SEARCH FOR THE DECAY $B^0 \rightarrow \tau^+ \tau^-$ WITH THE BELLE EXPERIMENT

Michael Ziegler

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Phys. Michael Ziegler

aus Reutlingen

Contents

1. Introduction

The Standard Model of particle physics [1] is the theory describing the known constituents of matter and the interactions between them. Despite the fact that many experimental observations are explained by the Standard Model up to a very high precision, it cannot explain all phenomena. One open question is, e.g. the existence of dark matter in the universe as none of the known particles has the properties to describe it. In order to get a better understanding of nature, an important task is the validation of the Standard Model by searching for discrepancies in observables predicted by the theory and the observed values in nature.

The Belle experiment in Tsukuba, Japan, recorded a large data sample in the years 1999 to 2010, which allows to study the properties of B mesons in great detail. Belle is located at the interaction point of the asymmetric-energy e^+e^- KEKB collider which provides a very clean experimental environment to study B mesons as exactly one pair is created with no further particles. One of the milestones of the experiment was the observation of time-dependent $\mathcal{CP}$-violation in the B meson system [2]. Other remarkable results, like first observations and best upper limits on branching ratios, have been achieved in searches for decays with small predicted branching ratios in the Standard Model [3, 4, 5].

The decay of a neutral B meson into two τ leptons, denoted by $B^0 \rightarrow \tau^+\tau^-$, offers a good way to search for new physics effects since the predicted branching ratio is very small and calculable with relatively small uncertainties. In addition, new physics models can have sizable effects on the observed branching ratio. From an experimental point of view $B^0 \rightarrow \tau^+\tau^-$ is very difficult to detect as the τ leptons decay within the detector and produce neutrinos which are not detectable. However, with the clean experimental environment of the Belle experiment it is possible to select $B^0 \rightarrow \tau^+\tau^-$ events. One of the two B mesons is reconstructed in a fully hadronic decay channel. With the detector signals not used in the reconstruction of the first B meson, the second B meson is recombined in the decay channel $B^0 \rightarrow \tau^+\tau^-$. After the event reconstruction the background level is very high. Therefore, within the scope of this thesis a multivariate selection procedure was

developed to suppress the huge amount of background and enhance the sensitivity of the search. For the first time the complete Belle data sample containing 772 million $B\bar{B}$ pairs was analyzed in order to search for $B^0 \to \tau^+ \tau^-$.

This thesis includes the description of the decay $B^0 \to \tau^+ \tau^-$ in the Standard Model and implications of new physics models are given in Chapter 2. In Chapter 3, the Belle experiment is shortly described, followed by an introduction to the tools and methods used in the analysis (Chapter 4). An overview of the analysis steps is given in Chapter 5, followed by the more detailed chapters about the reconstruction (Chapter 6), the description of the selection procedure (Chapter 7), and the validation of the selection (Chapter 8). The extraction of the branching ratio is described in Chapter 9 and the result of the measurement in Chapter 10. Various cross-checks and their results are discussed in Chapter 11.

2. The Decay $B^0 \to \tau^+\tau^-$

The Standard Model (SM) of particle physics was developed over the last 60 years and is phenomenologically very successful in the field of flavor physics. Still, there are arguments that the SM is not the complete model and only valid at low energies. In the complete model, new particles may enter the stage. Under the assumption that these new particles are heavier than the SM ones, the new particles and the heavy ones of the SM can be integrated out in order to get an effective theory. An effective theory has the advantage that effects from new physics (NP) can be described in terms of light SM fields [6]. Rare B decays are a good opportunity to look for such effects as they have to be small.

In this chapter, the rare decay $B^0 \to \tau^+\tau^-$ is described, where the SM is considered as an effective field theory at the B meson mass scale (Section 2.1). Afterwards, two different extensions of the SM and their effects on $B^0 \to \tau^+\tau^-$ are depicted in Section 2.2. The experimental status of the search for $B^0 \to \tau^+\tau^-$ is discussed in Section 2.3.

2.1. Calculation of the Branching Ratio

In the SM of particle physics flavor-changing neutral currents (FCNC) are forbidden at tree level and first occur via box or penguin Feynman diagrams. Hence, decays where a b quark converts into a d or s quark are highly suppressed. In the decay of a neutral $B^0 = |\bar{b}d\rangle$ meson into two oppositely charged leptons, written as $B^0 \to \ell^+\ell^-$, such FCNCs occur. The main contributions in the SM for the decays $B^0 \to \ell^+\ell^-$ come from the W box and Z penguin shown in Fig. 2.1. Contributions from diagrams where c or u quarks appear in the loop can be neglected due to their small masses m_c and m_u, respectively, compared to the top mass m_t [7, 8]. Calculating the branching ratio $\mathcal{B}(B^0 \to \ell^+\ell^-)$ using the full SM Lagrangian is problematic, since two highly separated energy scales are involved in the decay: the electroweak scale, characterized by the W boson mass M_W, and the scale of hadronic strong interactions λ_{QCD} [6]. The electroweak scale determines the flavor-changing transition at quark level, whereas the hadron formation is related to λ_{QCD}.

(a) $B^0 \to \ell^+\ell^-$ via a W box.
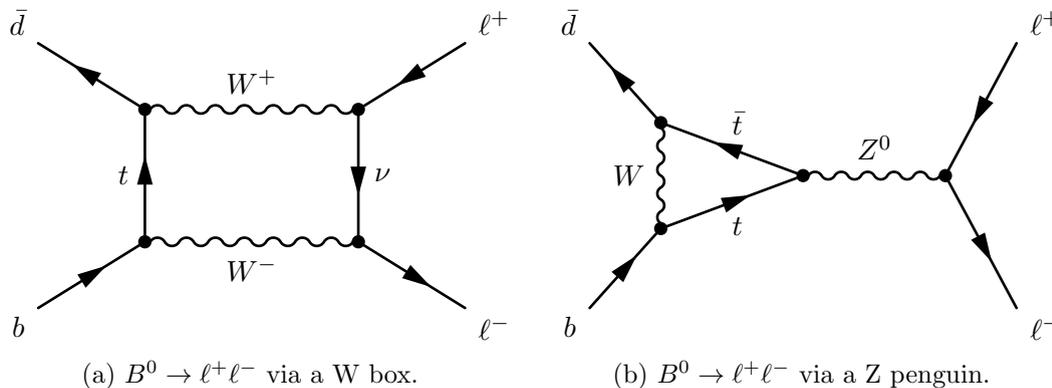

(b) $B^0 \to \ell^+\ell^-$ via a Z penguin.

Figure 2.1.: Dominant Feynman diagrams for the decay $B^0 \to \ell^+\ell^-$ in the SM with $\ell = e, \mu, \tau$.

A common tool set used in flavor physics to calculate decays like $B^0 \to \ell^+\ell^-$ is the framework of effective field theory. By integrating out the heavy SM fields at the electroweak scale, like W, Z, and the top quark, a low-energy effective theory can be constructed where the only degrees of freedom are the light SM fields [6]. A detailed introduction into the concept and construction of effective field theory can be found in [8]. The obtained effective Hamiltonian $\mathcal{H}_{\text{eff}}$ contains local operators, which describe the processes at low energies, and can be written as

$$\mathcal{H}_{\text{eff}} = \frac{G_{\text{F}}}{\sqrt{2}} \sum_i V_{\text{CKM}}^i C_i(\mu) Q_i, \tag{2.1}$$

with the index i denoting different operators, the Fermi constant G_{F}, the Cabibbo-Kobayashi-Maskawa (CKM) factors V_{CKM} [9], the local operators Q_i, and their effective couplings $C_i(\mu)$, called Wilson coefficients, at an energy scale μ. The operators Q_i themselves can be written in terms of the light SM fermions, photon and gluon fields [6]. The local operators describing the decay $B^0 \to \ell^+\ell^-$ allowed in the SM are

$$Q_S = m_b(\bar{b}P_L d)(\bar{\ell}\ell), \tag{2.2}$$

$$Q_P = m_b(\bar{b}P_L d)(\bar{\ell}\gamma_\mu \ell), \tag{2.3}$$

$$Q_A = (\bar{b}\gamma^\mu P_L d)(\bar{\ell}\gamma_\mu \gamma_5 \ell), \tag{2.4}$$

where $P_L = (1 - \gamma_5)/2$ is the left handed projection operator, m_b is the b quark mass, b and d are the quark fields, ℓ the lepton fields, and γ_μ and γ_5 the Dirac matrices [7]. The three operators Q_S, Q_P, and Q_A describe the scalar, pseudo-scalar, and axial coupling of the fermions, respectively. The effective Hamiltonian in the SM for $B^0 \to \ell^+\ell^-$ is

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \frac{\alpha_{\text{em}}}{\sin^2 \theta_W} V_{tb}^* V_{td} \left[C_S(\mu) Q_S + C_P(\mu) Q_P + C_A(\mu) Q_A \right], \tag{2.5}$$

with the fine structure constant α_{em} and the Weinberg angle θ_W. The Wilson coefficient C_S receives contributions from the Higgs boson exchange but only of the order $\mathcal{O}(M_B^2/M_W^2)$ (with the B^0 meson mass M_B) relative to the dominant contributions and can therefore be neglected [7, 10, 11]. C_P gets a contribution in the same order of magnitude from the

would-be neutral Goldstone boson [7] and can also be neglected. The W box and the Z penguin in Fig. 2.1 contribute to the Wilson coefficient C_A of the axial-current operator Q_A.

