Tagged analysis of semileptonic B decays with uncharmed vector mesons at Belle

Rieka Rittsteiger

Masterarbeit in Physik
angefertigt im Physikalischen Institut

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

July 2022

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. Gutachterin: Prof. Dr. Florian Bernlochner

Acknowledgements

I would like to thank Jochen Dingfelder for giving me the opportunity to write my thisis within the Belle group and giving me advice when needed. It was a great experience working with this wonderful bunch of physicists. Furthermore I would like to thank Peter Lewis for his supervision and guidance throughout the past year, I learned a lot. I also thank Florian Bernlochner for refereeing this thesis and for his input. Additionally, I thank Svenja for her countless advice and proof reading multiple drafts of my thesis, as well as going on coffee runs together.

I would also like to tank Jonathan and Stephan for their explanations and discussions at the beginning of this thesis in the old office. And of course everyone in the new office. Daniel for his calm advice, Tobias for coding tips and Ralf for making sure that we have always enough coffee in the Banana! As well as Lucas and Armin my fellow suffering Master students. Furthermore I would like to thank the rest of the Belle group. There were always interesting topics to talk about, work related or not.

Outside of the world of Physics, I would like to thank my family, for keeping me sane and supporting me over all the years of my studies. As well as my friends, especially Jan, Mitra and Simon for bringing variety and other topics than physics to my life. Whether it is over a cup of coffee or a glass of wine the conversations we have give me another point of few on life. Finally, my thanks go out to Micha, thank you for keeping up with me while I am doubting myself, and being my rock in turbulent waters.

Contents

CHAPTER 1

Introduction

Few Physics experiments are mentioned in pop culture but there is an odd fascination with CERN, especially the Large Hadron Collider (LHC). It appears in books, movies and even mangas. So where does this fascination come from? If we look back in time even philosophers in antiquity were occupied with fundamental questions, in particular "What is the universe made of?" and "How did the universe start?". Even today we still cannot fully answer these questions. Particle physics attempts to answer these question with the help of accelerators. Therefore, particle accelerators spark interest in authors of books or creators of TV series, especially the LHC at CERN. It appeared frequently in the media after the discovery of the Higgs Boson, and is therefore the most well known accelerator.

During the 20th century the particle physics experiments gained in precision and led to formulation of the Standard Model of particle physics (SM), in the 1970s, which will be discussed further in chapter 2. The SM has been tested thoroughly by hundreds of experiments and it is describing particles and their interactions successfully. However, the SM is still being challenged by physical observations that it cannot explain. It is unclear for example, why there are three generations of particles, and parameters like the masses of particles and the elements of the weak interaction mixing Cabibbo-Kobayashi-Maskawa (CKM) matrix are not explained by the SM. These have to be determined experimentally. The $|V_{ub}|$ matrix element is one of those parameters which can be obtained via two different methods that both analyse uncharmed $B \to X_u \ell \nu$ decays. The inclusive method, which does not identify the exact final state, and the exclusive method that only looks at a specific X_u final state. Currently a discrepancy of $\approx 3.5\sigma$ in the value of $|V_{ub}|$ [1] is observed between these two methods.

In this thesis data from the Belle experiment at KEK in Tsukuba is used. The motivation for the Belle experiment was the search for CP-violation. The KEKB accelerator is a B-factory, which means that electrons collide with positrons at a center of mass energy of 10.58 GeV, corresponding to the mass of the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ then decays nearly always into a B meson pair. KEKB produced the largest luminosity in the world during its run time. In this thesis B meson decays into either $\rho \ell \nu$ or $\omega \ell \nu$ are studied, where $\ell = e, \mu$.

One method to analyse these types of decays is to reconstruct both B mesons, this is called tagging. A new tagging algorithm is implemented which was not used in a previously conducted Belle analysis on the same dataset by Sibidanov et al. [2]. There, the branching fraction obtained for the $B^+ \to \rho^0 \ell \nu$ channel was $(1.84 \pm 0.10) \times 10^{-4}$, which is high compared to other measurements, and one of the reasons this analysis is re-performed using the new algorithm. With the start of the Belle II experiment, the Full event interpretation (FEI) algorithm [3] has been introduced. It achieves a higher efficiency

than the preciously used algorithm. Using the new setup a peaking background below the signal peak in the signal extration variable is studied. This was not observed in the analysis by Sibidanov, since the simulated data was modelled differently previously. Distinguish between signal and the peaking background component, is one of the goals of this analysis. Potentially this will allow for a simultaneous extraction of the two components eventually. In addition the matrix element $|V_{ub}|$ can be extracted in the future.

The analysis are presented mainly for the $B^+ \to \rho^0 e^+ \nu$ channel. The charged conjugate decays are always implied. The other channels are shown in more detail in the appendix.

CHAPTER 2

Theory

This chapter provides an overview of the theoretical concepts needed in this analysis. First the Standard Model of particle physics is discussed, followed by the weak interaction and semileptonic B decays.

2.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) is a theory that attempts to incorporate everything we know about the fundamental particles and interactions the universe consists of. It includes three quark and lepton generations, that contain two particles each. Therefore there are 12 such particles in total that are called fermions, they have half-integer spin. The quarks are the up-quark (u), down-quark (d), charm-quark (c), strange-quark (s), top-quark (t) and bottom-quark (b). One of the quarks in a generation has charge $+\frac{2}{3}$ and the other quark has charge $-\frac{1}{3}$. Furthermore, quarks carry a colour charge that is either red, blue or green. Additionally, each generation has a lepton with charge -1 and a massless neutrino. The leptons (ℓ) are the electron (e), muon (μ) and tauon (τ) and the corresponding neutrinos (ν) are the electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ).

Stable matter consists only of quarks and fermions of the first generation, due to the second and third generation being much heavier and accordingly unstable.

Each of the fermions has an antiparticle that has the same mass but opposite electric charge, and for quarks also opposite colour charge.
Next to fermions, the SM also includes bosons, which are the gluon (g), photon (γ), Z boson (Z) and W boson ($W^\pm$), as well as the Higgs boson (H). The bosons have an integer number spin. In this way the SM does not only incorporate the fundamental particles, but it also describes the fundamental forces in the universe, with the g being the gauge boson for the strong force, the γ being the gauge boson for the electromagnetic force and the Z and $W^\pm$ being the gauge bosons for the weak force. A visualisation of the SM can be seen in figure 2.1.

The strong force is described by quantum chromodynamics (QCD) being a non-abelian gauge theory with SU(3) symmetry between the three quark families. Its gauge bosons are the g which can not only interact with quarks but also with each other. Additionally gluons have a colour charge. Furthermore the strong force shows confinement. That means that the interaction strength growth with distance rather than decays, like in electromagnetism. This is the mechanism that explains why quarks can not be observed as free particles but only within bound states, also called hadrons. When quarks are separated the binding energy becomes large enough that another quark-antiquark pair is

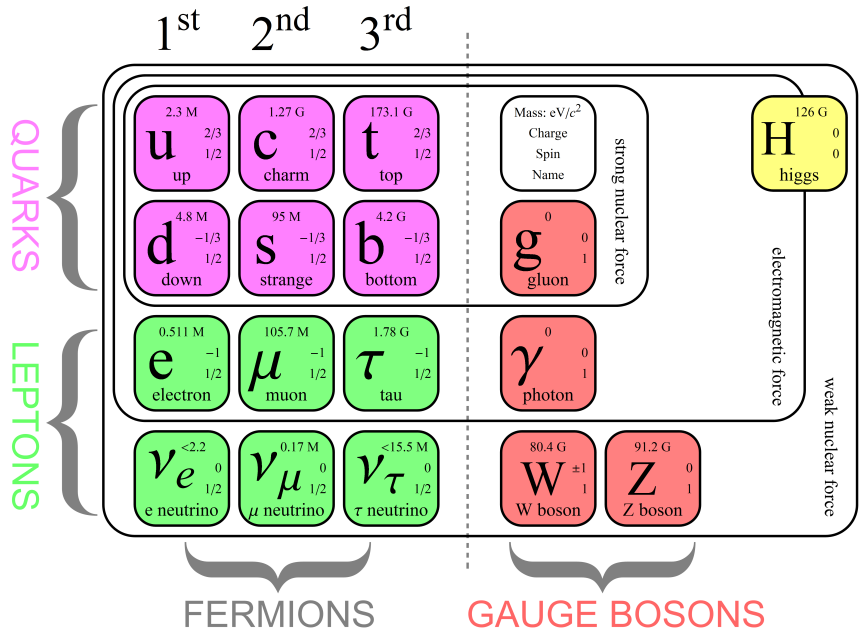


Figure 2.1: The SM of particle physics with 12 fermions, 5 bosons, 3 leptons and 3 quarks. They are each presented with their masses, charge and spin [4].

produced from vacuum [5].

All particles carrying electric charge interact via the electromagnetic interaction. This interaction is described by quantum electrodynamic (QED) which is a U(1) gauge theory, reflecting the fact of charge conservation.

The weak interaction is described by a SU(2) symmetry. The high mass of its gauge bosons are caused by spontaneous electroweak U(1)× SU(2) symmetry breaking involving the Higgs mechanism. Neutrinos carry neither colour nor electric charge therefore, they interact with other matter only via the weak interaction. More information about the weak interaction is given in chapter 2.1.1.

2.1.1 The weak interaction

In the weak interaction the fermions interact with each other by exchanging one of the gauge bosons $W^{\pm}$ or Z. Its gauge bosons have masses of $m_{W^{\pm}} \approx 80\,\text{GeV}$ and $m_{Z^0} \approx 91\,\text{GeV}$. The weak interaction is the only force that can violate parity and charge conservation in decays. Furthermore it is the only interaction that can change the flavour of quarks and leptons. This is the reason why the Cabbibo-Kobayashi-Maskawa (CKM) matrix was introduced. This matrix relates the mass eigenstates q to the weak interaction eigenstates q' [6]:

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

The elements describe the probability of a quark changing flavour from one flavour to another. The absolute value of the matrix elements and current world-averages of their magnitudes can be seen in

equation 2.2 [1].

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \qquad |V_{CKM}| = \begin{pmatrix} 0.974 & 0.224 & 0.004 \\ 0.221 & 0.997 & 0.042 \\ 0.008 & 0.039 & 1.019 \end{pmatrix} \tag{2.2}$$

Due to the diagonal matrix elements being close to unity, flavour changes within the same quark generation are favoured. Flavour changes between one generation are suppressed and flavour changes between two generations are doubly suppressed.

In this thesis, a doubly suppressed transition is studied, in this case the transituin of a b to a u quark. Studying these rare decays is difficult, due to their low branching fractions, but they are thus also a good probe for beyond the Standard Model (BSM) effects.

2.1.2 Semileptonic B decays

In this thesis, data from the Belle detector, which was located at the interaction point of the KEKB accelerator, is analysed. KEKB was a B-factory at which an electron-positron pair collides at the $\Upsilon(4S)$ resonance energy of 10.579 GeV. The $\Upsilon(4S)$ decays mainly into a B meson pairs. A Feynman-style diagram of this process is shown in figure 2.2.

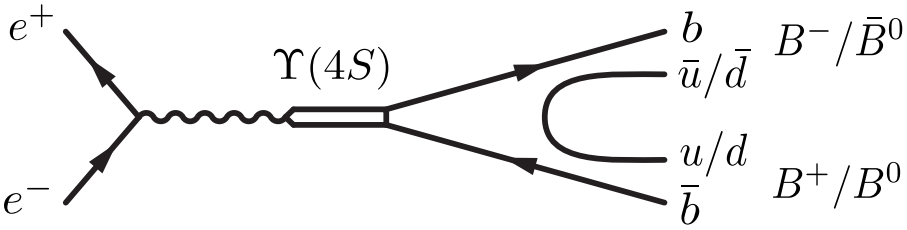


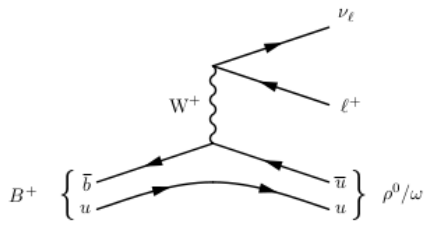
Figure 2.2: Feynman-style diagram of the process $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ [7]

The decays $\Upsilon(4S) \to B^+B^-$ and the $\Upsilon(4S) \to B^0\overline{B^0}$ have branching fractions of (51.4 ± 0.6) % of and (48.6 ± 0.6) % respectively [1]. The mass of the B meson is $m_B = 5.28\,\text{GeV}$ [1]. The collision of an electron-positron pair,can also result in the production of $q\bar{q}$ pairs. These events are called continuum background events.

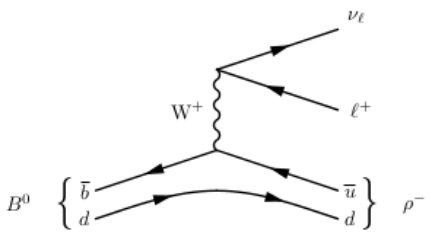
B mesons can decay semileptonically into a hadron, a lepton and a neutrino via the weak interaction. Under emission of $W^\pm$ from the b quark, the b quark changes its flavour, while the $W^\pm$ decays to a lepton and a neutrino. In this thesis the following four decay channel are analysed:

- $B^\pm \to \rho^0(\to \pi^\pm\pi^\mp)\ell^\pm\nu_\ell$
- $B^0 \to \rho^\pm(\to \pi^\pm\pi^0)\ell^\mp\nu_\ell$
- $B^\pm \to \omega(\to \pi^\pm\pi^\mp\pi^0)\ell^\pm\nu_\ell$
- $B^\pm \to \omega(\to \pi^0\gamma)\ell^\pm\nu_\ell$

The corresponding world-averages of the branching fractions are $BF(\rho^0) = (1.58 \pm 0.11) \times 10^{-4}$, $BF(\rho^\pm) = (2.94 \pm 0.21) \times 10^{-4}$ and $BF(\omega) = (1.19 \pm 0.09) \times 10^{-4}$ [1] respectively. The Feynman



(a) Feynman diagram $B^+ \to \rho^0/\omega\ell\nu$



(b) Feynman diagram $B^0 \to \rho^-\ell\nu$

Figure 2.3: Feynman diagrams for the processes $B^+ \to \rho^0/\omega\ell\nu$ and $B^0 \to \rho^-\ell\nu$

diagrams are shown in figure 2.3(a) and figure 2.3(b). The decays are charmless, meaning that charm quarks are not involved. Furthermore the hadrons in the decays are vector mesons, therefore the total angular momentum quantum number is 1.
From the decay rates of these processes the $|V_{ub}|$ matrix element of the CKM matrix can be determined. The decay rate consists of a strong and a weak component, which factorise. The weak constituent is proportional to $|V_{ub}|^2$. And the strong part can be described using angular distributions and form factors (FF). The FF can be parameterized as functions of the four-momentum transfer q^2 [7]:

$$\begin{aligned} q^2 = m_W^2 &= (P_\ell + P_\nu)^2 \\ &= (P_B - P_{\rho/\omega})^2 = M_B^2 + m_{\rho/\omega}^2 - 2M_B E_{\rho/\omega} \end{aligned} \tag{2.3}$$

Here m_W^2 is the squared mass of the virtual W boson, P_ℓ is the four-momentum of the lepton, P_ν is the four-momentum of the neutrino and P_B is the four-momentum of the B meson. M_B is the mass of the B meson and $m_{\rho/\omega}$, $E_{\rho/\omega}$, $P_{\rho/\omega}$ are the mass, energy and momentum of the ρ or ω meson.

The differential decay rates for weak components are a bit more tricky. It depends on the three helicity angles θ_l, θ_V and χ, which are defined as:

- θ_l is the angle between the W direction in the B rest frame and the lepton direction in the W rest frame,
- θ_V is the angle between the ρ or ω direction in the B rest frame and the π direction in the ρ or ω rest frame,
- χ is the angle between the ρ or ω and W decay planes.

A sketch showing the angles is shown in figure 2.4.

Since the parent meson has spin zero and the vector meson ρ/ω have total angular momentum one, the W must have the same helicity as the vector meson. In general this requires four different helicity amplitudes for all possible W configurations. But due to the small mass of the lepton the component, corresponding to the time-like W can be neglected. The remaining amplitudes only depend on q^2 and are H_0, H_+ and H_-. They are related to the axial form factors, which appear in the hadronic current and therefore are also the same which appear in the differential decay rates of the hadronic component. For the analysed case we are interest in the behavior wrt. $\cos\theta_V$, which is given after integration of

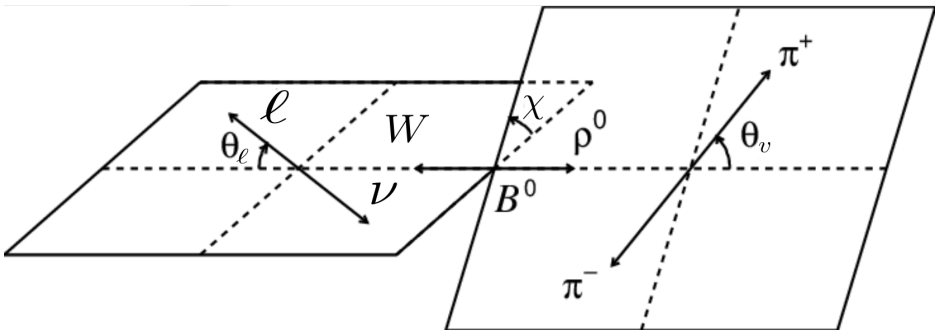


Figure 2.4: Sketch of the helicity angles

the other angles as [8]

$$\frac{d\Gamma(B \to \rho l\nu, \rho \to \pi\pi)}{dq^2 d\cos\theta_V} = |V_{ub}|^2 \frac{G_F^2}{128\pi^3} \frac{p_\rho q^2}{M_B^2} \left[\sin^2(\theta_V)\left[|H_+(q^2)|^2 + |H_-(q^2)|^2\right]/2 + \cos^2(\theta_V)|H_0(q^2)|^2\right] \tag{2.4}$$

With G_F being the Fermi constant and p_ρ being the four momentum of the ρ meson in the B meson rest frame.

$|V_{ub}|$ can now be extracted by measuring the differential decay rate and applying a theoretical FF model. The current world average of $|V_{ub}|$ from measurements [1]:

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

CHAPTER 3

The Belle Experiment

This chapter will provide a short overview of the KEKB accelerator and the Belle detector setup. Unless stated otherwise all information in this chapter is taken from ref. [9].

3.1 The KEKB accelerator

The KEKB accelerator was an asymmetric-energy, two-ring, electron-positron collider located at the KEK laboratory in Tsukuba, Japan. Positrons with energies of 3.5 GeV collided with electrons with energies of 8 GeV. It was also called a B-factory. The two rings had lengths of 3 016 m. They were installed next to each other, located 11 m below ground and crossed at the interaction point. The Belle detector is placed around the interaction point. The accelerator was built in the already existing tunnel of TRISTAN, which was an older electron-positron collider. In figure 3.1 the schematic layout of the KEKB accelerator can be seen. At KEKB electrons and positrons collided at a center of mass energy of 10.58 GeV, corresponding to the $\Upsilon(4S)$ resonance energy. The $\Upsilon(4S)$ nearly exclusively decays to B meson pairs.
The Belle detector collected the largest luminosity at its time. The full Belle dataset of 711 fb^{-1} contains $(772 \pm 11) \times 10^6$ $B\bar{B}$ pairs. The Belle detector was upgraded to the Belle II detector which started collecting data in 2019. Additionally the KEKB accelerator was upgraded to SuperKEKB.

3.2 The Belle Detector

The Belle detector surrounded the KEKB interaction point. A sketch of the side view of the Belle detector can be seen in figure 3.2. The Silicon Vertex Detector (SVD) measured the B meson decay vertices and directly surrounded the beam pipe. It was encompassed by the Central Drift Chamber (CDC) that supplied the charged particle identification with its measurements of track ionization losses (dE/dx). Its main purpose was the identification of tracks and measurement of their transverse momentum p_T. Outside the CDC, CsI crystals were placed that detected electromagnetic showers. The K_L and Muon Detector (KLM) supplied the muon and K_L meson identification. It was made of an array of resistive plate counters embedded in the iron yoke. Additional information for particle identification was supplied by the Time of Flight counters (TOF) that were located radially outside the CDC.

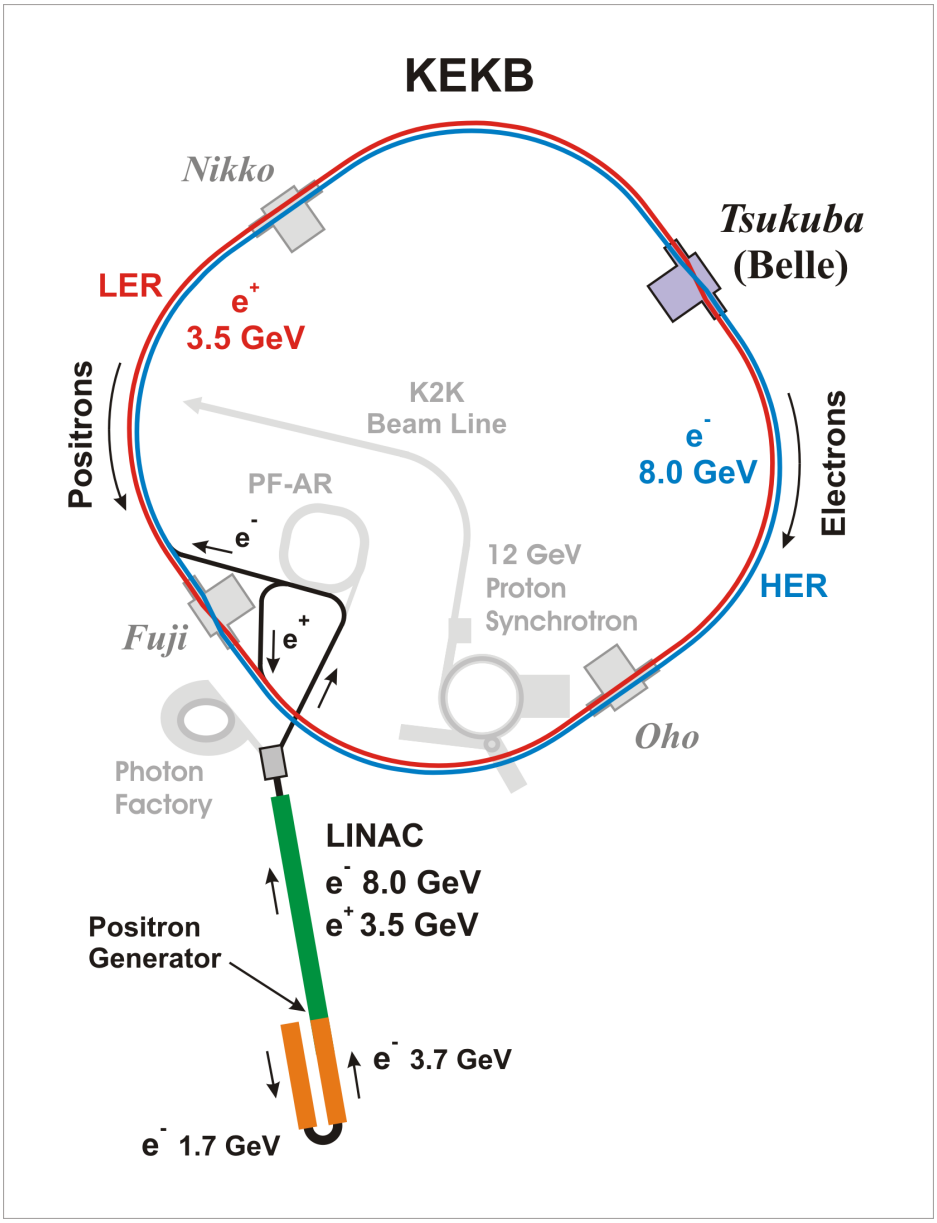


Figure 3.1: Schematic sketch of the KEKB accelerator [10]

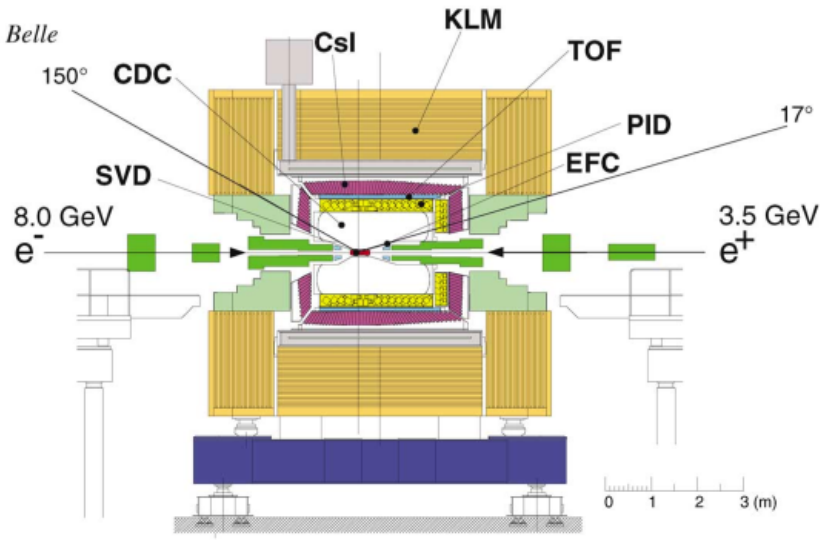


Figure 3.2: Side view of the Belle detector [7]

3.2.1 Silicon Vertex detector

The purpose of the Silicon Vertex Detector (SVD) was the precise measurement of B meson z-vertex positions and decay vertices of B and τ. The z axis in the detector is along the beamline. The SVD also contributed to tracking. The particles of interest for the Belle experiment typically had energies below 1 GeV, so the resolution of vertex measurements was mainly limited by multiple Coulomb scattering. To reduce this effect the innermost detector layer, the SVD, should be placed as close as possible to the interaction point. Therefore the support structure had to be light but must remain durable. The readout electronics were situated outside the tracking volume, to withstand large beam backgrounds. The SVD covered a polar angle of $23° < \theta < 139°$ and consisted of three layers. The three layers had diameters of 30 mm, 45.5 mm and 60.5 mm and consisted of independent ladders that each comprised double-sided silicon strip detector reinforced by boron-nitride support ribs.

3.2.2 Central Drift Chamber

The purpose of the Central Drift Chamber (CDC) was the efficient reconstruction of charged particle tracks and the precise determination of their momenta. As is the case for the SVD the minimization of multiple scattering is important also in the CDC. Therefore a gas with low proton number was used to minimize coulomb interaction. Furthermore, a good resolution in dE/dx was needed. The CDC was asymmetric in the z direction and had an angular coverage of $17° < \theta < 150°$. The longest wires were 2 400 mm long and the inner radius was 103.05 mm, while the outer radius was 874 mm. Due to these radii the CDC had a good tracking efficiency for low transverse momenta p_t by minimizing the material thickness. Its forward and backward small-r regions had conical shapes. To clear the detector components and simultaneously maximize the acceptance.
The chamber consisted of 50 cylindrical layers that each contained three to six axial or small-angle-stereo layers, and three cathode strip layers. The CDC had a total of 8400 drift cells, three layers for the innermost stereo super layer and four layers for the three outer stereo super layers. The stereo

angles in each super layer were determined by maximising its z-measurement capability.
In general, the drift cells were nearly square except for the three inner layers. The maximum drift distance was between 8 mm and 10 mm, and the radial thickness was between 15.5 mm and 17 mm. The drift cells in the innermost layer were smaller and they were read-out by cathode strips on the walls of the cylinders.
To resolve left-right ambiguities, neighbouring radial layers in a super layer were staggered by half a cell in the azimuth angle ϕ. The sense wires were made of gold plated tungsten and had diameters of 30 µm. The field wires were made of unplated aluminium and were 126 µm in diameter. They were arranged to produce high electric fields up to the edge of the cell wall and to simplify the functional relationship between drift time and distance. Three z-coordinate measurements in the innermost radii were made by cathode strips glued to the inner surface of the cylinder.

3.2.3 Particle Identification System

The Particle Identification (PID) system consisted of the Aerogel Cherenkov Counter (ACC) and the Time of flight (TOF) system. In the ACC, aerogel was used as the radiation material for the barrel part outside of the CDC and for the forward endcap of the detector, because it has minimal impact on the design of the accelerator and other components. The TOF system consisted of 128 scintillators that each were 4 cm thick. Photomultiplyer tubes detected scintillation light from charged particles traversing the scintillators. An additional layer of 64 counters were used to form a track trigger. The counters are each 4 mm thick, they form the Trigger Scintillation Counters (TSC). The velocities of charged particles could thus be determined by measuring the time difference between hits in the TSC and the corresponding hits in the TOF counter. The PID system could distinguish between kaons and pions of momenta below 1.2 GeV due to a time resolution of about 100 ps.

3.2.4 Electromagnetic Calorimeter

The main purpose of the Electromagnetic Calorimeter (ECL) was the detection of photons from B-mesons. Photons are detected vie electromagnetic showers that they produce in the ECL. Therefore a high efficiency and good resolution in energy and position was needed. Photons at the Belle experiment usually had low energies and thus a good performance at energies below 500 MeV was necessary. Additionally, a high resolution above 4 GeV was required to reduce backgrounds. The particle identification relies primarily on a comparison of the charged particle momentum and the energy deposit in the ECL. A high energy resolution also results in better hadron rejection. A precise determination of opening angles was needed, which led to a fine-grained segmentation of the ECL. Highly segmented arrays of CsI(TI) crystals with silicon photo diode readouts were used. Additionally, a superconducting solenoid magnet with a magnetic field of 1.5T was used.

3.2.5 K_L and Muon Detector

The K_L and Muon detector (KLM) was designed to identify K_L and μ with high efficiency. This was achieved for a broad momentum range greater than 600 MeV. It had a polar angle coverage of $45° < \theta < 125°$. The endcaps in the forward and backward directions had extended ranges of 20° and 155°. The KLM consisted of alternating layers of charged particle detectors and 4.7 cm thick iron plates. In total, it had 15 detector layers and 14 iron layers in each forward and backward endcap. The

iron plates provided 3.9 interaction lengths, while the ECL provided 0.8 interaction lengths of material to convert K_L. Interacting K_L produced hadronic showers of ionizing particles. The location of a shower determined the direction of the K_L. Only a portion of the total particle energy of the hadronic shower can be captured, therefore no precise energy measurements could be obtained with the KLM. Multiple layers of charged particle detectors and iron allowed to discriminate between muons and charged hadrons, because muons travel further and are deflected less.

3.3 Particle Identification

Using the information from the PID system and additiomal dE/dx information from the CDC a particle identification variable can be constructed based on likelihood ratios. The hadron identification likelihood for candidate particle α is calculated based on dE/dx measurements from the CDC, L_α^{CDC}, the time of flight information from the TOF L_α^{TOF}, and information about the number of photons from the ACC, L_α^{ACC}. The likelihood ratio is calculated, using equation 3.1 and used for identification. It is the likelihood ratio of particle α to particle β.

$$L(\alpha : \beta) = \frac{L_\alpha^{CDC} L_\alpha^{TOF} L_\alpha^{ACC}}{L_\alpha^{CDC} L_\alpha^{TOF} L_\alpha^{ACC} + L_\beta^{CDC} L_\beta^{TOF} L_\beta^{ACC}} \tag{3.1}$$

The electron identification (eID) uses L_α^{CDC}, L_α^{ACC} as well as information from the ECL about the matching of track positions, cluster energies, E/p and transverse shower shapes from the ECL are used to calculate likelihood ratios. For muon identification (muID) the extrapolated tracks from the CDC are compared to the reconstructed hits in the KLM. ΔR, the difference between the measured and expected range of a track and the statistic χ_r^2, reconstructed from the transverse deviations of all hits, is used to form likelihoods for the muon, pion and kaon hypotheses. The likelihood ratio $L_\mu/(L_\mu + L_\pi + L_K)$ is called muon ID and used as a discriminating variable [11].

CHAPTER 4

Reconstruction

This chapter describes the necessary steps to reconstruct signal decays. First, an overview of the Monte Carlo simulations and the different applied corrections is given, followed by an introduction to hadronic tagging. Next the selection steps used to reconstruct signal decays are explained. Afterwards, variables that are used later on in this analysis are introduced.

4.1 Monte Carlo Simulation

This analysis uses Monte Carlo (MC) simulated data. This data consists of events containing the same final states as the expected detector data. Even though the same final states are reconstructed, not all events reconstructed as signal actually correspond to signal decays, and these events are called background. Furthermore, final state particles can be misidentified during reconstruction. To understand the background and signal characteristics, MC simulated data is used, where the simulated decays are exactly known.

EvtGen [12] is used to simulate e^+e^- collisions and $B\overline{B}$ decays, as well as non-$B\overline{B}$ backgrounds, and returns the four-momentum vectors of the final state particles. Following this, information is passed on to GEANT3 [13], which simulates the bending of tracks, scattering and further decays. Next the detector response, such as the energy deposited by final state particles and their location is simulated. Furthermore, simulated beam-induced backgrounds are added. The simulated data is then passed to the reconstruction algorithm.

