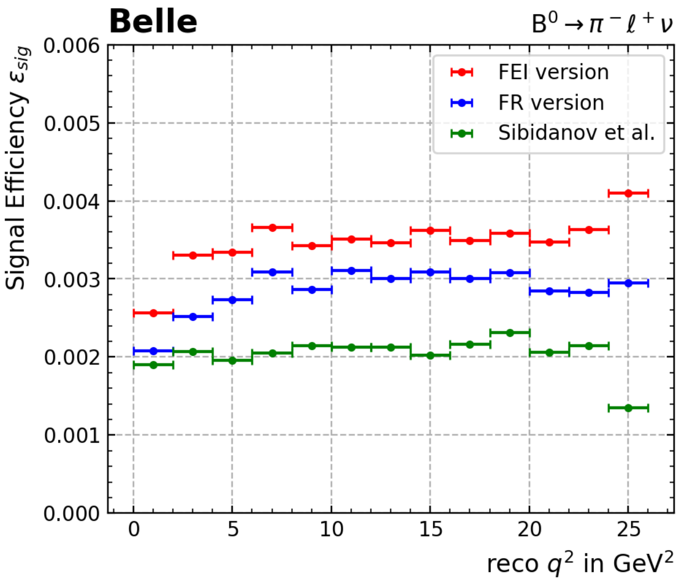


Figure 8.2: Signal efficiencies after all selections from the FR and FEI versions and the ones obtained by Sibidanov et al..

Table 8.2: Significance and signal yield over the full q^2 range for the FR and FEI version and the values obtained by Sibidanov et al..

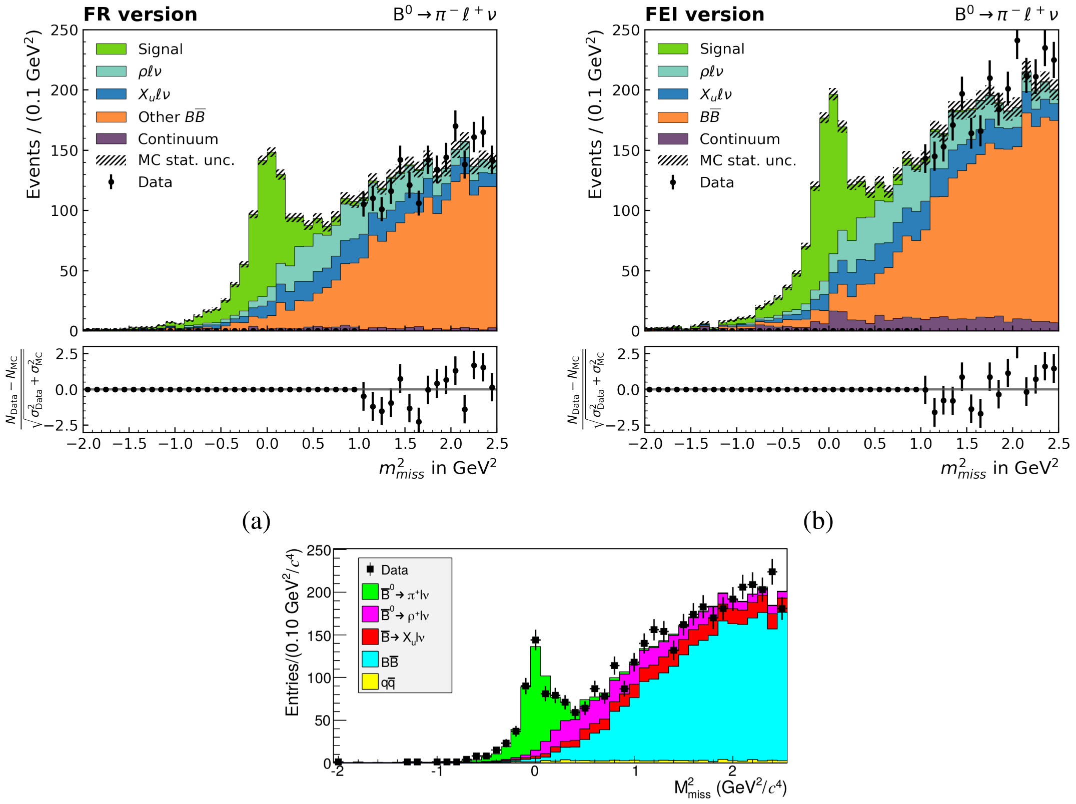
	Signal Yield	Significance
FEI version	754 ± 44	17.2
FR version	626 ± 39	16.0
Sibidanov et al.	463 ± 28	16.7

three cases are of the same order. The downward trend for the higher q^2 bins comes from the lower number of signal events inhabiting this region, as could be seen in Section 4.4 in Figure 4.4. At the same time one could expect the significance to be higher for the FEI and FR version because they result in higher signal efficiencies and the significance scales with the square root of the signal yield. Therefore, in these two versions the estimated standard deviation on the signal yield must be larger. That this is the case can be seen in Table 8.2.

Reasons for this could be the larger amount of $B^0 \to \rho^- \ell^+ \nu_\ell$ and $B^0 \to X_u \ell^+ \nu_\ell$ cross-feed in the low m^2_{miss} region, as could be seen in Figure 8.3.

8.5 Missing Mass Squared Resolution

One reason for the lack of significance improvement expected by the gain in signal efficiency in the FEI version could be due to a worse resolution of the signal component in the m^2_{miss} distribution. We can see in Figure 8.3 that in the m^2_{miss} distribution obtained by Sibidanov et al. there is a sharp peak of



(a)

(b)

(c)

Figure 8.3: Distributions of m^2_{miss} (a) for the FR version, (b) for the FEI version and (c) obtained by Sibidanov et al. [2] after applying all selections.

the signal at 0 and most signal events are contained within the region of $|m^2_{\text{miss}}| < 0.5\,\text{GeV}^2$. Also, almost no background events lie below $0\,\text{GeV}^2$. In both the FEI and FR version the signal component seems to be spread over a wider range of m^2_{miss} with tails to values higher than $0.5\,\text{GeV}^2$. In addition, there are more background events at $m^2_{\text{miss}} = 0$, resulting in the signal looking more background-like.

8.5.1 Significance vs. Resolution

To test the hypothesis that the m^2_{miss} resolution of signal events has an impact on the signal significance I perform a toy study by replacing the signal MC component with Gaussians of different widths σ, centered around $m^2_{\text{miss}} = 0\,\text{GeV}^2$. The integral of the Gaussians is fixed to the signal yield of 754 events from the FEI version. Illustrations of this are shown in Figure 8.5 for two different widths. The signal significances obtained from Asimov fits to these distributions can be seen in Figure 8.6. We can see that for an unchanged signal yield the significance increases for decreasing width of the signal peak. Therefore, in order to achieve a higher significance at the same signal yield one would need a better resolution in m^2_{miss}. The horizontal blue line in Figure 8.6 corresponds to the significance obtained by an Asimov fit to the m^2_{miss} distribution in the FEI version, as it was shown in Figure 8.3 (b). The vertical blue line corresponds to the root-mean-squared (RMS) of the signal component

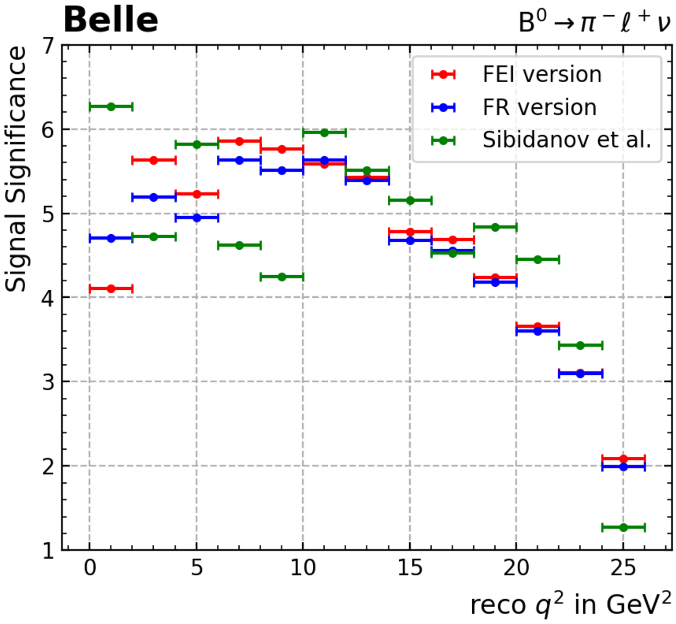
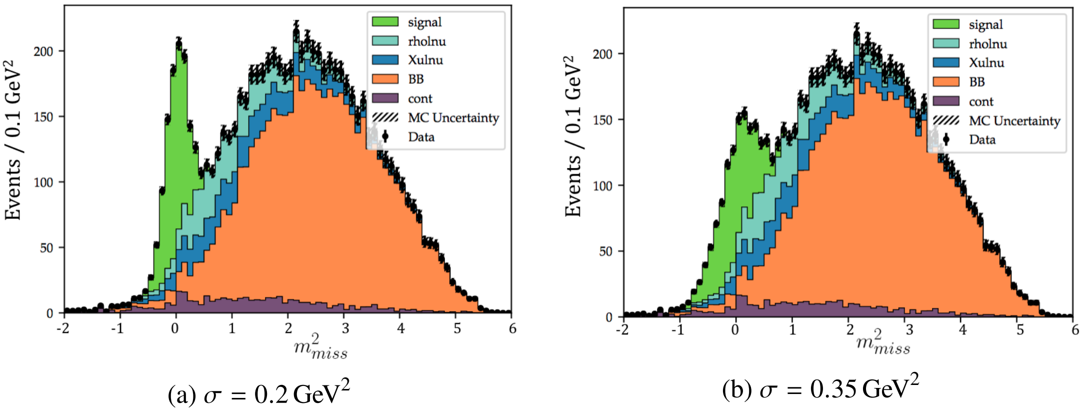


Figure 8.4: Signal significance in q^2 bins after all selections for the FR and FEI versions and the ones obtained by Sibidanov et al. [2].