Finally $\mathcal{H}_{\text{eff}}$ can be written as

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \frac{\alpha}{\sin^2 \theta_W} V_{tb}^* V_{td} C_A(\mu) (\bar{b} \gamma^\mu P_L d)(\bar{\ell} \gamma_\mu \gamma_5 \ell). \tag{2.6}$$

Using Eq. (2.6) the branching ratio $\mathcal{B}(B^0 \to \ell^+ \ell^-)$ is given by

$$\mathcal{B}(B^0 \to \ell^+ \ell^-) = \frac{G_F^4 M_W^4 M_B^3}{8\pi^5 \Gamma_B} \cdot \underbrace{f_B^2}_{\substack{\text{Decay} \\ \text{constant}}} \cdot \underbrace{|V_{tb}^* V_{td}|^2}_{\substack{\text{CKM} \\ \text{elements}}} \cdot \underbrace{\frac{4m_\ell^2}{M_B^2}}_{\substack{\text{Helicity} \\ \text{suppression}}} \cdot \underbrace{\sqrt{1 - \frac{4m_\ell^2}{M_B^2}}}_{\substack{\text{Phase} \\ \text{space factor}}} \cdot |C_A(\mu)|^2,$$

$$\tag{2.7}$$

with the decay width of the B^0 meson Γ_B, the lepton mass m_ℓ, the CKM matrix elements V_{tb}^* and V_{td}, and the B decay constant f_B. The decay constant f_B absorbs the strong interactions between the b and d quark in the B^0 meson. It is defined by the hadronic matrix element $\langle 0 | \bar{b} \gamma_\mu \gamma_5 d | B^0(p) \rangle = i p_\mu f_B$ [10, 6] (with the momentum p of the B meson) and can be calculated using lattice QCD [12]. Two factors in Eq. (2.7) depend on the lepton mass m_ℓ: the phase space factor and the helicity suppression. While the phase space factor decreases for large lepton masses, the helicity suppression factor behaves contrarily. The helicity suppression has its origin in the parity violating nature of the weak interaction. In Fig. 2.2 the possible momentum and spin configurations of the final state leptons in the decay $B^0 \to \ell^+ \ell^-$ in the rest frame of the B^0 are shown. As the B^0 is a pseudo-scalar
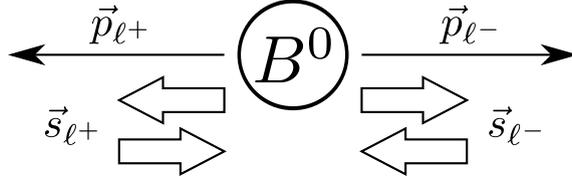

Figure 2.2.: Momentum $\vec{p}$ and spin $\vec{s}$ configuration of the final state leptons in the decay $B^0 \to \ell^+ \ell^-$ in the center-of-mass system of the B^0 meson. Either the upper or the lower configuration is possible.

particle and due to angular momentum conservation, the total spin of the leptons must be zero. Therefore, either the lepton or the anti-lepton must be generated with the wrong helicity. The higher the mass of the lepton, the weaker the suppression of the decay.

Currently the most precise theoretical values of $\mathcal{B}(B^0 \to \ell^+ \ell^-)$ in the SM are

$$\mathcal{B}(B^0 \to e^+ e^-)_{\text{SM}} = (2.48 \pm 0.21) \times 10^{-15}, \tag{2.8}$$

$$\mathcal{B}(B^0 \to \mu^+ \mu^-)_{\text{SM}} = (1.06 \pm 0.09) \times 10^{-10}, \tag{2.9}$$

$$\mathcal{B}(B^0 \to \tau^+ \tau^-)_{\text{SM}} = (2.22 \pm 0.19) \times 10^{-8}, \tag{2.10}$$

and have been calculated in [10]. The uncertainties on the CKM elements and the decay constant f_B are the dominant sources of uncertainty for the theoretical prediction [10].

Depending on the flavor of the final state lepton, the branching ratios increase by seven orders of magnitude from $\ell = e$ to $\ell = \tau$. This large violation of lepton flavor universality arises from the helicity suppression as described above.

2.2. Effects of New Physics

Particles appearing in NP models can replace, for example, the role of the W boson in the decay $B^0 \to \ell^+ \ell^-$. New contributions to the effective Hamiltonian in Eq. (2.7) can occur, either by enhancing C_A or by introducing sizable values for other Wilson coefficients. In this section, two important extensions of the SM are discussed. First, the two Higgs-doublet model (Section 2.2.1) and afterwards effects of models including leptoquarks (Section 2.2.2) are described.

2.2.1. Two Higgs-Doublet Model

In the two Higgs-doublet model (2HDM) of type II, the Higgs sector contains two doublet scalar fields. This leads to additional charged and neutral Higgs bosons ($H^\pm, A^0$) which can replace the role of the W and Z bosons in the decay [7]. In Fig. 2.3, a box and a penguin diagram are shown where the H^+ and A^0 replace the W and Z boson, respectively. By
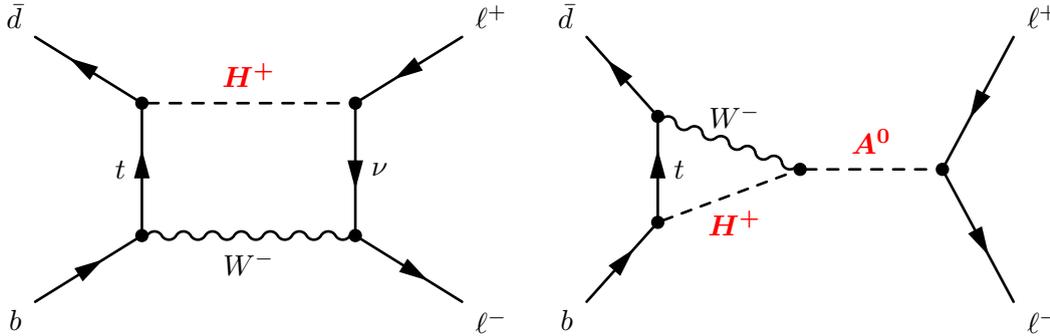

Figure 2.3.: Dominant diagrams for the decay $B^0 \to \ell^+ \ell^-$ in the 2HDM with large $\tan \beta$ ($\ell = e, \mu, \tau$).

including these new particles sizable effects on the Wilson coefficients C_S and C_P, defined in Eq. (2.5), can occur. The Wilson coefficient C_A is not influenced by the 2HDM and remains unchanged [7]. Including the effects from the 2HDM the branching ratio reads as follows

$$\mathcal{B}(B^0 \to \ell^+ \ell^-)_{2\text{HDM}} = \frac{G_F^4 M_W^4}{128\pi^5} \frac{M_B^3}{\Gamma_B} f_B^2 |V_{tb}^* V_{td}|^2 \sqrt{1 - \frac{4m_\ell^2}{M_B^2}}$$
$$\times \left[\left(M_B C_P(\mu) - \frac{2m_\ell}{M_B} C_A(\mu) \right)^2 + \left(1 - \frac{4m_\ell^2}{M_B^2} \right) M_B C_S(\mu) \right].$$
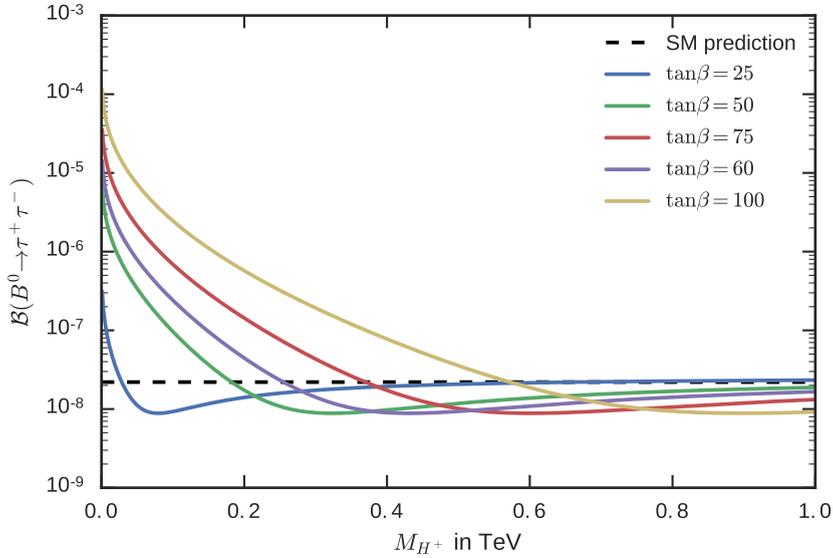$$(2.11)$$

Figure 2.4.: $\mathcal{B}(B^0 \to \tau^+\tau^-)$ in the 2HDM as a function of M_H^+ for different $\tan\beta$ values. The plot is generated using the results in Reference [7]. The SM prediction is given in Eq. (2.10).