The MC data is simulated to reproduce the data recorded by the detector but it has the advantage that the truth information is available. The Belle MC is produced separately for different physics processes. It is produced in streams, where one stream contains the expected amount of events of a certain process in the full Belle dataset of an integrated luminosity of 711 fb^{-1}. The different processes contained in the Belle MC can be seen in table 4.1.

MC Type	Physics Process
charged	$e^+e^- \to \Upsilon(4S) \to B^+B^-$
	B mesons decay generically
mixed	$e^+e^- \to \Upsilon(4S) \to B^0\overline{B^0}$
	B mesons decay generically
uds	$e^+e^- \to q\overline{q}$, $q = u, d, s$
charm	$e^+e^- \to c\overline{c}$
charged $X_u\ell^+\nu_\ell$	$e^+e^- \to \Upsilon(4S) \to B^+B^-$
	$B^+ \to X_u\ell^+\nu_\ell$
	$B^- \to$ generic
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

Table 4.1: Different processes contained in Belle MC.

4.2 Monte Carlo Corrections

The simulated data, however, does not represent the data recorded by the detector perfectly. Therefore, corrections to the simulated efficiency have to be applied, in order to improve agreement between data and MC. Furthermore, corrections for updated branching fraction measurements are necessary. Every correction results in an MC weight. The total weight of a MC event is the product of all weights that were assigned to the event by the different corrections.

$$w_{tot} = w_{hyb} w_{lep} w_{BF} \tag{4.1}$$

In the following, the necessary MC corrections are introduced. First, the hybrid model is explained, followed by the lepton ID corrections and the branching fraction corrections.

4.2.1 Hybrid Model

The hybrid model used in this analysis was implemented following references [14] and [15].

Charmless semileptonic decays are produced as a mixture of specific resonant and non-resonant contributions. The non-resonant part is overestimated which has to be corrected. For this purpose the hybrid model is used. There are measured rates for some resonant modes, and measured inclusive rates. There is a gap between these two rates, this gab below 1.4 GeV is filled with a model for the non-resonant production, this model is used above 1.4 GeV as well. The resonant modes are $B \to \pi\ell^+\nu_\ell$, $B \to \rho\ell^+\nu_\ell$, $B \to \omega\ell^+\nu_\ell$ and $B \to \eta^{(')}\ell^+\nu_\ell$. The non-resonant $B \to X_u\ell^+\nu_\ell$ modes are produced with at least two pions in the final state using the De Fazio-Neubert (DFN) [16] model. The triple differential rate $\Delta\mathcal{B}$ is regarded as a function of the squared four-momentum transfer (q^2), the lepton energy in the B rest-frame (E_ℓ^B) and the hadronic invariant mass squared (m_X^2). The $B \to X_u\ell\nu$ modes are combined in such a way, that the sum of the triple differential rates of the non-resonant ($\Delta\mathcal{B}_{ijk}^{non-res}$) and combined resonant ($\Delta\mathcal{B}_{ijk}^{res}$) modes reproduces the one obtained from inclusive measurements. Therefore, weights ($w_{ijk} = w_{hyb}$) are assigned to the non-resonant contribution such

that

$$\Delta\mathcal{B}_{ijk}^{incl} = \Delta\mathcal{B}_{ijk}^{res} + w_{ijk} \times \Delta\mathcal{B}_{ijk}^{non-res}$$

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

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

Figure 4.1 shows the effect of the hybrid model implementation, for charged B mesons, on the variables q^2, m_X, E_ℓ^B. The generator level distributions are shown, that means that the information used to generate the MC is used.

4.2.2 Lepton ID Corrections

MC simulations are produced to look as similar to recorded data as possible, but still particle identification performs slightly differently in MC and recorded data. The reconstruction efficiency is calculated based on the MC are not necessarily replicated in recorded data. To precisely estimate the branching fractions the differences have to be corrected. Therefore, correction factors are calculated in bins of lepton momentum ($|\vec{p}_\ell|$) and polar angle (θ) in the laboratory frame. The correction factors are provided by Belle, they were obtained by comparing the MC with data efficiencies for $\gamma\gamma \to e^+e^- \to e^+e^-\ell^+\ell^-$ [17]:

$$w_{lep} = \frac{\epsilon_{data}}{\epsilon_{MC}} \tag{4.2}$$

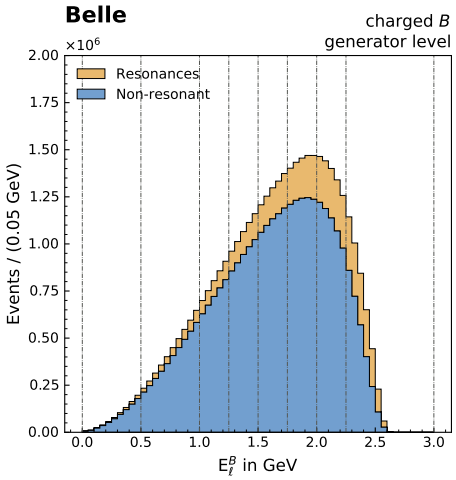
4.2.3 Branching Fraction Corrections

The branching fractions of the B decays used in the simulation of MC are stored in decay files, but some of their current world averages have changed. Therefore, a new weight (w_{BF}) has to be defined by dividing the current world average B_{PDG} by the branching fraction stored in the decay files B_{DEC}.

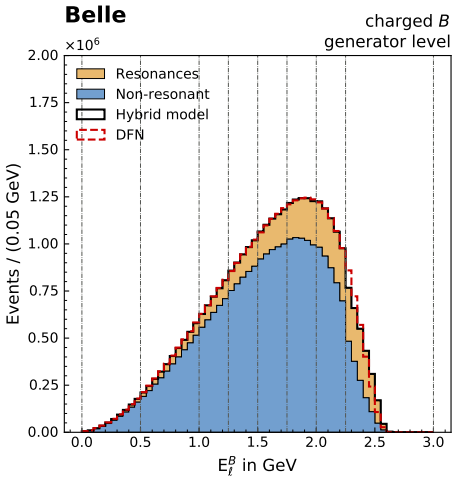
$$w_{BF} = \frac{B_{\text{PDG}}}{B_{\text{DEC}}}$$

In this analysis the decays of B mesons decays semileptonically to a D or D^* meson have to be corrected, while the decays B mesons to X_u already have the correct branching fractions in the decay files. The branching fractions of the following B decays are corrected:

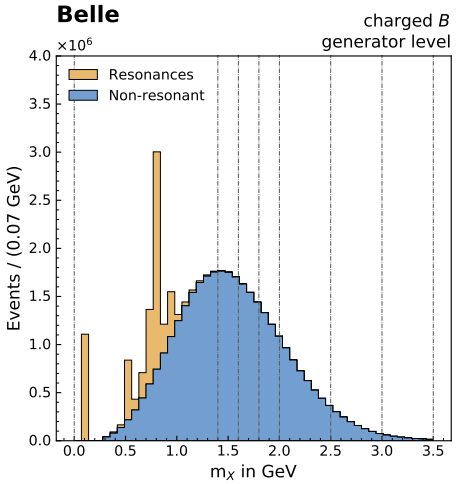
$$\begin{aligned} B^0 &\to D^-\ell^+\nu_\ell \\ B^0 &\to D^{*-}\ell^+\nu_\ell \\ B^+ &\to \overline{D}^0\ell^+\nu_\ell \\ B^+ &\to \overline{D}^{*0}\ell^+\nu_\ell \end{aligned}$$



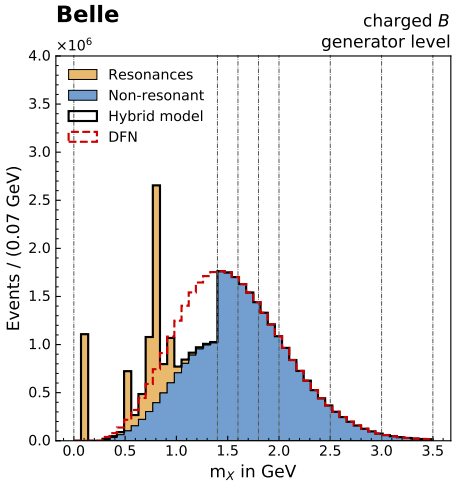
(a)



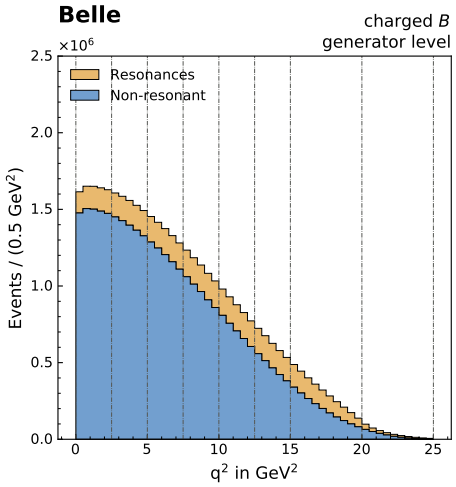
(b)



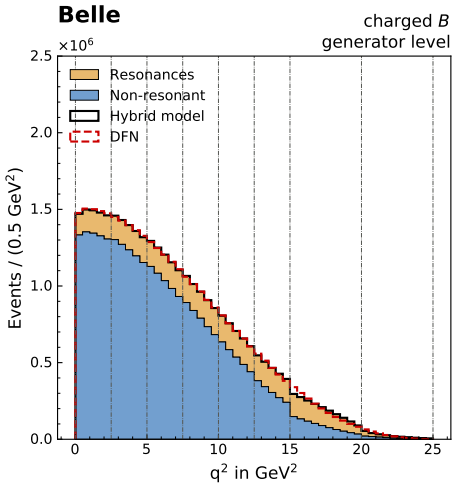
(c)



(d)



(e)



(f)

Figure 4.1: Generator level distributions of E_ℓ^B, m_X and q^2 for $B \to X_u \ell^+ \nu_\ell$ decays. On the left the distributions before and on the right the distributions after implementing the hybrid model are shown.

4.3 Hadronic Tagging and Full Event Interpretation

The Belle detector was designed to detect, reconstruct and identify different decays of B mesons. It can only detect stable particles that live long enough to interact with the different sub-detectors. These particles are called final state particles (FSP). They are: $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, p, $\overline{p}$, γ and K_L^0. The lifetimes of B mesons are of the order of 10^{-12} s, therefore, they decay close to the interaction point. To analyse B meson decays they have to be reconstructed from their final state particles.

Most $\Upsilon(4S)$ decay to two B mesons. In this analysis both are reconstructed. One of the B mesons is fully reconstructed in the hadronic mode, it is called B_{tag}. The signal decays are reconstructed as the second B meson, called B_{sig}. The signal decays contain neutrinos in the final states, which can not be detected. So the B_{sig} can not be completely reconstructed.

With the help of the B_{tag} the lost information from the signal side can be recovered, since the full momentum of the B_{tag} is known. In the $\Upsilon(4S)$ rest-frame the two B mesons are produced back-to-back and no other particles are created from the initial e^+e^- collision. Therefore, the four-momentum of the B_{sig} can be recovered through the precise knowledge of the initial four-momentum of the $\Upsilon(4S)$. Further measuring the four-momenta of the lepton and hadron candidate of the B signal decays allows the inference of the four-momentum of the undetected neutrino. This process is made possible by hadronic tagging. The Full Event Interpretation (FEI) algorithm reconstructs B_{tag} in $O(10000)$ hadronic channels using a hierarchical approach. Low level particles like K and π are reconstructed first from tracks and clusters. Clusters that are not associated with any tracks are used to construct photon candidates, and tracks to constructs FSP candidates. Then, low level particles are formed, for example, the π^0 candidates are reconstructed from two photons. Next, intermediate particles like D^0 mesons are reconstructed by combining K^-, π^+ and π^0 candidates. This process continues until a B_{tag} meson candidate is formed, for example, from D^* and π^+ candidates.

The FEI looks at all possible combinations of particles to reconstruct the B_{tag} candidate. For each reconstructed B_{tag} the FEI calculates a signal probability, which gives an estimate of how likely the B_{tag} candidate is reconstructed correctly. The signal probability is built from signal probabilities assigned to all particles in the reconstructed decay chain. It does not represent the actual probability of a B_{tag} candidate being correct though. Then, based on the FEI signal probability a best tag candidate selection is performed, and only the most likely B_{tag} candidates are selected. The B_{sig} is reconstructed from the remaining particles.

An illustration of hadronic tagging can be seen in figure 4.2. Electron positron pairs are collided and form $\Upsilon(4S)$, these than decay to two B mesons, one of them is B_{sig} and the other one is B_{tag}.

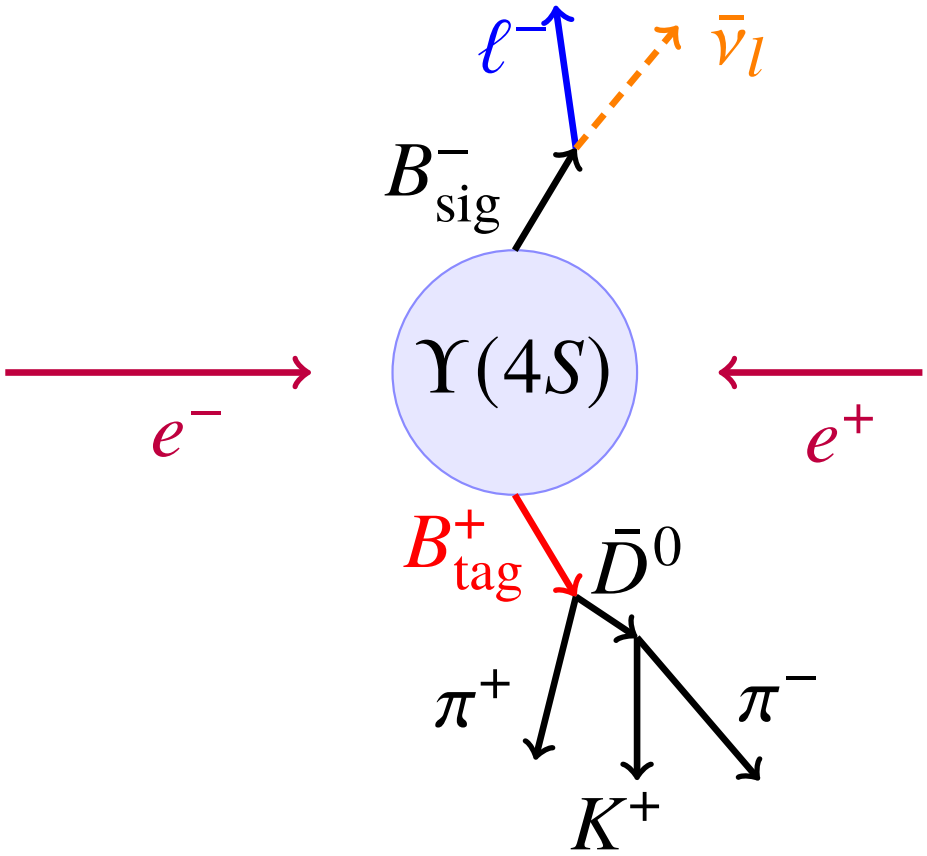


Figure 4.2: An illustration of hadronic tagging [18]

4.4 Reconstruction

During the reconstruction of the signal decays some preselections are set on both the B_{tag} and the B_{sig}. This is done to reduce the amount of background after the reconstruction, while keeping as much signal as possible. Furthermore, a *completeness constrain* is added. This constraint ensures that an event does not contain any additional charged tracks that are not associated with either the B_{tag}- or B_{sig} candidate. In the following sections the preselections are shown.

4.4.1 Tag-side Reconstruction

In table 4.2 the tag-side preselections are shown.

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

Table 4.2: Preselections on the tag-side.

M_{bc} is the beam-energy constrained mass and defined as [7]:

$$M_{bc} = \sqrt{E^*_{\text{beam}} - p^*_B} \tag{4.3}$$

E^*_{beam} is the beam energy in the center of mass (CM) frame, which is equal to half of the CM energy. p^*_B is the reconstructed momentum of the B_{tag} in the CM frame. If the B_{tag} candidate is reconstructed perfectly then its CM energy is equal to E^*_{beam}. Therefore, M_{bc} is similar to the invariant mass of the B_{tag} meson if B_{tag} is reconstructed correctly. Consequently M_{bc} is sensitive to the reconstructed momentum of B_{tag}.

ΔE is the difference between the beam energy in the CM frame and the reconstructed energy, E^*_B, of the B_{tag} in the CM frame.

$$\Delta E = E^*_B - E^*_{\text{beam}} \tag{4.4}$$

Both B mesons have the same energy in the CM frame. Since the $\Upsilon(4S)$ decays into two B mesons without any additional particles, their energies are equal to the beam energy. That results in the ΔE distribution peaking around zero, if the energies of all final state particles are measured correctly and the B_{tag} was correctly reconstructed. Hence, ΔE is sensitive to the reconstructed energy including the mass hypothesis assignment of final state particles.

The FEI signal probability is an indicator for how likely the B_{tag} candidate to be correctly reconstructed.

Multiple B_{tag} candidates per event, are possible. Consequently, a best candidate selection is performed and only the B_{tag} candidate with the highest FEI signal probability is accepted. This is done after the selections listed in table 4.2 are applied.

4.4.2 Signal-side Reconstruction

For the signal side the preselections can be seen in table 4.3. In a cylindrical system with the origin at the interaction point, z is the distance along the electron beam direction and the axis of symmetry along the electron beam and r is the radial coordinate. To ensure that the pion and lepton are originating near the interaction point, cuts on r_{PCA} and z_{PCA} are used, which are the coordinates of the point of closest approach (PCA) of a reconstructed charged particle to the z axis. The variables muID and eID are explained in chapter 3.3. $|p_{\text{lab}}|$ is the lepton momentum in the laboratory frame. A vertex fit is performed for B_{sig}, using `TreeFitter` [19], and the B_{sig} candidate. The vertex fit has to converge for the B_{sig} candidate to be considered.

Variable	$\pi^\pm$	$\mu^\pm$	$e^\pm$
r_{PCA} in cm	< 2	< 2	< 2
$\lvert z_{\text{PCA}}\rvert$ in cm	< 4	< 4	< 4
muID	< 0.98	> 0.8	
eID	< 0.98		> 0.8
$\lvert p_{\text{lab}}\rvert$ in GeV		> 0.6	> 0.3

Table 4.3: Signal-side preselections.

After reconstructing B_{sig} and B_{tag} any tracks and clusters that are left are called rest of event (ROE). In table 4.4 the preselections for the ROE are shown. The ROE should only contain tracks and clusters

from beam backgrounds, if B_{tag} and B_{sig} are correctly reconstructed. If the reconstruction is missing a final state particle on the other hand the ROE contains non-beam tracks and clusters.

Variable	cut value
nCDCHits	> 0
p^*	≤ 3.2 GeV
r_{PCA}	< 2 cm
$\lvert z_{\text{PCA}} \rvert$	< 4 cm

Table 4.4: ROE preselections.

Only tracks with momenta smaller than 3.2 GeV are kept in the ROE to remove high energy tracks from non B decays. The r_{PCA} and $|z_{\text{PCA}}|$ cuts ensure that only tracks coming from the interaction point are considered. The cut on the number of hits in the CDC ($nCDCHits$) serves the purpose of ensuring that there actually are hits in the Central Drift Chamber (CDC).
To improve the selection of signal B candidates, additionally a cut on $|\cos\theta_{BY}| < 6$ is implemented. This is the cosine of the angle between the direction of the nominal particle Y and the B_{sig} candidate in the CM frame. Y is a pseudoparticle that is obtained by combining the four-momentum vectors of the hadron and the lepton candidates on the signal-side [7].

$$p_Y = p_{\text{hadron}} + p_\ell, \tag{4.5}$$

$$\cos\theta_{BY} = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2p_B p_Y}, \tag{4.6}$$

where m_Y is the invariant mass of the pseudoparticle Y, P_Y is its momentum and E_Y is its energy. E_B is the energy of the B candidate, p_B its momentum and m_B is its mass. p_ℓ is the momentum of the lepton and p_{hadron} the momentum of the hadron. This cosine, however, is not limited to the range of -1 to 1 for background events. Even signal events can be reconstructed outside of this range due to finite detector resolution. For now a loose cut on this variable is chosen.

4.5 Reconstruction by Channels

In this thesis four different channels are reconstructed. Each channel has slightly different properties and therefore different additional preselections have to be applied. In the following, the additional preseletions for the channels $B^+ \to \rho^0 \ell^+ \nu$, $B^0 \to \rho^- \ell^+ \nu$, $B^+ \to \omega(\to \pi^+\pi^-\pi^0)\ell^+\nu$ and $B^+ \to \omega(\to \pi^0\gamma)\ell^+\nu$, with ℓ being an e or μ, are discussed.

4.5.1 ρ^0 Channel

The reconstruction of the channel $B^+ \to \rho^0(\to \pi^+\pi^-)\ell^+\nu$ is started by reconstructing the ρ^0 from π^+ and π^- candidates. Next the B^+ meson is reconstructed from the ρ^0 and either an e or a μ candudate as a lepton. No further selections are necessary for this channel.

4.5.2 $\rho^{\pm}$ Channel

To reconstruct $B^0 \to \rho^-(\to \pi^0\pi^-)\ell^+\nu$, first the ρ^- is reconstructed from a π^0 and a π^- candidate. After that, the B^0 is reconstructed from the ρ^- and again either an e or a μ candidate. For this mode a loose cut on the invariant mass of $\pi^0\pi^+$ is necessary to reduce the amount of combinatorial background. A selection of $(0.47730 < M_{\pi^0\pi^+} < 1.07370)$ GeV is used.

4.5.3 $\omega \to \pi^+\pi^-\pi^0$ Channel

To reconstruct $B^+ \to \omega(\to \pi^+\pi^-\pi^0)\ell^+\nu$ first the ω is reconstructed from $\pi^+\pi^-\pi^0$ candidates. This is followed by the reconstruction of the B^+ meson by adding either an e or μ candidate to the ω candidate. A cut on the invariant mass of $\pi^+\pi^-\pi^0$ is necessary. A window of $(0.74265 < M_{\pi^+\pi^-\pi^0} < 0.81265)$ GeV is selected.

4.5.4 $\omega \to \pi^0\,\gamma$ Channel

In the second ω channel, $B^+ \to \omega(\pi^0\gamma)\ell^+\nu$, the ω decays to a π^0 and a γ. Again, the ω is reconstructed first and next the B^+ meson is reconstructed from the ω and either an e or a μ candidate. The same invariant mass window is set for both ω channels.

4.6 Squared Four Momentum Transfer

In this analysis the signal will be extracted in momentum transfer squared q^2 bins for the $|V_{ub}|$ measurement. q^2 can be calculated as [2]:

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

$E^*_{B_{\text{tag}}}$ and $\vec{p}^{\,*}_{B\text{tag}}$ are the energy and momentum, respectively, of the reconstructed B_{tag}, in the center of mass (CM) frame. E^*_{hadron} and $\vec{p}^{\,*}_{\text{hadron}}$ are the reconstructed energy and momentum, respectively of the corresponding hadron both in the CM frame.

To not only measure $|V_{ub}|$ but aslo get information about the form factors, the signal extraction, which is explained in chapter 8, is done in q^2 bins. A previous analysis performed by A. Sibidanov et al. [2] divided the q^2 spectrum for the $B \to \rho^0\ell\nu$ channel into 11 bins between $q^2 = 0$ GeV and $q^2 = 22$ GeV. The distribution of the reconstructed q^2 against the true q^2 for $B \to \rho^0\ell\nu$ is shown in figure 4.3 also in 11 bins to make a comparison possible. It can be seen that the reconstructed q^2 and the true q^2 are not identical which they would be in the ideal case. An event can fall into different q^2 bins of reconstructed and true q^2. This is called bin migration. The corrections for this are shown in section 8.3.

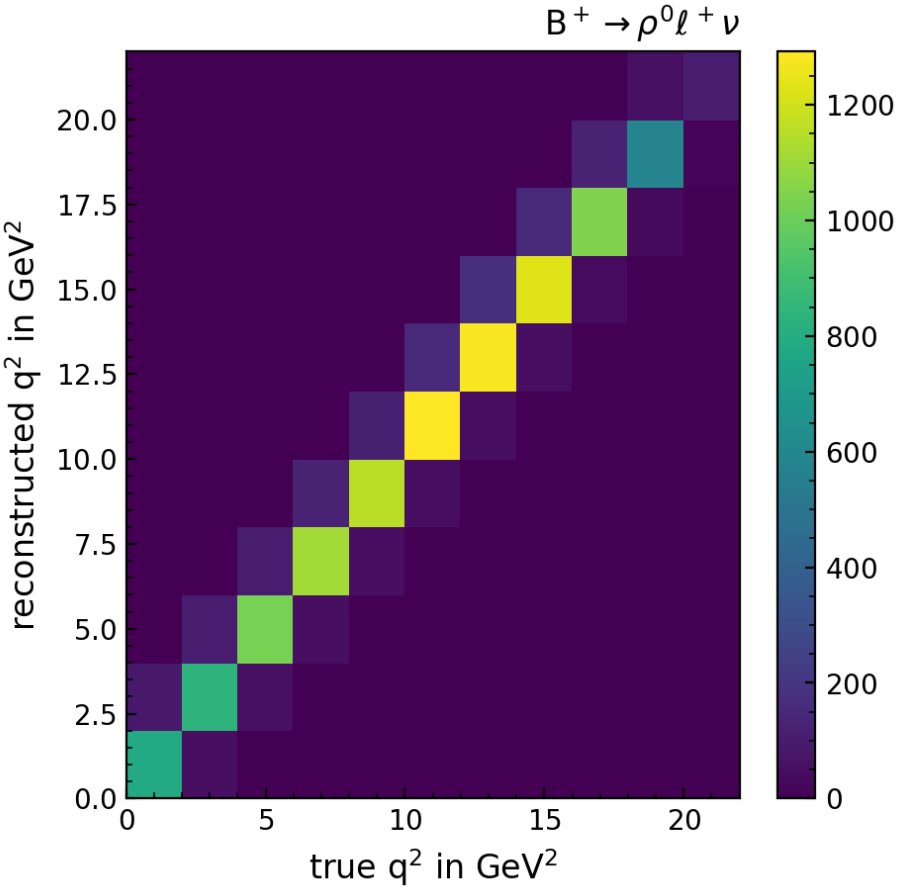


Figure 4.3: Reconstructed q^2 against true q^2 for $B^+ \to \rho^0 \ell^+ \nu$.

4.7 Missing Mass squared

To derive the four-momentum of the B_{sig} meson in the $\Upsilon(4S)$ rest-frame, the reconstructed four-momentum of the B_{tag} meson is required. Additionally, the constraints can be used that the invariant mass of the B candidate has to be the B meson mass, and that the energy of the B_{tag} has to be half the CM collision energy $m_{\Upsilon(4s)}$ [2]. The $p_{B_{\text{sig}}}$ can then be defined as:

$$p_{B_{\text{sig}}} = (E_{B_{\text{sig}}}, \vec{p}_{B_{\text{sig}}}) = \left(\frac{m_{\Upsilon(4S)}}{2}, -\frac{\vec{p}_{B_{\text{tag}}}}{|\vec{p}_{B_{\text{tag}}}|} \sqrt{\left(\frac{m_{\Upsilon(4S)}}{2}\right)^2 - m_B^2} \right) \tag{4.8}$$

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

$$p_{\text{miss}} = p_{B_{\text{sig}}} - p_{\text{hadron}} - p_\ell \tag{4.9}$$

For a correctly reconstructed tag-side B candidate and a correctly reconstructed semileptonic decay on the signal-side, p_{miss} corresponds to the four-momentum of a single neutrino. Since the neutrino is

effectively massless, the missing mass squared $m^2_{\text{miss}} = p^2_{\text{miss}}$ is therefore zero. This distribution can be seen in figure 4.4 for the $B \to \rho^0 \ell \nu$ channel, as an example.

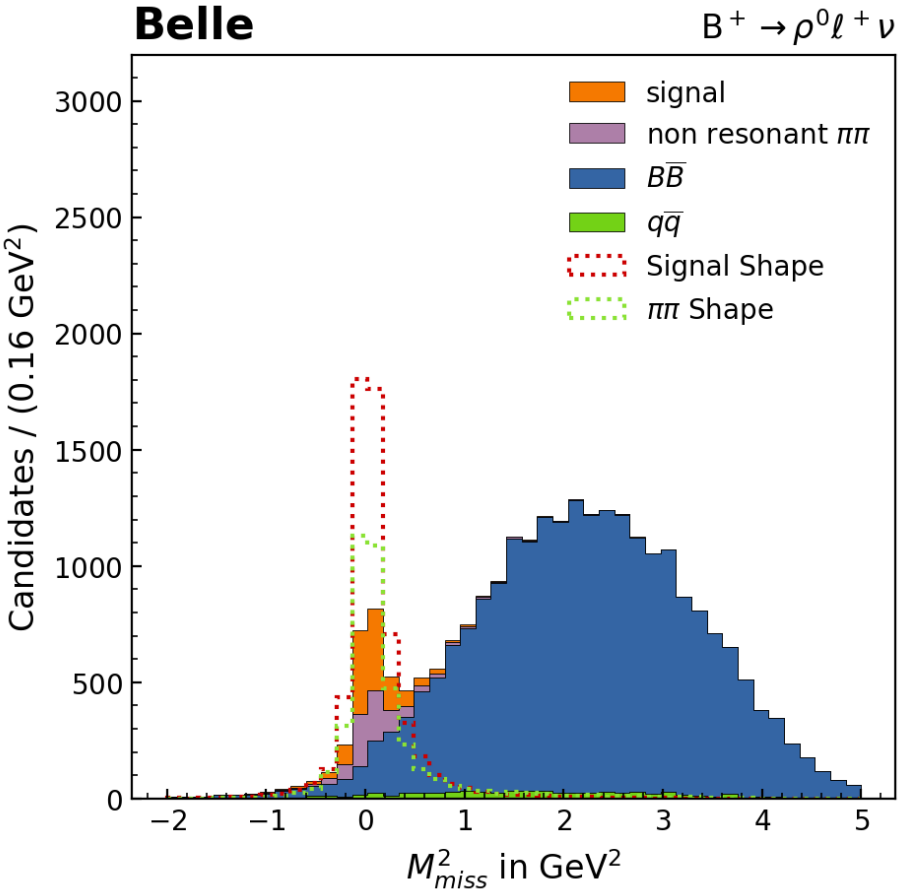


Figure 4.4: M^2_{miss} distribution for $B^+ \to \rho^0 \ell^+ \nu$ from simulation.

As expected for correctly reconstructed signal events M^2_{miss} peaks at zero. The main event categories can be seen in the M^2_{miss} distribution in figure 4.4. The orange called signal here represents correctly reconstructed signal events. In this case, the $B^+ \to \rho^0 \ell^+ \nu_\ell$ channel was reconstructed correctly where the lepton and ρ^0 candidates were correctly identified as the daughters of the signal-side B^+ meson. Even though B_{sig} was reconstructed correctly, it does not necessarily mean that B_{tag} was reconstructed accurately. Therefore, signal events can also have M^2_{miss} values larger than zero, due to the M^2_{miss} calculation relying on the reconstructed four-momentum of the B_{tag} meson. For background events M^2_{miss} peaks mainly at larger values due to additional particles that were not reconstructed in the event. The missing particles contribute to lager M^2_{miss} values. However, a small component in the $B\overline{B}$ category can also be seen to peak around zero in figure 4.4. This is discussed further in chapter 9.

4.8 Background Classification

The classification of the background into combinatorial $B\overline{B}$ and continuum $q\overline{q}$ background is explained in the following section.

4.8.1 Combinatorial background

In combinatorial background events, also called $B\overline{B}$ background, one or more particles are wrongly assigned to the B_{sig} decay. This background is depicted in blue in figure 4.4. Typically, the particles originate from the other B decay. Variables like ΔE, M_{bc} and $\cos\theta_{BY}$ can be used to reduce the amount of $B\overline{B}$ background.

This background can be further split into different categories. First a check is performed to see whether or not the lepton is the daughter of the B meson. If this is the case, it is called primary lepton, if not it is a secondary lepton. For primary leptons the identity of the hadron produced in the same B decay can be checked and the evnets can be sorted into further categories. The M^2_{miss} distribution for $B^+ \to \rho^0 e\nu$ is shown in fig 4.5.