(a) $\sigma = 0.2\,\mathrm{GeV}^2$

(b) $\sigma = 0.35\,\mathrm{GeV}^2$

Figure 8.5: Distributions of m^2_{miss} where the signal component is replaced by a Gaussian of width (a) $0.2\,\mathrm{GeV}^2$ and (b) $0.35\,\mathrm{GeV}^2$.

m^2_{miss}. The two lines approximately cross on the curve formed by the red dots, which means that the modeling of the signal component by Gaussians appears valid.

In the toy study I only change the width of the signal component, whereas a general improvement of the m^2_{miss} resolution would not effect signal events but also background events. This would then lead to a decrease of $B \to \rho\ell\nu$ and $B \to X_u\ell\nu$ cross-feed in the low m^2_{miss} region and better separation between signal and background which could in turn further increase the significance.

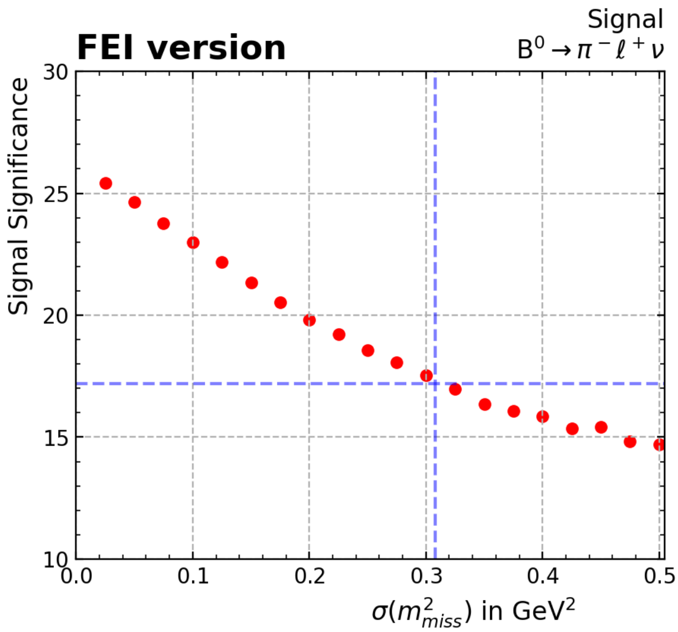


Figure 8.6: Signal significance for different widths of the signal component. The blue lines correspond to the significance obtained in the FEI version and the RMS of the signal component in this version.

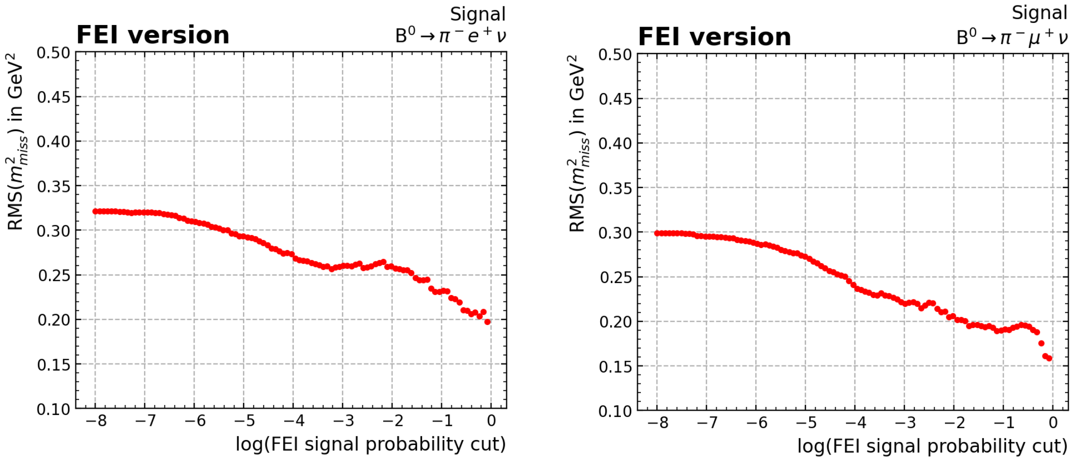
8.5.2 FEI Signal Probability

The worse signal resolution in the FEI version could be due to the loose selection on the FEI signal probability on the B_{tag} candidates. I require B_{tag} candidates to have values larger than around 10^{-8} in the electron and muon channel, respectively. This loose selection was suggested by the significance optimization described in Section 8.2. The effect of a tighter FEI signal probability selection on the resolution of m^2_{miss} for signal events can be seen in Figure 8.7. Here the resolution is measured again in form of the RMS of the signal component. We can see that for both the electron and muon channel the RMS decreases with tightening selections on the FEI signal probability, which corresponds to an increase in resolution. The calculation of m^2_{miss} in an event relies on the reconstructed four-momentum of the B_{tag} candidate. Cutting loosely on the FEI signal probability allows for B_{tag} candidates of poor quality to pass the selections. This results in a poor reconstruction of m^2_{miss}.

This could explain why the significance is not much better when using the FEI instead of the FR for tagging. Also, this has implications for the performance of the FEI in other analyses that also rely on the kinematic resolution on the tag-side, where we could see similar effects.

8.5.3 Tag Mode ID

The same effect of a decrease in m^2_{miss} resolution caused by poor reconstructions on the tag-side can be observed by looking at the resolution for different decay channels reconstructed on the tag-side. Each B decay channel that is reconstructed on the tag-side is assigned a *decay mode ID*. This decay mode ID only applies to the decay channel the B_{tag} candidate was reconstructed as, without taking into account the channels from which the daughter particles of the B_{tag} were reconstructed. The FR used



(a) (b)

Figure 8.7: Resolution of the signal component in the FEI version for different cuts on the FEI signal probability for the (a) electron channel and (b) muon channel.

15 different decay channels for reconstructing B^0_{tag} candidates. The FEI uses the same 15 channels plus an additional 10 modes, exclusive to the FEI. The specific channels that are reconstructed by the FR and FEI are shown in Table 8.3. In Figure 8.8 and Figure 8.9 the resolution of m^2_{miss} in the FEI analysis version is shown for the different decay modes reconstructed on the tag-side. The blue circles correspond to the decay modes reconstructed by the and FEI and the FR, and the red circles to the ones exclusively used by the FEI. The area of each circle is proportional to the number of correctly reconstructed signal events where the corresponding decay channel was reconstructed on the tag-side. We can see that the modes reconstructed by the FR generally deliver a better m^2_{miss} resolution than the ones that are only reconstructed by the FEI. The m^2_{miss} resolution from all FR channels combined is represented by the blue horizontal line, the resolution from the FEI exclusive channels by the red horizontal line.

The lower m^2_{miss} resolution from the FEI channels could come from the fact that the FEI includes more modes with one or more π^0. As explained in Section 4.1, π^0 are reconstructed from two clusters in the ECL identified as originating from photons. Light leakage leads to a poor energy resolution, which in return results in a poor reconstruction of the four-momentum of B_{tag} candidates. The poor kinematic resolution on the tag-side leads to the same effect described in the previous section, where B_{tag} candidates of poor quality result in a decrease in m^2_{miss} resolution.