The Wilson coefficients C_S and C_P depend on the mass of the charged Higgs (M_{H^+}) and on the ratio of the vacuum expectation values of the two Higgs fields ($\tan\beta$). In Fig. 2.4, the theoretical values of the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$ in the 2HDM as a function of M_{H^+} are shown. Depending on $\tan\beta$ and M_{H^+}, the influence on $\mathcal{B}(B^0 \to \tau^+\tau^-)$ from the 2HDM can be large.

2.2.2. Leptoquarks

In the theoretical models of extended technicolor [13] or quark- and leptoncomposite models [14] so-called scalar leptoquarks appear [15]. They carry both baryon and lepton number, and therefore they can couple leptons directly to quarks. Including the leptoquarks, Wilson coefficients negligible in the SM get sizable contributions from interactions between SM particles and new particles. For example, the Wilson coefficient of the local operator

$$\mathcal{O} = (\bar{b}\gamma_\mu(1-\gamma_5)d)(\bar{\ell}\gamma^\mu(1+\gamma^5)\ell), \qquad (2.12)$$

which describes the coupling of left-handed quarks to right-handed lepton fields, can get a value different from zero. The new coefficients depend on the coupling of the leptoquarks to quarks and leptons λ_{ij}, where i and j are the indices of the quark and lepton families, respectively. By comparing measured values of $\mathcal{B}(B^0 \to \ell^+\ell^-)$ to the SM predictions it is possible to constrain the couplings of the leptoquarks to the different quark and lepton generations. In Table 2.1, the constraints on the couplings, using the latest measurements of $B^0 \to \mu^+\mu^-$ and $B_s \to \mu^+\mu^-$ from [16], and the upper limits on $\mathcal{B}(B^0 \to \tau^+\tau^-)$ in [17], are shown. Searching for the decay $B^0 \to \tau^+\tau^-$ with the complete Belle data sample can set tighter constraints on the couplings of the leptoquarks to the third lepton generation.

Table 2.1.: Constraints on the leptoquarks couplings obtained from $B^0 \rightarrow \ell^+\ell^-$ decays. The values are taken from [15].

Decay	Involved couplings	Constraints on the couplings (in GeV^{-2})
$B_d \rightarrow e^+e^-$	$\frac{|\lambda^{31}\lambda^{11*}|}{M_S^2}$	$< 1.73 \times 10^{-5}$
$B_d \rightarrow \mu^+\mu^-$	$\frac{|\lambda^{32}\lambda^{12*}|}{M_S^2}$	$(1.5, 3.9) \times 10^{-9}$
$B_d \rightarrow \tau^+\tau^-$	$\frac{|\lambda^{33}\lambda^{13*}|}{M_S^2}$	$< 1.28 \times 10^{-6}$

2.3. Previous Searches for $B^0 \rightarrow \tau^+\tau^-$

In the past, one search for the decay $B^0 \rightarrow \tau^+\tau^-$ was performed by the *BABAR* collaboration [18]. The analysis is based on a data sample of 210 fb^{-1} at the center-of-mass energy of $\Upsilon(4S)$ corresponding to $(232 \pm 3) \times 10^6$ $B\bar{B}$ events, which is approximately half of the final *BABAR* sample. An exclusive hadronic tagging method was used to reconstruct the accompanying B meson [19]. Using τ decay channels with only one charged track in the final state, signal candidates have been recombined out of the remaining particles in the event. Due to the large background component, no significant signal was found and an upper limit was set on $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)$ at the 90% confidence level (CL) of

$$\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)_{BABAR} < 4.1 \times 10^{-3}. \tag{2.13}$$

The estimated limit in Eq. (2.13) is five orders of magnitude above the SM expectation. Constraints derived from $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)$ benefit from a search performed on the full Belle data sample. The data sample collected with the Belle experiment corresponds to roughly three times the number of $B\bar{B}$ events used in *BABAR*'s analysis. Hence, it can be expected to lower the limit by a factor of $\sqrt{3}$.

3. The Belle Experiment

In this chapter, the technical setup of the Belle experiment is described. It consists of the KEKB accelerator (Section 3.2) and the Belle detector (Section 3.3), both located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. The KEKB accelerator is often called a B factory. This is due to the chosen center-of-mass energy corresponding to the invariant mass of the $\Upsilon(4S)$ resonance (Section 3.1). It has a relatively large production cross section and decays in almost all cases into a pair of B mesons. These two facts make it possible to collect a large data sample of B decays in a very clean environment.

3.1. The $\Upsilon(4S)$ Resonance

The Υ resonances are bound states of a b and an anti-b quark with the quantum numbers $J^{PC} = 1^{--}$. They can be produced directly in e^+e^- collisions. In Fig. 3.1, the cross section for $e^+e^- \to$ hadrons is shown. There are four resonances in the depicted mass region: the $\Upsilon(1S)$, the $\Upsilon(2S)$, the $\Upsilon(3S)$, and the $\Upsilon(4S)$ resonance. The cross section producing other hadrons in $e^+e^- \to q\bar{q}$ processes (with $q = u, d, s, c$) is flat in the same mass window (grey area in Fig. 3.1). These processes are called continuum background. Fig. 3.1 also shows that the cross section for continuum processes is about three times higher than the cross section of $\Upsilon(4S)$ at the energy corresponding to the mass of $\Upsilon(4S)$ $M_{\Upsilon(4S)} = (10.5794 \pm 0.0012)$ GeV/c^2 [17]. The widths of $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ are small compared to the width of $\Upsilon(4S)$, as their decays are suppressed by the Okubo-Zweig-Iizuka (OZI) rule [21, 22, 23]. The mass of the $\Upsilon(4S)$ resonance is only 20 MeV/c^2 over the B meson threshold. Therefore, $\Upsilon(4S)$ decays in $> 96\%$ of cases solely into $B^0\bar{B}^0$ and B^+B^- pairs. The B mesons originating from a $\Upsilon(4S)$ decay evolve coherently. This allows for the measurement of time-dependent CP violation in the B system [24]. Another outstanding feature of the $\Upsilon(4S)$ is the fact that, due to the small mass difference between $\Upsilon(4S)$ and its decay products of $M_{\Upsilon(4S)} - 2M_B \approx 20$ MeV/c^2, only two B mesons are produced in the decay and nothing more. Using special analysis techniques (see Section 4.2) it is possible to search for B meson decays with more than one neutrino in the final state.
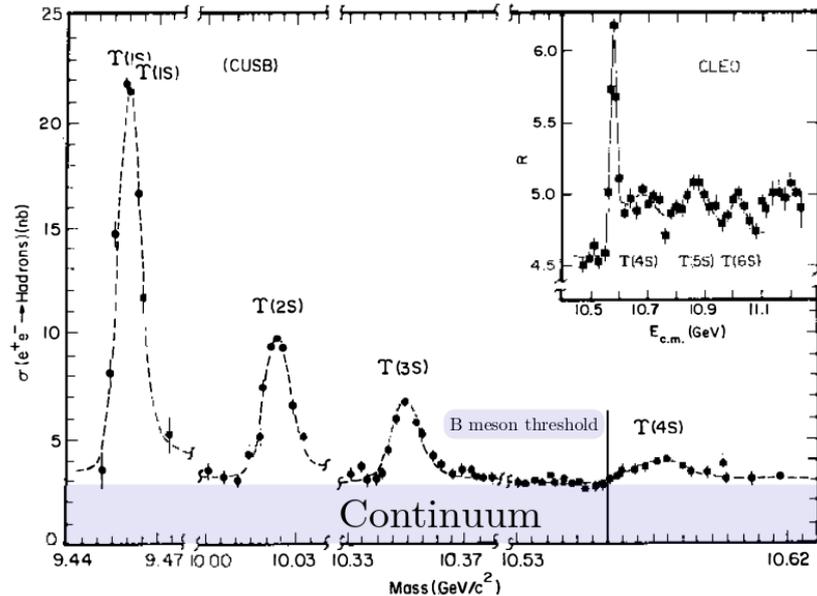
Figure 3.1.: Cross section for $e^+e^- \to$ hadrons in the mass region of $\Upsilon(1S) - \Upsilon(4S)$. The vertical line marks the B meson threshold. The continuum describes the $e^+e^- \to q\bar{q}$ events with $q = u, d, s, c$. The figure is taken from [20].