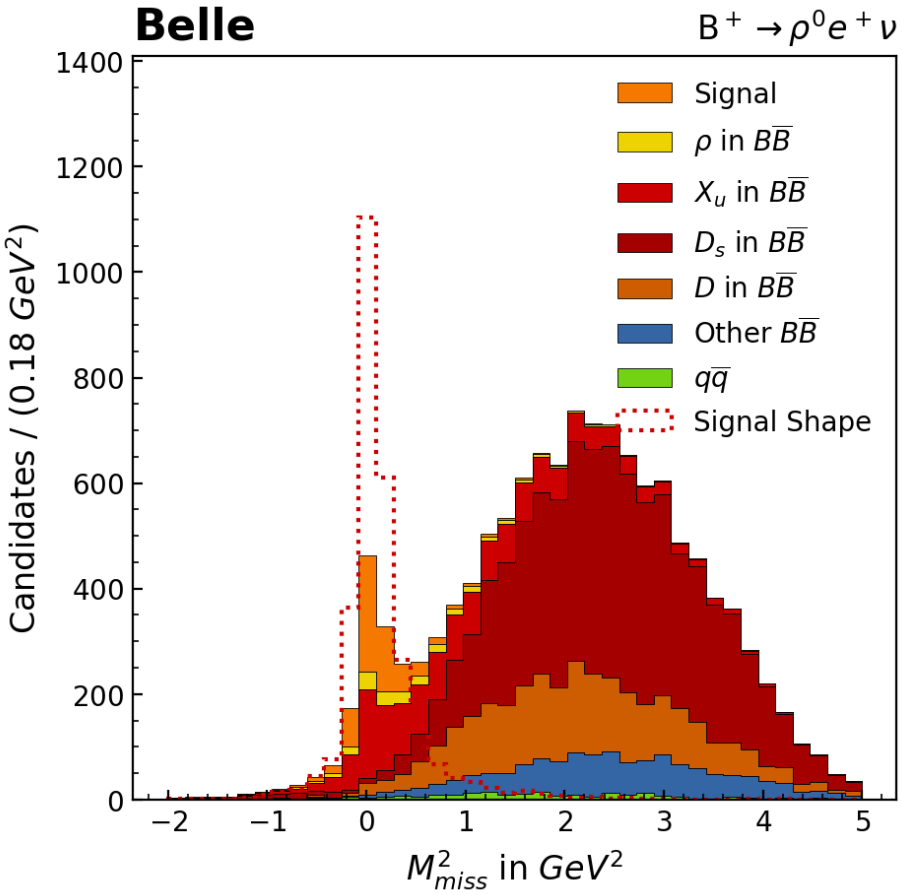


Figure 4.5: M^2_{miss} distribution for the decay $B^+ \to \rho^0 e^+ \nu$ with the combinatorial background split into further categories.

The background is split into the categories "ρ in $B\overline{B}$", "X_u in $B\overline{B}$", "D_s in $B\overline{B}$", "D in $B\overline{B}$" and "Other $B\overline{B}$" background. The last category is composed of all events with secondary leptons, mis-identified leptons or where the hadron produced in the B decay does not belong to the above categories.

4.8.2 Continuum background

Electron-positron collisions at the $\Upsilon(4S)$ resonance do not only produce B meson pairs but also quark pairs involving lighter quarks than the bottom quark (u, d, s, c). These events are called continuum or $q\overline{q}$ background. Continuum background is shown in green in Figure 4.4 and 4.5.

The event shapes of these type of events differ drastically from the $B\overline{B}$ event shape. This happens because B mesons are produced only slightly above the kinematic $B\overline{B}$ pair threshold. Therefore, the momentum of the B meson in the $\Upsilon(4S)$ rest-frame is only about 320 MeV, which results in a nearly isotropic, spherical decay in the $\Upsilon(4S)$ rest-frame. The continuum backgrounds have jet-like event shapes because the quarks are lighter than the b quark, therefore they have a very high momentum available. A machine learning algorithm is used to exploit these event shape difference, and therefore reduce this type of background. This is described in more detail in chapter 5.

CHAPTER 5

Continuum Suppression

For the suppression of the continuum background a multivariate analysis (MVA) is used. Currently two separate trainings are used, where one is performed for the $B^0 \to \rho^- \ell^+ \nu$ channel and the other for the $B^+ \to \rho^0 \ell^+ \nu$ channel. The ρ^0 training is used for the ω channels as well. This is valid because the signals have similar shapes. In figure 5.1 the M^2_{miss} distribution for the $B^+ \to \rho^0 e^+ \nu_e$ channel before continuum suppression can be seen. The continuum background is here still a major background.

5.1 Multivariate Analysis

Multivariate Analysis (MVA) methods map a multidimensional input vector to a single output classifier to separate one variable from another. MVA algorithms are based on multivariate statistics. A common task these algorithms are used for is the separation of a dataset into signal and background. They are used for this purpose in this thesis as well. Machine learning offers an effective way to learn a statistical model using an appropriate data-set. The knowledge is extracted from experience, and therefore this approach is a simple form of artificial intelligence. Machine learning algorithms include a fitting phase, during which the statistical model is learned by studying the provided training data-set. In the course of the interference phase the statistical model is used to infer the desired information for a new independent test data-set. There are three different types of learning: supervised learning, where the target information is provided during the fitting-phase, unsupervised learning, here no additional information is given during the fitting-phase, and reinforced learning, where rewards and punishments are provided in a dynamic environment. In this thesis the target information is given by Monte Carlo simulated data, therefore supervised learning is employed. In supervised learning the desired information is predicted by a statistical model. This model has internal degrees of freedom, called weights. These can be adapted by the algorithm to minimize the discrepancy between the true and predicted value. The discrepancy is defined by a loss function [20].

In this analysis boosted decision trees (BDT) are used. They construct a more robust classification model by sequentially constructing shallow decision trees. Decision trees perform classifications using a number of consecutive cuts. The maximum number of cuts is called the depth of the tree. The cuts are determined during the fitting-phase. The decision tree is often dominated by statistical fluctuations in the training data-set. In a BDT the decision trees are fitted so that the expectation value of a negative binomial log-likelihood loss function is minimized. The depth of the individual decision trees is strongly limited to avoid over fitting. A single decision tree is able to separate signal and

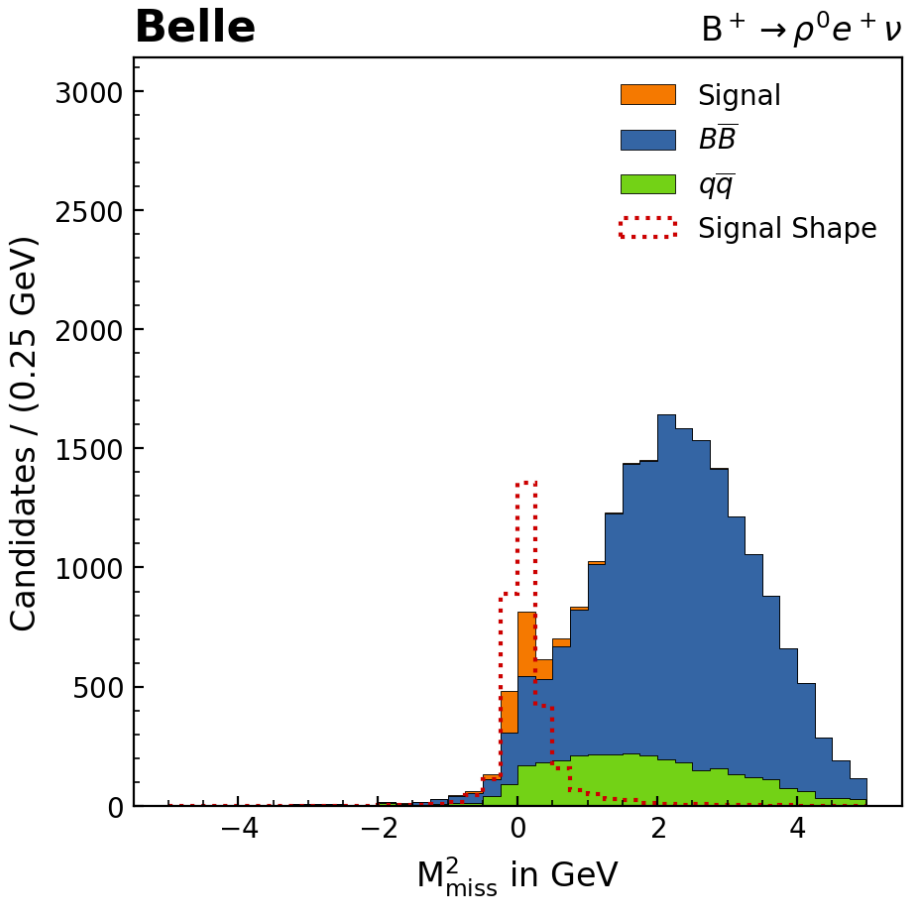


Figure 5.1: M^2_{miss} distribution for $B^+ \to \rho^0 e^+ \nu$ channel before continuum suppression

background only approximately (so called weak learner). Using many weak learners a well regularized classifier with large separation power is constructed. The number of trees is equivalent to the number of boosting steps [21]. A schematic drawing of a decision can be seen in Figure 5.2.

5.2 Event shape Variables

The $\Upsilon(4S)$ energy is just above the $B\overline{B}$ production energy threshold, therefore the B mesons in $B\overline{B}$ events are produced almost at rest in the $\Upsilon(4S)$ rest-frame. This leads to a nearly spherical, isotropical distribution of the B decay products. The quarks in continuum events, on the other hand, are produced with a large initial momentum, which results in back-to-back jets of light hadrons, due to fragmentation. These event shape differences visible in the angular distributions are used to distinguish $B\overline{B}$ events from continuum events. An illustration of the different event shapes can be seen in Figure 5.3.

The event shape variables describe the phase-space distribution of particles contained in an event. In the following sections the event shape variables that are used to distinguish between continuum and $B\overline{B}$ events, are described. The explanations are adapted from Ref. [11] if not stated differently.

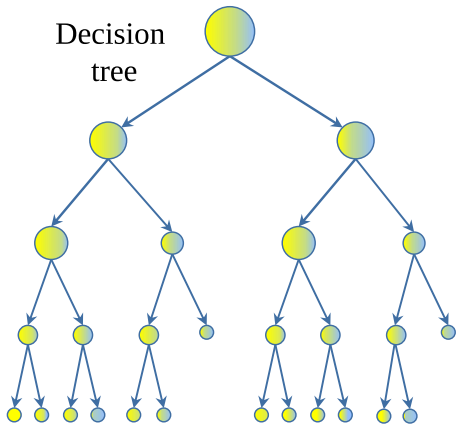


Figure 5.2: Predictive models of decision trees Ref. [22].

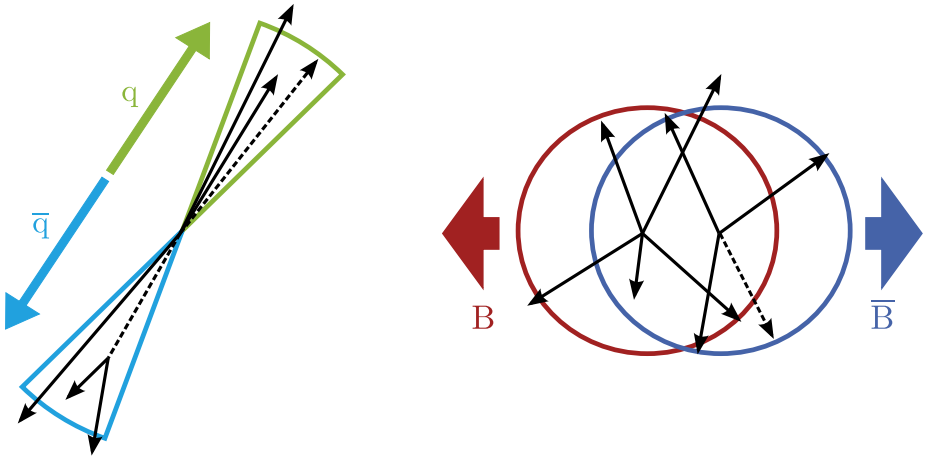


Figure 5.3: Schematic drawing of the jet-like structure of continuum events (left) and the more spherical distribution of $B\overline{B}$ events (right) Ref. [23].

5.2.1 Thrust

Thrust was originally used to describe jets. The thrust axis is the axis that maximizes the sum of longitudinal momenta. It is defined as the unit vector along the total momentum projection. The thrust is defined as:

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

Two different thrust axes are calculated for each event. One is calculated for the visible decay products of the signal side B decay and the other for the rest of the event. These two axes are uncorrelated for $B\overline{B}$ events due to the isotropic distributions of the B decays. The thrust axis of the B candidate and that of the rest of the event are collinear, while in continuum events the particles are aligned along the direction of the two jets. Four features of the thrust are used in continuum suppression. The magnitude of the thrust axis of the B candidate, as well as the magnitude of the thrust axis of the rest of the event is helpful. Additionally, the cosine of the angle between the two thrust axes is used, due to

the fact that an angle between two vectors in three dimensions is flat in $\cos\theta$. The magnitude of the cosine of the angle between the two thrust axes $|\cos\Delta\theta_T| = |\hat{T}_{\text{cand}} - \hat{T}_{\text{ROE}}|$ is uniformly distributed between 0 and 1 for signal, while it strongly peaks at 1 for continuum events. Furthermore, the cosine of the angle between the thrust axis of the signal B candidate and the beam axis can be used [7].

5.2.2 CLEO Cones

CLEO cones were introduced by the CLEO collaboration, they are cones that measure the scalar momentum flow around the thrust axis. They are concentric cones in angular intervals of 10° resulting in 9 cones in total. They measure the magnitude of the particle momenta, falling into the different cones around the thrust axis. The flight direction of the decay products for jet-like continuum events are correlated, while in $B\overline{B}$ events they are uncorrelated.

5.2.3 Fox-Wolfram Moments

The Fox-Wolfram moments are a parameterization of the phase-space distribution of momentum and energy flow in an event. The k-th order Fox-Wolfram moment H_k, for a collection of N particles with momenta p_i is defined as

$$H_k = \sum_{i,j}^{N} |p_i||p_j| P_k(\cos\theta_{i,j}).$$

$\theta_{i,j}$ is here the angle between p_i and p_j and P_k is the k-th order Legendre polynomial. In the limit of vanishing particle masses $H_0 = 1$, therefore the normalized ratio $R_k = H_k/H_0$ is often used. For events with two strongly collimated jets, R_k then takes values close to zero for odd, and values close to one for even values of k. R_2 provides a good discrimination power between $B\overline{B}$ and continuum events. The normalized Fox-Wolfram moments are defined as

$$h_l^k = \frac{\sum_{m,n} |\vec{p_m}||\vec{p_n}| P_l(\cos\theta_{mn})}{\sum_{m,n} |\vec{p_m}||\vec{p_n}|}.$$

k categorizes the order of Fox-Wolfram moment, m indicates the signal B daughter, while n indicates the rest of event [11].

5.3 Continuum Suppression Study

In this thesis three different versions of continuum suppression are presented in total. One has been performed with all of the cuts already tightened. The specific cuts can be seen in the following chapters. A second one were only the loose online cuts are applied, and for the third version only a charged π track was reconstructed.
For continuum suppression only variables describing the whole event are used. A BDT is used to distinguish between continuum and signal events. The test and training samples contain a 50 : 50 combination of signal and continuum events. The ratio of event in the training to the test sample is 70 : 30. First a large number of event shape variables is used in the training. Then, the ones with the highest discrimination powers are used. A FastBDT is used, for all the mentioned trainings. As

hyperparameters the number of trees is set to 200, the depth of trees is 4, the number of cuts is 8, the shrinkage is 0.05 and the sampling rate is 0.5.

5.3.1 Continuum Suppression on Samples with Tight Cuts

For the first continuum suppression (CS) version tight cuts is applied. That means, for example, that for the $B^+ \to \rho^0 \ell^+ \nu$ channel in addition to applying all cuts mentioned in chapter 4.4 the cuts shown in table 5.1 are applied. The individual variables are explained later.

Variable	Selection
M_{bc}	> 5.27 GeV
$\lvert\Delta E\rvert$	$< 0.\ 1$GeV
FEI signal probability	$> 10^{-4}$
$\lvert\cos\theta_{BY}\rvert$	< 3 GeV
E_{ECL}	< 1.5
$0.4773\ \text{GeV} < M_{\pi^+\pi^-} < 1.0737\ \text{GeV}$	

Table 5.1: Tightened cuts for CS.

Separate trainings are performed for the all channels, but the same variables are used for training all channels. The input variables are:

- $\cos\theta_{TB}$
- h_{20}^{so}
- R_2
- E_t
- $|\vec{T}_B|$

The normalized distributions of these variables for signal $q\overline{q}$ and $B\overline{B}$ background events can be seen in figure 5.4. For the $B^+ \to \rho^0 e^+ \nu$ channel. The other channels are shown in the appendix. A separation between signal and continuum events can be seen in all variables. E_t is the transverse energy of an event, $\cos\theta_{TB}$ is the cosine of the angle between the two thrust axes, h_{20}^{so} is one of the Fox-Wolfram moments, $|\vec{T}_B|$ is the magnitude of the thrust axis of the B candidate, and R_2 is the normalized ratio R_k with $k = 2$.

The performance of the BDT on the training and testing sample can be seen in figure 5.5. The Kolmogorov-Smirnov overtraining test plot is included as well. This test results in p-values of 0.41 for signal and 0.16 for continuum. These p-values suggest that the sample is not overtrained and that both test and training samples are drawn from the same data-set.

In figure 5.6 the receiver operating characteristic (ROC) curve for the BDT training is shown. The ROC curve shows background versus signal efficiency. For the test sample the area under the ROC curve is 0.97. An area under the ROC curve of 1 would mean that all of the signal can be retained, while 100% of the background can be rejected. Therefore, a value of 0.97 shows that the BDT can distinguish well between signal and continuum events.

However, with the tight cuts, it is only possible to train a BDT for the ρ^0, $\rho^\pm$ and the $\omega \to \pi^+\pi^-\pi^0$ channels. The statistics in the $\omega \to \pi^0\gamma$ channel are too low to train the mode separately, and no reasonable p-values could be attained.

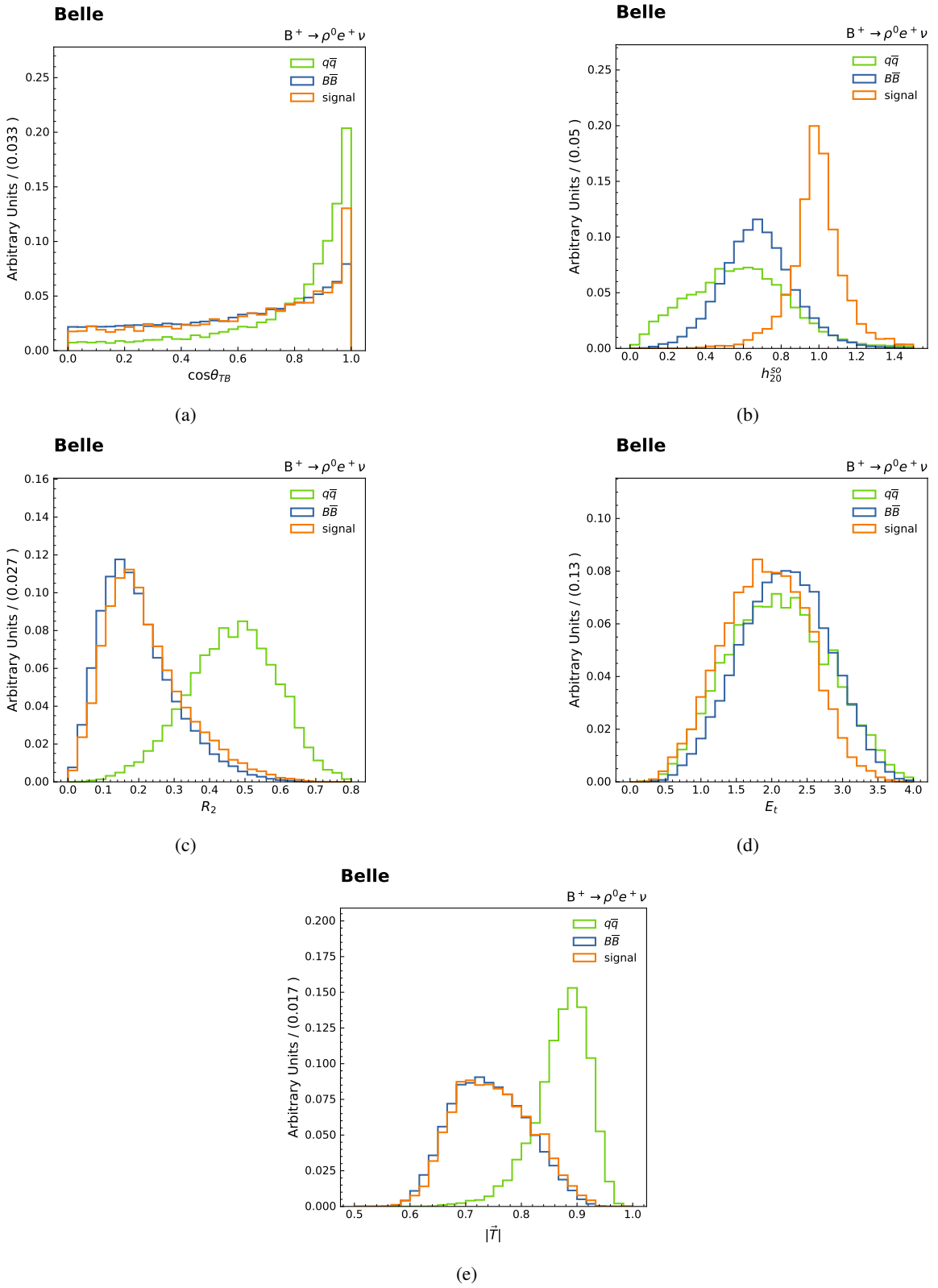


Figure 5.4: Normalized distributions of the CS variables for the $B^+ \to \rho^0 e^+ \nu$ channel in $q\overline{q}$, $B\overline{B}$ and signal categories from MC.

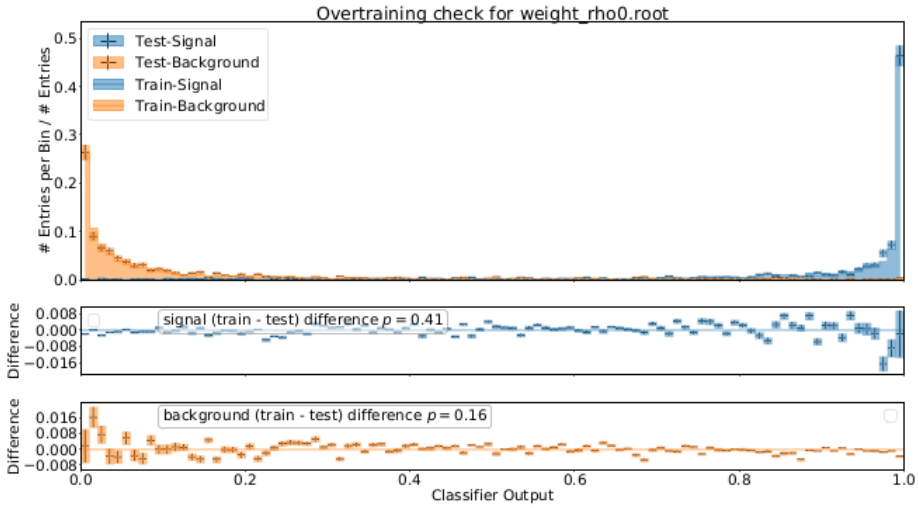


Figure 5.5: BDT classifier output for continuum and signal events or the test and training samples. The overtraining check is shown in the bottom.

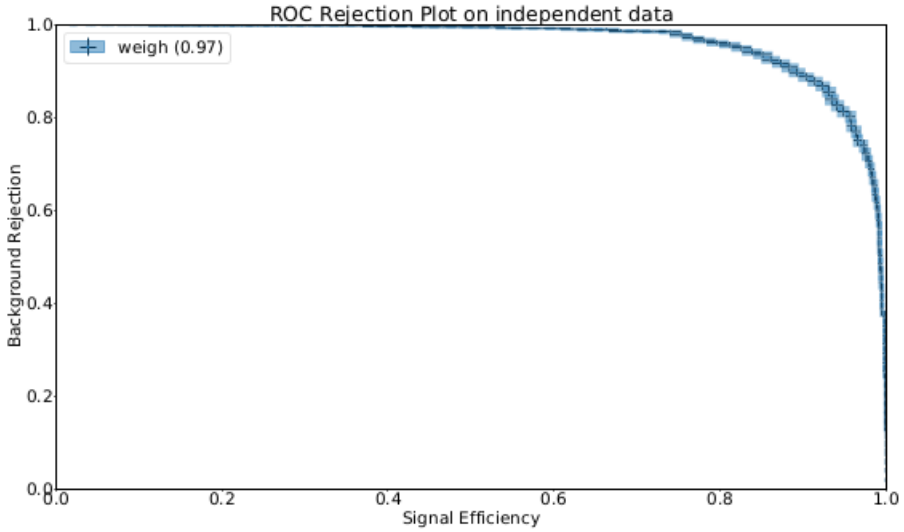


Figure 5.6: ROC curve from independent test sample.

5.3.2 Continuum Suppression on Sample with Loose Cuts

Since the statistics are too low after applying the tight cuts, a training with only the preselections described in chapter 4.4.2 is done. Again, separate trainings for each decay channel are performed. This time, however, slightly different input variables are chosen, depending on the channel. The $B^+ \to \rho^0 \ell^+ \nu$ channel is trained on the following variables:

- $|\vec{T_{ROE}}|$
- $cc8$
- $cc9$
- $\cos\theta_{TB}$
- h_{14}^{so}
- $cc7$
- $cc3$
- h_{12}^{so}
- $cc6$
- $cc2$
- h_{10}^{so}
- $cc5$
- $cc4$
- $\cos\theta_{Bz}$
- $cc1$
- E_t
- h_{20}^{so}
- $|\vec{T_B}|$
- R_2

Here $|\vec{T_{ROE}}|$ is the magnitude of the thrust axis of the ROE, the cc variables are the Cleo cones and $\cos\theta_{Bz}$ is the cosine of the angle between the thrust axis of the B candidate and the beam axis. In figure 5.7 the performance of the BDT on the test and training samples, as well as the overtraining test can be seen. The p-values are now 0.37 for signal and 0.34 for continuum. Therefore the BDT is not overtrained.

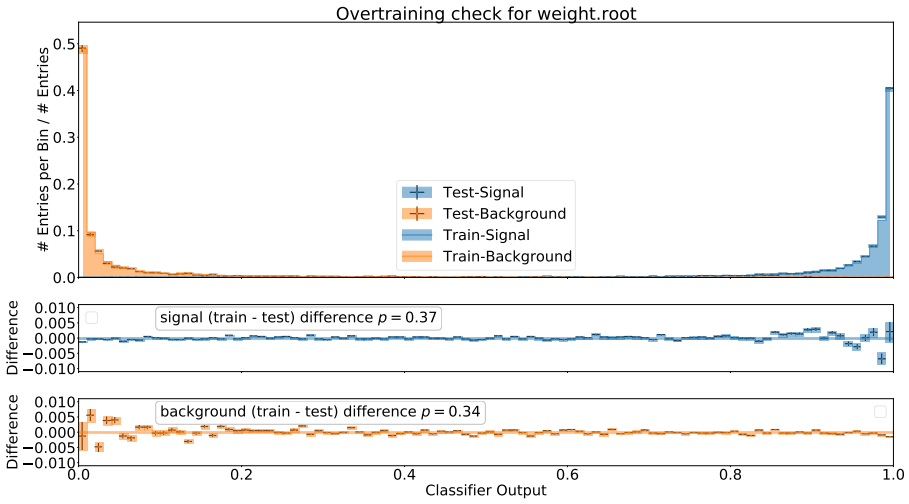


Figure 5.7: BDT classifier output for continuum and signal events or the test and training samples. The overtraining check is shown in the bottom.

In figure 5.8 the ROC curve for the training with loose cuts is shown. The area under the curve is 0.96, and the BDT can differentiate well between signal and background. However even without the tight

cuts it is not possible to obtain enough statistics to do a separate training for the $\omega \to \pi^0\gamma$ channel, so no reasonable p-values can be obtained.

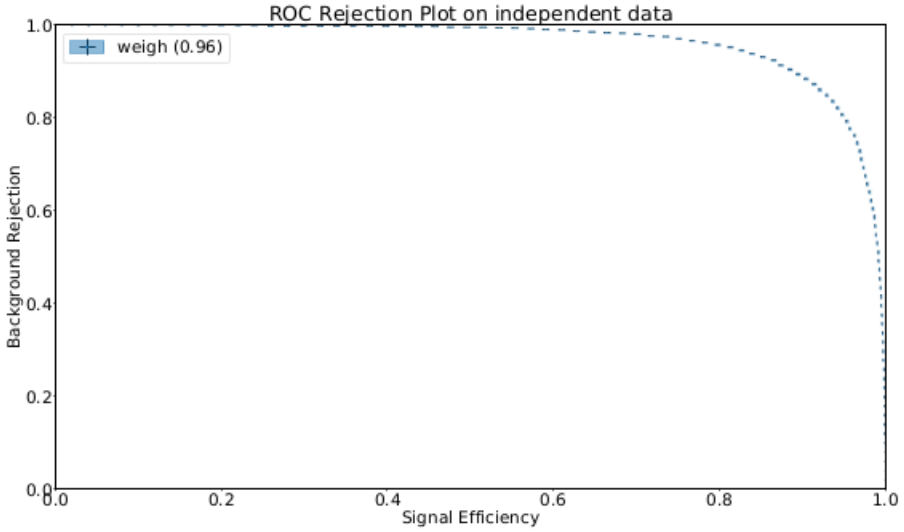


Figure 5.8: ROC curve from independent test sample.

5.3.3 Continuum Suppression on Sample with 1 Reconstructed Track

Due to insufficient statistics the next approach is to reconstruct only one charged track, where all tracks are assigned the π mass hypothesis. Therefore a π^- track is reconstructed. The idea is that signal for $X_u\ell\nu$ still has the $B\overline{B}$ characteristic and the continuum a jet-like event shape. Therefore a BDT can still be trained on the event shape variables. This has the advantage that a very high statistics sample is available and one training would suit all channels with a $B^\pm$ meson and another for all channels with a B^0 meson. The variables trained on for the $B^\pm$ decays are:

- h_2^{oo}
- R_2
- E_t
- $cc1$
- $\cos\theta_{TB}$
- $\cos\theta_{Bz}$
- $cc3$
- $cc4$
- $cc2$
- $cc5$
- $cc6$
- $cc7$

In figure 5.9 the performance of the BDT on the test and training sample can be seen, together with the overtraining test. The p-value for signal is 0.95 and the p-value for background is 0.31. Therefore, overtraining is not a problem.
Taking a closer look at the ROC curve shown in figure 5.10, however, it can be seen that this training does not perform as well as the decay specific trainings do. Here, the area under the ROC curve is only 0.85, while the signal specific training has an area under the ROC curve greater than 0.97. That means that this training does not perform as well as the others in differentiating signal from background events. Therefore this training is not chosen for the further analysis.

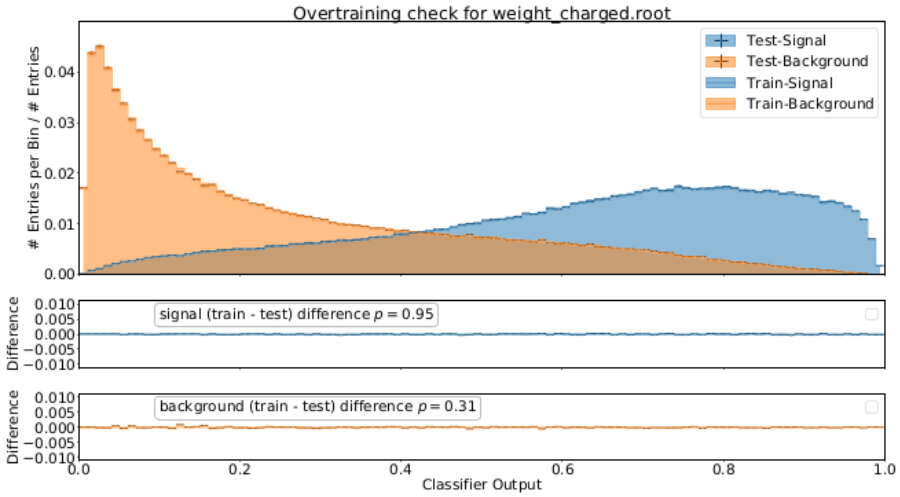


Figure 5.9: BDT classifier output for continuum and signal events or the test and training samples. The overtraining check is shown in the bottom

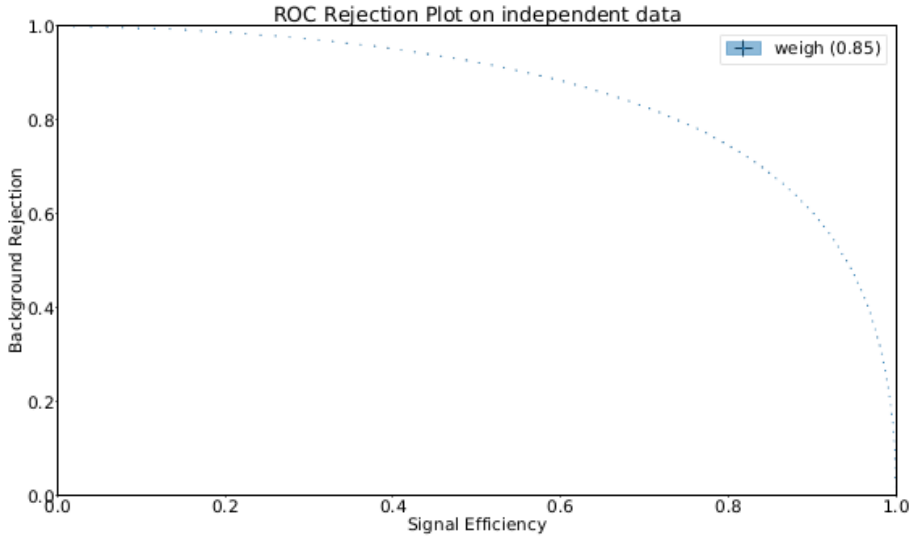


Figure 5.10: ROC curve from independent test sample.