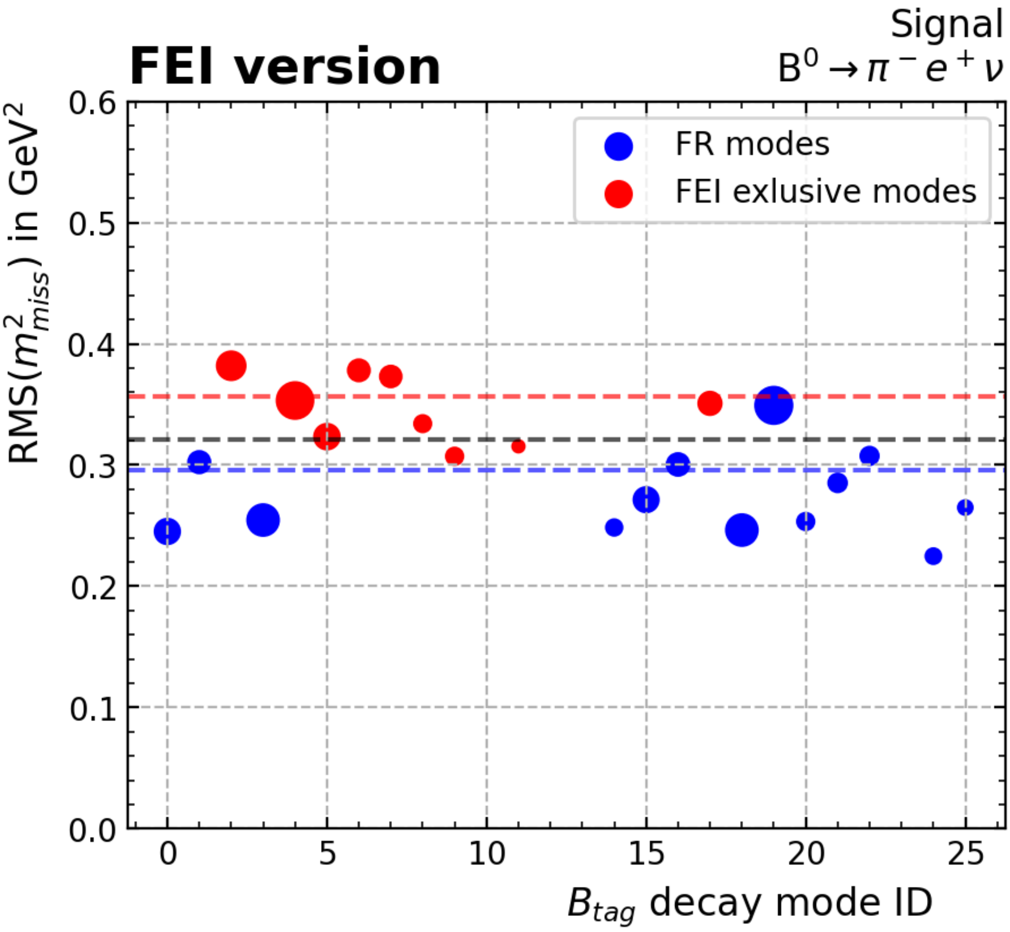


Figure 8.8: Signal component m^2_{miss} resolution in the FEI analysis version (electron channel) for different decay modes reconstructed on the tag-side. Blue circles correspond to decay modes reconstructed by both the FR and FEI, red circles to the ones exclusively reconstructed by the FEI. The area of each circle is proportional to the number of correctly reconstructed signal events with the corresponding decay channel reconstructed on the tag-side.

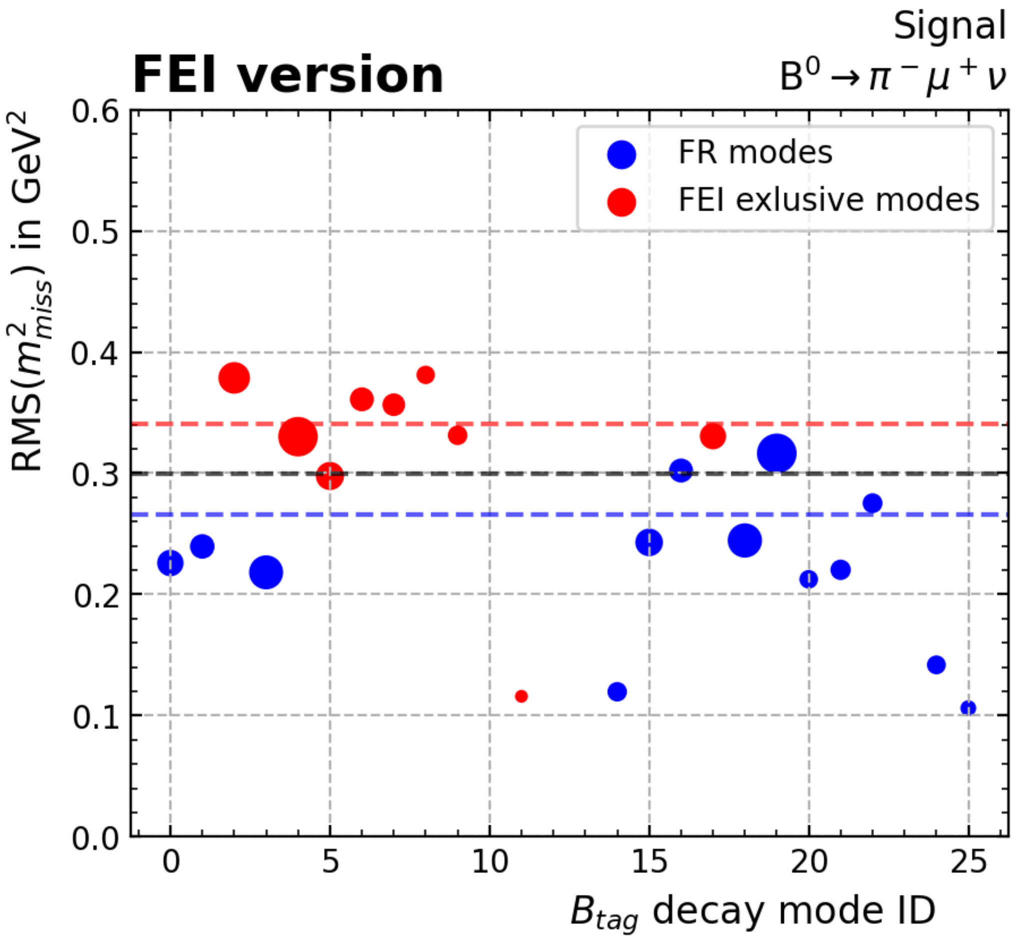


Figure 8.9: Signal component m^2_{miss} resolution in the FEI analysis version (muon channel) for different decay modes reconstructed on the tag-side. Blue circles correspond to decay modes reconstructed by both the FR and FEI, red circles to the ones exclusively reconstructed by the FEI. The area of each circle is proportional to the number of correctly reconstructed signal events with the corresponding decay channel reconstructed on the tag-side.

Table 8.3: Decay mode IDs of B^0 decay channels reconstructed on the tag-side by the FR and FEI. Blue modes are used by both the FR and FEI, red modes are only used by the FEI. One mode is reconstructed by the FR but is not reconstructed by the FEI. This mode is colored in black.

		Reconstructed by	
Decay Mode ID	Reconstructed B^0_{tag} Decay	FR	FEI
0	$D^-\pi^+$	✓	✓
1	$D^-\pi^+\pi^0$	✓	✓
2	$D^-\pi^+\pi^0\pi^0$		✓
3	$D^-\pi^+\pi^+\pi^-$	✓	✓
4	$D^-\pi^+\pi^+\pi^-\pi^0$		✓
5	$\overline{D}^0\pi^+\pi^-$		✓
–	$\overline{D}^0\pi^0$	✓	
6	$D^-D^0K^+$		✓
7	$D^-D^{*0}K^+$		✓
8	$D^{*-}D^0K^+$		✓
9	$D^{*-}D^{*0}K^+$		✓
10	$D^-D^+K^0_S$		✓
11	$D^{*-}D^+K^0_S$		✓
12	$D^-D^{*+}K^0_S$		✓
13	$D^{*-}D^{*+}K^0_S$		✓
14	$D^+_sD^-$	✓	✓
15	$D^{*-}\pi^+$	✓	✓
16	$D^{*-}\pi^+\pi^0$	✓	✓
17	$D^{*-}\pi^+\pi^0\pi^0$		✓
18	$D^{*-}\pi^+\pi^+\pi^-$	✓	✓
19	$D^{*-}\pi^+\pi^+\pi^-\pi^0$	✓	✓
20	$D^{*+}_sD^-$	✓	✓
21	$D^+_sD^{*-}$	✓	✓
22	$D^{*+}_sD^{*-}$	✓	✓
23	$J/\Psi K^0_S$	✓	✓
24	$J/\Psi K^+\pi^-$	✓	✓
25	$J/\Psi K^0_S\pi^+\pi^-$	✓	✓

CHAPTER 9

Conclusion and Outlook

Using the FEI algorithm with simulated data for hadronic tagging in the reconstruction of $B^0 \longrightarrow \pi^- \ell^+ \nu_\ell$ results in a signal yield of 626 ± 39 when applying the same selections as Sibidanov et al., who used the FR algorithm and obtained a signal yield of 463 ± 28. However, this increase in efficiency does not result in an increase of significance when performing a fit to the m^2_{miss} distribution. Sibidanov et al. obtained a significance of 16.7 from a fit to experimental data, while an Asimov fit to simulated data reconstructed with the FEI yields a significance of 16.0. By relaxing the event selection I was able to further increase the signal efficiency and obtain a signal yield of 754 ± 44 events. This also results in slight gain in significance of 17.2.