3.2. KEKB Accelerator

In 1994, the construction of the KEKB accelerator started and was completed at the end of 1998 [25]. It is an asymmetric-energy e^+e^- collider mainly operating at a center-of-mass energy of $\sqrt{s} = 10.58$ GeV, which corresponds to the mass of the $\Upsilon(4S)$ resonance $M_{\Upsilon(4S)}$. The collider consists of two rings with a circumference of 3016 m, installed eleven meters below ground level. The ring in which the electrons are stored is denoted as the High Energy Ring (HER), the one for the positrons as the Low Energy Ring (LER), due to the fact that electrons and positrons have an energy of 8 GeV and 3.5 GeV, respectively. In Fig. 3.2, a schematic view of the KEKB accelerator is shown. At the interaction point (IP) (at the top of Fig. 3.2) the two rings cross and the electrons and positrons collide. The IP is surrounded by the Belle detector in order to detect the particles generated in the collision. For optimizing the luminosity only one IP is installed [19, p. 5]. The design instantaneous luminosity of $\mathcal{L} = 1 \times 10^{-34}\text{cm}^{-2}\text{s}^{-1}$ was reached in 2003. During the runtime between 1999 and 2010 several improvements like crab cavities [26] were installed. All these enhancements led to a new world record luminosity of $\mathcal{L} = 2.1 \times 10^{-34}\text{cm}^{-2}\text{s}^{-1}$. In 2010, KEKB was shut down after successfully operating for over ten years. Since then, KEKB has been upgraded to the SuperKEKB with up to 40 times increased luminosity for the Belle II experiment [27, p. 20].

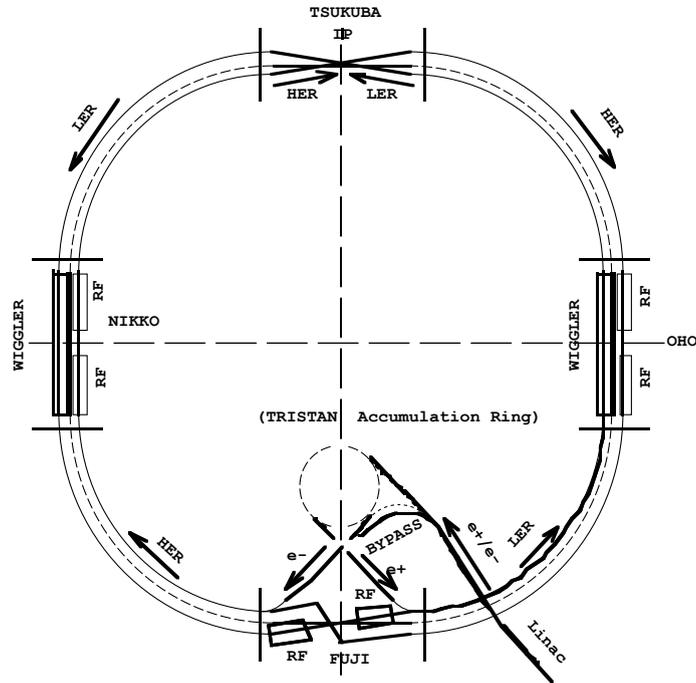A detailed description of the KEKB accelerators can be found in [25, 28].

Figure 3.2.: Schematic view of the KEKB accelerator. Taken from [25].

3.3. The Belle Detector

The Belle detector was designed and built to fulfill the requirements set by the task to measure quantities in B decays at an e^+e^- collider. Belle surrounds the IP of the KEKB accelerator. It has the same structure as other general-purpose detectors and consists of several subdetectors. Two different tracking detectors, the silicon vertex detector (SVD) and the central drift chamber (CDC), are used to reconstruct trajectories of charged particles. Neutral particles can be detected in the electromagnetic calorimeter (ECL). Information from the CDC and ECL is used for particle identification in addition to subdetectors optimized for this task. These detectors are the aerogel Cherenkov counters (ACC) and the time of flight counters (TOF). A superconducting solenoid magnet, providing a magnetic field of 1.5 T, is located around all the detector parts mentioned before. After the magnet and therefore the outermost part of the Belle detector is the KLM system to identify neutral K_L mesons and muons. It consists of arrays of resistive plate counters placed in the iron yoke that provides the magnetic flux return [24]. In Fig. 3.3, the longitudinal (top) and transverse (bottom) cross sections of the Belle detector are shown. A detailed description of the different subdetectors can be found in [29].
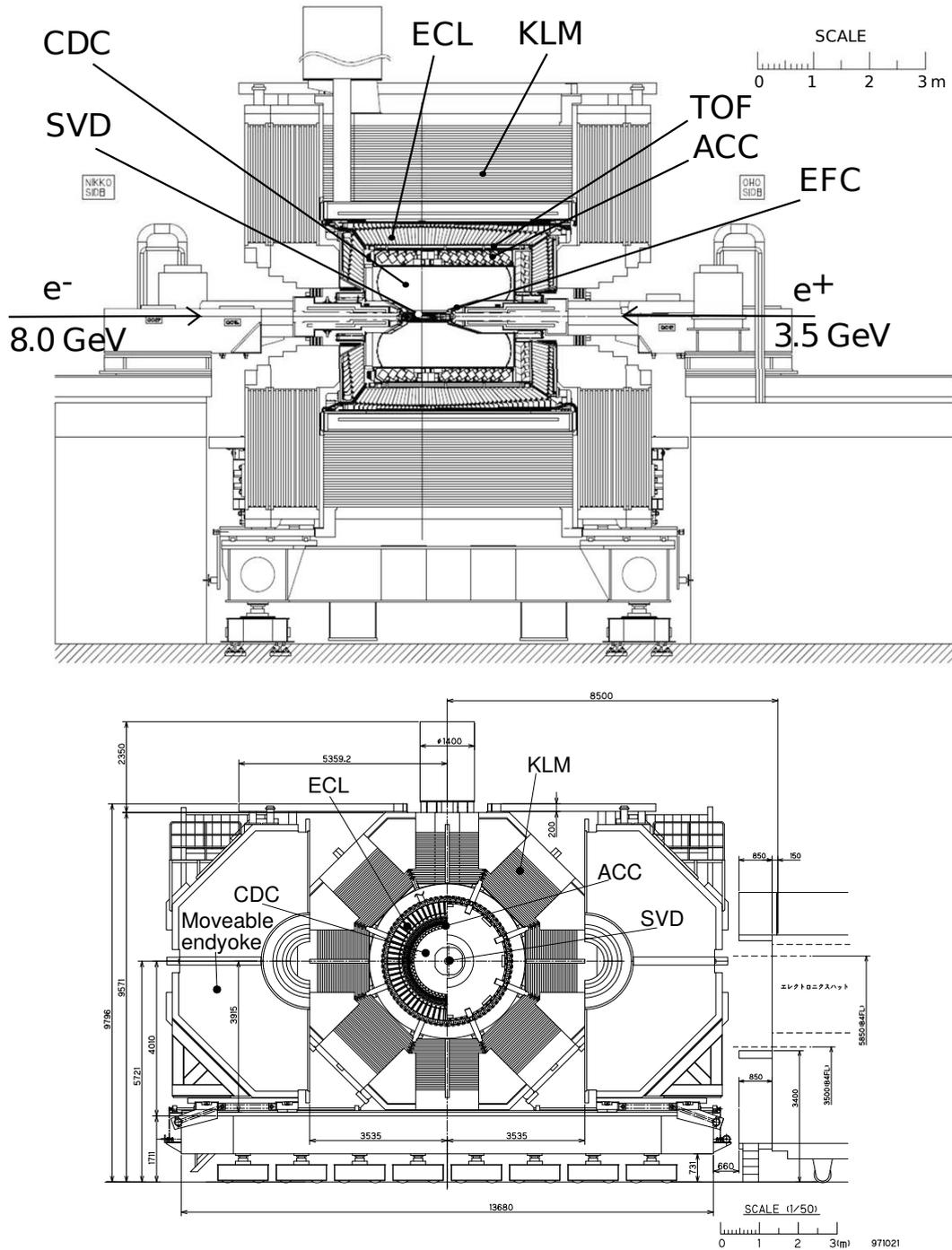
CDC

SVD

ECL KLM

TOF
ACC

EFC

e⁻

8.0 GeV

e⁺

3.5 GeV

SCALE

0 1 2 3 m

KLM

ECL

ACC

CDC
Moveable
endyoke

SVD

Figure 3.3.: Longitudinal (top) and transverse (bottom) cross sections of the Belle detector. Taken from [19].