5.3.4 Comparison

To decide whether to use the tight or loose training for each mode, the different trainings are applied to MC simulated data and the figure of merit (FOM) is calculated as:

$$\text{FOM} = \frac{N_{\text{signal}}}{\sqrt{N_{\text{signal}} + N_{\text{background}}}}$$

The training which gives the highest FOM for each channel is used. This results in the use of the loose $\rho^{\pm}$ training for the $\rho^{\pm}$ channel and the loose ρ^{0} training for the ρ^{0}, as well as the ω channels.

In figure 5.11 the M^2_{miss} distribution for the $B^+ \to \rho^0 e^+ \nu$ channel, after applying CS, is shown. 94.29% of continuum background are rejected, while only 9.80% of signal events are removed. Additionally, 27.41% of $B\overline{B}$ background events are rejected.

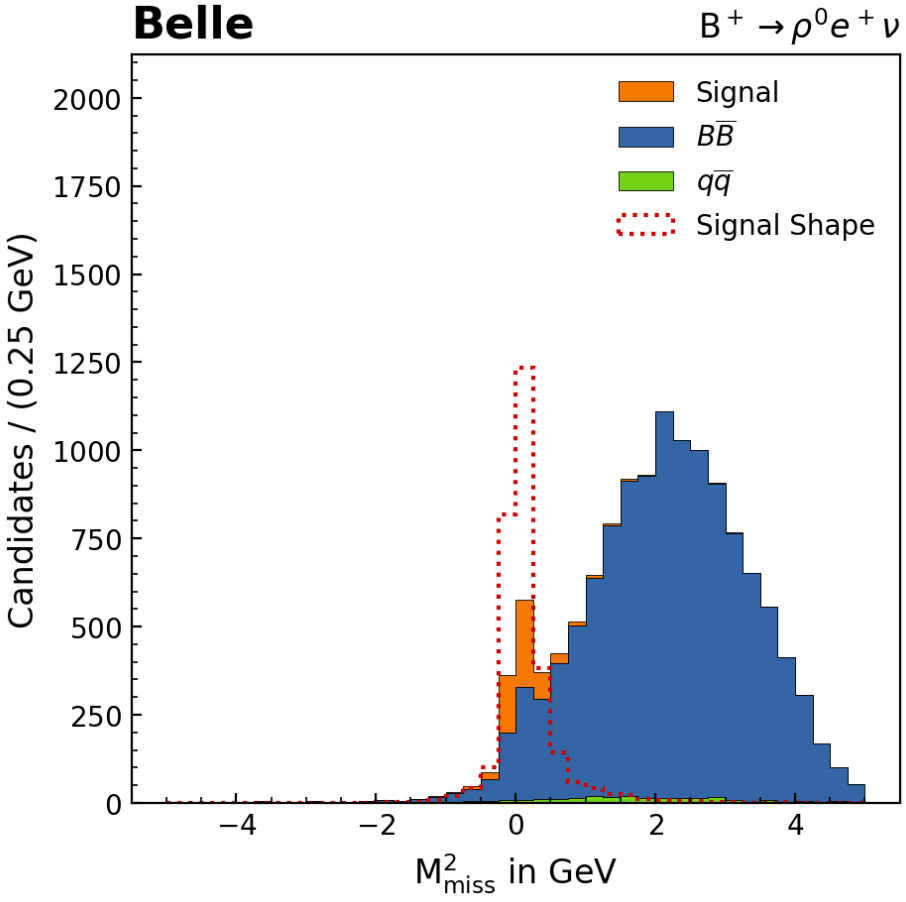


Figure 5.11: M^2_{miss} distribution for the $B^+ \to \rho^0 e^+ \nu$ channel after continuum suppression.

CHAPTER 6

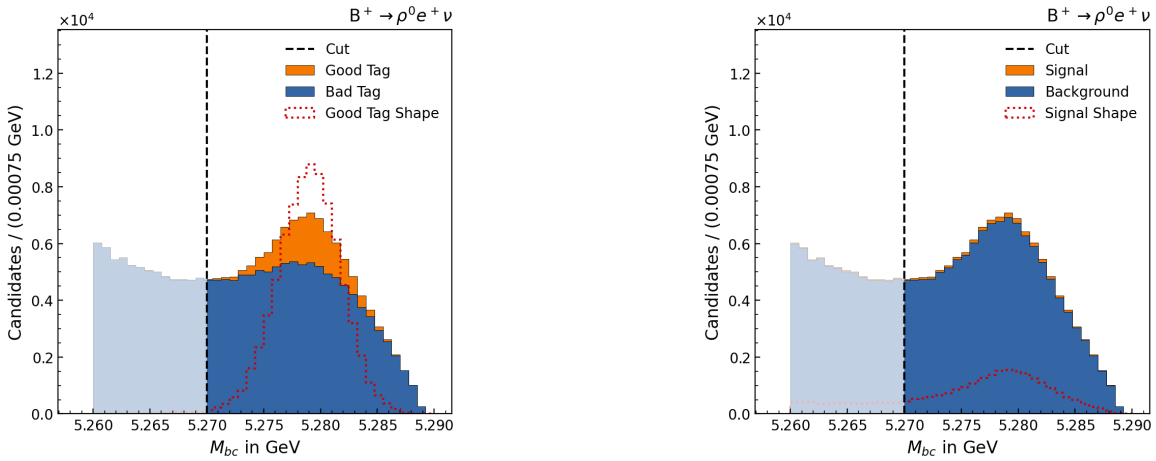
Tag Selections

In this chapter the selections on the tag-side variables are revisited. The variables are M_{bc}, ΔE and the FEI signal probability. The data set is split up in good and bad tag categories. Good tag candidates are events where the B_{tag} meson is correctly reconstructed, while bad tag candidates are events where the B_{tag} is not correctly reconstructed. The cuts are optimised by eye. A variable that gives information whether or not the particle is correctly reconstructed, is used to determine whether the events fall into the good tag or bad tag category. It has a value of one if the B meson on the tag side is correctly reconstructed and a value of zero if it is not correctly reconstructed. The goal of the tag-side cuts is to purify the data by removing as many bad tag candidates and keeping as many good tag candidates as possible. That means increasing the purity inside the peak of the M^2_{miss} distribution by decreasing the number of events where particles from the signal and tag side are mixed. Furthermore, B particles with faulty reconstruction are removed. This M^2_{miss} distribution is going to play an important role in the further analysis. In this chapter the distributions are shown for the $B^+ \to \rho^0 e^+ \nu$ channel, but the cuts are applied to all channels. The plots for the other channels can be found in the appendix.

6.1 Beam Constrained Mass

The beam constrained mass (M_{bc}) should peak around the B meson mass of 5.28 GeV, as discussed in chapter 4.4.1. The M_{bc} distribution, split into good and bad tag categories, can be seen in figure 6.1(a), for the $B^+ \to \rho^0 e^+ \nu$ channel. The distribution is peaking around the B meson mass for the good tag candidates. In figure 6.1(b) the M_{bc} distribution for the $B^+ \to \rho^0 e^+ \nu$ channel, split into signal and background categories, is shown. Both signal and background events are peaking around the B meson mass. The signal distribution is flat below 5.27 GeV. Therefore, events with M_{bc} smaller than 5.27 GeV are removed.

By implementing this cut 99.45% of good tags are retained, while 63.58% of bad tag candidates are retained. For the signal and background categories, this leads to keeping 80.32% of signal and 66.47% of background events.



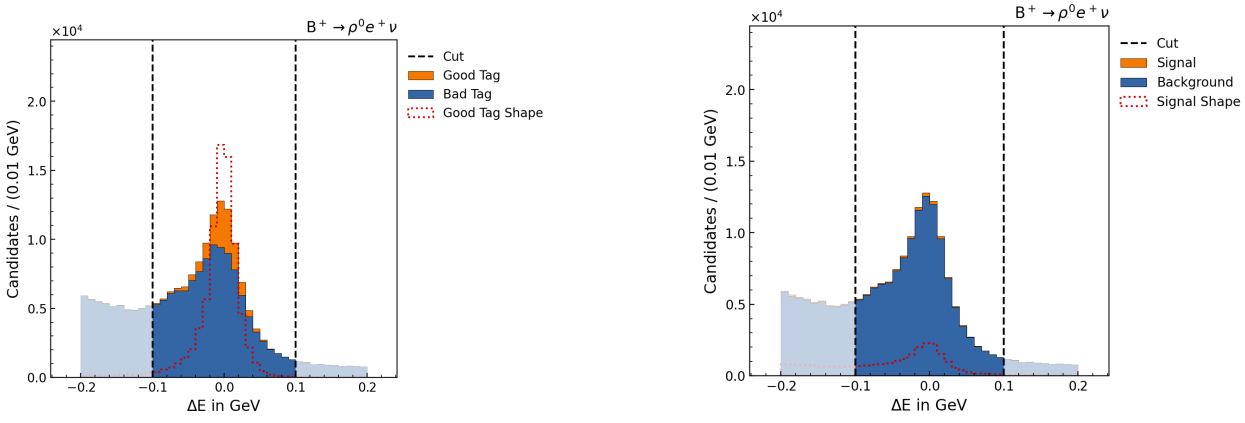
(a) (b)

Figure 6.1: M_{bc} distributions for the channel $B^+ \to \rho^0 e^+ \nu$. In a) the distribution is divided into good and bad tag categories and in b) the distribution is divided into signal and background categories.

6.2 Energy difference

The energy difference (ΔE), between the energy of the B meson and the beam energy should be close to zero for a correctly reconstructed B meson as discussed in chapter 4.4.1. The ΔE distribution can be found in figure 6.2. In figure 6.2(a) the distribution is shown split into the good and bad tag categories. The good tag category as well as the bad tag category is peaking around zero. The split into signal and background categories can be seen in 6.2(b), and signal and background events peak around zero as well. Therefore a window of $|\Delta E| < 0.1$ is selected.

By implementing this cut window 97.93% of the good tag category is kept, as well as 65.23% of the bad tag category is retained. This selection keeps 73.79% of signal and 70.19% of background.



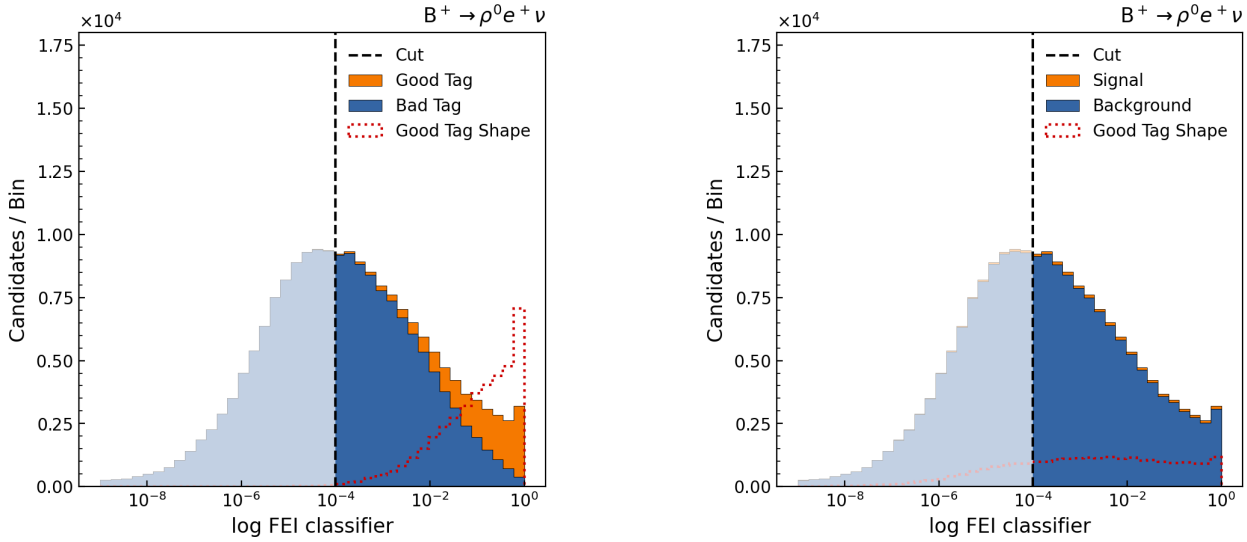
(a) (b)

Figure 6.2: ΔE distributions for the channel $B^+ \to \rho^0 e^+ \nu$. a) divided into good and bad tag categories and b) divided into signal and background categories.

6.3 FEI Signal Probability

The signal probability is a variable that indicates the confidence that the FEI Has in its tag reconstruction, and takes values between 0 and 1. It is expected to peak at one for correctly reconstructed tag candidates as discussed in chapter 4.4.1. The distribution for the good and bad tag categories can be seen in figure 6.3(a). The good tag candidates peak around one and the tail ends at 10^{-4}, while the bad tag candidates peak at 10^{-4}. In figure 6.3(b) the distribution for signal and background candidates can be seen. For signal candidates the distribution peaks around one with a long tail to smaller values and for background events the distribution peaks again around 10^{-4}. For this reason events with signal probability greater 10^{-4} are selected.

This cut keeps 99.73 % of the good tag category and 73.73 % of the bad tag category. 89.06 % for signal and 78.58 % of background events are kept.



(a) (b)

Figure 6.3: FEI signal probability distributions for the channel $B^+ \to \rho^0 e^+ \nu$. a) divided into good and bad tag categories and b) divided into signal and background categories.

CHAPTER 7

Signal Selection

After selecting tighter cuts on the tag side, cuts on the signal side are applied to further reduce background events. Here the simulated data is split into the signal, $B\overline{B}$ background and $q\overline{q}$ continuum background categories. In this chapter the distributions are shown for the $B^+ \to \rho^0 e^+ \nu$ channel. When additional cuts are applied to other channels, the corresponding distributions are shown for the other channels,

7.1 $\cos\theta_{BY}$

$\cos\theta_{BY}$ is the cosine of the angle between the B candidate and the visible system Y, as discussed in chapter 4.4.2. A sketch of the angle is shown in figure 7.1.

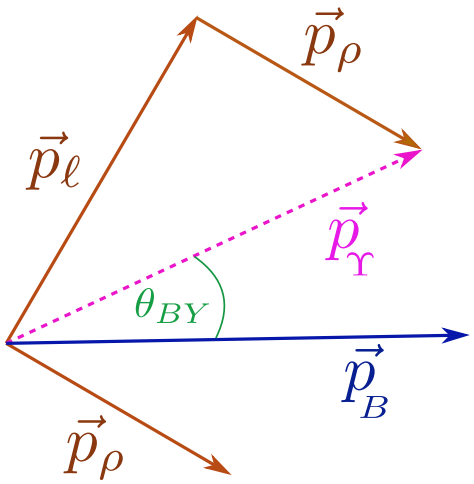


Figure 7.1: Illustration of the angle θ_{BY} of $\cos\theta_{BY}$.

For incorrectly reconstructed B mesons and other background, this variable does not correspond to the cosine of a physical angle. Therefore, the distribution can spread out over the boundaries of -1 and 1. Even for correctly reconstructed B candidates the distribution can contain events slightly outside the window between -1 and 1. This is due to detector resolution and photon radiation effects.

The distribution is shown in figure 7.2 for the $B^+ \to \rho^0 e^+ \nu$ channel. It can be seen that the signal mainly populates the window between −2 and 2. The background, on the other hand, peaks around −4. A window between −3 and 3 is selected to keep nearly all correctly reconstructed events, while rejecting a large amount of background, and allowing for resolution effects. This cut is applied to all channels. The percentages of each event category within each channel that are kept after the cut can be seen in table 7.1

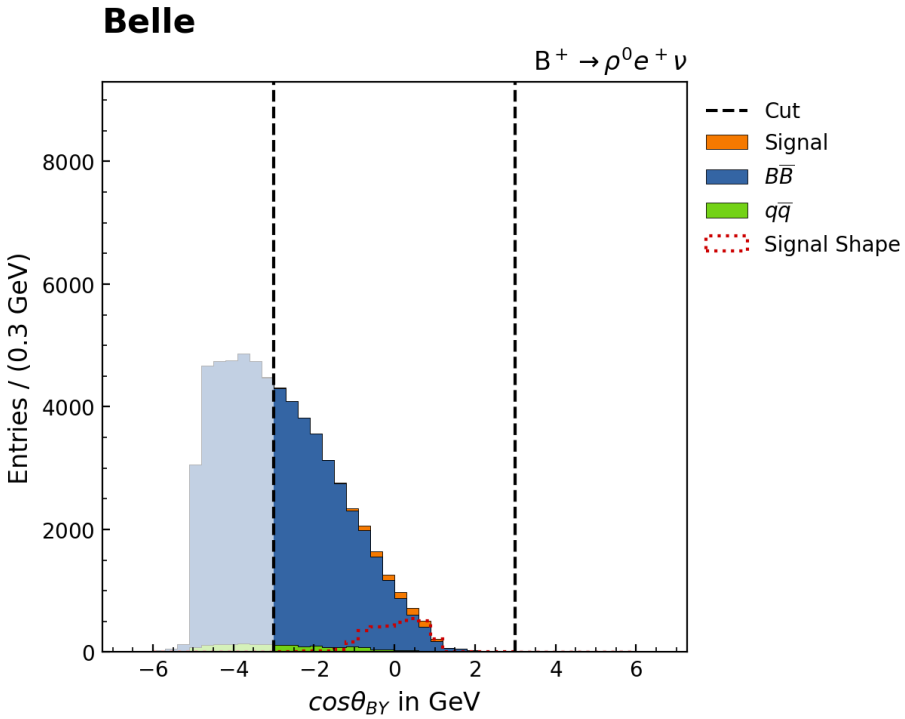


Figure 7.2: $\cos\theta_{BY}$ distribution for the $B^+ \to \rho^0 e^+ \nu$ channel.

Channel	Signal	$B\overline{B}$	$q\overline{q}$
ρ^0	98.58%	63.82%	52.10%
ρ^-	99.66%	55.90%	51.55%
$\omega \to \pi^0\pi^+\pi^-$	99.31%	47.38%	37.07%
$\omega \to \pi^0\gamma$	100.00%	48.59%	35.02%

Table 7.1: Percentages per category and channel that survive the $\cos\theta_{BY}$ selection.

7.2 Extra Energy in the Electromagnetic Calorimeter

The extra energy in the electromagnetic calorimeter (E_{ECL}) originates from clusters that are not assigned to the reconstructed final state particles. If the signal and tag-side particles are correctly reconstructed, there are no unreconstructed particles in the electromagnetic calorimeter (ECL). In that

case, there are only photons from material interaction and beam background photons left. Therefore, the distribution has to peak around zero for correctly reconstructed signal events. Background events, on the other hand, can have higher E_{ECL} values due to unreconstructed particles that deposite energy in the calorimeter. The distribution of E_{ECL} is given in figure 7.3.

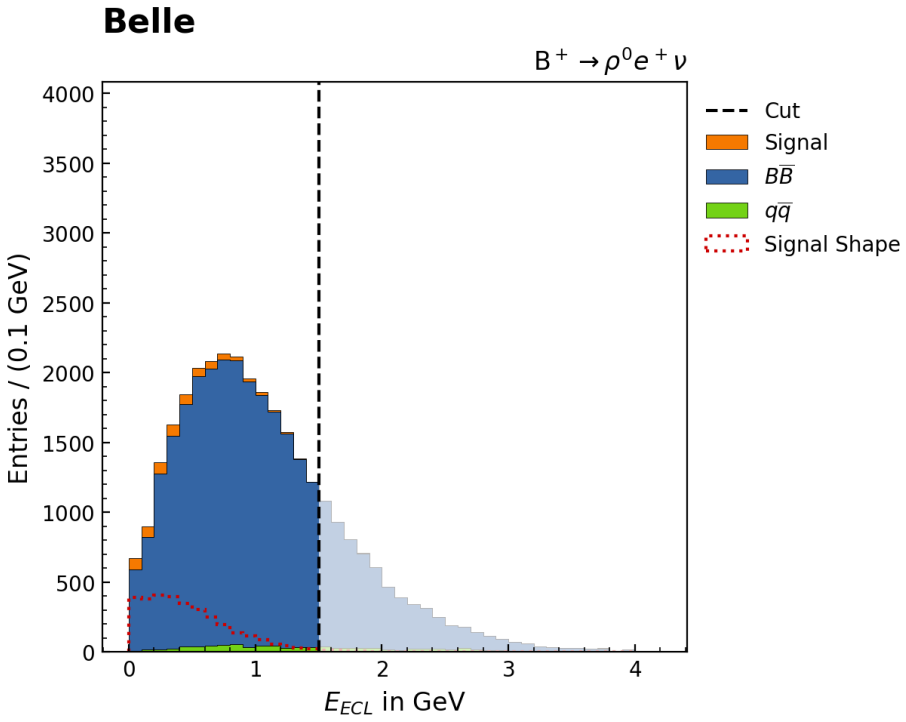


Figure 7.3: E_{ECL} distribution for the $B^+ \to \rho^0 e^+ \nu$ channel.

Looking at the signal it can be seen that it peaks around zero, while the background peaks around one. Due to beam backgrounds, and incorrectly assigned energies the signal peaks around but not directly at zero. To keep nearly all correctly reconstructed events a cut of $E_{\mathrm{ECL}} < 1.5$ GeV is implemented. This selection is introduced for all channels. The percentages of the event categories within each channel after the selection, can be seen in table 7.2.

channel	signal	$B\overline{B}$	$q\overline{q}$
ρ^0	97.36%	84.40%	55.47%
ρ^-	100.00%	91.60%	60.19%
$\omega \to \pi^0\pi^+\pi^-$	100.00%	83.03%	58.53%
$\omega \to \pi^0\gamma$	100.00%	83.78%	84.09%

Table 7.2: Percentages per category and channel that survive the E_{ECL} selection.

7.2.1 Invariant Mass

The invariant mass of the daughters of the signal hadron ($M_{\pi^+\pi^-}$) should peak around the mass of the hadron if it has been correctly reconstructed. For the $B^+ \to \rho^0 \ell^* \nu$ channel this is the case for signal events and true ρ^0 in $B\overline{B}$ events. There is an additional peak in the background, which lies at the K_s mass of 0.497 GeV, in the distribution for the ρ^0 channel. The $M_{\pi^+\pi^-}$ distribution for ρ^0 is shown in figure 7.4, here true ρ^0 in $B\overline{B}$ background is shown as well.

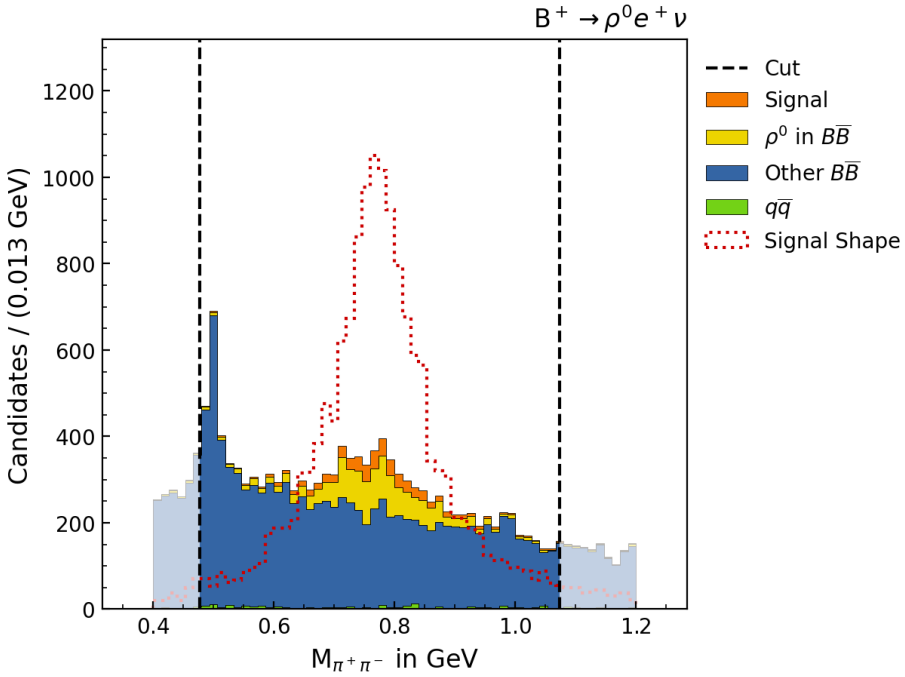


Figure 7.4: $M_{\pi^+\pi^-}$ distribution for the $B^+ \to \rho^0 e^+ \nu$ channel.

The cut window for the $B^+ \to \rho^0 \ell^+ \nu$ and $B^0 \to \rho^- \ell^+ \nu$ channels is set between 0.48 GeV and 1.07 GeV.

In the $\omega \to \pi^0\pi^+\pi^-$ and the $\omega \to \pi^0\gamma$ channels, a mass window around the ω meson mass is introduced between 0.74 GeV and 0.81 GeV. This can be seen in figure 7.5 for the channel $\omega \to \pi^0\pi^+\pi^-$.

A table with the surviving candidates per category within each channel in percent is shown in table 7.3.

channel	signal	$B\overline{B}$	$q\overline{q}$
ρ^0	92.71%	56.90%	41.73%
ρ^-	99.88%	70.62%	48.80%
$\omega \to \pi^0\pi^+\pi^-$	89.42%	34.56%	28.75%
$\omega \to \pi^0\gamma$	100.00%	38.88%	45.00%

Table 7.3: Percentages per category and channel that survive the invariant mass cuts.

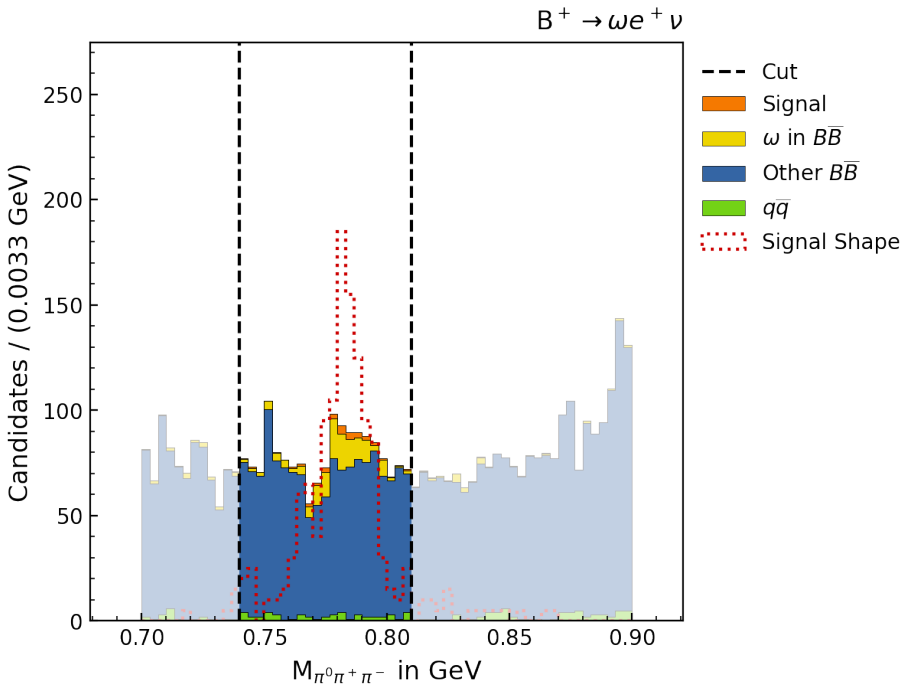


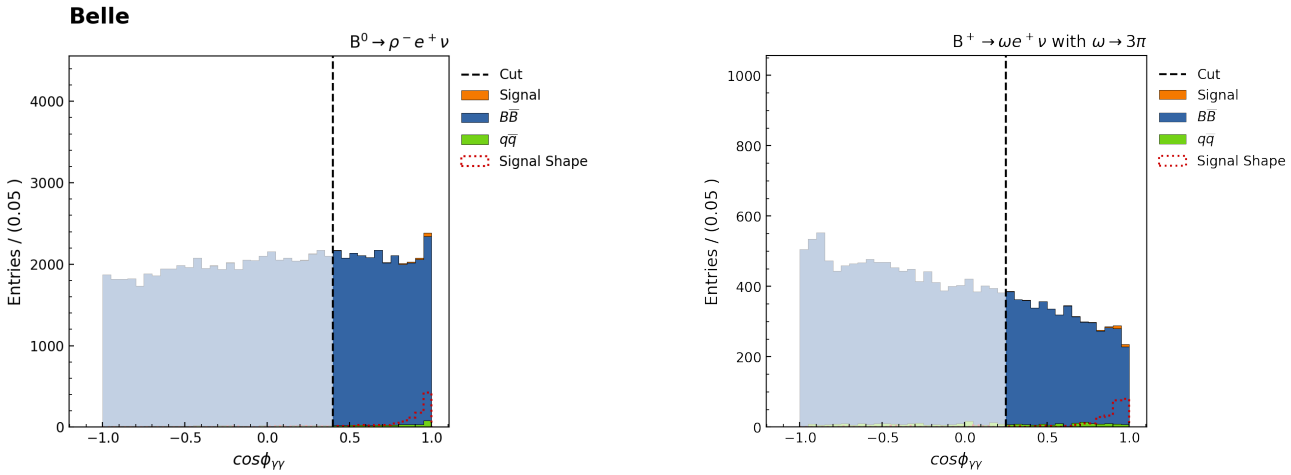
Figure 7.5: $M_{\pi^0\pi^+\pi^-}$ distribution for the channel $B^+ \to \omega(\to \pi^0\pi^+\pi^-)e^+\nu$.

7.3 Best Candidate Selection

Expect for the $B^+ \to \rho^0\ell^+\nu$ channel, there are multiple candidates per event for all channels. This makes a best candidate selection necessary. For the $B^0 \to \rho^-\ell^+\nu$ channel the event containing the $\pi^-\pi^0$ pair with the largest energy is kept as the best candidate. This results in a signal efficiency of 0.79 for the signal. For the $B^+ \to \omega\ell^+\nu$ channels a random best candidate selection is performed. This leads to a signal efficiency of 0.47 for the $\omega \to \pi^0\pi^+\pi^-$ channel and for the $\omega \to \pi^0\gamma$ channel to a signal efficiency of 0.59. The signal efficiencies here are that low, because the best candidate selection is the second cut applied to the channels.

7.4 $\cos\phi_{\gamma\gamma}$

An additional cut is applied for the $B^0 \to \rho^-\ell^+\nu$ channel and both the $B^+ \to \omega\ell^+\nu$ channels to reduce further background. This is a cut on the cosine of the angle between the photons from the π^0 decay $\cos\phi_{\gamma\gamma}$ in the lab frame. Since there is no reconstructed π^0 in the $B^+ \to \rho^0(\to \pi^+\pi^-)e^+\nu$ channel, this cut is not necessary. The π^0 have high momenta, therefore their daughters, which are photons, are produced boosted. So they are mostly colliner, consequently they peak at 1 in the cosine, while random combinations of photons are uniform in the cosine. That is the reason why event with $\cos\phi_{\gamma\gamma} < 0.4$ are removed for the $B^0 \to \rho^-\ell^+\nu$ channel. The distribution is shown in figure 7.6(a). For the $B^+ \to \omega\ell^+\nu$ channels the selection is loosened to 0.25 to keep more events in total. This can be seen in figure 7.6(b).



(a) (b)

Figure 7.6: The $\cos\phi_{\gamma\gamma}$ distributions for the $B^0 \to \rho^- e^+ \nu$ channel (a) and the $B^+ \to \omega(\to \pi^0\pi^+\pi^-)e^+\nu$ channel (b)

The retained percentages of each category in each channel after the cut can be seen in table 7.4.

channel	signal	$B\overline{B}$	$q\overline{q}$
ρ^-	97.00%	88.78%	96.15%
$\omega \to \pi^0\pi^+\pi^-$	95.14%	50.09%	52.98%
$\omega \to \pi^0\gamma$	100.00%	83.78%	84.09%

Table 7.4: Percentages per category and channel that survive the $\cos\phi_{\gamma\gamma}$ cut.

7.5 $\cos\theta_{\omega\gamma}$

For the $\omega \to \pi^0\gamma$ channel an additional cut is applied on the absolute value of the cosine of the angle between the photon from the $B^+ \to \omega\ell^+\nu$ decay in the ω rest frame and the ω direction in the laboratory frame $|\cos\theta_{\omega\gamma}|$. A sketch of the angle can be found in figure 7.7. The reason for this cut is that the angle is bound by momentum and energy conservation due to the nature of the two body decay of the ω. The distribution is shown in figure 7.8. The signal is peaking around zero as expected. The selection is set to $|\cos\theta_{\omega\gamma} < 0.5|$. 90.91 % of signal, 59.85 % of $B\overline{B}$ background and 54.05 % of continuum events survive this cut.

7.6 Summary of the Signal Selections

A summary table for all applied signal selections for all channels is shown in table 7.5.