That the gain in efficiency by using the FEI does not result in a higher significance can be attributed to the worse resolution in m^2_{miss} by the FEI compared to the FR. Additional decay channels that are reconstructed on the tag-side by the FEI are responsible for the gain in efficiency. Some of these new channels involve the reconstruction of π^0 and lead to a poor resolution in the reconstruction of the B_{tag} four-momentum. This again leads to a poor resolution of m^2_{miss}.

In order to still benefit from these lower quality B_{tag} modes one could individually optimize the FEI signal probability selection for each decay channel reconstructed on the tag-side. One could choose a loose selection in high quality channels and therefore achieve a good m^2_{miss} resolution, while retaining a high signal efficiency. In low quality channels one could choose a tighter selection on the FEI signal probability to still use these additional channels, which the FR does not reconstruct, while only keeping the candidates with a high kinematic resolution on the tag-side.

Bibliography

[1] C. Bouchard, L. Cao, and P. Owen,
Summary of the 2018 CKM working group on semileptonic and leptonic b-hadron decays,
2019, arXiv: `1902.09412 [hep-ex]` (cit. on p. 1).

[2] A. Sibidanov et al., *Study of Exclusive $B- > X_u l$ nu Decays and Extraction of $|V_{ub}|$ using Full Reconstruction Tagging at the Belle Experiment*, Physical Review D **88** (2013)
(cit. on pp. 1, 21, 22, 24, 31, 49, 53, 54).

[3] T. Keck et al.,
The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment,
Computing and Software for Big Science **3** (2019) (cit. on pp. 1, 14).

[4] F. Halzen and A. D. Martin,
Quark & Leptons: An Introductory Course In Modern Particle Physics,
John Wiley & Sons, 2008 (cit. on p. 3).

[5] C. B. David Galbraith, *Standard model of particle physics*, CERN Webfest (2012) (cit. on p. 4).

[6] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001,
URL: `https://link.aps.org/doi/10.1103/PhysRevD.98.030001` (cit. on pp. 5, 17).

[7] 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 p. 5).

[8] E. Kou et al., *The Belle II physics book*,
Progress of Theoretical and Experimental Physics **2019** (2019) 123C01 (cit. on p. 6).

[9] PDG, *89. Semileptonic b-Hadron Decays, Determination of Vcb, Vub*, 2019 (cit. on p. 6).

[10] T. Keck,
Machine learning algorithms for the Belle II experiment and their validation on Belle data,
PhD thesis: KIT-Bibliothek, 2017 (cit. on p. 7).

[11] D. Ignatov and A. Ignatov, “Decision stream: Cultivating deep decision trees,”
2017 ieee 29th international conference on tools with artificial intelligence (ictai), IEEE, 2017
905 (cit. on p. 8).

[12] D. Jacobi, *Investigation of B->mu nu with inclusive tagging at Belle II*,
MA thesis: Universitaet Bonn, 2021 (cit. on p. 8).

[13] A. Abashian et al., *The Belle detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **479** (2002) 117, Detectors for Asymmetric B-factories, ISSN: 0168-9002, URL: `https://www.sciencedirect.com/science/article/pii/S0168900201020137` (cit. on pp. 9, 10).

[14] A. Bevan et al., *The physics of the B factories*, The European Physical Journal C **74** (2014) 1 (cit. on pp. 10, 30).

[15] A. Abashian et al., *The KL/μ detector subsystem for the BELLE experiment at the KEK B-factory*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **449** (2000) 112, ISSN: 0168-9002, URL: `https://www.sciencedirect.com/science/article/pii/S0168900299013832` (cit. on p. 11).

[16] W. Sutcliffe, *Early results from Full Event Interpretation at Belle II*, Proceedings of Science (2020) (cit. on p. 14).

[17] M. Feindt et al., *A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654** (2011) 432, ISSN: 0168-9002, URL: `https://www.sciencedirect.com/science/article/pii/S0168900211011193` (cit. on p. 14).

[18] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002, URL: `https://www.sciencedirect.com/science/article/pii/S0168900201000894` (cit. on p. 15).

[19] R. Brun et al., *GEANT: detector description and simulation tool*, tech. rep., CERN, 1993 (cit. on p. 15).

[20] C. Ramirez, J. F. Donoghue, and G. Burdman, *Semileptonic $b \to u$ decay*, Physical Review D **41** (1990) 1496 (cit. on p. 16).

[21] M. T. Prim et al., *Search for $B^+ \to \mu^+\nu_\mu$ and $B^+ \to \mu^+N$ with inclusive tagging*, Phys. Rev. D **101** (3 2020) 032007, URL: `https://link.aps.org/doi/10.1103/PhysRevD.101.032007` (cit. on p. 16).

[22] F. D. Fazio and M. Neubert, *B→Xulbar nul decay distributions to order αs*, Journal of High Energy Physics **1999** (1999) 017, URL: `https://doi.org/10.1088/1126-6708/1999/06/017` (cit. on p. 16).

[23] L. Hinz, *Lepton ID efficiency correction and systematic error*, Belle Note 954 (2006) (cit. on p. 17).

[24] J. Schwab, *Calibration of the Full Event Interpretation for the Belle and the Belle II Experiment*, MA thesis: Karlsruhe Institute of Technology (KIT), 2017 (cit. on p. 19).

[25] F. Tenchini, *PoS (Vertex 2016) 059 Vertex Fitting in the Belle II Analysis Framework*, (2016) (cit. on p. 20).

[26] D. Weyland, "Continuum Suppression with Deep Learning techniques for the Belle II Experiment," 2017 (cit. on p. 29).

[27] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (23 1978) 1581 (cit. on p. 30).

[28] D. M. Asner et al., *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53** (3 1996) 1039, URL: `https://link.aps.org/doi/10.1103/PhysRevD.53.1039` (cit. on p. 31).

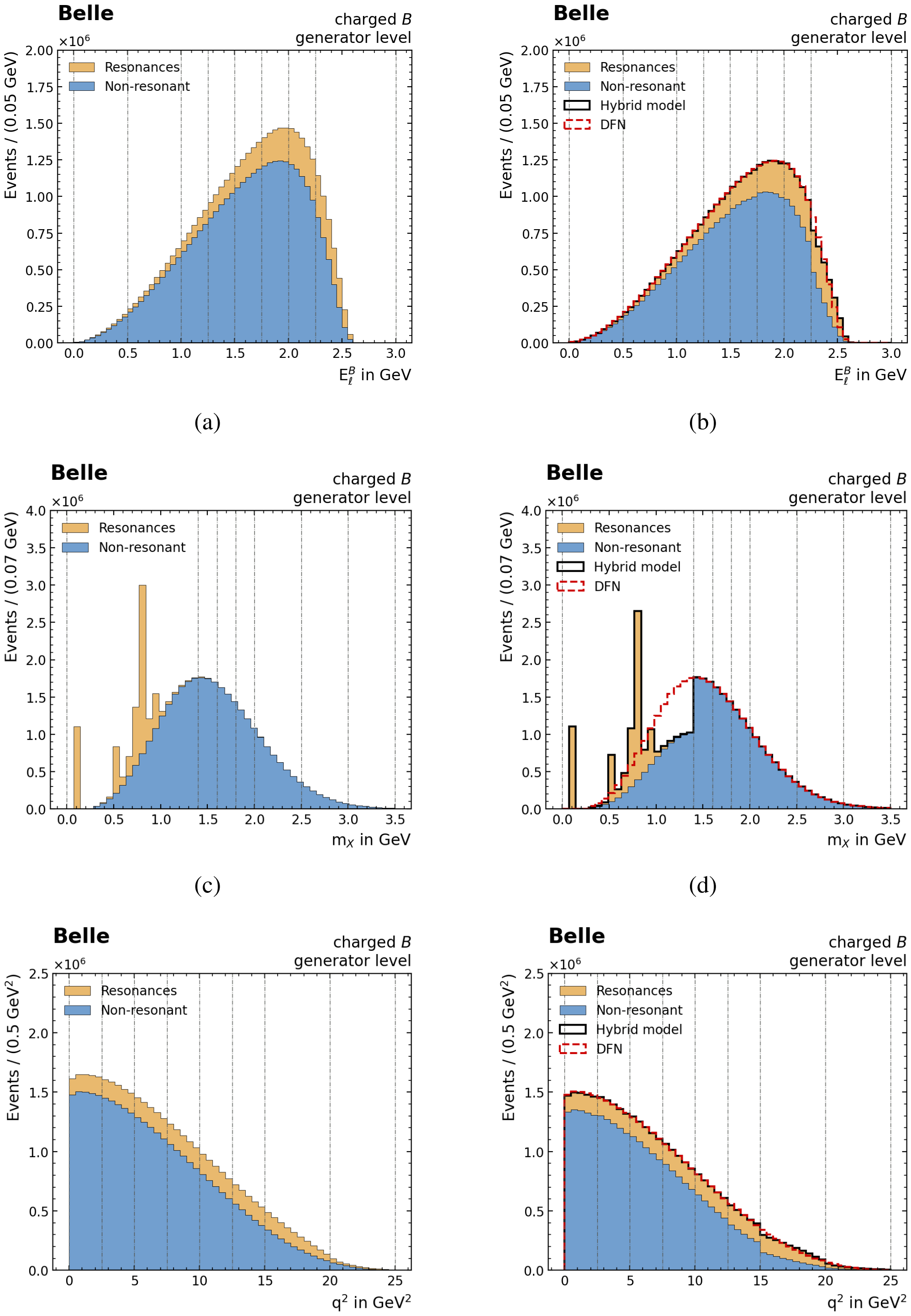
[29] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Computer Physics Communications **77** (1993) 219, ISSN: 0010-4655, URL: `https://www.sciencedirect.com/science/article/pii/001046559390005W` (cit. on p. 49).