12

4. Tools and Methods

4.1. Artificial Neural Networks

A very common problem in particle physics and statistical data analysis is binary classification, e.g. the determination of whether or not the reconstructed event is a signal event. For these kind of binary classifications an algorithm which is able to project a multidimensional input vector onto a single scalar output variable is needed. One such algorithm is the artificial neural network. It can learn arbitrary functions which depend on a large number of input variables.

Starting with a graphical description, a neural network is built of several layers of nodes. The first layer is the input layer, the last one the output layer. Intermediate layers are called hidden layers. The nodes represent the input, hidden, and output variables. Links between the nodes of neighboring layers represent weight parameters [30, p. 228]. In Fig. 4.1, the structure of an artificial neural network with one hidden layer, also called a two-layer network, is illustrated.

The output value for a two-layer neural network is defined as

$$y(\mathbf{x}, \mathbf{w}) = f \left[\sum_{j=0}^{L} w_j^{(2)} h \left(\sum_{i=0}^{M} w_{ji}^{(1)} x_i \right) \right], \tag{4.1}$$

with the input variable vector $\mathbf{x}$, and the weight vector $\mathbf{w}$. The activation function h is typically a sigmoid function. For classification problems f is also a sigmoid function, e.g. tanh. Neural networks can also be used to perform a regression. Then f is the identity [30, p. 227-228].

In order to use a neural network (to decide whether a reconstructed event is signal or background) it has to be trained. The input data used in the training can be historical or simulated data, for which the correct target values are known. Given a set of N input vectors $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ and the corresponding target values $\mathbf{T} = \{t_1, \cdots, t_N\}$, during
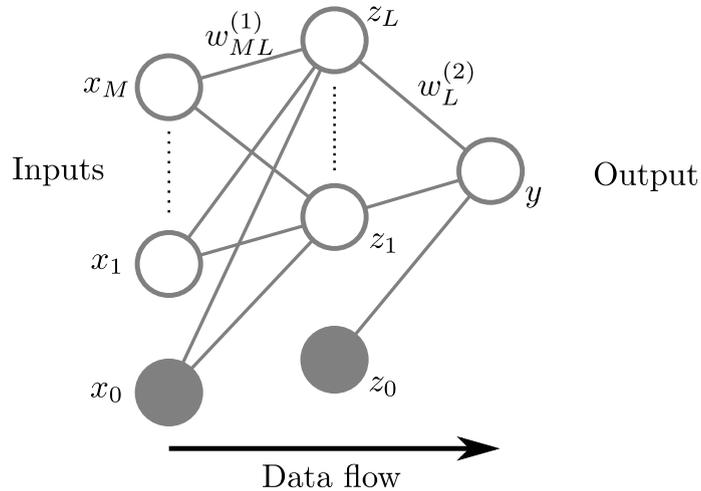
Figure 4.1.: Structure of a two-layer neural network. The numerical values of the input, hidden, and output layer are represented by nodes. The links between the different nodes represent the weights. x_0 and z_0 are so-called bias nodes and are set to one. The arrow shows the data flow.

the training process a loss function $L(\mathbf{w})$ is minimized by adapting the weights $\mathbf{w}$. The choice of the loss function depends on the specific task. In general, the loss function is a measure of the distance between the output values of the neural network and the target values. The optimal set of weights $\mathbf{w}_o$ minimizes this distance. A detailed description of the training procedure can be found in [30].

With the optimal weights, estimated during the training of the network, for data sets with the same input vector $\mathbf{X}$ and unknown target, a prediction can be calculated.

There are many implementations of artificial neural networks. In the scope of this work, the NeuroBayes [31, 32] package is used. It combines a sophisticated preprocessing of the input variables with a neural network. One example of the options that users can choose for the preprocessing is the linear decorrelation of the input variables. Some preprocessing steps are applied to all variables, but there is also the possibility to adjust the preprocessing steps to the variables individually. Another advantage of the NeuroBayes package are the implemented regularization algorithms. With these techniques, the risk of overtraining is reduced and the generalization ability of the networks is enhanced [32].

4.2. Hadronic Full Reconstruction

One advantage of a lepton collider like KEKB is the knowledge of the initial state. This is used in the Full Reconstruction (FR) [33] algorithm. In an $\Upsilon(4S)$ decay, exactly two B mesons are generated. The FR module developed as part of the official **B**elle **A**nalysis **F**ramework (BASF) aims to reconstruct one of these two B mesons in one of many different hadronic decay modes. The B meson reconstructed by the FR is called B_{tag}. Since the

B_{tag} candidate is reconstructed in fully hadronic modes, its 4-momentum p_{tag} is known. Using momentum conservation

$$p_{\mathrm{tag}} + p_{\mathrm{sig}} = p(e^+) + p(e^-) \tag{4.2}$$

one can calculate the 4-momentum p_{sig} of the second B meson, called B_{sig}, without further analysis. A fully reconstructed B_{tag} leads to another advantage: all charged tracks in the tracking system and all electromagnetic clusters in the ECL which have been used in the reconstruction of B_{tag} can be removed from the event. Assuming that the B_{tag} is correctly reconstructed, the remaining tracks and clusters have to originate from the second B meson B_{sig}, since the $\Upsilon(4S)$ solely decays into a pair of $B\bar{B}$ mesons. This provides the possibility to search for B decays with more than one non-detectable particle, like neutrinos, in the final state. For these reasons, the FR is used in the search for $B^0 \to \tau^+\tau^-$ as the τ leptons decay within the detector volume into a ν_τ and other particles.

In order to reconstruct B mesons in about 1000 different exclusive decay channels, a hierarchical structure was developed for the FR. The technical details of the FR are described in [33]. In the analysis developed in the scope of this work, the final discriminator variable $\mathcal{N}_{\mathrm{tag}}$ of the FR is used. It is calculated for each B_{tag} candidate. This discriminator is the result of a sophisticated combination of several NeuroBayes [31, 32] neural networks and preselections within the FR algorithm. Hence, $\mathcal{N}_{\mathrm{tag}}$ can be interpreted as the probability of a B_{tag} candidate being a correctly reconstructed B meson.

4.3. Continuum Suppression

At a center-of-mass energy of 10.58 GeV, continuum events have a large cross section compared to a $\Upsilon(4S)$ resonance. In order to reduce the number of events where no $\Upsilon(4S)$ resonance is created, the so-called Continuum Suppression method is used. The event shape of a continuum event is different from the one of a $B\bar{B}$ event as can be seen in Fig. 4.2. While a continuum event has a jet-like structure (Fig. 4.2a), the shape of a $B\bar{B}$ event (Fig. 4.2b) is more spherical. Variables which describe the shape of an event and which are used in the Continuum Suppression method are listed below.

Thrust angle The thrust angle θ_{Thrust} is defined as the angle between the thrust axis of the reconstructed B_{tag} candidate and the one of all remaining particles in the event not used in the B_{tag} reconstruction. The thrust axis is the vector $\vec{t}$ that maximizes

$$T = \frac{\sum_i |\vec{p}_i \cdot \vec{t}|}{\sum_i |\vec{p}_i|}, \tag{4.3}$$

with the 3-momentum $\vec{p}_i$ of particle i and the condition $|\vec{t}| = 1$. For the thrust axis of the B_{tag} candidate, i runs over all particles used in the B_{tag} reconstruction. The thrust axis of the remaining event is calculated with i running over the remaining particles not used in B_{tag}. Since the B mesons in an $\Upsilon(4S)$ event are almost at rest in the center-of-mass system (CMS) and since there is no directional preference of the decay products, the distribution of the cosine of the polar angle of the thrust axes

(a) Continuum event.
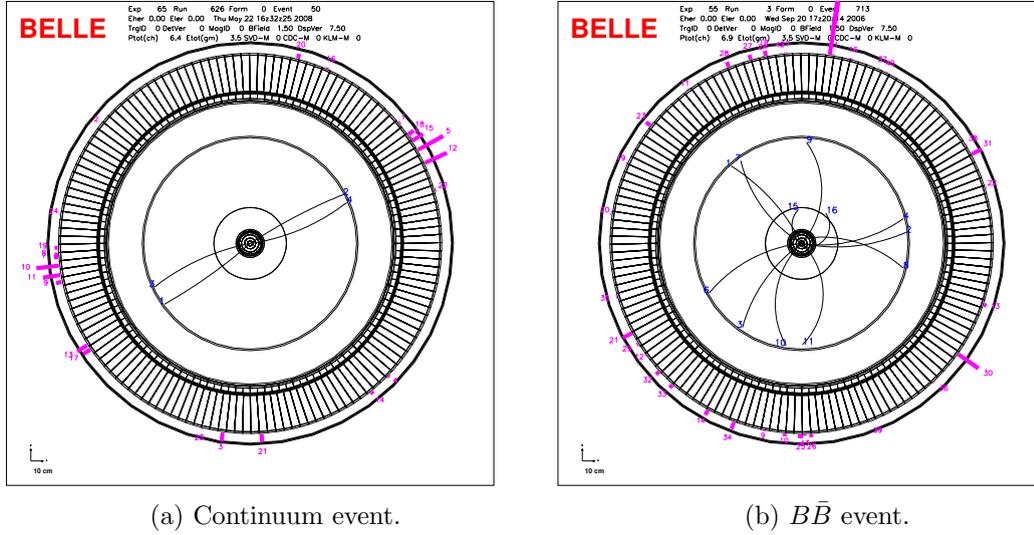

(b) $B\bar{B}$ event.