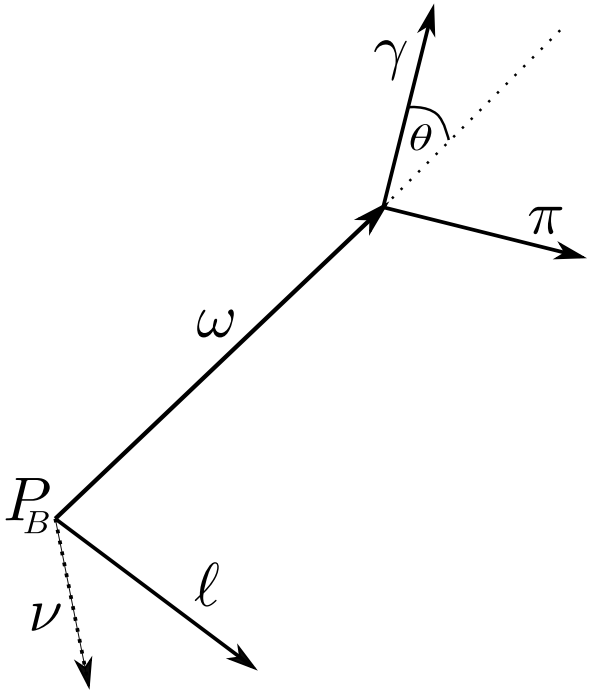


Figure 7.7: A sketch of the $\theta_{\omega\gamma}$ angle of $\cos\theta_{\omega\gamma}$.

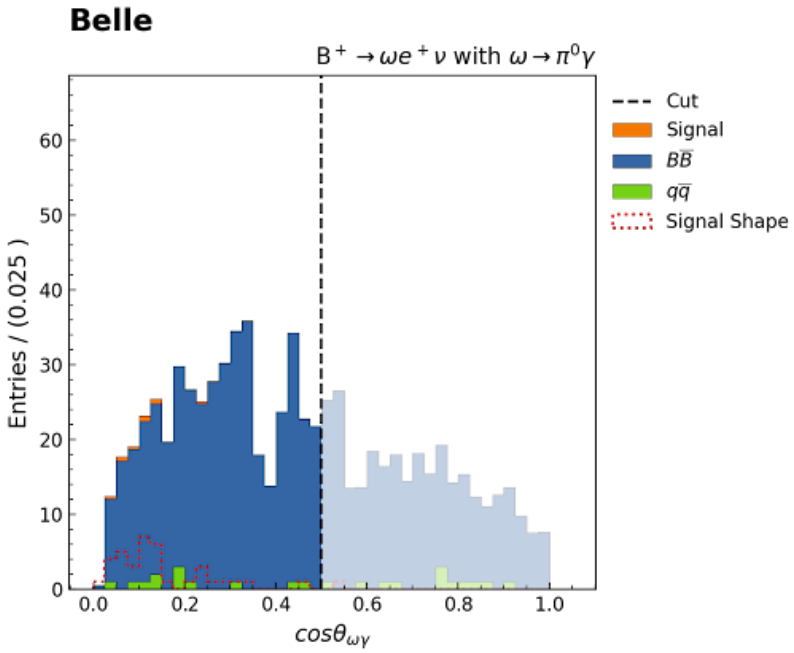


Figure 7.8: The $|\cos\theta_{\omega\gamma}|$ distribution.

Channel	Variable	Selection	Signal e^+	$B\overline{B}$ e^+	Cont e^+	Signal μ^+	$B\overline{B}$ μ^+	Cont μ^+
ρ^0	BDT classifier	< 0.5	0.90	0.73	0.06	0.91	0.73	0.15
	M_{bc}	> 5.27	0.9	0.69	0.49	0.90	0.67	0.50
	$\lvert\Delta E\rvert$	< 0.1	0.74	0.70	0.64	0.75	0.71	0.67
	FEI signal Probability	$> 10^{-4}$	0.92	0.79	0.49	0.92	0.79	0.60
	$\lvert\cos\theta_{BY}\rvert$	< 3	0.99	0.63	0.52	0.99	0.63	0.54
	E_{ECL}	< 1.5 GeV	0.97	0.84	0.55	0.97	0.83	0.51
	$0.47\text{ GeV} < M_{\pi^+\pi^-} < 1.07$ GeV		0.93	0.57	0.42	0.93	0.57	0.41
	total		0.0050			0.0054		
ρ^-	BDT classifier	< 0.5	0.83	0.55	0.04	0.82	0.54	0.03
	BCS		0.79	0.17	0.28	0.78	0.17	0.27
	M_{bc}	> 5.27	1.00	0.64	1	1	0.61	1
	$\lvert\Delta E\rvert$	< 0.1	1	0.73	0.30	1	0.73	0.31
	FEI signal Probability	$> 10^{-4}$	0.74	0.68	0.43	0.75	0.69	0.49
	$\lvert\cos\theta_{BY}\rvert$	< 3	0.99	0.56	0.52	0.99	0.56	0.56
	E_{ECL}	< 1.5 GeV	1	0.92	0.60	1	0.91	0.68
	$\cos\theta_{\gamma\gamma}$	> 0.4	0.97	0.89	0.96	0.97	0.89	0.95
	$0.47\text{ GeV} < M_{\pi^0\pi^-} < 1.07$ GeV		1	0.71	0.49	1	0.71	0.39
	total		0.00038			0.00043		
$\omega \to \pi^0\pi^+\pi^-$	BDT classifier	< 0.5	0.83	0.70	0.07	0.83	0.70	0.07
	BCS		0.47	0.16	0.24	0.49	0.15	0.21
	M_{bc}	> 5.27	1	0.64	0.50	1	0.65	0.50
	$\lvert\Delta E\rvert$	< 0.1	1	0.75	0.65	1	0.75	0.68
	FEI signal Probability	$> 10^{-4}$	0.80	0.73	0.56	0.82	0.74	0.64
	$\lvert\cos\theta_{BY}\rvert$	< 3	0.99	0.48	0.36	1	0.48	0.37
	E_{ECL}	< 1.5 GeV	1	0.83	0.56	1	0.82	0.49
	$\cos\theta_{\gamma\gamma}$	> 0.25	0.95	0.50	0.57	0.94	0.51	0.50
	$0.74\text{ GeV} < M_{\pi^0\pi^+\pi^-} < 0.81$ GeV		0.91	0.35	0.35	0.95	0.35	0.41
	total		0.0070			0.00080		
$\omega \to \pi^0\gamma$	BDT classifier	< 0.5	0.79	0.69	0.06	0.82	0.68	0.06
	BCS		0.59	0.24	0.30	0.48	0.24	0.27
	M_{bc}	> 5.27	1	0.66	0.27	1	0.66	0.49
	$\lvert\Delta E\rvert$	< 0.1	1	0.76	0.65	1	0.76	0.63
	FEI signal Probability	$> 10^{-4}$	0.72	0.69	0.50	0.80	0.69	0.58
	$\lvert\cos\theta_{BY}\rvert$	< 3	1	0.48	0.33	1	0.50	0.36
	E_{ECL}	< 1.5 GeV	1	0.88	0.51	1	0.86	0.51
	$\cos\theta_{\gamma\gamma}$	0.97	0.84	0.77	1	0.84	0.82	
	$\lvert\cos\omega\gamma\rvert$	< 0.5	0.97	0.58	0.61	0.95	0.59	0.61
	$0.74\text{ GeV} < M_{\pi^0\gamma} < 0.81$ GeV		0.97	0.38	0.60	0.86	0.38	0.30
	total		0.034			0.036		

Table 7.5: Signal selections for all channels.

CHAPTER 8

Signal extraction

In this chapter the fitting method and the fit procedure for the M^2_{miss} distribution is described. Furthermore the signal extraction using a one dimensional fit integrated over q^2 and in separate q^2 bins is described.

8.1 Fitting Method

The aim of the fit is to extract the signal events that pass all selection criteria from the full Belle data set of 711 fb^{-1}. The results are shown in chapter **??**. An extended binned maximum likelihood technique is used [24]. Therefore, a likelihood function for the agreement between data and Monte Carlo components is built by maximizing the likelihood. This takes the finite MC statistics in the template histograms into account [2]. Each of the components has a yield, which is the number of events within that component. The yields of the individual components can be fixed to a certain value or they can float within a range. The MC components are split up into the following categories:

- signal: signal component
- $B\overline{B}$: all $B\overline{B}$ decays excluding signal
- continuum: all $q\overline{q}$ decays

The fit yields are free parameters in the likelihood optimization. The binned maximum likelihood fit returns a fit yield for each category. Here, Asimov fits are performed. This means that a fitting model based on the MC data is fitted to the MC data. This is done to obtain information about the yield uncertainties, which are independent of the statistical fluctuations of the specific yields.

The signal significance is a measure of how well signal is separated from background. It is defined as:

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

Here N_{sig} is the signal yield and σN_{sig} is the uncertainty on the signal yield. Both of these numbers are obtained from the Asimov fit.

8.2 One-Dimensional Fit of M^2_{miss}

In general with the one dimensional fit the total branching fraction can be measured, for this analysis however the FEI corrections are missing. Still a one dimensional fit to the M^2_{miss} distribution is fitted for all channels, to test the fit and obtain significance estimates, Asimov fits are performed. Here the post-fit M^2_{miss} distribution for the $B^+ \to \rho^0 \ell \nu$ channel can be seen in figure 8.1. The fit is performed in the range of $-2 < M^2_{\text{miss}} < 2.5$.

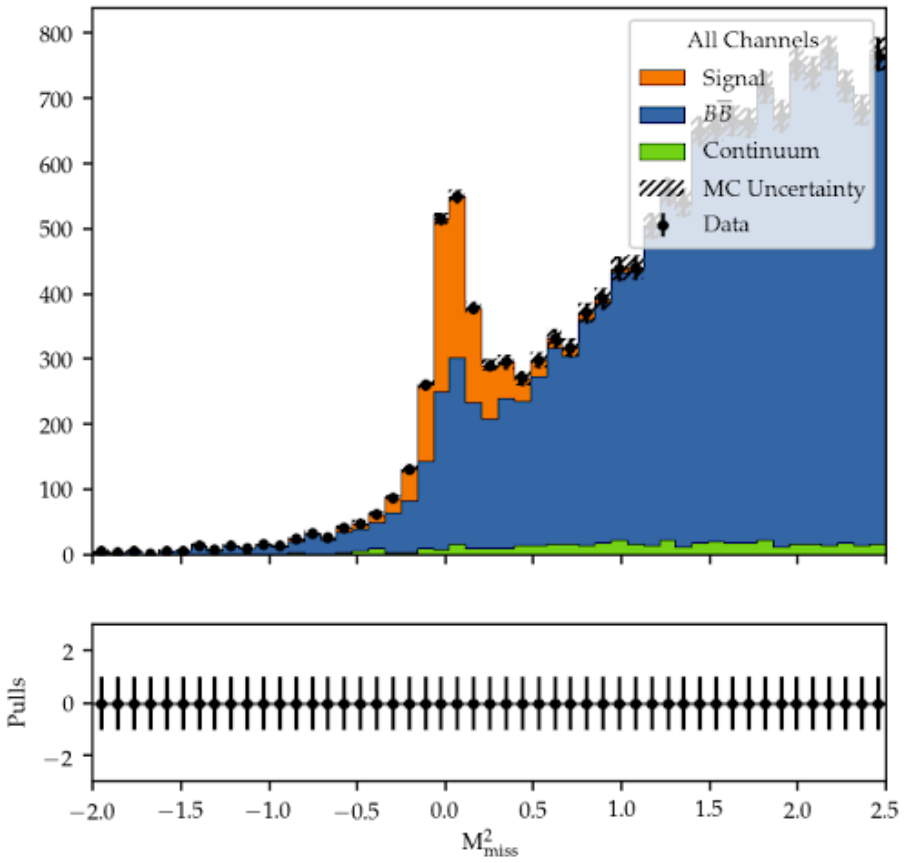


Figure 8.1: Post fit of M^2_{miss} distribution for the channel $B^+ \to \rho^0 \ell^+ \nu$ from Asimov fit.

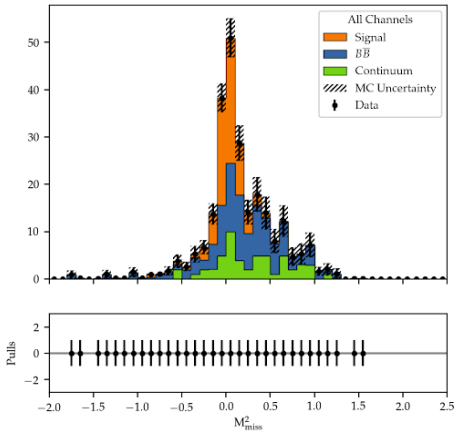
In table 8.1 the fit yields with their errors and the significances for each template are shown.

category	yield	significance
signal	1205 ± 63	19
$B\bar{B}$	26551 ± 177	150
continuum	1161 ± 34	34

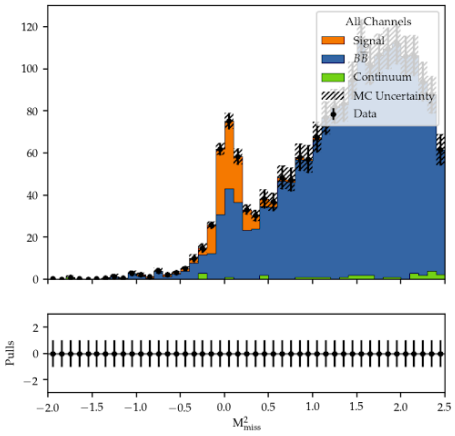
Table 8.1: Yields and significance for the fit of the $B^+ \to \rho^0 \ell^+ \nu$ channel.

8.3 Fit in q^2 Bins

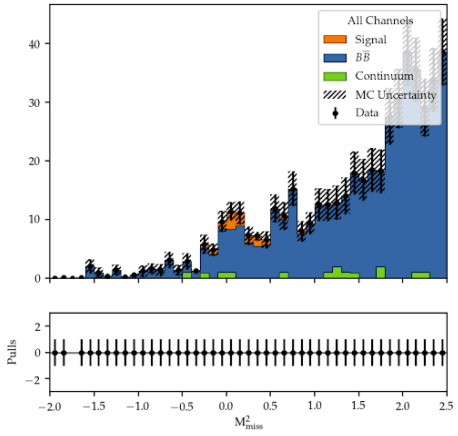
In order to set up a possible extraction of the branching fraction, fits to the M^2_{miss} distribution in separate q^2 bins are done. For the $B^+ \to \rho^0 \ell^+ \nu$ channel the M^2_{miss} distribution is split into 11 bins. For the $B^0 \to \rho^- \ell^+ \nu$ channel six q^2 bins are used and for the $\omega \to \pi^0 \pi^+ \pi^-$ channel three q^2 bins are used. The separate q^2 bin fits are not performed for the channel where ω decays to π^0 and a photon due to insufficient statistics. In figure 8.2 the fits for the 1st, 6th and 11th q^2 bin are shown. In table 8.2 the signal yields with their uncertainties and the signal significances in eachq^2 bin can be found.



(a) $q^2 = [0 : 2)$



(b) $q^2 = [10 : 12)$



(c) $q^2 = [20 : 22)$

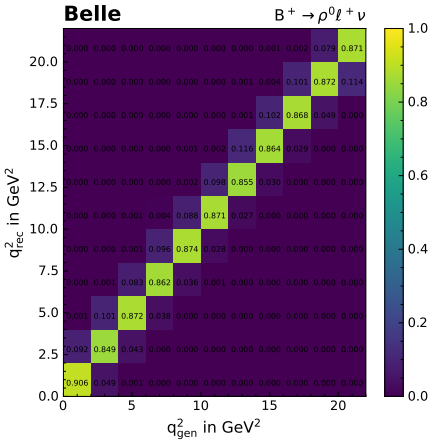
Figure 8.2: M^2_{miss} post fit distributions in the 1st, 6th and 11th q^2 bin for the channel $B^+ \to \rho^0 \ell \nu$ from Asimov fits.

To correct for finite detector resolution and bremsstrahlung effects in the q^2 distribution, an unfolding procedure is used. The migration matrix is calculated using the true and reconstructed q^2. To obtain the unfolded yields bin-by-bin correction factors are determined from the different q^2 variables and multiplied with the fit yields. The migration matrix for the ρ^0 channel is shown in figure 8.3(a). The migration matrix shows the reconstructed q^2 versus the true q^2. Ideally, the diagonal matrix

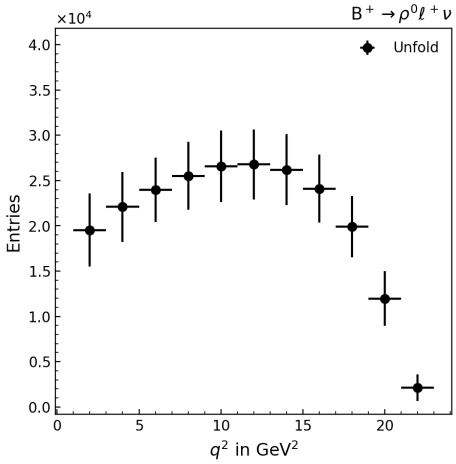
q^2 bin	Signal Yield	Significance
$[0;2)$	81 ± 17	4.84
$[2;4)$	95 ± 17	5.72
$[4;6)$	114 ± 17	6.74
$[6;8)$	122 ± 18	6.77
$[8;10)$	129 ± 19	6.72
$[10;12)$	141 ± 20	6.91
$[12;14)$	142 ± 21	6.68
$[14;16)$	138 ± 22	6.41
$[16;18)$	117 ± 20	5.87
$[18;20)$	69 ± 17	3.96
$[20;22]$	15 ± 10	1.46

Table 8.2: Signal yields with errors and signal significance for the fit in q^2 bins for the $B^+ \to \rho^0 \ell \nu$ channel.

elements would be one meaning that reconstructed and generated q^2 would be the same. Additionally the reconstruction efficiencies have to be considered. The reconstruction efficiency is calculated by calculating generated q^2 is calculated, then the true q^2 is divided by the generated q^2. To get the unfolded q^2 the corrected fit yields are then divided by the efficiencies. In figure 8.3(b) the unfolded q^2 spectrum is shown. It depicts the corrected yields in the q^2 bins.



(a) Migration Matrix for the $B^+ \to \rho^0 \ell^- \nu$ channel.



(b) Unfolded q^2 spectrum for the channel $B^+ \to \rho^0 \ell^- \nu$.

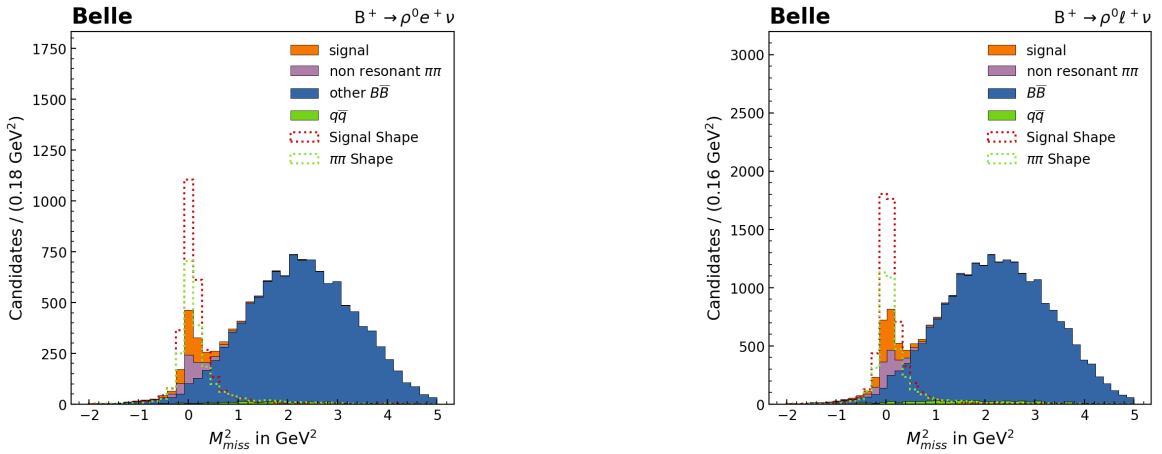
Figure 8.3: The migration matrix and the unfolded q^2 spectrum.

The significances and yields are comparable to a previous analysis by Sibidanov et al. [2]. However here a peak in the M^2_{miss} contribution can be seen in the $B\overline{B}$ component around zero. The previous analysis does not observe this peak. This issue is adressed in chapter 9

CHAPTER 9

Non-Resonant $\pi\pi$ Contribution

Splitting up the $B\overline{B}$ background into components and taking a look at its origins, it becomes apparent that the peak around zero in the M^2_{miss} distribution is caused by a non-resonant $\pi\pi$ contribution. The M^2_{miss} distributions with the non-resonant $\pi\pi$ contribution separated from the $B\overline{B}$ background for the $B^+ \to \rho^0 e^+ \nu$ and the combined $B^+ \to \rho^0 \ell^+ \nu$ channels are shown in figure 9.1.



(a) (b)

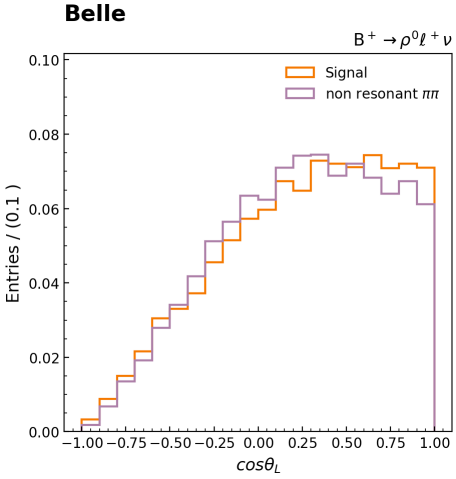
Figure 9.1: The M^2_{miss} distributions with the non-resonant $\pi\pi$ contribution separated from the $B\overline{B}$ background (a) for the $B^+ \to \rho^0 e^+ \nu$ channel and (b) for the $B^+ \to \rho^0 \ell^+ \nu$ channel.

The non-resonant $\pi\pi$ peak is right underneath the signal peak and has a very similar shape as well. This makes the signal extraction harder, because the yields are floating in a range. When the same shape is at the same position twice stacked on top of each other and a fit is attempted the yields cut be pulled in the direction of the wrong peak. That would lead to a wrong yield. Sibidanov did not see this in his analysis because a different model for the Monte Carlo was used, which was modeled in a way, that the non-resonant $\pi\pi$ contribution could not be seen. That means that the MC he used did not include a significant amount of it. If the non-resonant $\pi\pi$ component in the data was added to the signal component, this miss-modeling could explain why the branching fraction Sibidanov obtained was too high compared to other results. Therefore a method to distinguish between the signal and the

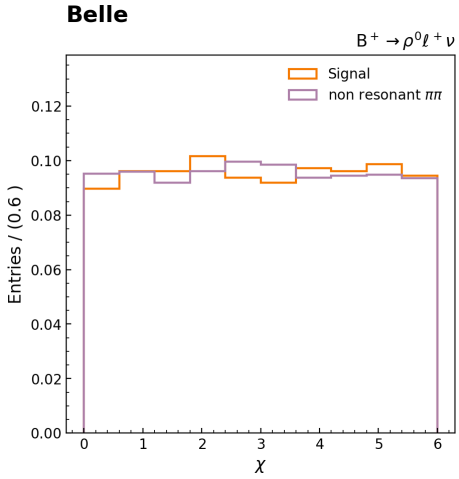
non-resonant $\pi\pi$ component is required.

9.1 Helicity Angles

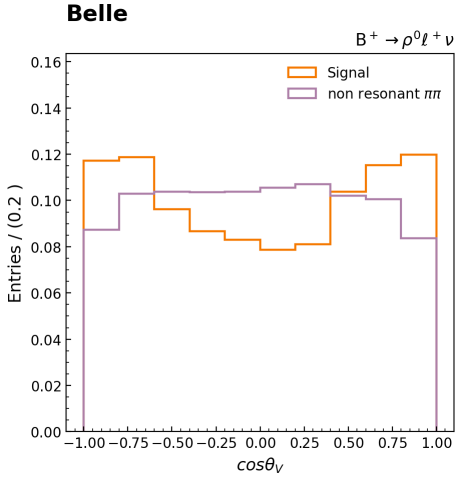
One idea to separate the signal from the non-resonant $\pi\pi$ component is taking a closer look at the helicity angles explained in section 2.1.2. The goal is to uniquely identify the non-resonant $\pi\pi$ component as having the same spin state as the ρ meson, instead of measuring non-resonant $\pi\pi$ and assuming that it has the same properties as the ρ meson. The helicity angles are measured in reference to the momentum of a particle. θ_V is the angle that is measured in reference to the hadron. This is illustrated in figure 2.4. Here θ_V is the only angle that takes the direction of the hadron directly into account. The normalised helicity angle distributions for signal and non-resonant $\pi\pi$ are shown in figure 9.2.



(a) $\cos\theta_\ell$ for signal and non-resonant $\pi\pi$



(b) χ for signal and non-resonant $\pi\pi$

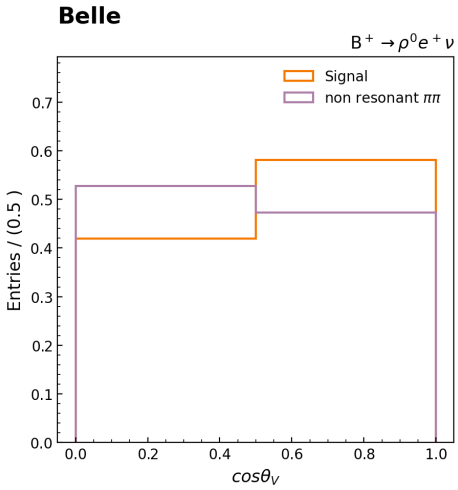


(c) $\cos\theta_V$ for signal and non-resonant $\pi\pi$

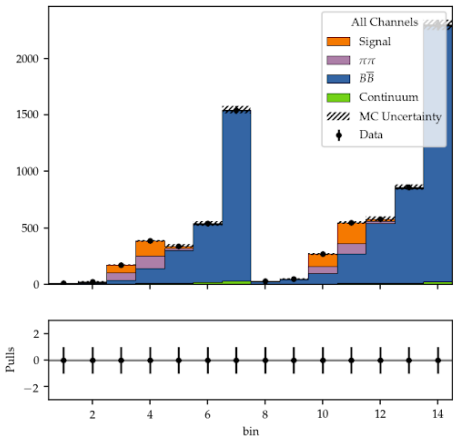
Figure 9.2: The three helicity angles for $B^+ \to \rho^0 \ell^+ \nu$.

Looking at the plots in figure 9.2 it becomes apparent that only $\cos\theta_V$ in figure 9.2(c) might be useful to distinguish between signal and the non-resonant $\pi\pi$ contribution. $\cos\theta_V$ is the only helicity angle distribution showing a clear shape difference, as expected.

One idea to benefit from this is to perform a two-dimensional fit in the M^2_{miss} and $\cos\theta_V$ distribution. The absolute value of $\cos\theta_V$ is chosen and at first $\cos\theta_V$ is split into only two bins. The distribution can be found in figure 9.3(a). For M^2_{miss} for now 7 bins are used. Adding the two bins in $\cos\theta_V$, leads to a signal significance of 2.91 and a significance of the non-resonant $\pi\pi$ contribution of 1.89. Furthermore the correlation of these two categories could be decreased. From a full anti-correlation of -1 to an anti-correlation of -0.84. The M^2_{miss} post fit distribution using two $\cos\theta_V$ bins is shown in figure 9.3(b) for the $B^+ \to \rho^0 e^+ \nu$ channel.



(a) Normalized distribution of $|\cos\Theta_V|$ in two bins.



(b) Two-dimensional post fit distribution in 2 bins of $\cos\theta_V$ and 9 bins of M^2_{miss}

Figure 9.3: The two-dimensional Asimov post fit distribution in $\cos\theta_V$ and M^2_{miss} for $B^+ \to \rho^0 e^+ \nu$

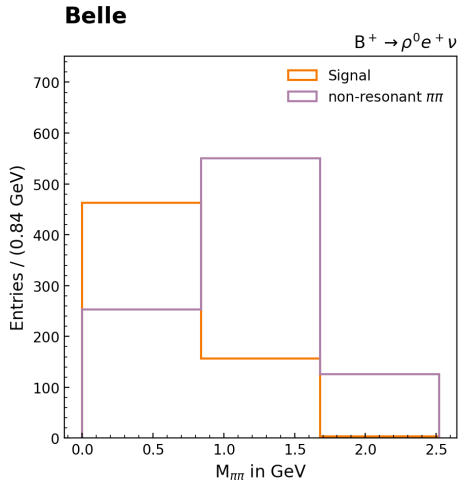
With the two dimensional fit in $\cos\theta_V$ and M^2_{miss} the correlation between the non-resonant $\pi\pi$ contribution and the signal peak could be brought down. Additionally the yields of the two components could be obtained with a significance of 2.91 and 1.89, and it can be distinguished between the singal and non-resonant $\pi\pi$ contribution. Therefore with this analysis not only the branching fraction for the ρ^0 meson but also the branching fraction for the non-resonant $\pi\pi$ contribution could be measured.

9.2 $M_{\pi\pi}$

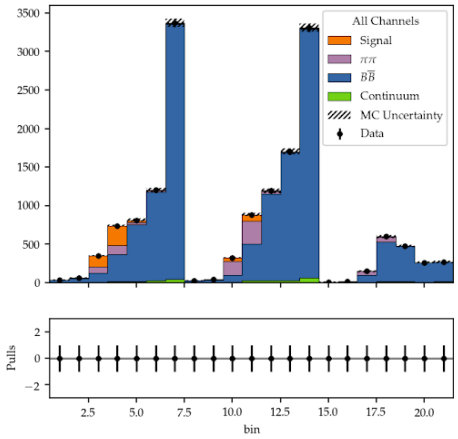
Another approach to separate signal from non-resonant $\pi\pi$ component is to take a closer look at its invariant mass distribution. The ρ^0 is a resonance, therefore a peak in the invariant mass of its daughters can be seen at the mass of the ρ^0 meson. Non-resonant decays are not associated with a resonance. Therefore they do not form narrow peak like structures in the phase space of the decay. Therefore no prominent peak can be seen in the invariant mass distribution for the non-resonant $\pi\pi$ contribution. The invariant mass plot can be found in figure 9.4(c). The shape of $M_{\pi\pi}$ differs greatly from the shape of signal. This can be used to distinguish between the signal peak and the non-resonant

$\pi\pi$ peak as well.

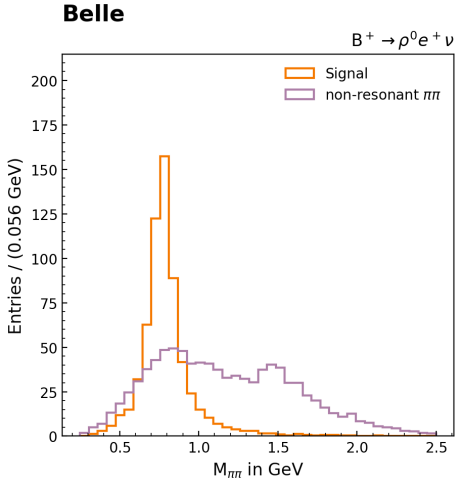
A first two dimensional fit was performed and can be seen in figure 9.4(b). Here the significance for the signal is 9.87 and the $\pi\pi$ significance is 11.29. The correlation is brought down to -0.81 as well. This method appears to work slightly better than the two dimensional fit with $\cos\theta_V$, however it does depend on the modeling of $M_{\pi\pi}$, which is incorrect. The peak around 1.5 GeV that can be seen in figure 9.4(c), originates from the hybrid model re-weighting, which is not done on data. Therefore no conclusive statement on how well this method actually performs can be made. However on MC this method was able to distinguish between the signal component and the non-resonant $\pi\pi$ contribution. Therefore this method could in theory also be used to perform a measurement of the non-resonant $\pi\pi$ contribution. Maybe if more bins in $M_{\pi\pi}$ are used the correlation between the two components could be brought down more, however this would also make the fit more dependent on the modeling.



(a) Invariant mass $M_{\pi\pi}$ distribution in three bis.



(b) M^2_{miss} fit in 3 $M_{\pi\pi}$ bins.



(c) Invariant mass $M_{\pi\pi}$ distribution.

Figure 9.4: $M_{\pi\pi}$ distribution on the left in three bins and M^2_{miss} fit in 3 $M_{\pi\pi}$ bins on the right. On the bottom the full $M_{\pi\pi}$ distribution is shown.

CHAPTER 10

Conclusion

The full 711 fb^{-1} of data from the Belle experiment were analysed. Therefore B meson decays, that were reconstructed using the FEI tagging algorithm, were investigated. In total four different channels were analysed in this thesis. The $B^+ \to \rho^0 \ell\nu$, $B^0 \to \rho^- \ell\nu$, $B^+ \to \omega(\to \pi^+\pi^-\pi^0)\ell\nu$ and $B^+ \to \omega(\to \pi^0\gamma)\ell\nu$. After reconstructing the signal, first, preselections are performed. Next, continuum background is suppressed, by training fastBDTs. Furthermore additional selections are performed and the signal extraction variable M^2_{miss} is studied.