[30] S. K. Granderath, *Untagged Exclusive Analysis of the Semileptonic Decay $B \to \pi\ell\nu$ from Belle II Data in Preparation for $|Vub|$ Extraction*, MA thesis: Universitaet Bonn, 2019 (cit. on p. 49).

APPENDIX A

Appendix

A.1 Reconstruction

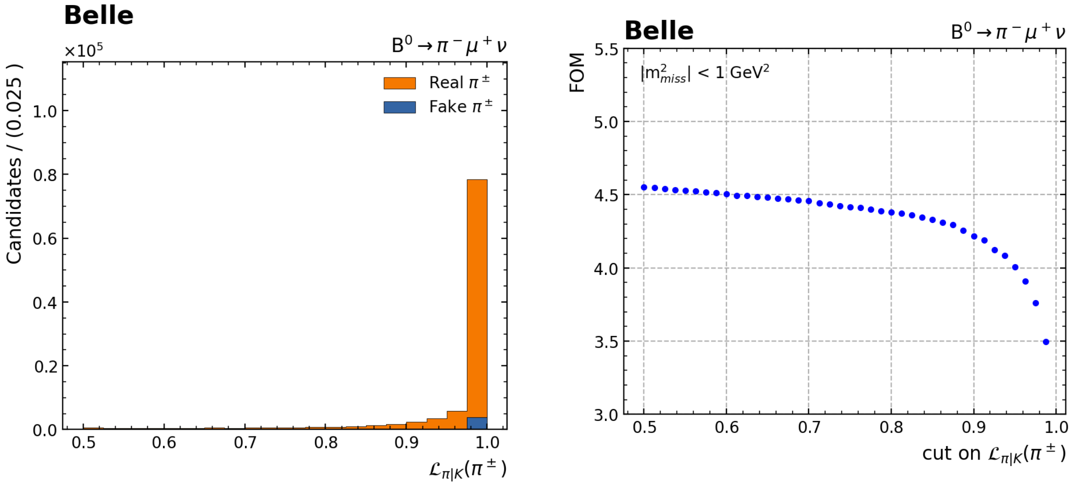


(a) (b)

(c) (d)

(e) (f)

Figure A.1: Generator level $B \to X_u \ell^+ \nu_\ell$ distributions of E_ℓ^B, m_X and q^2 before (left) and after (right) re-weighting for charged B mesons. The black line on the right shows the hybrid model, composed of resonant and non-resonant contributions.

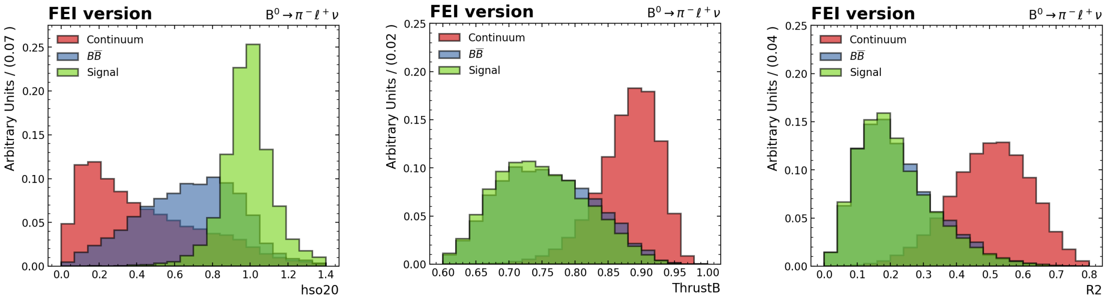


(a)

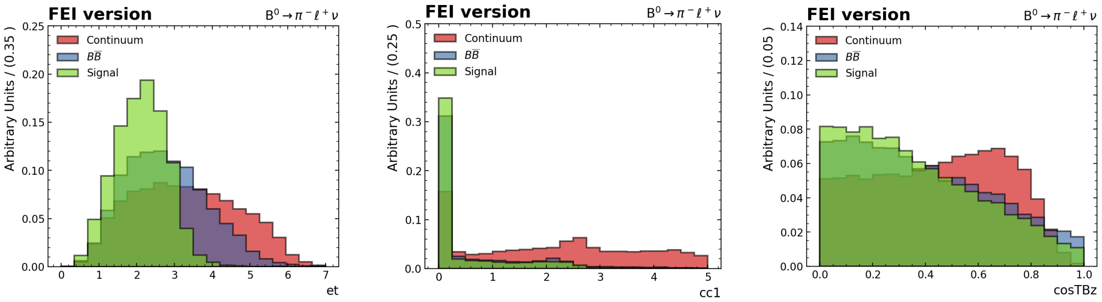
(b)

Figure A.2: (a) shows the distribution of $\mathcal{L}_{\pi|K}$ for reconstructed pions in $B^0 \longrightarrow \pi^- e^+ \nu_e$ split into real and fake pions. (b) shows the FOM for different cuts on $\mathcal{L}_{\pi|K}$. The FOM is calculated for events in the m^2_{miss} signal region.

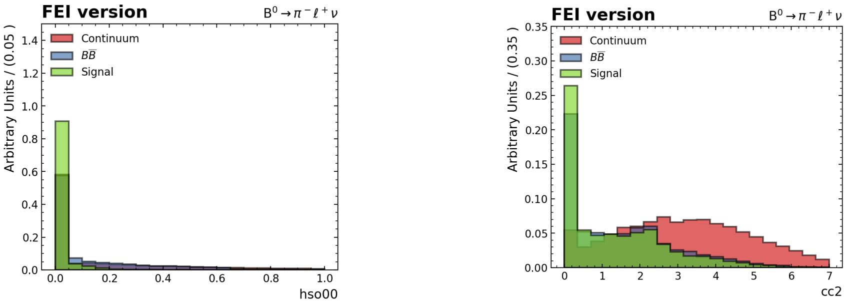
A.2 Continuum Suppression



(a) h_{20}^{so} (b) $|\vec{T}|$ (c) R_2



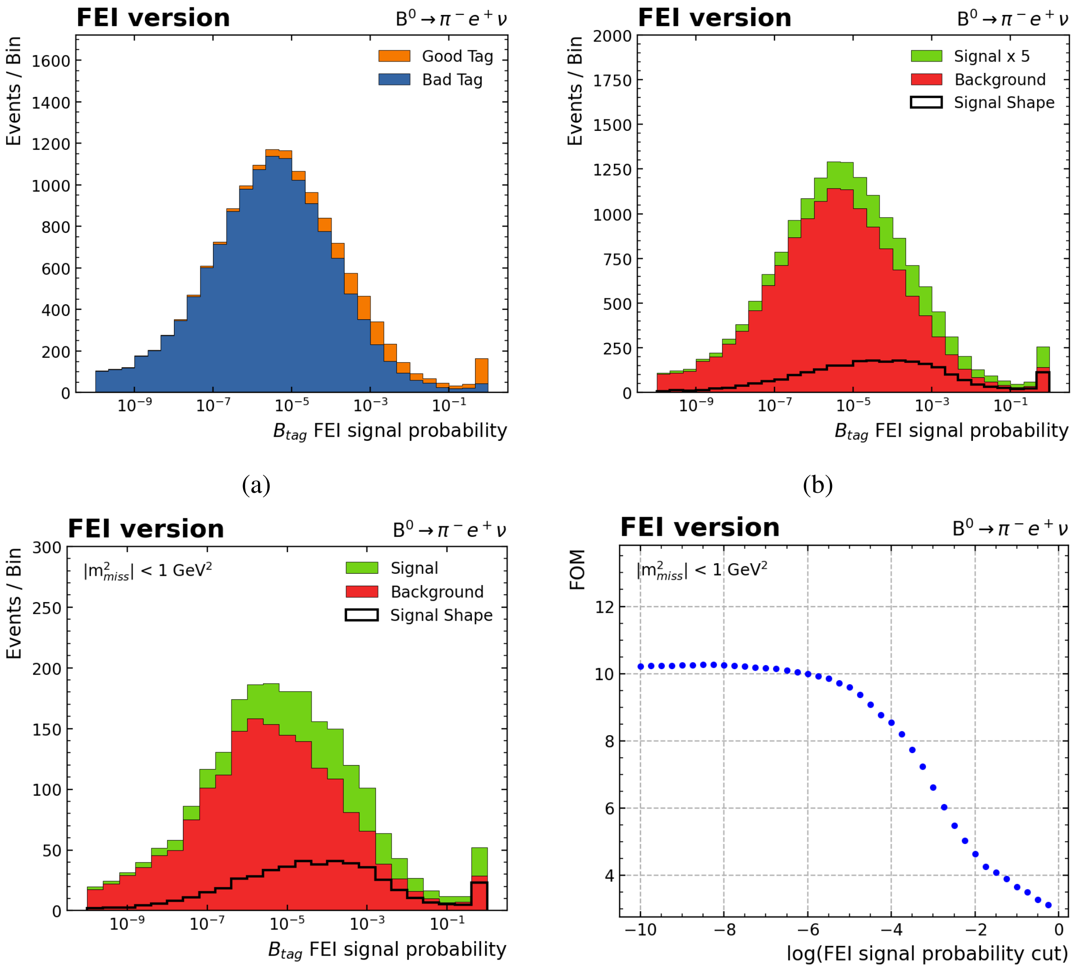
(d) E_t (e) $cc1$ (f) $\cos\theta_{\text{T,B}}$



(g) h_{00}^{so} (h) $cc2$

Figure A.3: Distributions used to train the continuum suppression BDT for $B^0 \to \pi^-\ell^+\nu_\ell$ in the FEI version.