Figure 4.2.: Illustration of the event shape of a continuum and a $B\bar{B}$ event in the Belle detector.

$|\cos\theta_{\text{Thrust}}|$ is uniform for $B\bar{B}$ events. In a continuum event, it is likely that the B_{tag} candidate is recombined from particles of one single jet and the jets are back-to-back in the CMS. Therefore, $|\cos\theta_{\text{Thrust}}|$ peaks at one for continuum events [24].

$\cos\theta_B$ This is the angle between the reconstructed momentum of the B_{tag} candidate and the direction of the electron beam in the CMS. As $\Upsilon(4S) \to B\bar{B}$ is a decay of a vector meson into two pseudo-scalar mesons, θ_B follows a $1 - \cos^2\theta_B$ distribution for $B\bar{B}$ events. For continuum events the corresponding distribution is uniform [24].

Fox-Wolfram Moments The Fox-Wolfram Moments (FWM) [34] are defined as

$$H_\ell \equiv \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_\ell(\cos\phi_{ij}), \tag{4.4}$$

with the momenta $\vec{p}_i$ of the particle i, the angle ϕ_{ij} between the particles i and j, and the Legendre polynomials P_ℓ. The indices i and j run over all charged particles in the event. The reduced FWM are normalized to the FWM of order zero: $R_\ell = H_\ell/H_0$. In the Continuum Suppression used in this analysis, only the second reduced FWM R_2 is used, since it describes the deviation from a spherical shape, which is a strong discriminator between $B\bar{B}$ and continuum events.

For Continuum Suppression, all variables described above are used as input to a NeuroBayes neural network to obtain an optimal discrimination between continuum and $B\bar{B}$ events. This algorithm can be used together with the FR. The network output for the B_{tag} candidate is recalculated including the information from the Continuum Suppression. In the following analysis, $\mathcal{N}_{\text{tag}}$ refers to the network output of the FR algorithm including the Continuum Suppression, unless explicitly stated otherwise.

4.4. Particle Identification

Only a few charged particles, namely electrons, muons, pions, kaons, and protons, have a lifetime large enough to reach the tracking detectors and leave measurable signals. To distinguish between the different particles more information than just the trajectory is required. Therefore, different measurements from the CDC, TOF, ACC, ECL, and KLM are used:

- Energy loss per flight distance, denoted as dE/dx, measured in the CDC

- Time of flight measured in the TOF

- Number of Cherenkov photons measured by ACC

- Energy depositions measured by the ECL

- Shape of electromagnetic showers in the ECL

- Matching between charged tracks and clusters in the ECL

- Matching of hits in the KLM and charged tracks

Depending on the particle hypothesis, sub-sets of the different measurements mentioned above are used to calculate likelihood ratios. The following particle identification (PID) variables are used:

Kaon vs pion ID (`kpiID`) The `kpiID` is used to separate kaons and pions. For this purpose the likelihoods $L_\alpha^{\mathrm{CDC}}, L_\alpha^{\mathrm{ACC}}$, and L_α^{TOF} measured by the sub-detectors CDC, ACC, and TOF, respectively, are used, with the different particle types α. The likelihood ratio to distinguish between K and π is defined as

$$L(K:\pi) = \frac{L_K^{\mathrm{CDC}} L_K^{\mathrm{ACC}} L_K^{\mathrm{TOF}}}{L_K^{\mathrm{CDC}} L_K^{\mathrm{ACC}} L_K^{\mathrm{TOF}} + L_\pi^{\mathrm{CDC}} L_\pi^{\mathrm{ACC}} L_\pi^{\mathrm{TOF}}}. \tag{4.5}$$

electron ID (`eID`) The `eID` uses additional measurements about the shape of the electromagnetic shower measured by the ECL and a measure for the matching between a cluster in the ECL and a charged track. No time-of-flight measurement is used. The shape of a shower is the ratio of the measured energy in 3×3 and 5×5 ECL crystals surrounding the one with the deposit [35]. Furthermore, the ratio of the measured energy in the ECL and the measured momentum in the CDC of a charged track is used to discriminate between electrons and other particle types in the high momentum region [35]. Complementary information is given by the dE/dx measurement. The last information used for the final `eID` is the number of photo-electrons measured in the ACC. The `eID` is then defined as

$$L_{\mathrm{eID}} = \frac{\prod_i L_e^i}{\prod_i L_e^i + \prod_i L_x^i}, \tag{4.6}$$

with the likelihoods of each measurement L_e^i assuming the electron hypothesis and the measurements L_x^i with all other hypotheses for the track. A detailed description of the `eID` can be found in [36].

muon ID (`muID`) To separate muons from the other charged particle hypotheses, solely the KLM information is used, i.e. reconstructed hits in the KLM are compared to extrapolated tracks measured in the CDC using the difference in the expected and measured range in the KLM [19]. The discriminating `muID` is defined as

$$L_{\mathrm{muID}} = \frac{L_\mu}{L_\mu + L_\pi + L_K}. \tag{4.7}$$

A detailed description of the `muID` and its performance can be found in [37].

5. Analysis Procedure

This chapter gives a short introduction into the analysis procedure as the analysis itself is subdivided in three main steps:

- event reconstruction (Chapter 6)
- event selection(Chapter 7)
- branching ratio extraction (Chapter 9)

Using the measured and processed detector data, both B mesons are recombined from final state particles during the event reconstruction. One B candidate is called B_{tag} and is reconstructed by applying the Full Reconstruction method (FR) described in Section 4.2. The second B meson, denoted B_{sig}, is reconstructed in the decay mode of interest, $B^0 \rightarrow \tau^+\tau^-$. A detailed description of the event reconstruction is given in Chapter 6.

After the reconstruction, most of the candidates are wrongly reconstructed combinations. Either the B_{tag} candidate is not correctly reconstructed or the B_{sig} candidate is not a true $B^0 \rightarrow \tau^+\tau^-$ decay but another B decay with the same detectable final state particles. In order to suppress the amount of background events an event selection is developed. Properties of the B_{tag} and B_{sig} candidate are used to distinguish between signal and background events. In a first step all events not fulfilling certain simple requirements are discarded. Afterwards, a neural network based selection is trained and optimized on simulated Monte Carlo (MC) events to further reduce the number of background events. An in-depth description of the selection is given in Chapter 7.

In the final step, the branching ratio $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)$ is extracted in a one-dimensional fit. The probability density function (PDF) of the distributions for background and signal components are determined on MC and used in the fit on real data in order to measure $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)$. The fit procedure is described in detail in Chapter 9 and the final fit on data in Chapter 10.