In the M^2_{miss} distribution a peak in the $B\overline{B}$ background, directly below the signal peak was observed. This was the most prominent in the $B^+ \to \rho^0 \ell^+ \nu$ channel, but it was observed in the other channels as well. It was discovered that this peak arises from a non-resonant $\pi\pi$ background contribution.

Before discovering this peaking background the goal was extracting signal from a one-dimensional fit of M^2_{miss}. In that the non-resonant $\pi\pi$ contribution had to be added to the background template, otherwise the Asimov fit was not possible, due to large correlations. The problem with this approach is that the assumption holds that the non-resonant $\pi\pi$ component is modeled perfectly. Originally the plan was also to perform signal extraction in q^2 bins, to prepare for $|V_{ub}|$ extraction. In addition to the peaking background problem, unfortunately the FEI efficiencies are unknown, therefore neither the branching fraction nor $|V_{ub}|$ can be obtained, but the measurement of the shape of the FF was set up.

The thesis then focused on the separation between the non-resonant $\pi\pi$ contribution and signal. Two methods were studied both relying on two-dimensional fits. First a two-dimensional Asimov fit in one of the helicity angles, $\cos\theta_V$ and M^2_{miss} is performed. On MC data a separation was possible, and the signal yield could be determined with a significance of 2.91 while the non-resonant $\pi\pi$ contribution was extracted with a significance of 1.89. This resulted from decreased correlations between these two components of -0.84. A second method to distinguish signal from the non-resonant $\pi\pi$ component is performing a two-dimensional Asimov fit in the hadronic invariant mass $M_{\pi\pi}$ and M^2_{miss}. Here the signal and non-resonant $\pi\pi$ contribution yields could be determined with significances of 9.87 and 11.29, respectively. Furthermore the correlation between the two components decreased to -0.81. The second method seems to perform better than the first method on MC data, but a measurement of the branching fraction of the non-resonant $\pi\pi$ contribution should be possible with both methods.

Next, the full Belle dataset of 711 fb^{-1} was analysed using the setup tested on MC data and described before. Data plots of the $B^+ \to \rho^0 \ell^+ \nu$ channel were shown in chapter **??**. Outside the signal region the data was in relatively good agreement with MC data. Nevertheless, in the signal region some discrepancies were observed. Taking a closer look at the M^2_{miss} distributions with the non-resonant $\pi\pi$

component separated from the $B\overline{B}$ background, the MC to data agreement improved significantly when setting the non-resonant $\pi\pi$ component to zero. This suggests that the non-resonant $\pi\pi$ component was mismodeled by the MC data. Looking at the $\cos\theta_V$ distribution, this was not that clear. The discrepancy could therefore result from mismodeling of the signal as well as the non-resonant $\pi\pi$ contribution.

Finally a two-dimensional fit of the M^2_{miss} and the $\cos\theta_V$ distributions was attempted, however on data the fit was not able to separate the signal and non-resonant $\pi\pi$ contribution and correlations of -0.96 were observed.

It is therefore not possible to state definitely what component is mismodeled. The discrepancies could either come from mismodeling of only the non-resonant $\pi\pi$ contributions or from the combination of a mismodeling of the $\rho\ell\nu$ and non-resonant $\pi\pi$ contribution. Therefore, further studies to distinguish between these processes are needed. This includes an in-depth study of the effect of the number of bins in the different distributions for the two-dimensional fit on the discrimination power of the fit. A three dimensional fit in the M^2_{miss}, $\cos\theta_V$ and the $M_{\pi\pi}$ distributions can also be performed.

Bibliography

[1] M. T. et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) (cit. on pp. 1, 5, 7).

[2] A. S. et al., *Study of Exclusive $B \to X_u \ell \nu$ Decays and Extraction of $|V_{ub}|$ using Full Reconstruction Tagging at the Belle Experiment*, Physics Review **D 88** (2013) (cit. on pp. 1, 23, 24, 53, 56).

[3] T. Keck et al., *The Full Event Interpretation*, Computing and Software for Big Science **3** (2019), URL: https://doi.org/10.1007%2Fs41781-019-0021-8 (cit. on p. 1).

[4] *Standard Model University of Zuerich*, https://www.physik.uzh.ch/groups/serra/StandardModel.html, Accessed: 2022-06-17 (cit. on p. 4).

[5] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013 (cit. on p. 4).

[6] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weka Interaction*, 49th ed., Progress of Theoretical Physics, 1973 (cit. on p. 4).

[7] T. M. Jochen Dingfelder, *Leptonic and semileptonic decays of B mesons*, RevModPhys **88** (2016) (cit. on pp. 5, 6, 11, 21, 22, 32).

[8] J. D. Richman and P. R. Burchat, *Leptonic and semileptonic decays of charm and bottom hadrons*, Rev. Mod. Phys. **67** (4 1995) 893, URL: https://link.aps.org/doi/10.1103/RevModPhys.67.893 (cit. on p. 7).

[9] A. et al., *The Belle detector*, Nuclear Instruments and Methods in Physics Research **A 479** (2002) 117 (cit. on p. 9).

[10] *KEKB accelerator*, https://en.wikipedia.org/wiki/KEKB_%28accelerator%29, Accessed: 2022-07-17 (cit. on p. 10).

[11] B. et al., *The Physics of B Factories*, Eur. Phys. J. **C 74** (2014) 3026 (cit. on pp. 13, 30, 32).

[12] 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).

[13] R. Brun et al., *GEANT: Detector Description and Simulation Tool; Oct 1994*, CERN Program Library, Long Writeup W5013, Geneva: CERN, 1993, URL: https://cds.cern.ch/record/1082634 (cit. on p. 15).

[14] J. D. C.Ramirez and G. Burdman, *Semileptoic $b \to u$ decay*, Physical review **D 41** (1990) 1496 (cit. on p. 16).

[15] M. T. P. et al, *Search for $B^+ \to \mu^+ \nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging*, Phys. rev. **D 101** (3 2020) 032007 (cit. on p. 16).

[16] F. D. Fazio and M. Neubert, *$B \to X_u \ell \nu$ decay distributions to order $\alpha_s u$*, Journal of High Energy Physics **1999** (1999) 017, URL: https://doi.org/10.1088%2F1126-6708%2F1999%2F06%2F017 (cit. on p. 16).

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

[18] W. Sutcliffe, *Early results from Full Event Interpretation at Belle II*, Proceedings of Science (2020) (cit. on p. 20).

[19] F. Techini, *Vertex Fitting in the Belle II Analysis Framework*, Proceedings of Science **059** (2016) (cit. on p. 21).

[20] T. Keck, *Machine learning algorithms for the Belle II eperiment and their validation on Belle data*, PhD thesis: KIT-Bibliothek (2017) (cit. on p. 29).

[21] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, arXiv:1609.06119 (2016) (cit. on p. 30).

[22] D. Ignatov and A. Ignatov, "Decision Stream: Cultivating Deep Decision Trees", 2017 905 (cit. on p. 31).

[23] D. Weyland, *Continuum Suppression with Deep Learning techniques for the Belle II Experiment*, (2017) (cit. on p. 31).

[24] R. Barlow and V.Beeston, *Fitting using finite Monte Carlo samples*, Computer Physics Communications **77** (1993) (cit. on p. 53).

APPENDIX A

Useful information

category	yield	significance
signal	170 ± 30	6
$B\overline{B}$	6322 ± 85	74
continuum	188 ± 14	14

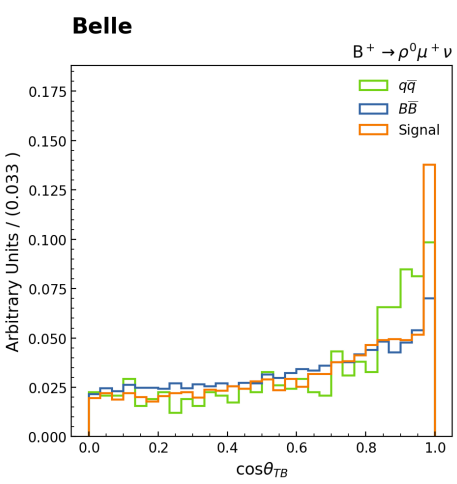
Table A.1: Yields and significance for the fit of the $B^0 \to \rho^- \ell^+ \nu$ channel.

q^2 bin	Signal Yield	Significance
$[0; 4)$	33 ± 12	3
$[4; 8)$	39 ± 13	3
$[8; 12)$	41 ± 15	3
$[12; 16)$	35 ± 14	2
$[16; 20)$	20 ± 13	2
$[20; 24)$	1 ± 4	0

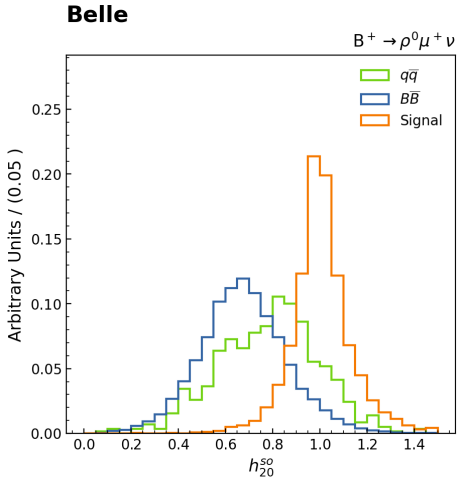
Table A.2: Signal yields with errors and signal significance for the fit in q^2 bins for the $B^0 \to \rho^- \ell \nu$ channel.

category	yield	significance
signal	49 ± 17	3
$B\overline{B}$	1660 ± 45	37
continuum	87 ± 9	9

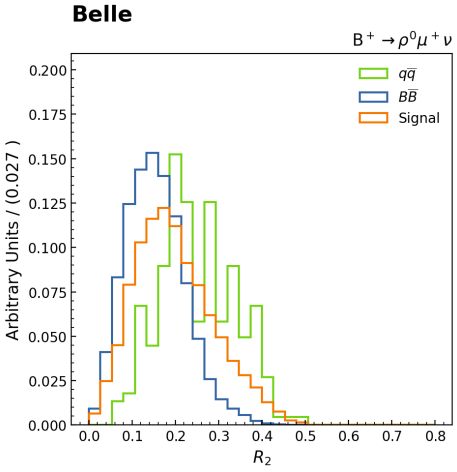
Table A.3: Yields and significance for the fit of the $\omega \to \pi^0 \pi^+ \pi^-$ channel.



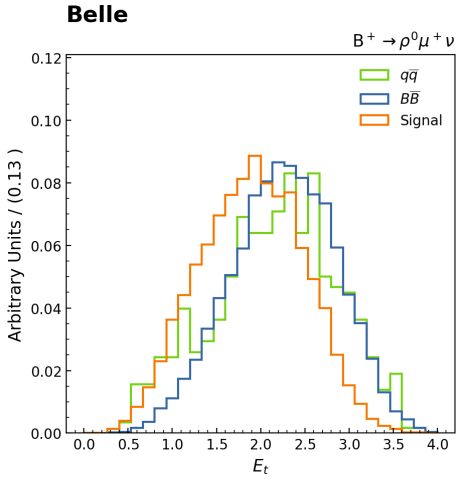
(a)



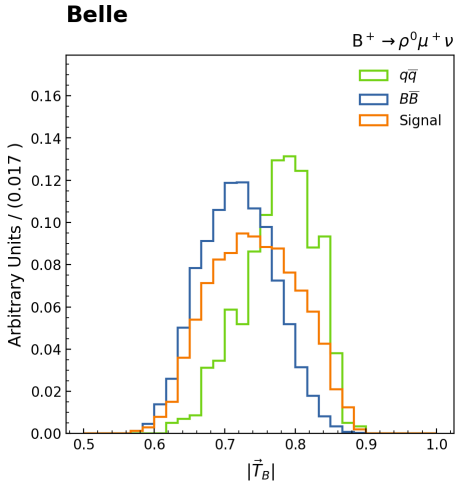
(b)



(c)

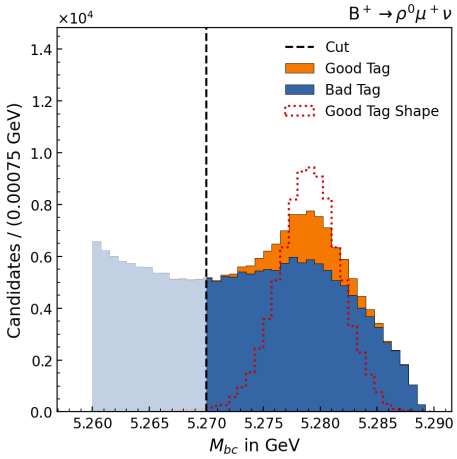


(d)

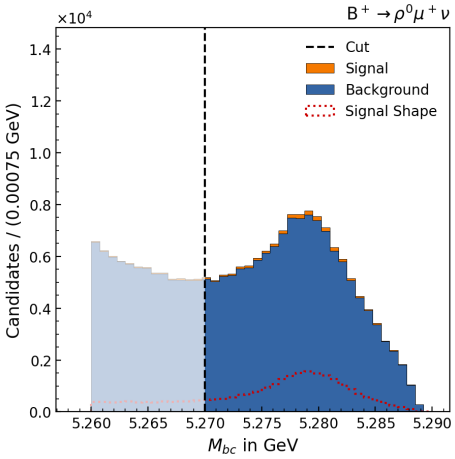


(e)

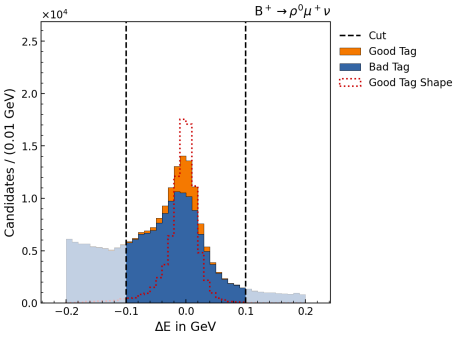
Figure A.1: Normalized distributions of the CS variables for the $B^+ \to \rho^0 \mu^+ \nu$ channel in $q\overline{q}$, $B\overline{B}$ and signal categories from MC.



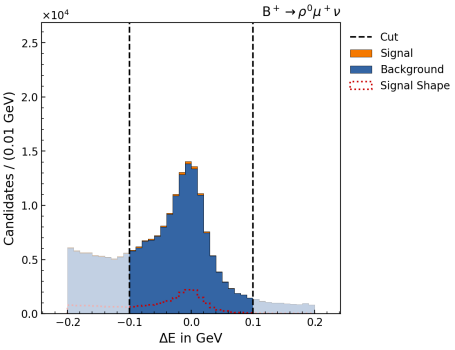
(a)



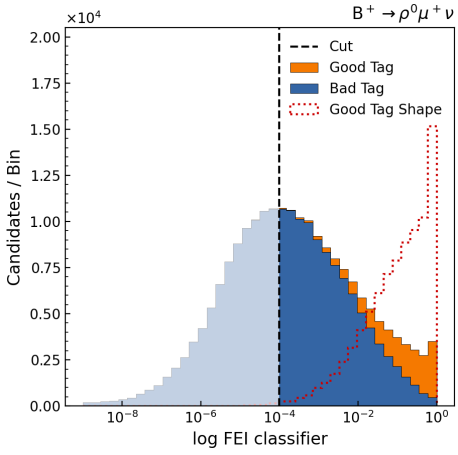
(b)



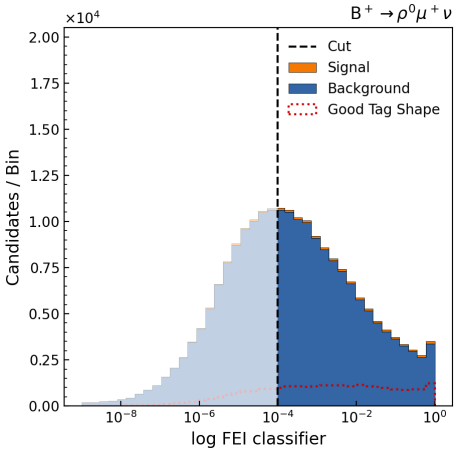
(c)



(d)



(e)



(f)

Figure A.2: Tag-side variables for the $B^+ \to \rho^0 \mu^+ \nu$ channel.

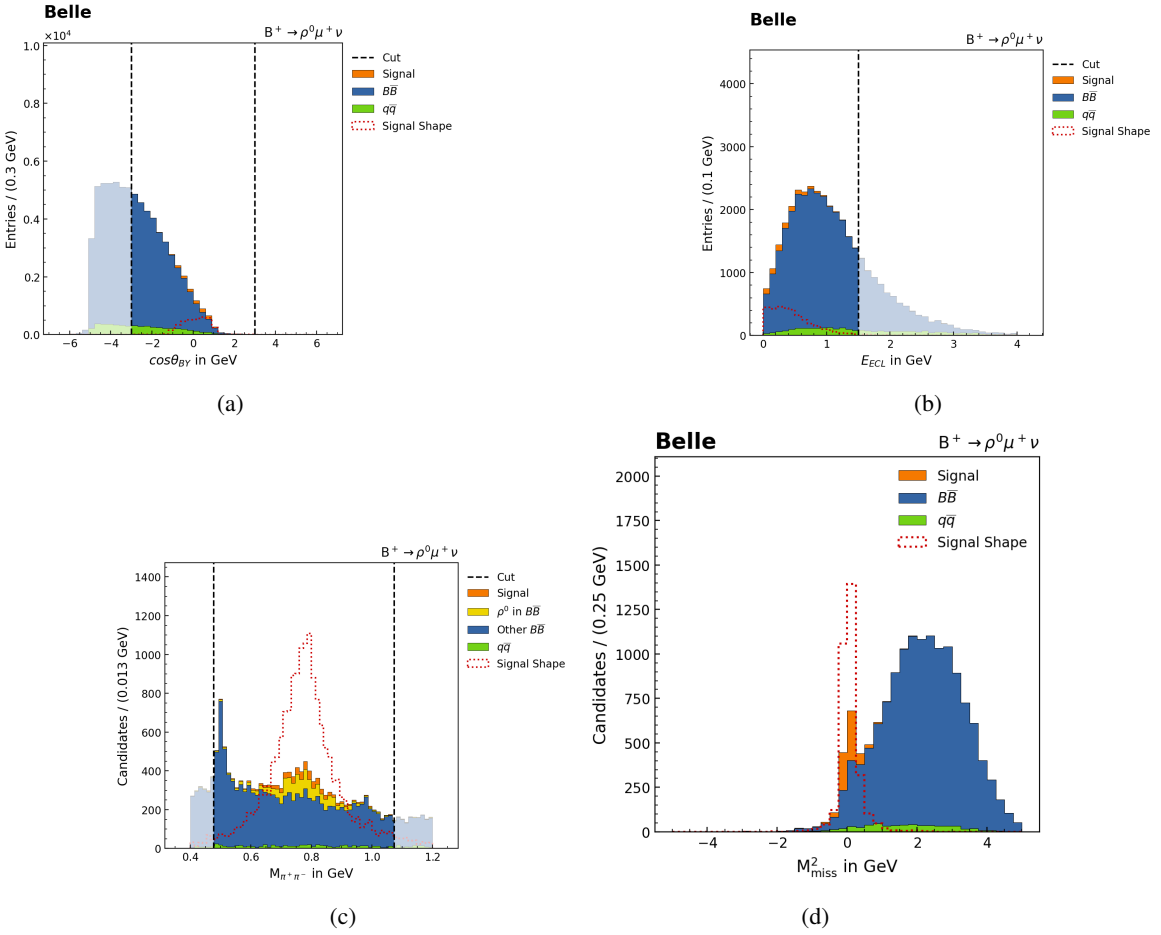


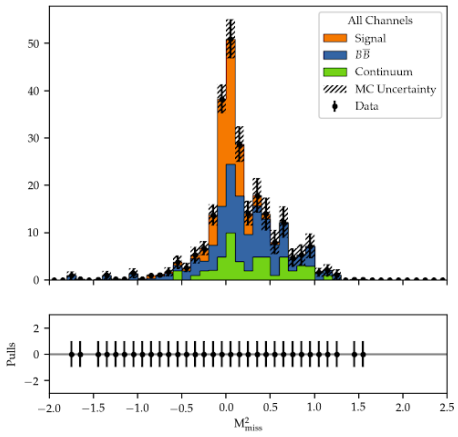
Figure A.3: Signal-side variables for the $B^+ \to \rho^0 \mu^+ \nu$ channel.

q^2 bin	Signal Yield	Significance
$[0;7)$	26 ± 10	2
$[7;14)$	16 ± 10	2
$[14;21)$	7 ± 9	0

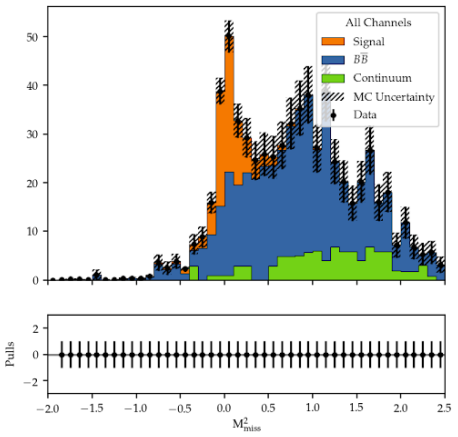
Table A.4: Signal yields with errors and signal significance for the fit in q^2 bins for the $\omega \to \pi^0\pi^+\pi^-$ channel.

category	yield	significance
signal	6 ± 5	1
$B\overline{B}$	360 ± 21	17
continuum	28 ± 11	3

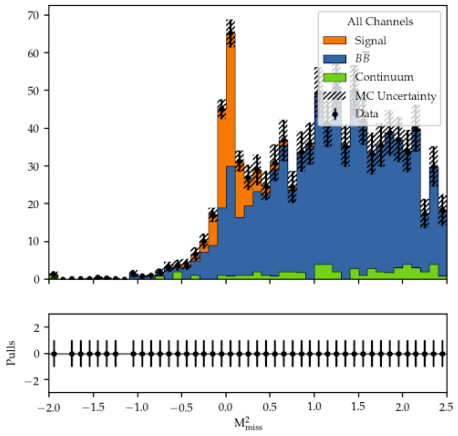
Table A.5: Yields and significance for the fit of the $\omega \to \pi^0\gamma$ channel.



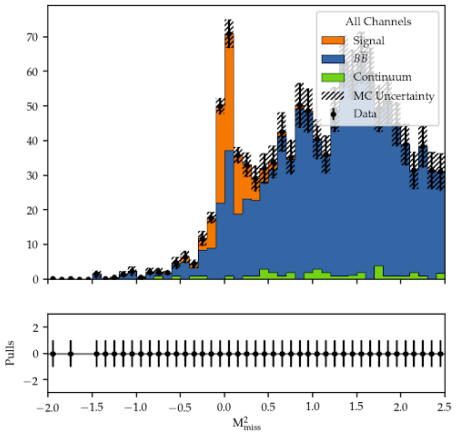
(a)



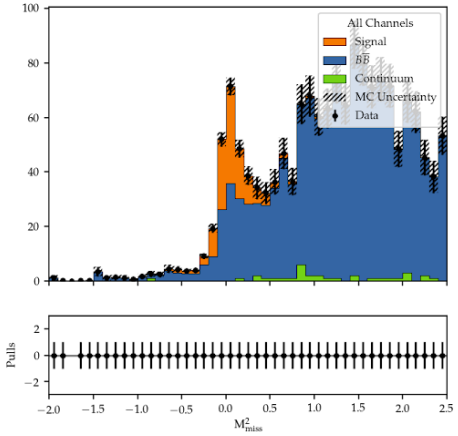
(b)



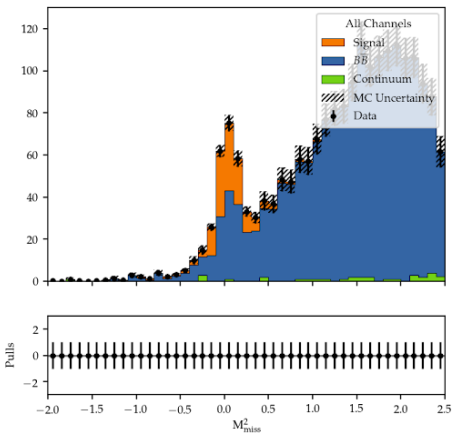
(c)



(d)

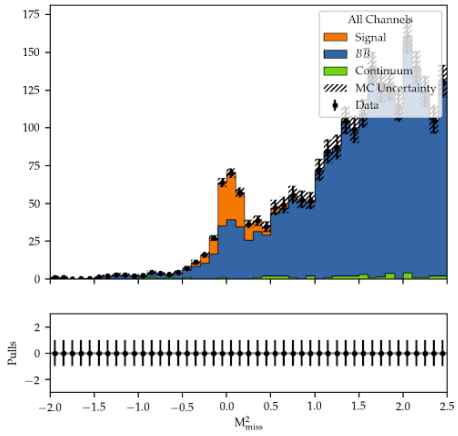


(e)

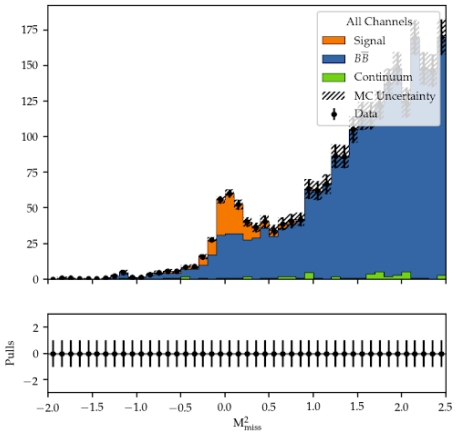


(f)

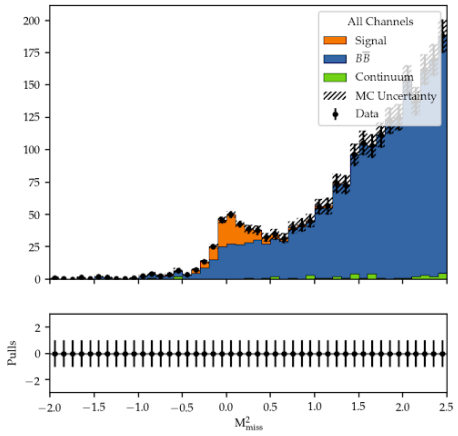
Figure A.4: Fit in 11 q^2 bins for the $B^+ \to \rho^0 \ell^+ \nu$ channel.



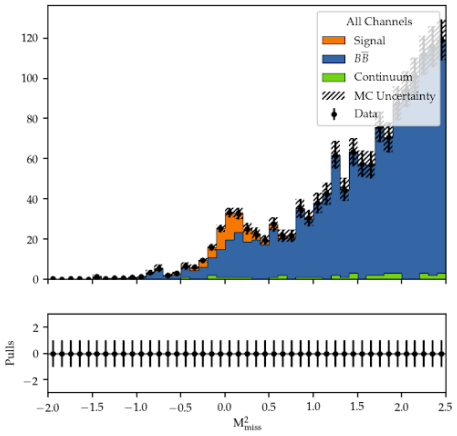
(a)



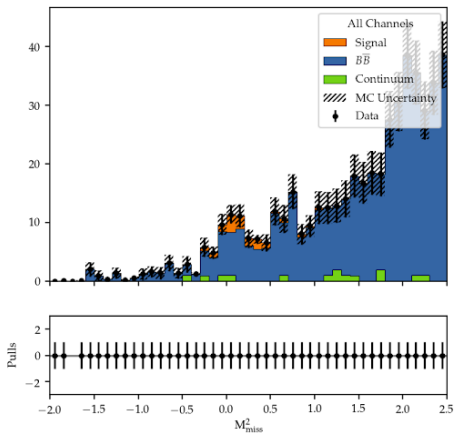
(b)



(c)

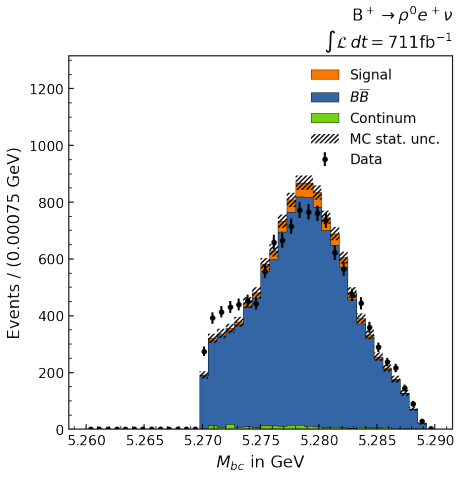


(d)

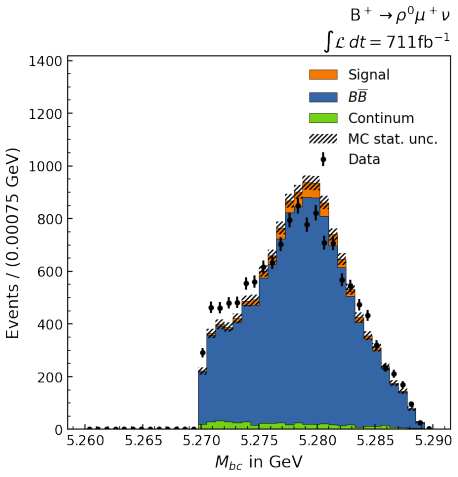


(e)

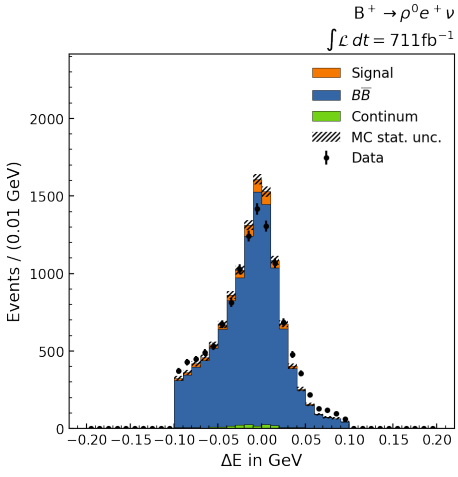
Figure A.5: Fit in 11 q^2 bins for the $B^+ \to \rho^0 \ell^+ \nu$ channel.



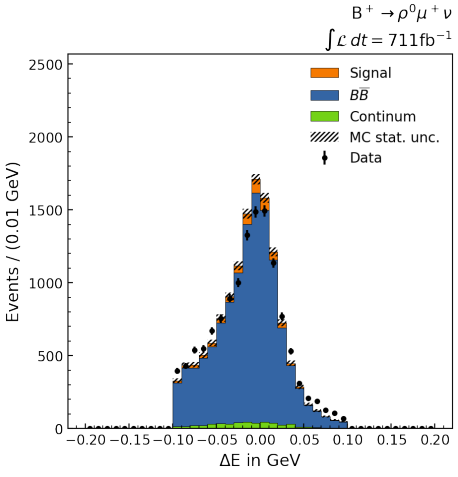
(a)



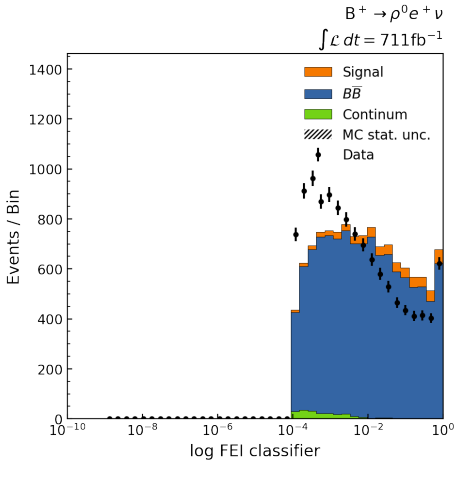
(b)



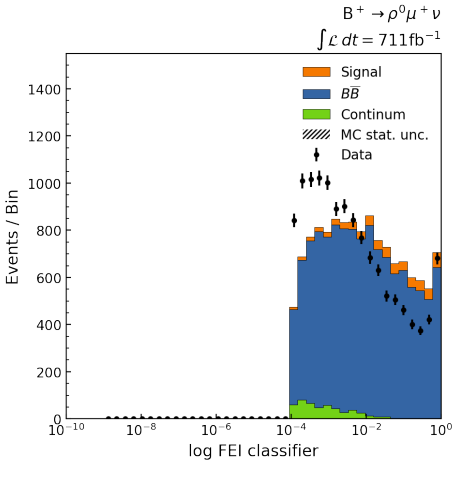
(c)



(d)

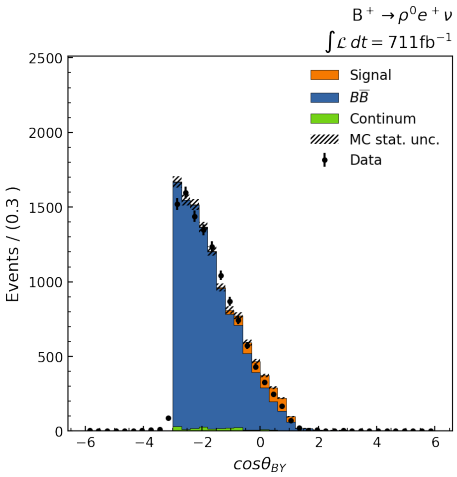


(e)

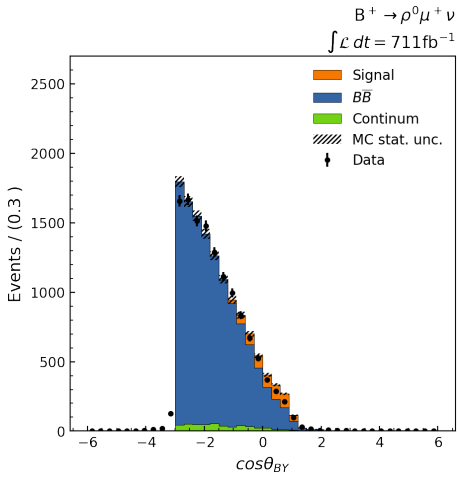


(f)

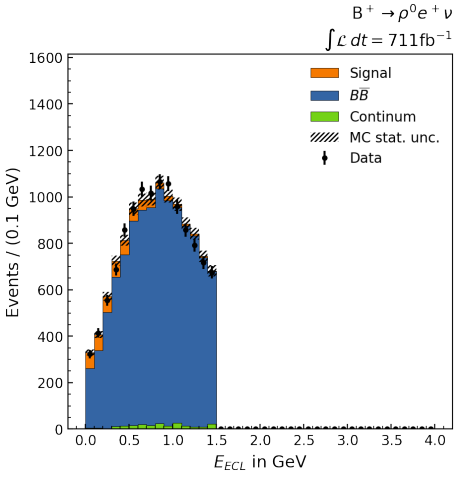
Figure A.6: Tag-side variables for the channel $B^+ \to \rho^0 \ell \nu$ with data.