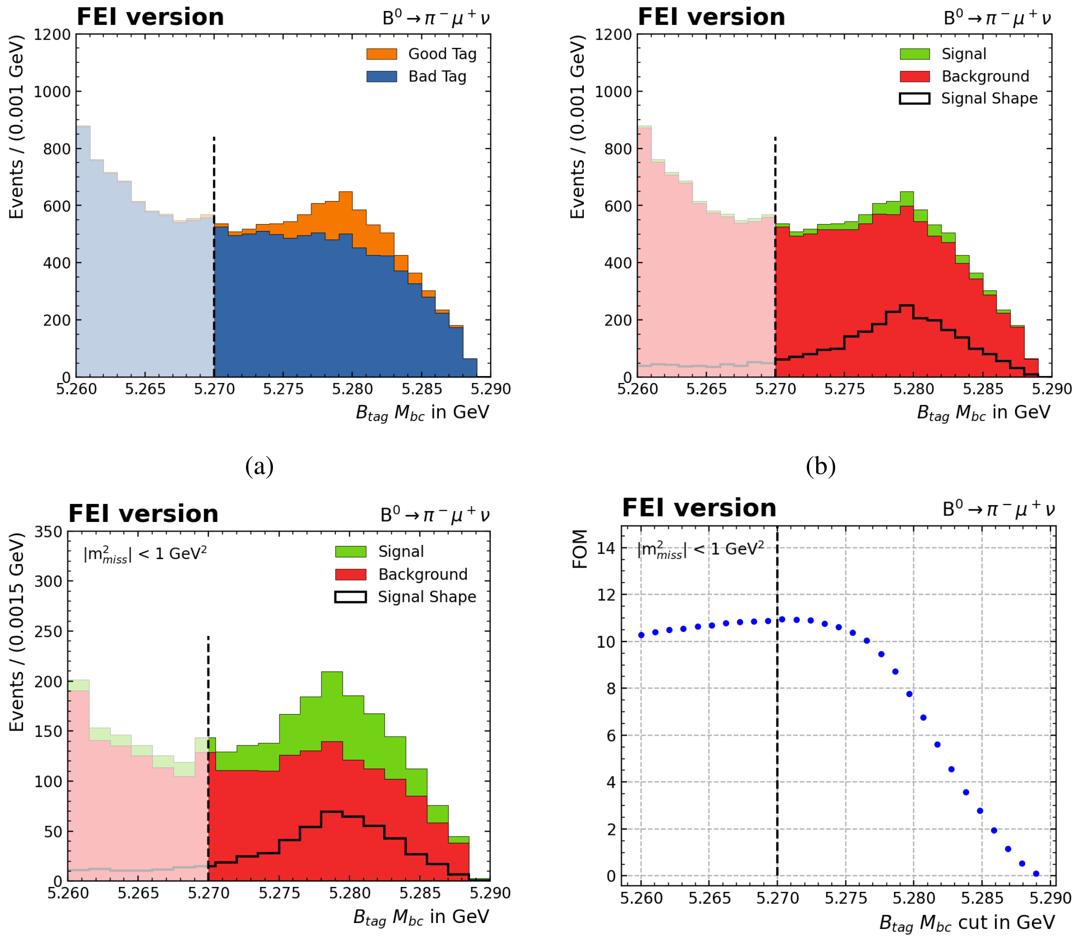
A.3 Tag Selection



(a) (b)

(c) (d)

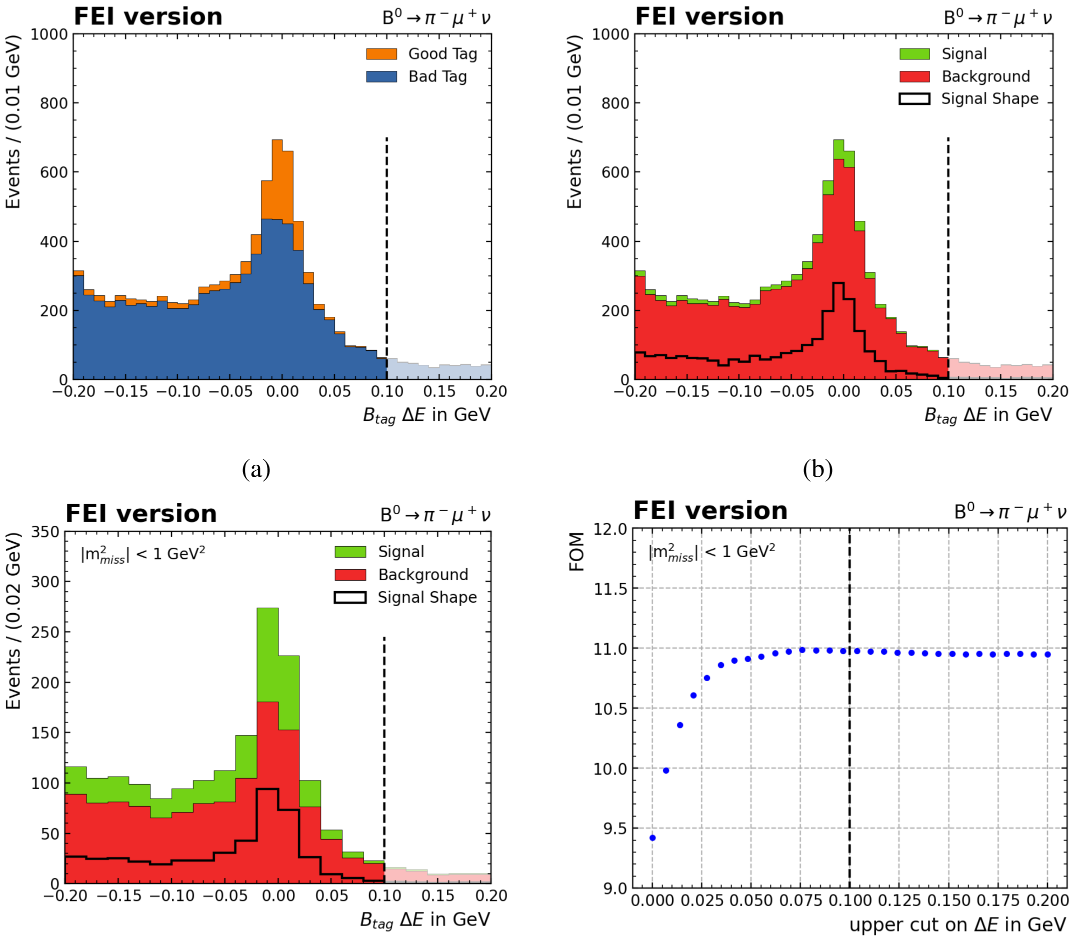
Figure A.4: Distribution of FEI signal probability for $B^0 \to \pi^- \mu^+ \nu_\mu$ after applying the continuum suppression. (a) shows the distribution split into good tags and bad tags. (b) shows the same distribution split into signal and background decays. The signal component is scaled up by a factor of 5. (c) Distribution of FEI signal probability for $B^0 \to \pi^- \mu^+ \nu_\mu$ in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region split into signal and background events. (d) FOM in the signal region for different cuts on FEI signal probability.



(a) (b)

(c) (d)

Figure A.5: (a) Distribution of B_{tag} M_{bc} for $B^0 \to \pi^- \mu^+ \nu_\mu$ split into good and bad tags. (b) The same distribution split into signal and background events. (c) M_{bc} distribution in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region split into signal and background events. (d) FOM in the signal region for different cuts on M_{bc}. The black dashed line indicates the cut I choose.