A flowchart of the analysis procedure is shown below in Fig. 5.1. The three steps of the analysis described are visualized for MC and measured data.
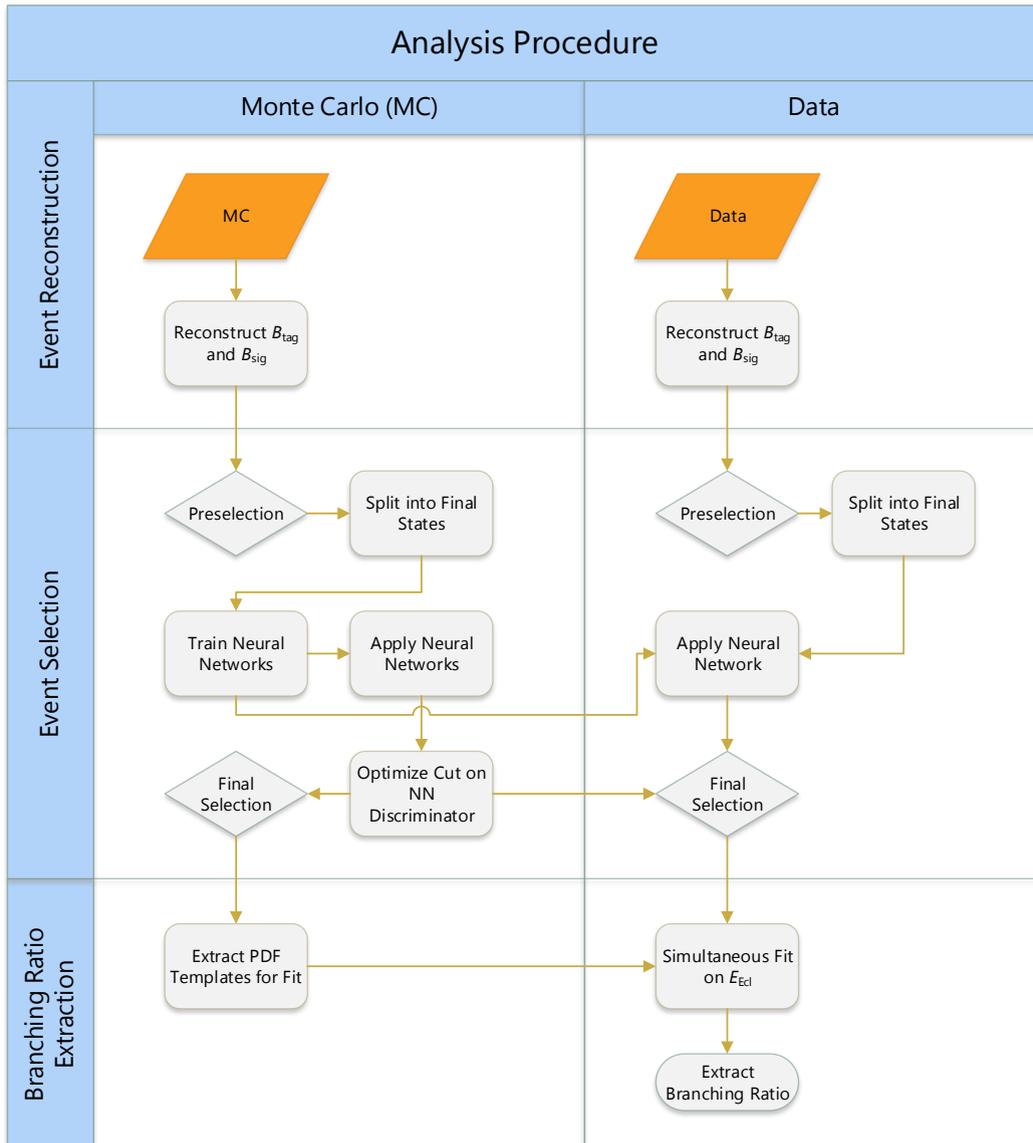
Figure 5.1.: Flowchart of the analysis procedure. Processing steps are shown as rectangles, selections steps in which events are vetoed as diamonds. The cut on the network output is optimized on MC and applied on MC and data. The branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is extracted using a PDF template fit. The templates are determined on MC.

6. Event Reconstruction

In this chapter, the detector data, MC samples, and the reconstruction of B^0 mesons are discussed.

6.1. Data Samples

6.1.1. Data

The analysis is performed on the complete Belle data sample recorded from 1999 to 2010 at a center-of-mass (CM) energy of $\sqrt{s} = 10.58\,\text{GeV}$. This energy corresponds to the mass of the $\Upsilon(4S)$ resonance. The data sample consists of an integrated luminosity of $\mathcal{L}_{\text{int}} = 711\,\text{fb}^{-1}$ equivalent to $772 \times 10^6\ B\bar{B}$ events.

6.1.2. Simulated Data

The analysis method is developed and optimized on large MC samples. All decay processes are simulated using the program packages `EvtGen` [38] and `PYTHIA` [39]. The detector response is simulated using `GEANT3` [40].

For the development of the analysis procedure two samples are important: signal and background MC. In the signal MC the process of interest $B^0 \to \tau^+\tau^-$ is simulated. Other processes that can occur are bundled in the background samples. They can be grouped into different types, namely, generic, $BD\ell\nu_\ell$, continuum, rare, and $u\ell\nu_\ell$. In the following, detailed descriptions of the individual samples are given:

Signal In the signal sample one B meson decays into the signal channel $B^0 \to \tau^+\tau^-$, while the other B decays via a $b \to c$ transition[a]. Depending on the τ decays, different samples are generated:

- 50 million events where both τ can decay in all possible final states.

---

[a]Throughout the whole document, charge-conjugated decays are implied, unless stated explicitly.

Table 6.1.: Branching ratios used for the scaling of the $BD\ell\nu_\ell$ component and the values used in the simulation of the generic MC sample. Values are taken from [17].

Process	Simulated $\mathcal{B}$ in %	$\mathcal{B}$ in %
$\Upsilon(4S) \to B^0\bar{B}^0$		48.6 ± 0.6
$B^0 \to D^- e^+ \nu_e$	2.13	2.19 ± 0.12
$B^0 \to D^- \mu^+ \nu_\mu$	2.13	2.19 ± 0.12
$D^- \to K^0 e^- \bar{\nu}_e$	6.8	8.83 ± 0.22
$D^- \to K^0 \mu^- \bar{\nu}_\mu$	6.8	9.3 ± 0.7

- Six times 10 million and six times 5 million events, where both τ leptons have a specific final state. For example, both τ decay into an electron and two neutrinos. The different final states are described later in Section 7.1.4. The samples with 10 million events are used in the training of the neural networks, the ones with 5 million events are used for the optimization of the cut on the neural network output.

Generic The generic samples contain events where both B mesons decay via a $b \to c$ transition. They are divided into ten separate samples called streams. The integrated luminosity of one stream MC corresponds to the recorded amount of data. As the $\Upsilon(4S)$ decays either into a B^+B^- or a $B^0\bar{B}^0$ pair, the generic sample is subdivided into one with the charged B mesons decaying via a $b \to c$ transition and one with the neutral B mesons decaying via the same process, respectively. In the simulated sample, the ratio between neutral and charged B meson decays is 1:1.

$BD\ell\nu_\ell$ The $BD\ell\nu_\ell$ sample consists of $\Upsilon(4S) \to B^0\bar{B}^0$, where one of the neutral B mesons decays via $b \to c$ transition, while the other one decays explicitly into $B^0 \to D^-(\to K^0\ell'^-\bar{\nu}_{\ell'})\ell^+\nu_\ell$, with $\ell^{(\prime)} = e, \mu$. 50 million events are simulated and scaled to the world average values of the corresponding branching ratios. To avoid double counting, events with the same simulated decay chain in the neutral generic sample are discarded. The number of expected events of the type $BD\ell\nu_\ell$ in the complete Belle data sample can be calculated as

$$N_{BD\ell\nu_\ell}^{\text{expected}} = N_{B\bar{B}} \cdot 2 \cdot \mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0) \cdot (\mathcal{B}(B^0 \to D^- e^+ \nu_e) + \mathcal{B}(B^0 \to D^- \mu^+ \nu_\mu)) \cdot$$
$$(\mathcal{B}(D^- \to K^0 e^- \bar{\nu}_e) + \mathcal{B}(D^- \to K^0 \mu^- \bar{\nu}_\mu)), \tag{6.1}$$

with the number of $B\bar{B}$ events $N_{B\bar{B}}$, the probability that a $\Upsilon(4S)$ decays into a pair of neutral B mesons $\mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0)$, the branching ratios of the B^0 and D^- decays $\mathcal{B}(B^0 \to D^-\ell^+\nu_\ell)$ and $\mathcal{B}(D^- \to K^0\ell^-\bar{\nu}_\ell)$, respectively, with $\ell = e, \mu$. The factor 2 is due to the fact that both B mesons in the event can decay via the specific decay. The values for the branching ratio are stated in Table 6.1 and are taken from [17].

The normalization of the $BD\ell\nu_\ell$ component in each final state i is estimated as

$$N^i_{BD\ell\nu_\ell} = \underbrace{\frac{N^{\text{expected}}_{BD\ell\nu_\ell}}{N_{\text{generated}}}}_{\equiv f_{BD\ell\nu_\ell}} \cdot N^{i,\text{selected}}_{BD\ell\nu_\ell}, \tag{6.2}$$

with the number of generated $BD\ell\nu_\ell$ events $N_{\text{generated}}$ and the number of selected $BD\ell\nu_\ell$ events $N^{i,\text{selected}}_{BD\ell\nu_\ell}$ in the final state i. Hence, $f_{BD\ell\nu_\ell}$ is the scaling factor and it is applied as an event-by-event weight to $BD\ell\nu_\ell$ events. For 50 million generated $BD\ell\nu_\ell$ events $f_{BD\ell\nu_\ell}$ is 0.12.

Continuum The continuum sample contains events of the type $e^+e^- \rightarrow q\bar{q}$ with $q = u, d, s, c$. There are six streams of continuum MC.

Rare In the rare sample $\Upsilon(4S)$ decays are simulated, in which one of the B mesons decays via processes which have a small branching ratio compared to $b \rightarrow c$ decays. The other B decays generically. Examples for such a rare process are $B \rightarrow K\nu\nu$ and $B^+ \rightarrow \tau^+\nu_\tau$. Not only semi-leptonic and leptonic decays are included but also decays like $B^0 \rightarrow \rho^0\rho^0$. The signal decay $B^0 \rightarrow \tau^+\tau^-$ is simulated in the rare sample as well, but is removed as it is treated in the signal MC samples. The rare samples' size correspond to 50 times the amount of rare decays in the recorded data and is divided into neutral and charged B meson decays.