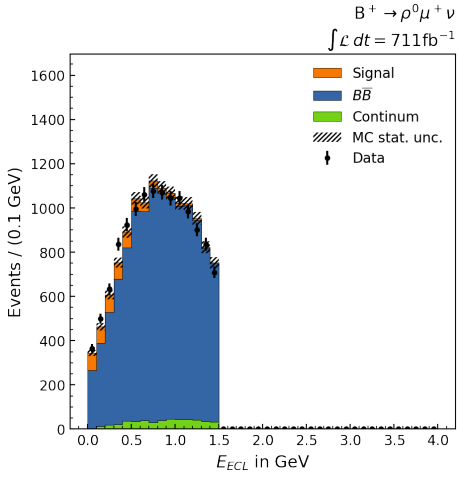
(a)



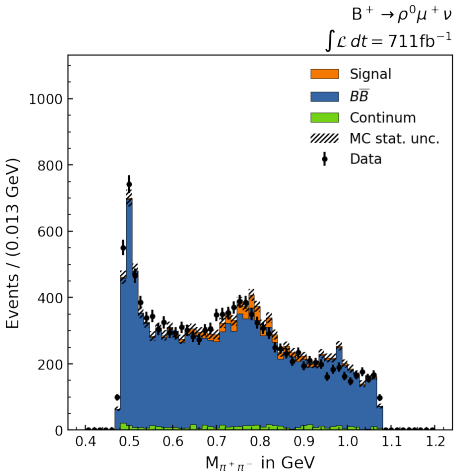
(b)



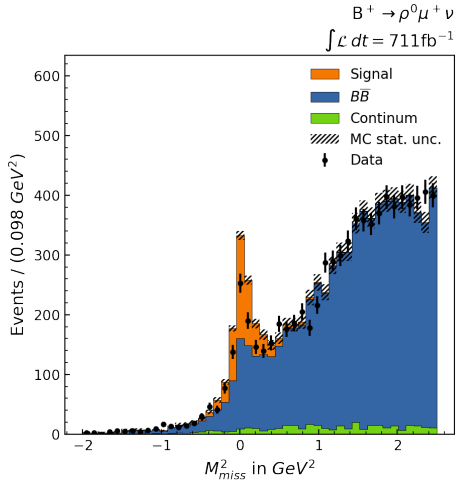
(c)



(d)

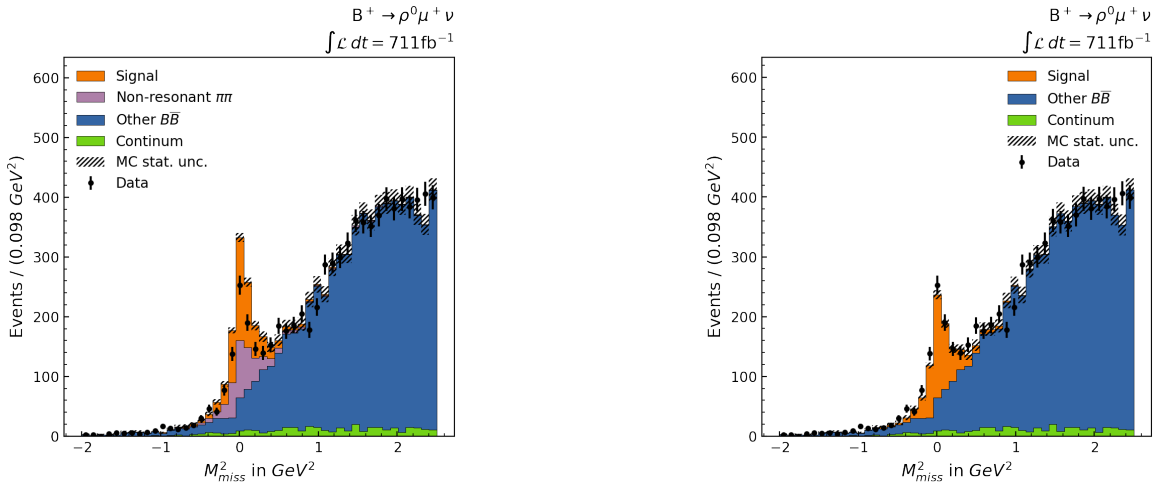


(e)



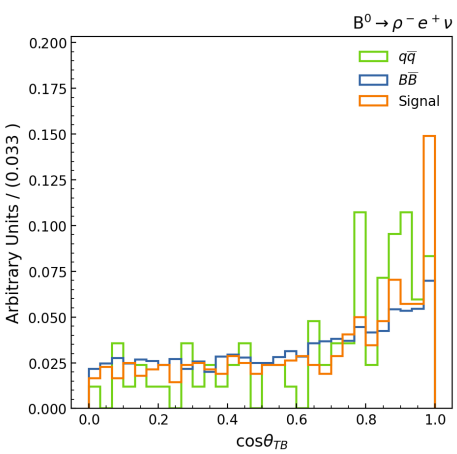
(f)

Figure A.7: Signal-side variables for the channel $B^+ \to \rho^0 \ell \nu$ with data.

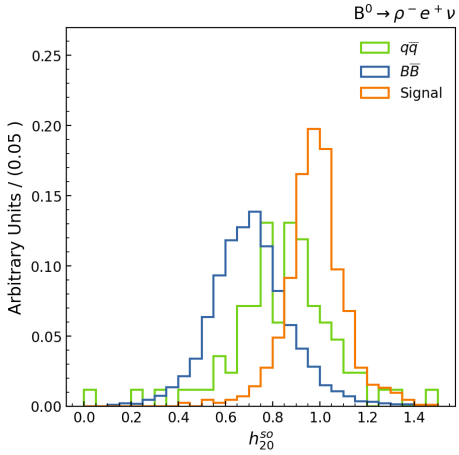


(a) (b)

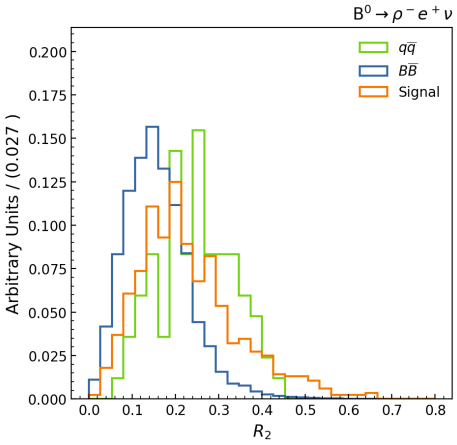
Figure A.8: M^2_{miss} distributions for the channel $B^+ \to \rho^0 \mu \nu$ with data.



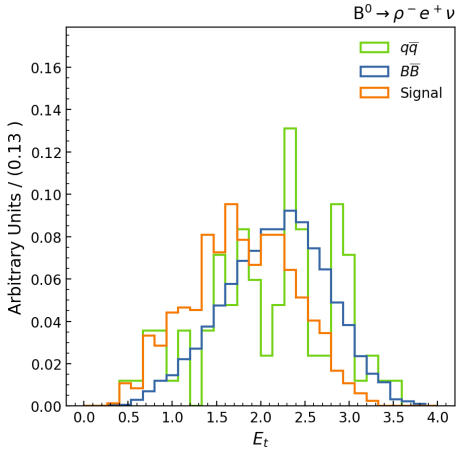
(a)



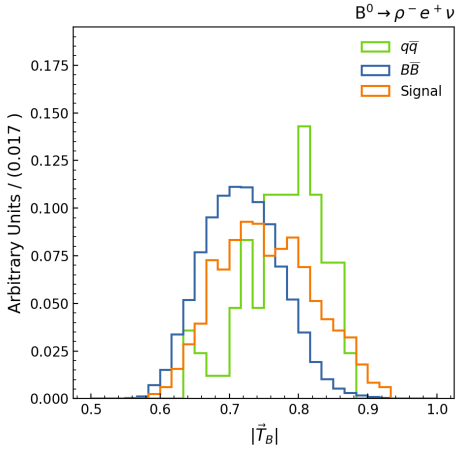
(b)



(c)

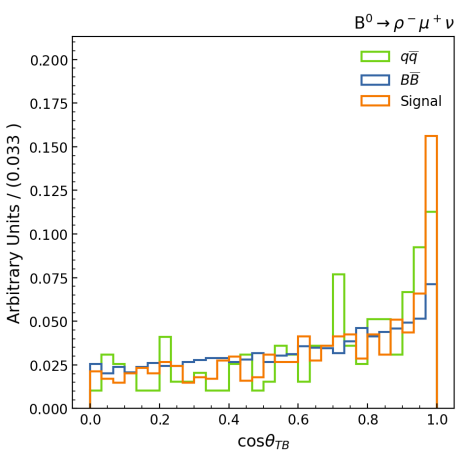


(d)

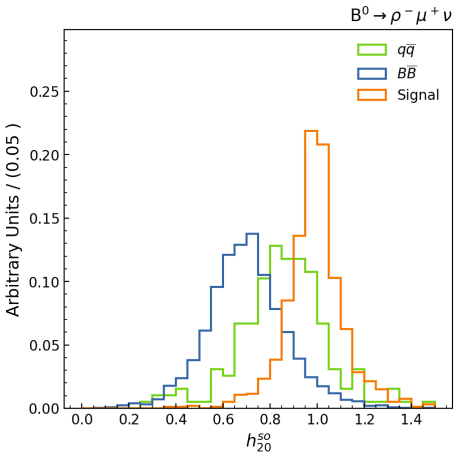


(e)

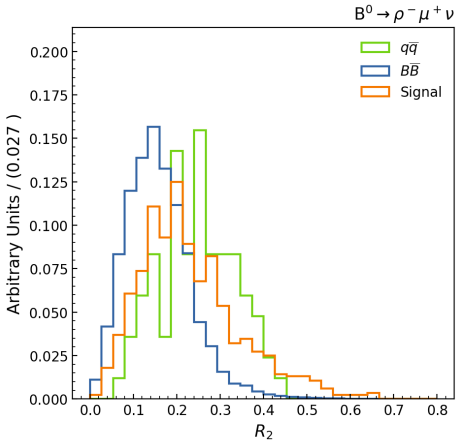
Figure A.9: Normalized distributions of the CS variables for the $B^+ \to \rho^- e^+ \nu$ channel in $q\overline{q}$, $B\overline{B}$ and signal categories from MC.



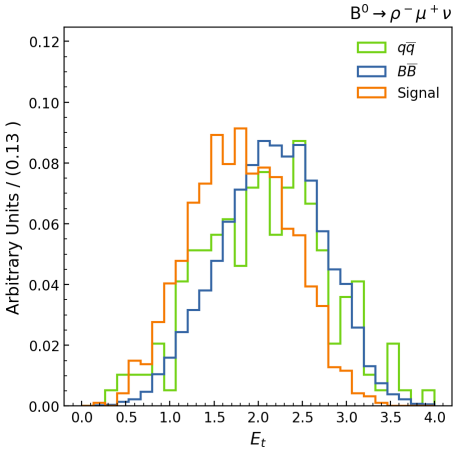
(a)



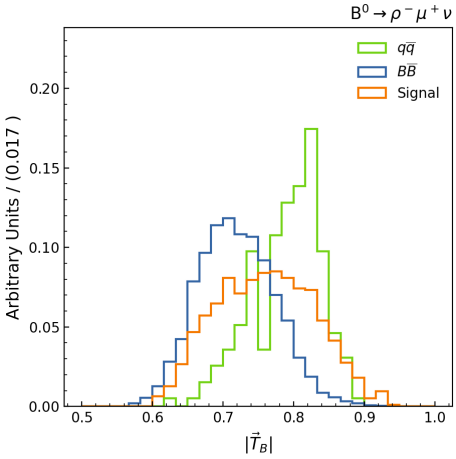
(b)



(c)

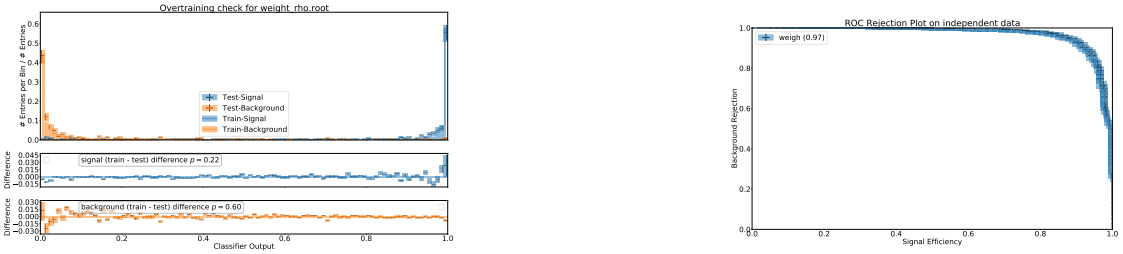


(d)



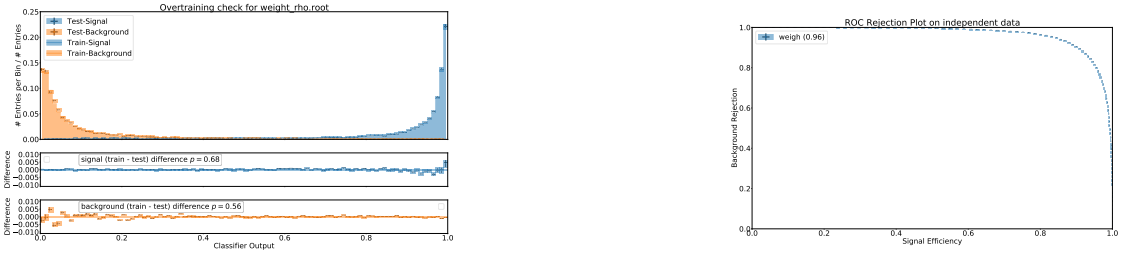
(e)

Figure A.10: Normalized distributions of the CS variables for the $B^+ \to \rho^- \mu^+ \nu$ channel in $q\overline{q}$, $B\overline{B}$ and signal categories from MC.



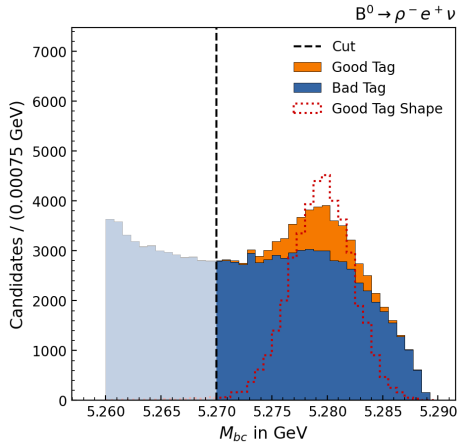
(a) (b)

Figure A.11: Overtraining plot and Roc curve for the tight training for the $B^+ \to \rho^- \mu^+ \nu$ channel.

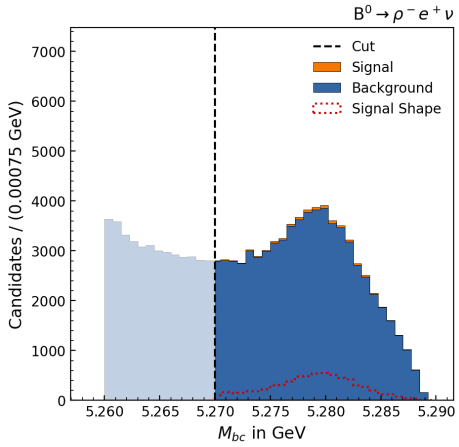


(a) (b)

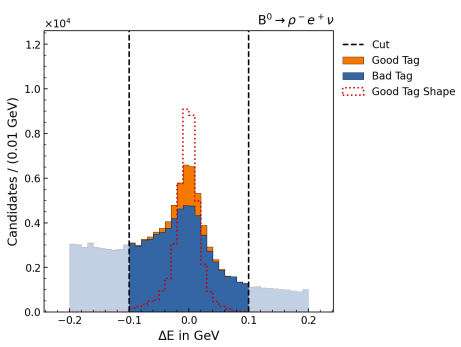
Figure A.12: Overtraining plot and Roc curve for the loose training for the $B^+ \to \rho^- \mu^+ \nu$ channel.



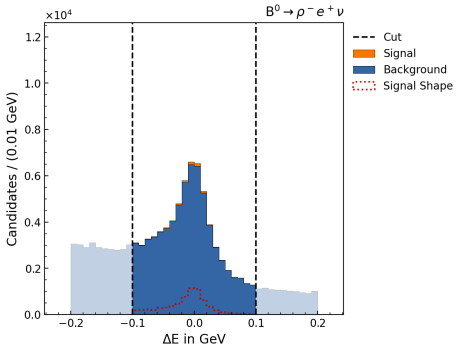
(a)



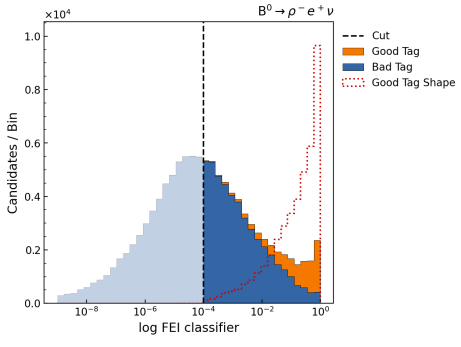
(b)



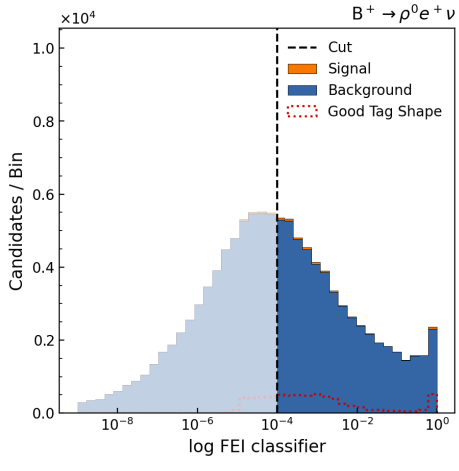
(c)



(d)

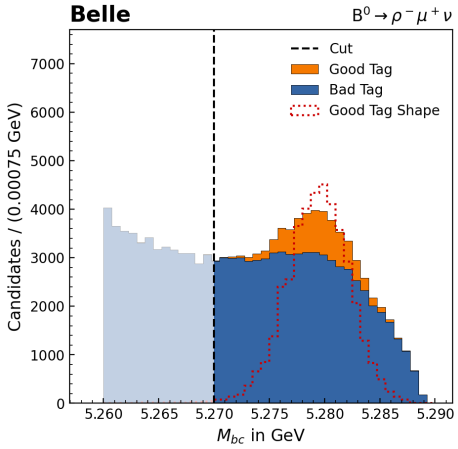


(e)

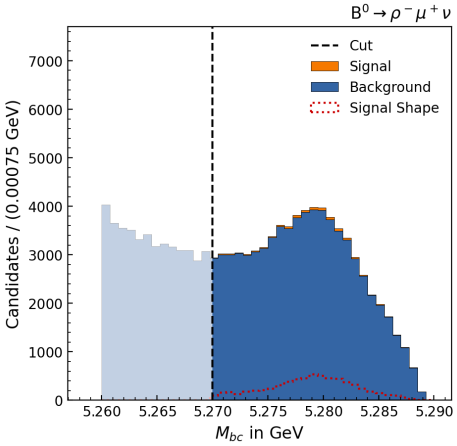


(f)

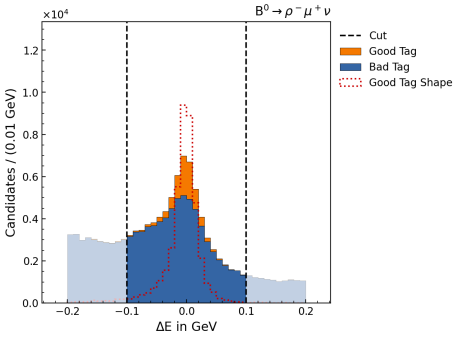
Figure A.13: Tag-side variables for the $B^+ \to \rho^- e^+ \nu$ channel.



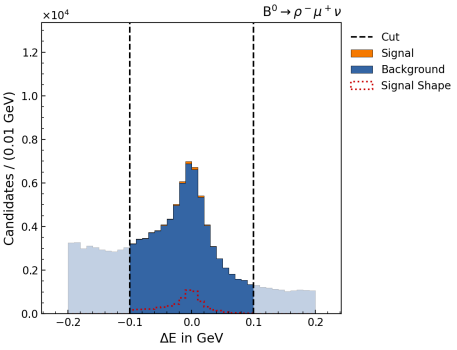
(a)



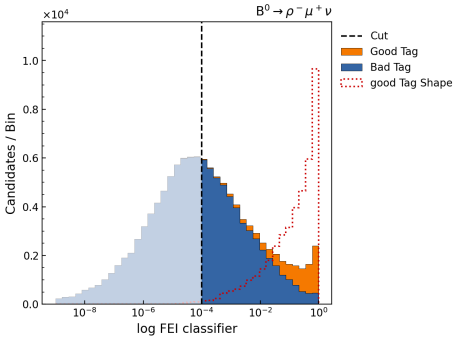
(b)



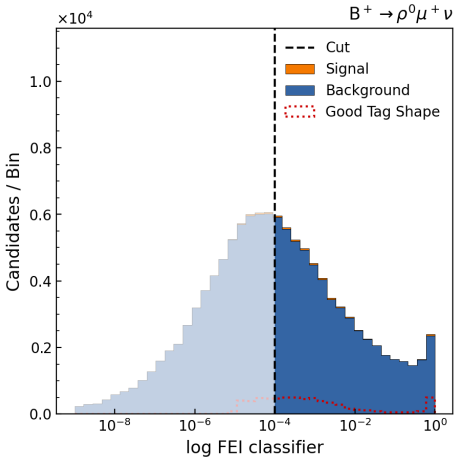
(c)



(d)



(e)



(f)

Figure A.14: Tag-side variables for the $B^+ \to \rho^- \mu^+ \nu$ channel.

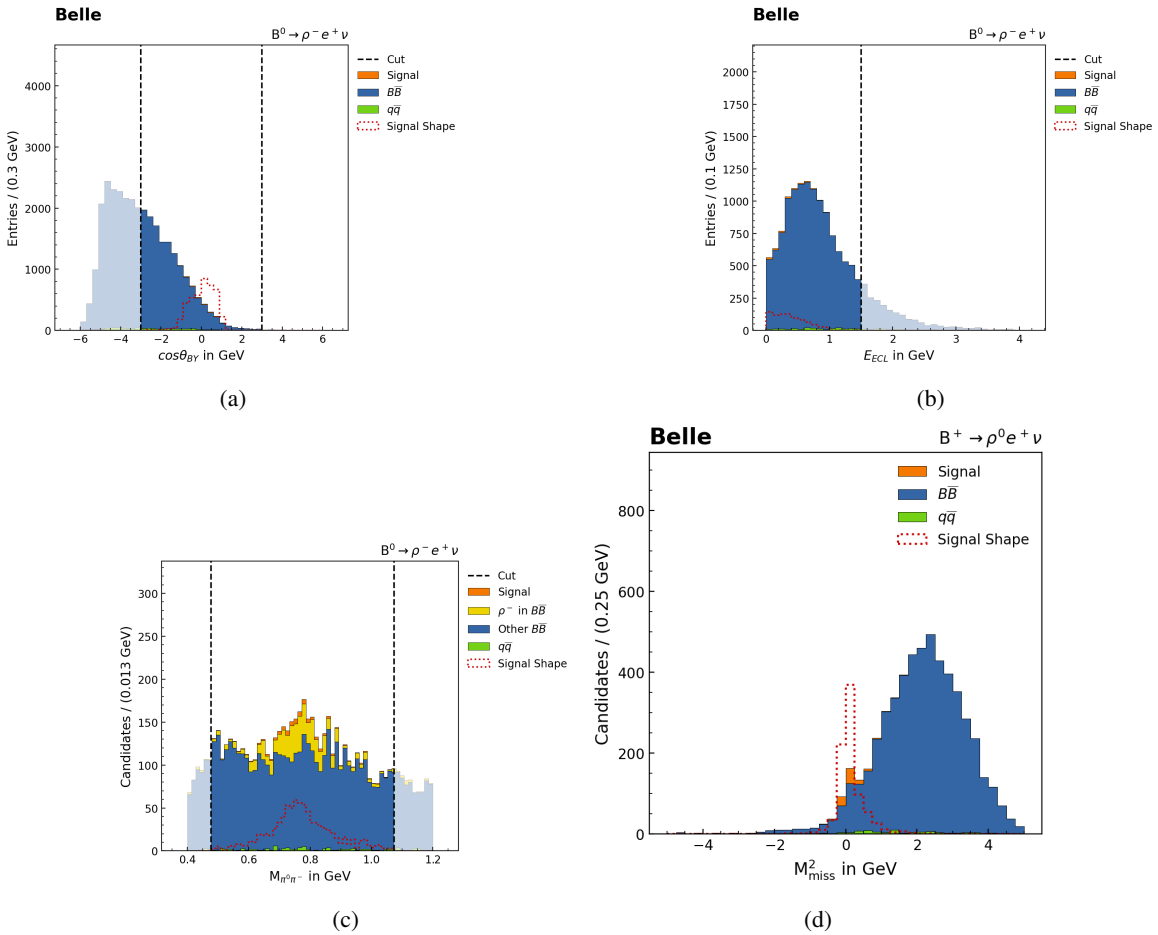


Figure A.15: Signal-side variables for the $B^+ \to \rho^- e^+ \nu$ channel.

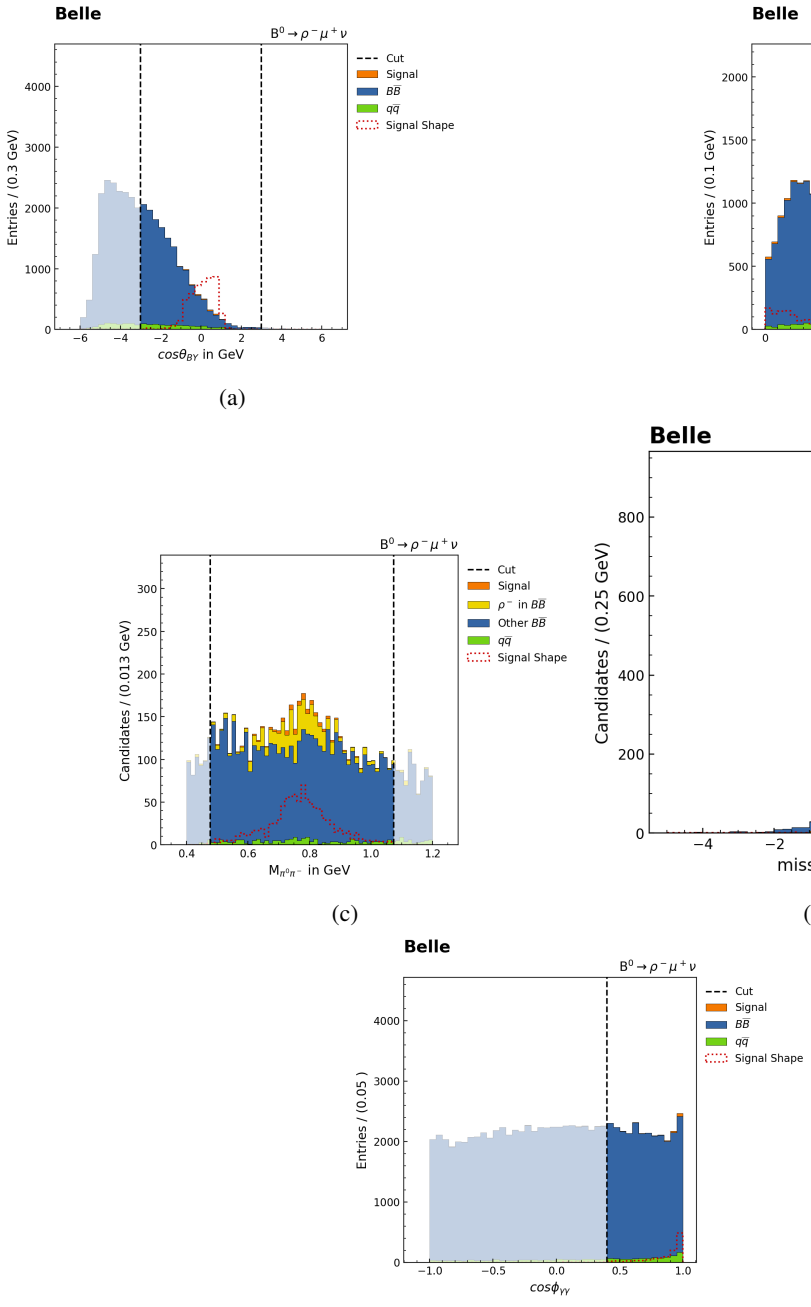


Figure A.16: Signal-side variables for the $B^+ \to \rho^- \mu^+ \nu$ channel.

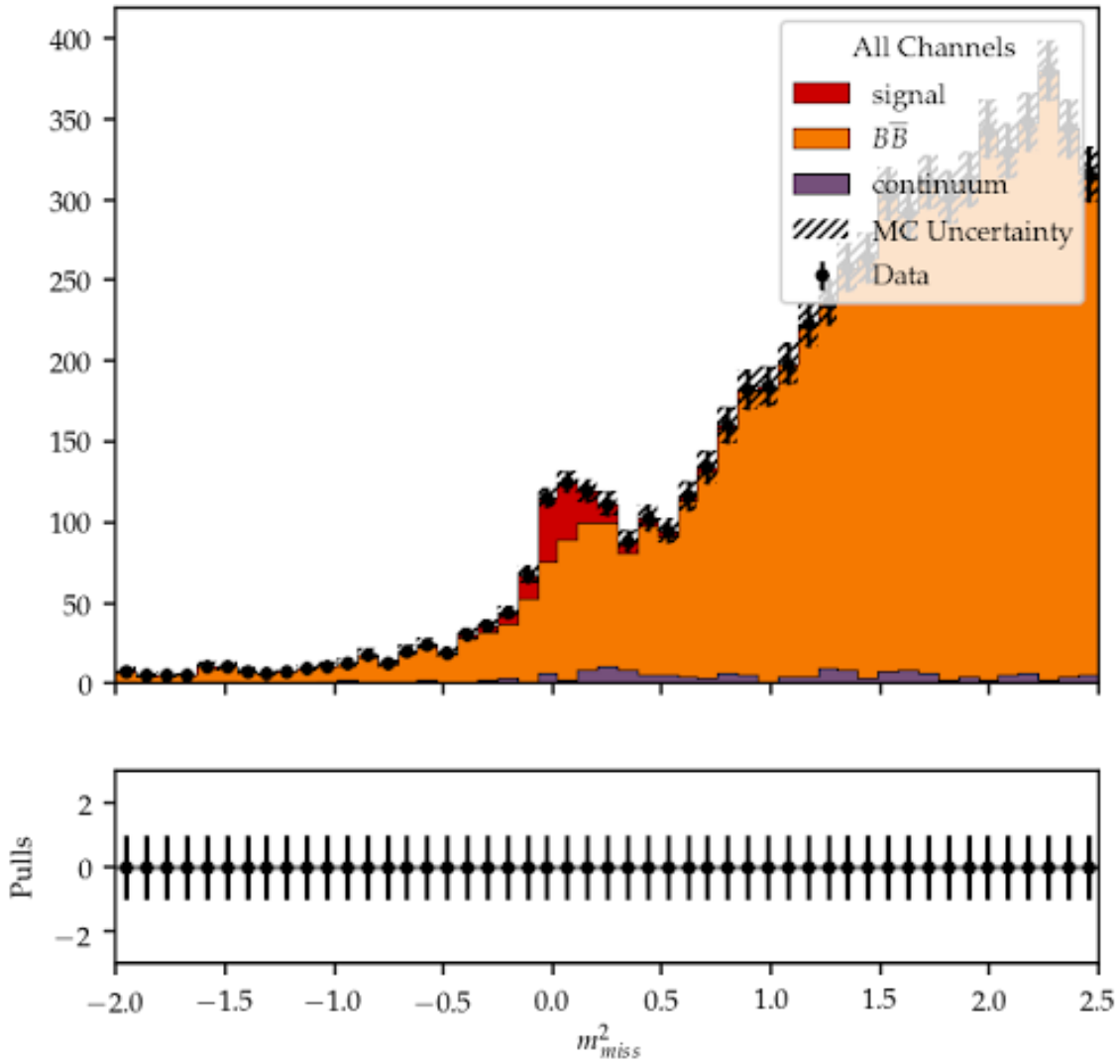
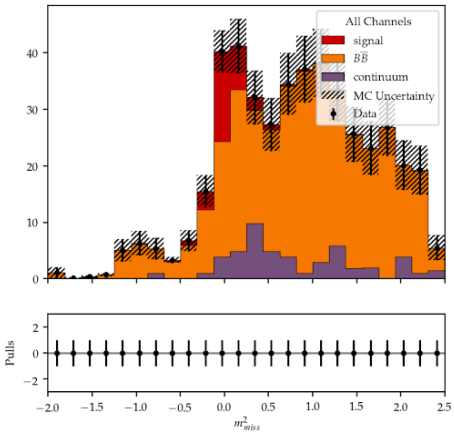
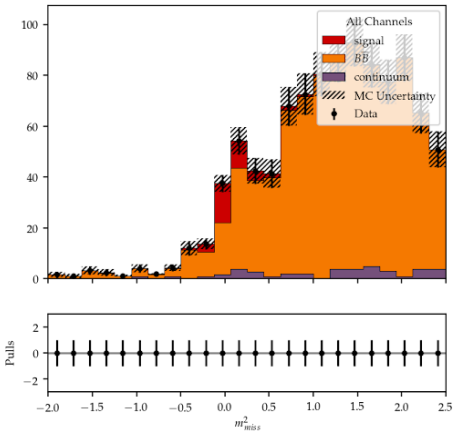


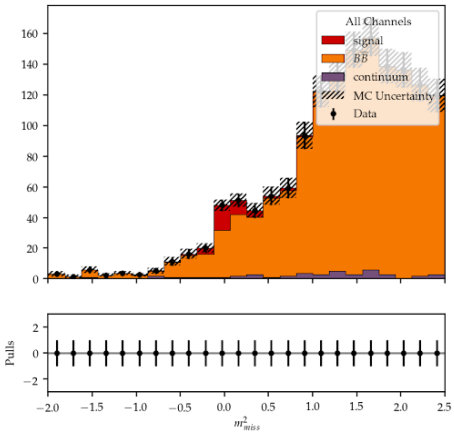
Figure A.17: Fit for the channel $B^+ \to \rho^- \ell^+ \nu$.