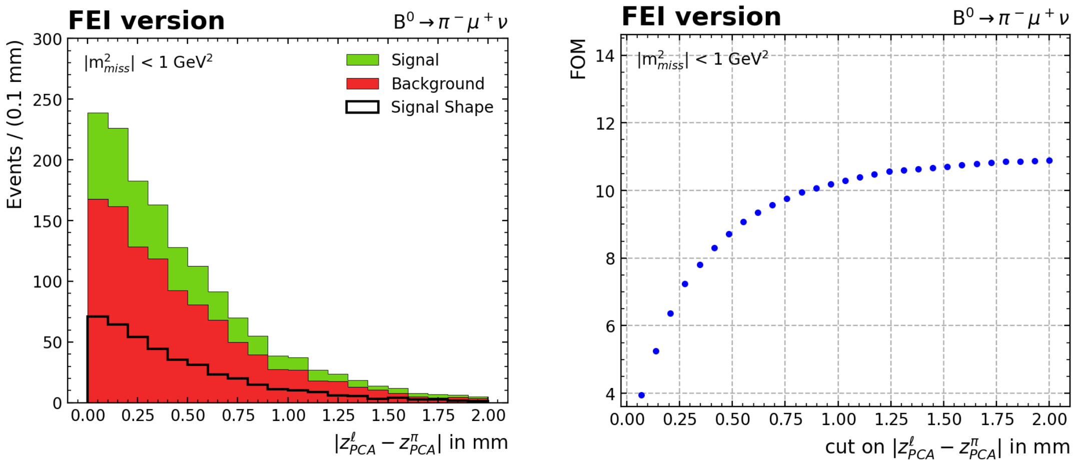


(a) (b)

(c) (d)

Figure A.6: (a) Distribution of B_{tag} ΔE for $B^0 \to \pi^- \mu^+ \nu_\mu$ split into good and bad tags. (b) The same distribution split into signal and background events. (c) M_{bc} distribution in the $|m^2_{\text{miss}}| < 1\,\text{GeV}^2$ signal region split into signal and background events. (d) FOM in the signal region for different cuts on ΔE. The black dashed line indicates the cut I choose.

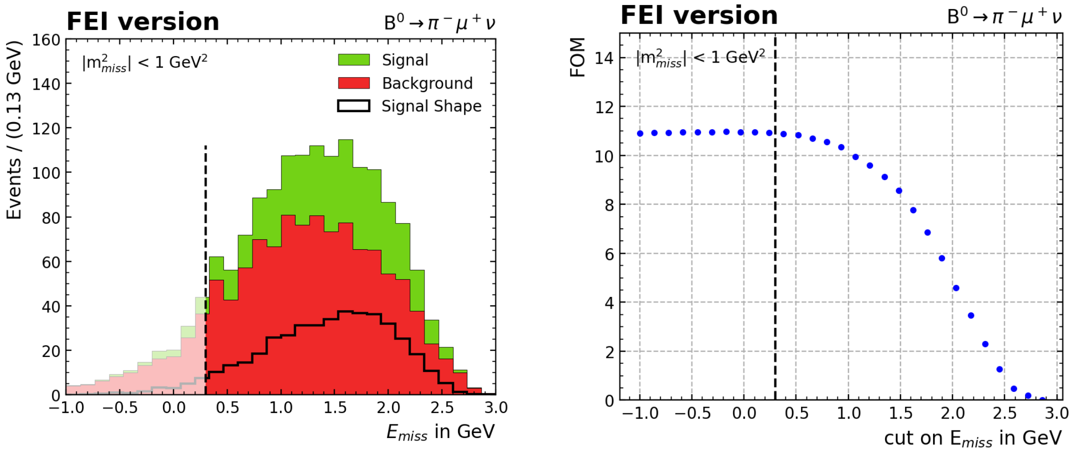
A.4 Signal Selection



(a)

(b)

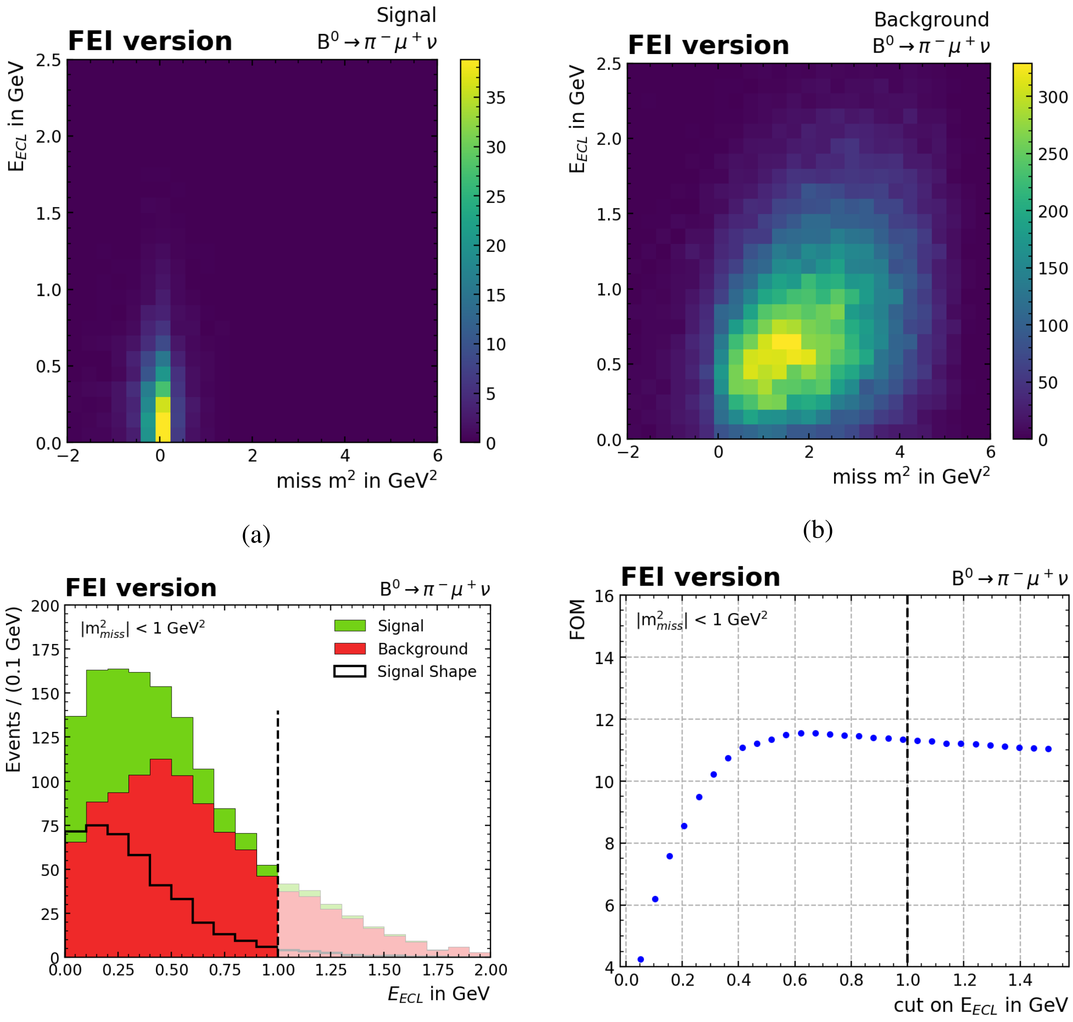
Figure A.7: (a) $|z^\ell_{\mathrm{PCA}} - z^\pi_{\mathrm{PCA}}|$ distribution in the $|m^2_{\mathrm{miss}}| < 1\,\mathrm{GeV}^2$ signal region for signal and background events. (b) FOM in the signal region for different cuts. Both are shown for the muon channel of the FEI version.



(a)

(b)

Figure A.8: (a) Distribution of E_{miss} for signal and background events in the m^2_{miss} signal region for the muon channel of the FEI version. (b) FOM for different lower cuts on E_{miss}.



(a) (b)

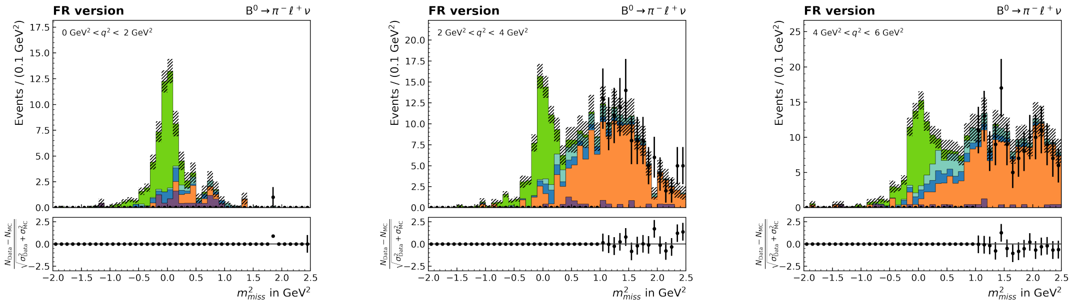
(c) (d)

Figure A.9: E_{ECL} vs. m^2_{miss} for (a) signal events and (b) background events. (c) Distribution of E_{ECL} for signal and background events in the m^2_{miss} signal region for the muon channel of the FEI version. (d) FOM for different lower cuts on E_{ECL}.