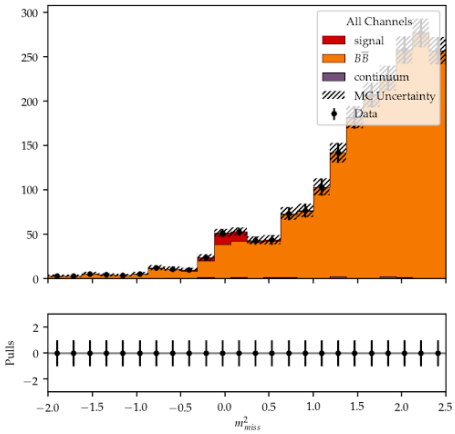
(a) $[0, 4)$



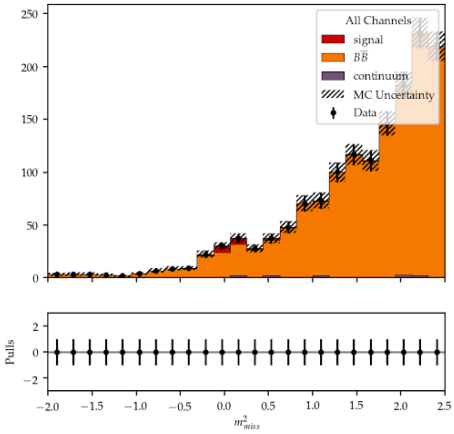
(b) $[4, 8)$



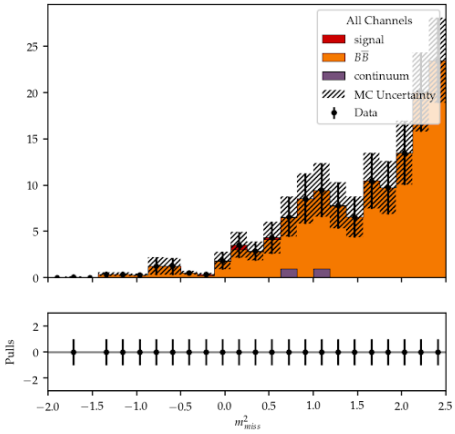
(c) $[8, 12)$



(d) $[12, 16)$

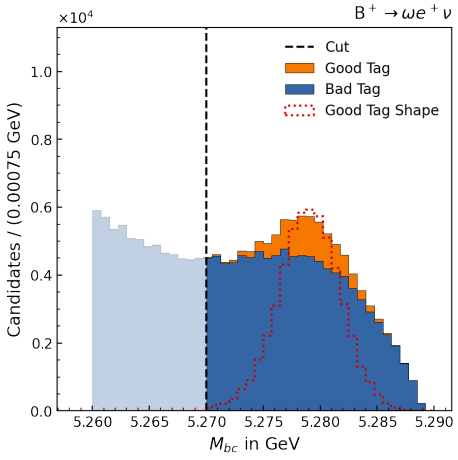


(e) $[16, 20)$

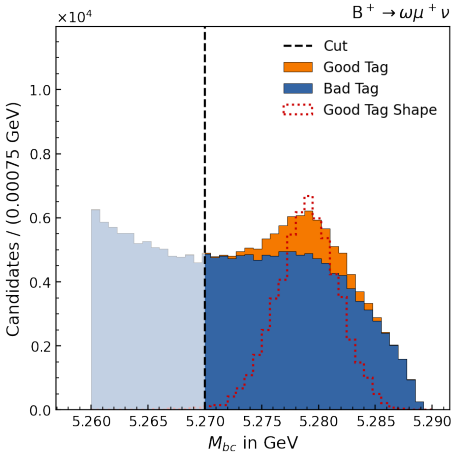


(f) $[20, 24)$

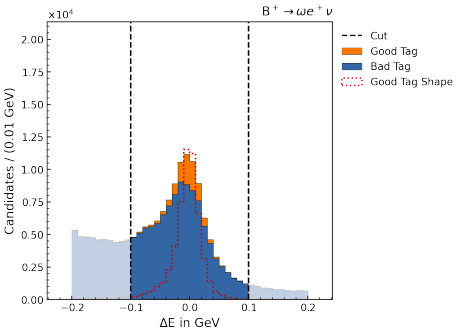
Figure A.18: Fit in 6 q^2 bins for the $B^+ \to \rho^- \ell^+ \nu$ channel.



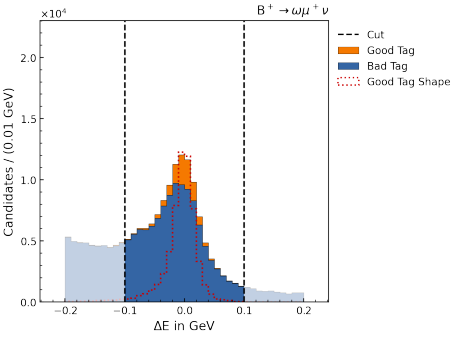
(a)



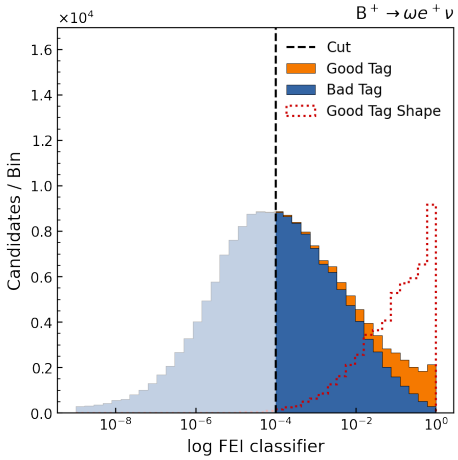
(b)



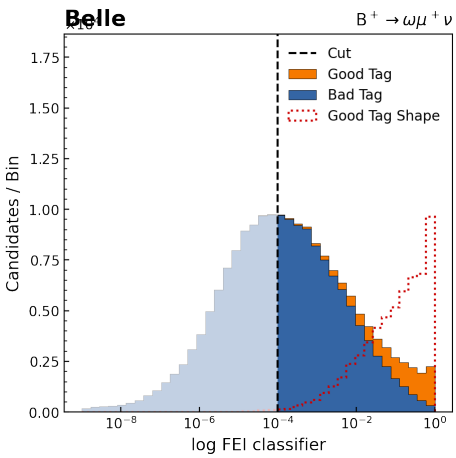
(c)



(d)



(e)



(f)

Figure A.19: Tag-side variables for the $\omega \to \pi^0\pi^+\pi^-$ channel.

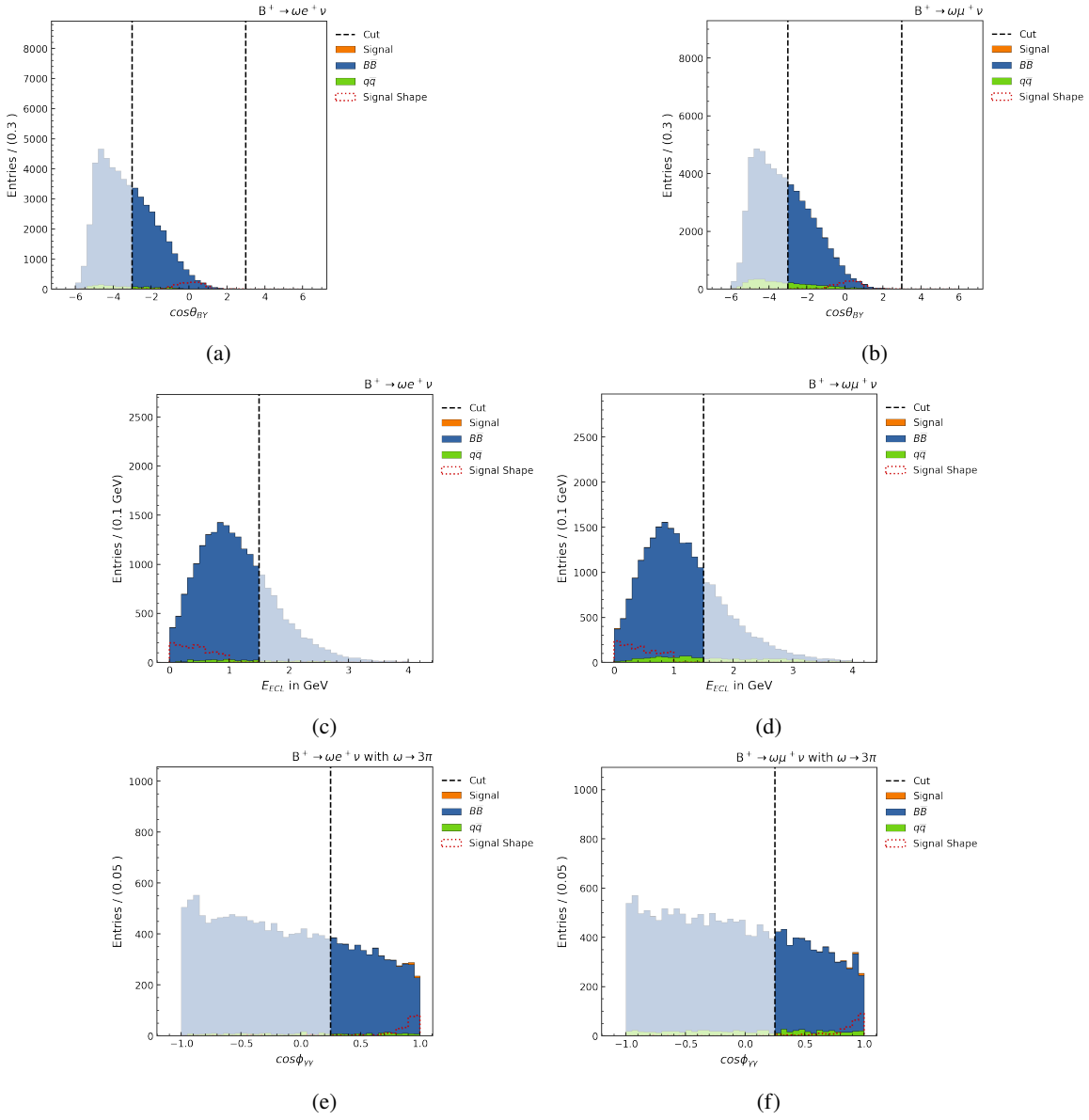


Figure A.20: Signal-side variables for the $\omega \to \pi^0\pi^+\pi^-$ channel.

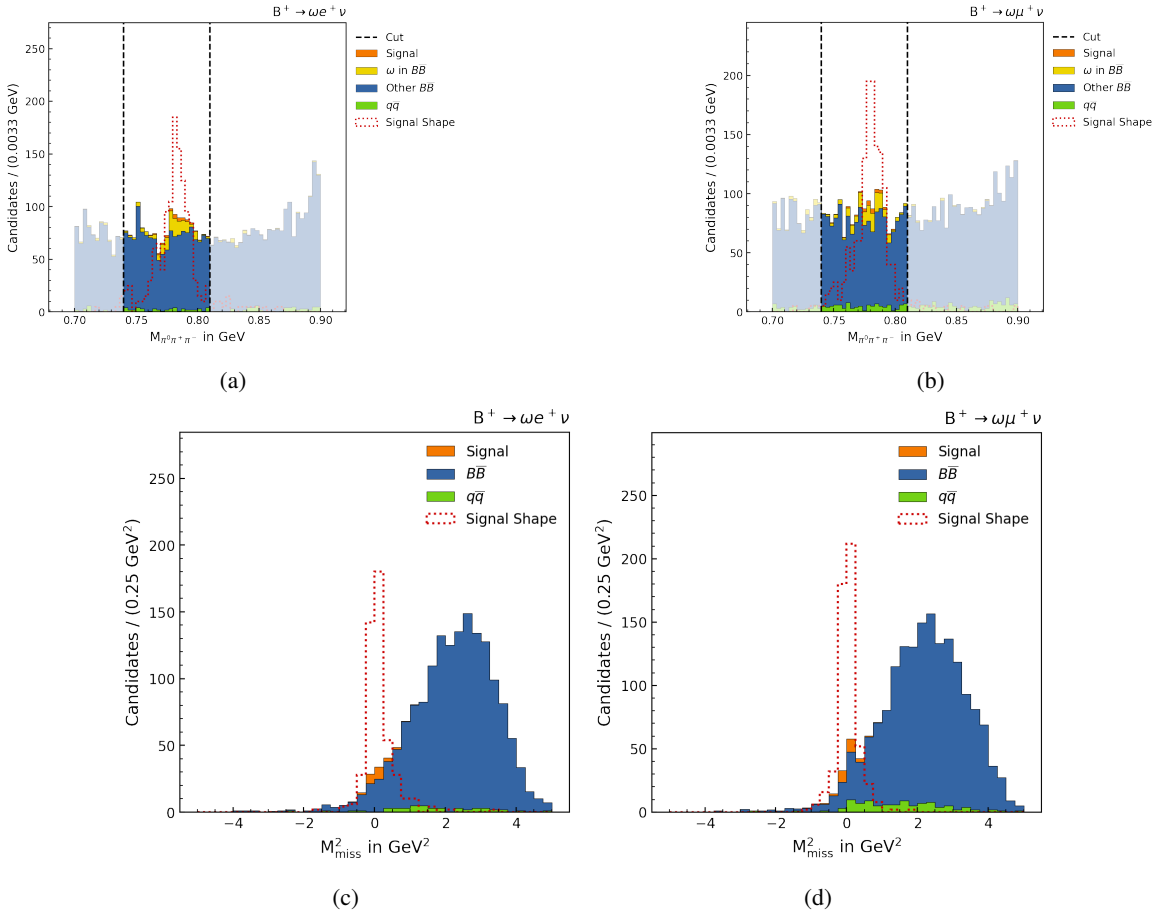
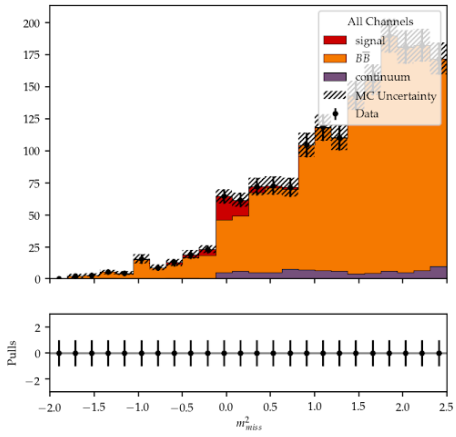
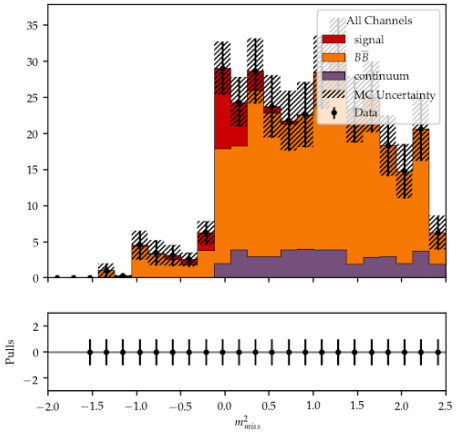


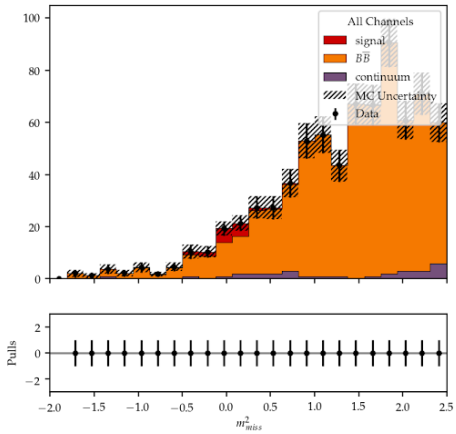
Figure A.21: Signal-side variables for the $\omega \to \pi^0\pi^+\pi^-$ channel.



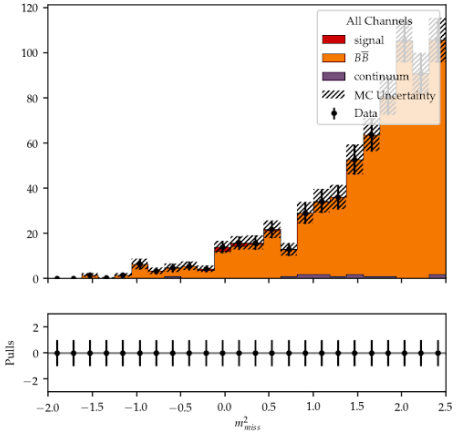
(a) Fit in 1 q^2 bin.



(b) $[0, 7)$

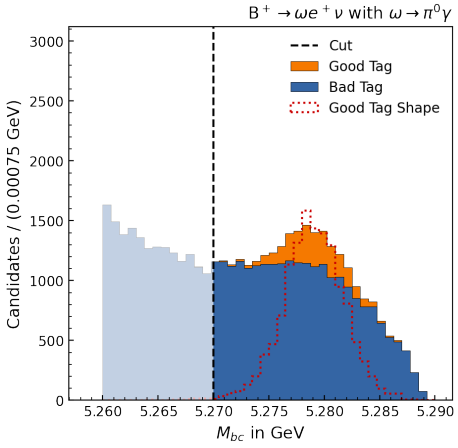


(c) $[7, 14)$

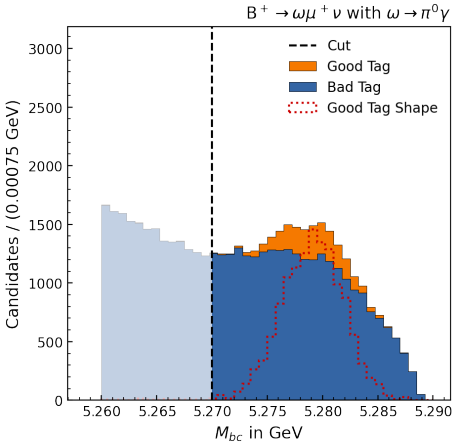


(d) $[14, 21)$

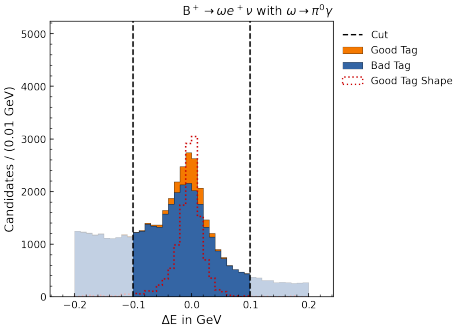
Figure A.22: Fit in 3 q^2 bins for the $\omega \to \pi^0\pi^+\pi^-$ channel.



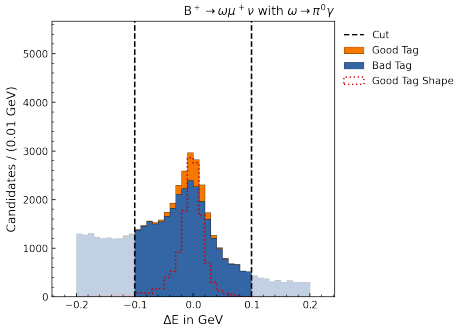
(a)



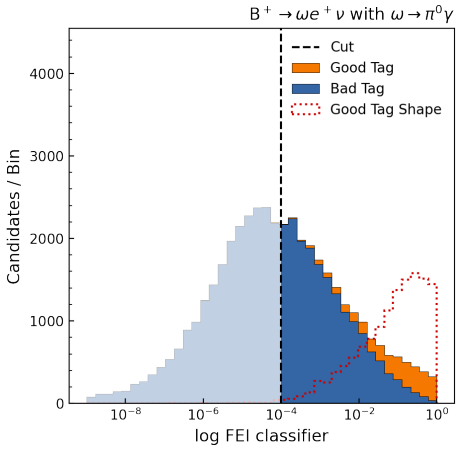
(b)



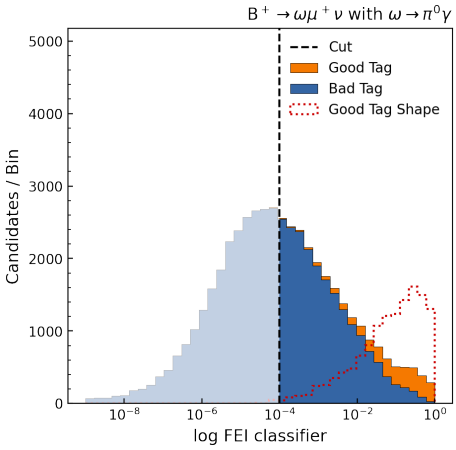
(c)



(d)



(e)



(f)

Figure A.23: Tag-side variables for the $\omega \to \pi^0\gamma$ channel.

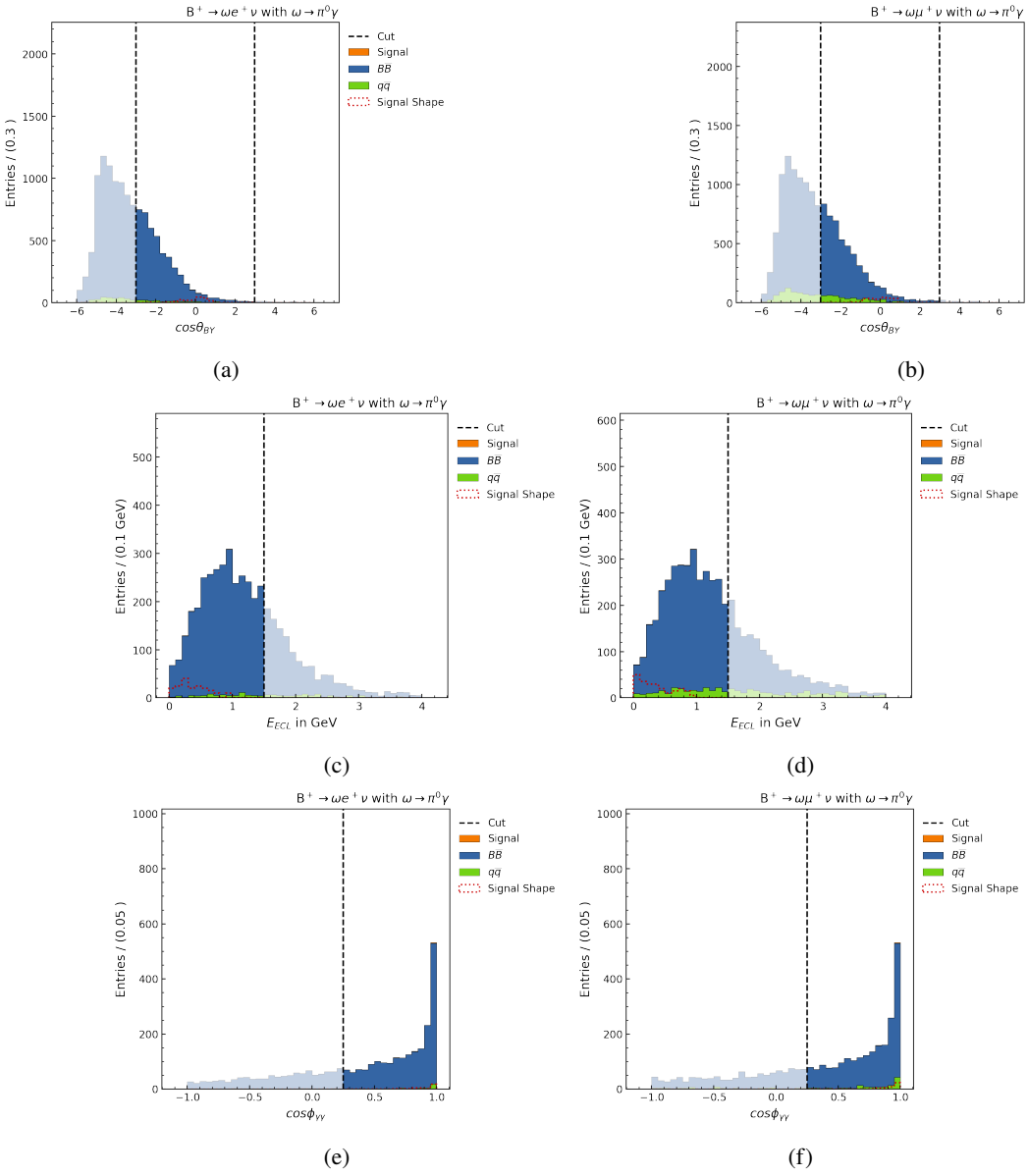


Figure A.24: Signal-side variables for the $\omega \to \pi^0\gamma$ channel.

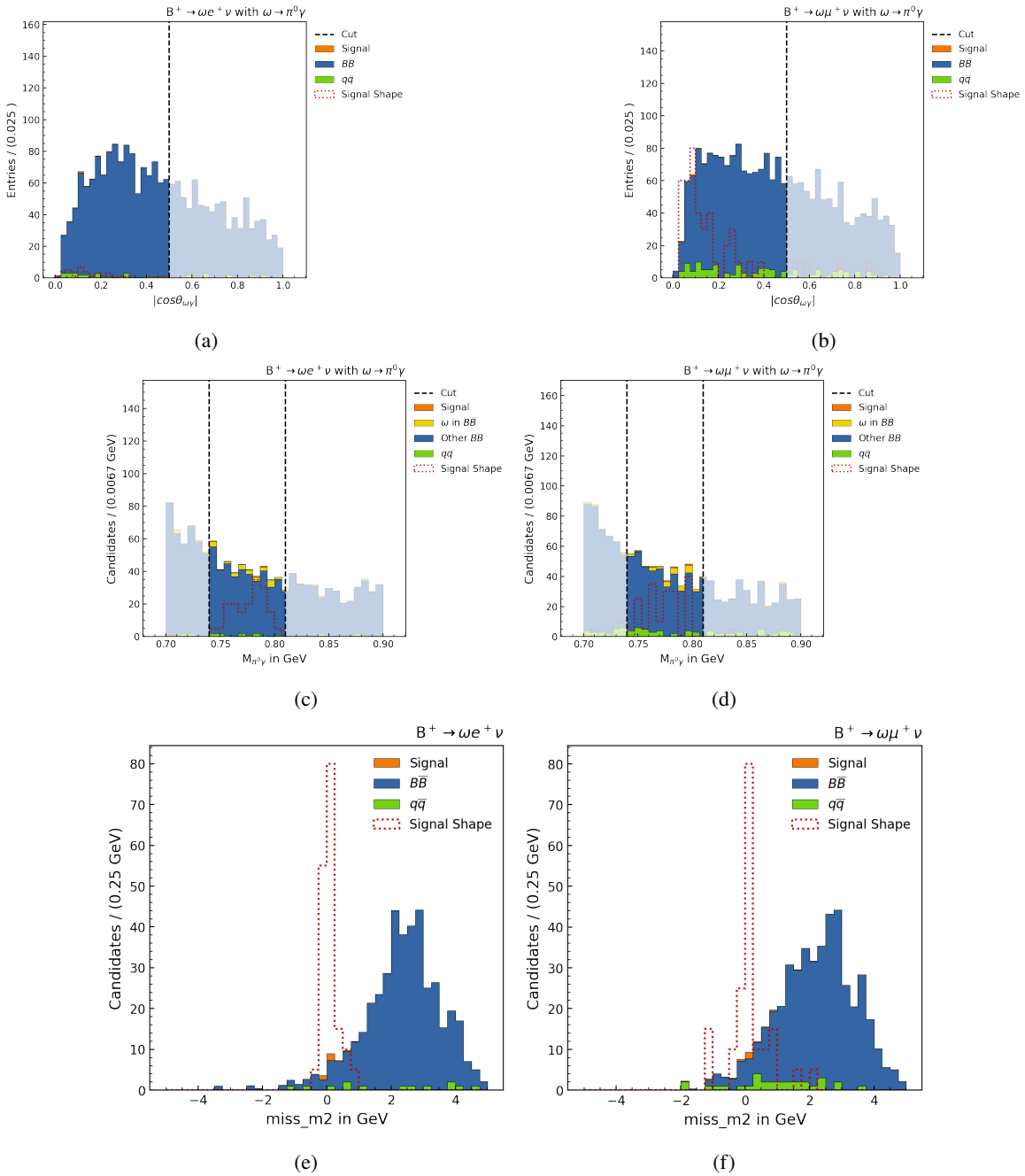


Figure A.25: Signal-side variables for the $\omega \to \pi^0\gamma$ channel.

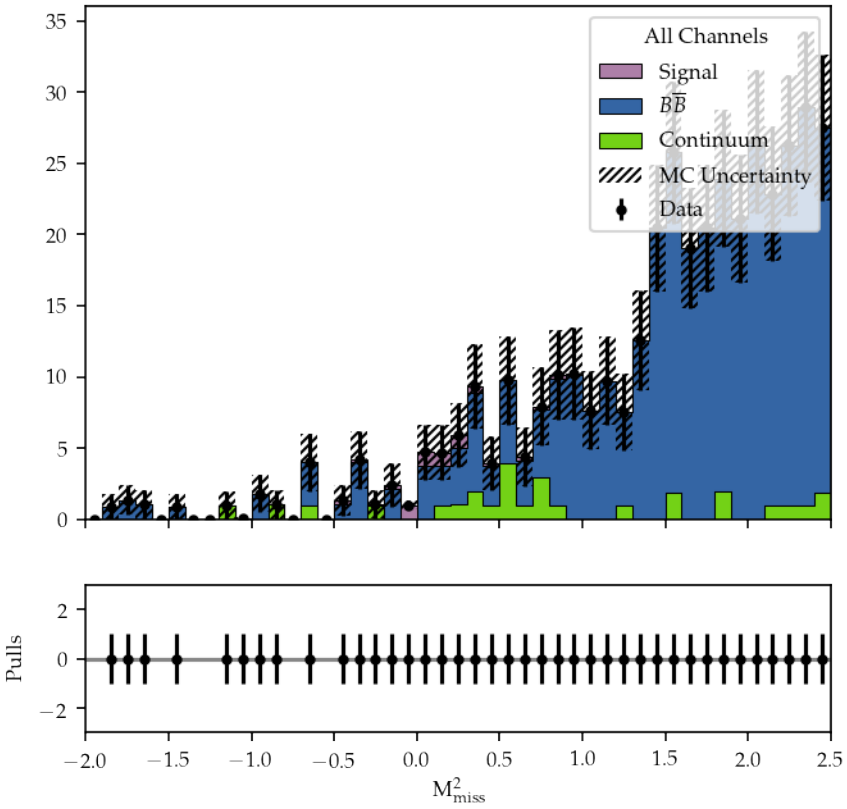


Figure A.26: Fit in 1 q^2 bin for the channel $\omega \to \pi^0\gamma$

List of Figures

List of Tables