A.5 Signal Extraction

Table A.1: Signal efficiency and significance for the FR and FEI version and obtained by Sibidanov et al. in bins of q^2.

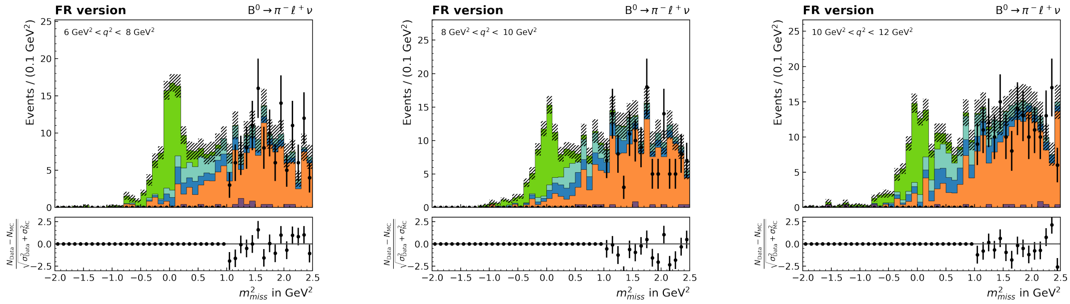
Δq^2	ϵ_{sig} in 10^{-3}			significance		
GeV^2	FEI	FR	Sib.	FEI	FR	Sib.
0 – 2	2.57	2.08	1.90	4.11	4.71	6.27
2 – 4	3.31	2.52	2.07	5.63	5.19	4.72
4 – 6	3.34	2.73	1.96	5.23	4.95	5.82
6 – 8	3.66	3.09	2.05	5.86	5.63	4.62
8 – 10	3.42	2.86	2.14	5.76	5.51	4.25
10 – 12	3.51	3.11	2.13	5.58	5.63	5.96
12 – 14	3.46	3.00	2.13	5.43	5.39	5.51
14 – 16	3.62	3.09	2.02	4.78	4.68	5.15
16 – 18	3.49	3.01	2.16	4.69	4.56	4.53
18 – 20	3.58	3.08	2.31	4.24	4.18	4.84
20 – 22	3.47	2.84	2.06	3.66	3.60	4.45
22 – 24	3.63	2.82	2.14	3.11	3.10	3.43
24 – 26	4.10	2.95	1.35	2.09	1.99	1.27



(a) $0\,\mathrm{GeV}^2 < q^2 < 2\,\mathrm{GeV}^2$

(b) $2\,\mathrm{GeV}^2 < q^2 < 4\,\mathrm{GeV}^2$

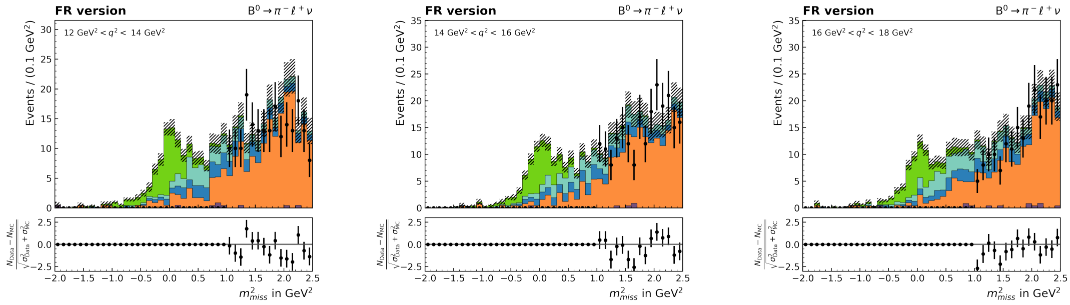
(c) $4\,\mathrm{GeV}^2 < q^2 < 6\,\mathrm{GeV}^2$



(d) $6\,\mathrm{GeV}^2 < q^2 < 8\,\mathrm{GeV}^2$

(e) $8\,\mathrm{GeV}^2 < q^2 < 10\,\mathrm{GeV}^2$

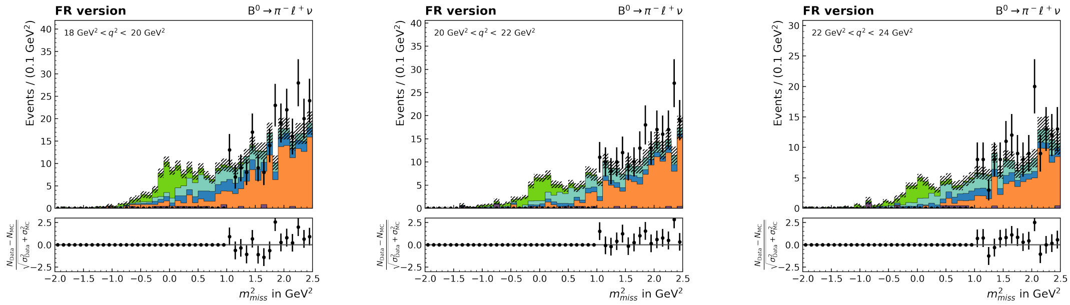
(f) $10\,\mathrm{GeV}^2 < q^2 < 12\,\mathrm{GeV}^2$



(g) $12\,\mathrm{GeV}^2 < q^2 < 14\,\mathrm{GeV}^2$

(h) $14\,\mathrm{GeV}^2 < q^2 < 16\,\mathrm{GeV}^2$

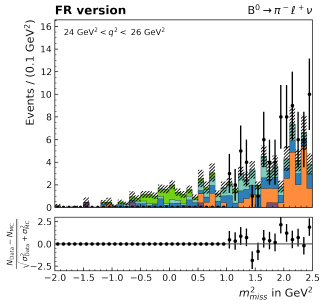
(i) $16\,\mathrm{GeV}^2 < q^2 < 18\,\mathrm{GeV}^2$



(j) $18\,\mathrm{GeV}^2 < q^2 < 20\,\mathrm{GeV}^2$

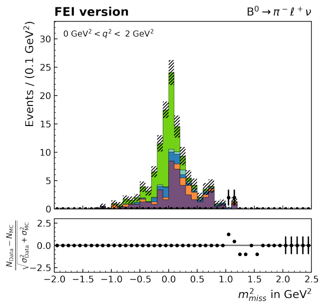
(k) $20\,\mathrm{GeV}^2 < q^2 < 22\,\mathrm{GeV}^2$

(l) $22\,\mathrm{GeV}^2 < q^2 < 24\,\mathrm{GeV}^2$

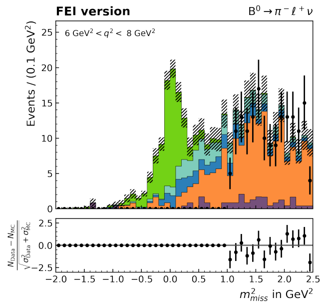


(m) $24\,\mathrm{GeV}^2 < q^2 < 26\,\mathrm{GeV}^2$

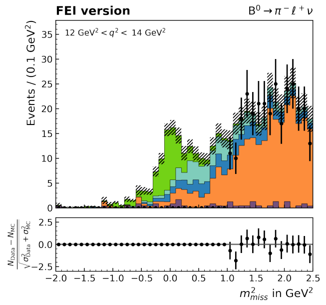
Figure A.10: Distribution of m^2_{miss} in bins of q^2 in the FR version after all selections.



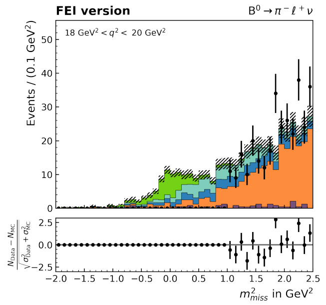
(a) 0 GeV^2 < q^2 < 2 GeV^2 (b) 2 GeV^2 < q^2 < 4 GeV^2 (c) 4 GeV^2 < q^2 < 6 GeV^2



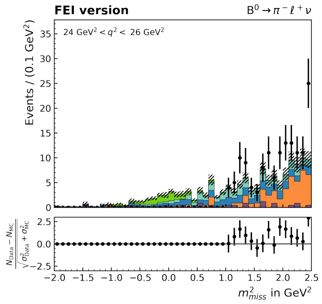
(d) 6 GeV^2 < q^2 < 8 GeV^2 (e) 8 GeV^2 < q^2 < 10 GeV^2 (f) 10 GeV^2 < q^2 < 12 GeV^2



(g) 12 GeV^2 < q^2 < 14 GeV^2 (h) 14 GeV^2 < q^2 < 16 GeV^2 (i) 16 GeV^2 < q^2 < 18 GeV^2



(j) 18 GeV^2 < q^2 < 20 GeV^2 (k) 20 GeV^2 < q^2 < 22 GeV^2 (l) 22 GeV^2 < q^2 < 24 GeV^2



(m) 24 GeV^2 < q^2 < 26 GeV^2

Figure A.11: Distribution of m^2_{miss} in bins of q^2 in the FEI version after all selections.

List of Figures

List of Tables