$u\ell\nu_\ell$ As for the rare sample, in the $u\ell\nu_\ell$ sample one B meson decays generically, while the other decays via a $b \rightarrow u\ell\nu_\ell$ transition, resulting in $B \rightarrow X_u\ell\nu_\ell$ final states. X_u denotes light mesons including a u quark, like π and ρ. The branching ratios for such decays are small in comparison to $b \rightarrow c$ decays. They are not included in the rare sample, since they have a special and common signature. The amount of events corresponds to 20 times the amount in data and the sample is also split into neutral and charged B meson decays.

6.2. Reconstruction

As stated in Chapter 5, one of the three main steps of the analysis is the reconstruction of the event. Each event, in MC and data, is reconstructed using the same procedure. Since the τ leptons in the decay $B^0 \rightarrow \tau^+\tau^-$ have a short lifetime, they decay inside the detector volume. At least two neutrinos, which only interact weakly and do not leave measurable signals within the detector, are in the final state. Reconstructing only the detectable parts of the τ decays and combining them to a B meson would lead to a very high background due to missing kinematic information, and thus discriminating a signal from a background candidate would be almost impossible. Therefore, the B_{tag} candidate is first reconstructed using the FR algorithm. If more than one B_{tag} candidate are found in an event, only the candidate with the highest $\mathcal{N}_{\text{tag}}$ value will be used in the further event reconstruction. After removing all tracks, calorimeter clusters, and signals in the K_L and muon system (KLM) used in the B_{tag} reconstruction, the remaining tracks, clusters etc. have to originate from the second B meson in a signal event. These remaining objects are

Table 6.2.: Decay modes of the τ lepton used in the signal side reconstruction. Values are taken from [17].

Decay channel	Branching ratio in %
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$	17.83 ± 0.04
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$	17.41 ± 0.04
$\tau^- \to \pi^- \nu_\tau$	10.83 ± 0.06
Sum	46.07 ± 0.08

called the rest of the event (ROE). In the ROE, a signal candidate B_{sig} is reconstructed. Only single-prong decays are used as decay channels of the τ leptons, i.e. τ decays with only one charged particle in the final state. The used decay channels of the τ are listed in Table 6.2 with their branching fractions. The charged, stable particles ($e^\pm$, $\mu^\pm$, $\pi^\pm$, p, $K^\pm$) create hits in the detector volume along their trajectories. Using a track fit to the measured hit coordinates based on the Kalman filter the parameters of the track are determined [19]. Charged particles as used in the analysis are formed of reconstructed charged tracks. A B_{sig} candidate is a combination of two oppositely charged particles. As τ leptons have a small mean flight length of $c\tau_\tau = 87.03$ µm [17] (with the lifetime τ_τ of the τ lepton) and the B^0 mesons' is $c\tau_{B^0} = 455.4$ µm [17], only charged tracks with a distance from the point of closest approach of the track to the interaction point (IP) in the plane perpendicular to the beam axis (x-y-plane) of $|dr| < 2$ cm and in the beam direction (z-axis) of $|dz| < 4$ cm are used, where the electron beam defines the positive z-direction (see Fig. 3.3 for a geometrical comparison).

At this stage of the analysis, no separation between the different particle hypotheses (electron, muon, and pion) for the charged particles forming the B_{sig} candidate has been performed. In order to determine whether a reconstructed event is formed by a correctly reconstructed B_{tag} and B_{sig} candidate, several variables, which are described in Section 7.1, are used.

7. Event Selection

In order to suppress the huge amount of background events, a cut-based preselection (Section 7.1) and a neural-network-based selection (Section 7.3) are trained and applied. As data and MC have known discrepancies, MC events are reweighted. The weighting procedure is described in Section 7.2. The optimization of the cut on the network discriminator is described in Section 7.3.2. In Section 7.4 and Section 7.5 the final signal efficiency and the background composition after the final selection are shown, respectively.

7.1. Preselection

In the final state of the decay $B^0 \to \tau^+ \tau^-$, there are two to four neutrinos, which are not detectable. Hence, there are no strong constraints on kinematic variables for the B_{sig} candidate which allows to distinguish between signal and background events. But as the full event is reconstructed, vetoes and variables related to the whole event can be used. In the flowchart in Fig. 5.1 the preselection is the first step in the event selection.

7.1.1. Vetoes

For an event in which B_{tag} and B_{sig} are correctly reconstructed, all charged and neutral final state particles measured in the detector should be used in forming either B_{tag} or B_{sig}. Therefore, in a perfect detector and with perfect reconstruction methods, no further charged tracks or neutral particles should be present in the event. This fact can be used to reject events in which additional particles are reconstructed.

Charged Track Veto

The charged track veto rejects events where charged tracks are reconstructed but not used in the combination of either B_{tag} or B_{sig}. Only charged tracks fulfilling the same conditions as the ones used to reconstruct the B_{sig} candidate, namely $|dr| < 2$ cm and $|dz| < 4$ cm, are used to reject an event. The veto itself is physically motivated, as in an $\Upsilon(4S)$ decay

exactly two B mesons are generated. If the B_{tag} is correctly reconstructed in a hadronic mode with the FR and if B_{sig} is a true signal candidate, no more charged particles should be present in the event. Therefore, only events with no additional tracks are selected.

π^0 Veto

Events with additional good π^0 candidates are neglected. π^0 candidates are formed of two γ candidates. A good π^0 candidate is a combination of two γ clusters fulfilling the following criteria:

Photon energy The energy of each γ has to be $E_{\gamma_i} > 50$ MeV ($i = 0, 1$).

Energy asymmetry The energy asymmetry of the two photons used in the π^0 reconstruction is defined as

$$A_{\text{energy}} = \frac{|E_{\gamma_0} - E_{\gamma_1}|}{E_{\gamma_0} + E_{\gamma_1}}, \tag{7.1}$$

with the measured energies E_{γ_i} ($i = 0, 1$). For a good π^0, the energy asymmetry has to be $A_{\text{energy}} < 0.9$.

Invariant mass The invariant mass of $M(\gamma\gamma)$ has to satisfy the condition

$$117.8 \text{ MeV} < M(\gamma\gamma) < 150.2 \text{ MeV}.$$

The world average of the π^0 mass is $m_{\pi^0} = (134.9766 \pm 0.0006)$ MeV [17].

For this veto, the same reasons hold as for the charged track veto described above. If B_{tag} and B_{sig} are correctly reconstructed, no additional π^0 candidates will be present in the event.

K_S Veto

K_S candidates are reconstructed from two oppositely charged π candidates. Events are vetoed in which at least one K_S candidate fulfills certain selection criteria, called *goodKs* selection. The *goodKs* selection is commonly used in Belle analyses and is described in [41]. Using the *goodKs* selection, cuts on the distance between the IP and the point-of-closest approach of the pions dr and dz, the flight length of the K_S candidate, and the angle $d\phi$ between the K_S momentum and the direction of the decay vertex of the K_S are applied. The flight length d of a K_S candidate is the distance between the IP and the fitted vertex of the K_S daughters. A rough sketch of the flight length is depicted in Fig. 7.1.

The cut values depend on the momentum of the K_S candidate and are summarized in Table 7.1. As for the other vetoes in a correctly reconstructed $\Upsilon(4S)$ event, no additional K_S candidates should be present after removing all particles used in the reconstruction.

K_L Veto

In contrast to the vetoes described above, the K_L veto behaves differently. K_L mesons are neutral hadrons with a long lifetime. Hence they typically do not decay inside the Belle
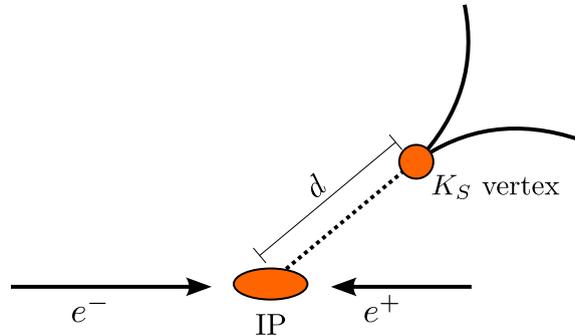
Figure 7.1.: Definition of the K_S flight length, denoted by d.

Table 7.1.: Selection cut for good K_S candidates [42].

Momentum in GeV/c	dr in cm	$d\phi$ in rad	dz in cm	d in cm
< 0.5	> 0.05	< 0.3	< 0.8	-
0.5 − 1.5	> 0.03	< 0.1	< 1.8	> 0.08
> 1.5	> 0.02	< 0.03	< 2.4	> 0.22

detector. As they do not interact electromagnetically they cannot be detected in either the tracking detectors nor in the ECL. The KLM was intended to detect muons and K_L, but the probability that the K_L interact with the material of the KLM system is very low. Another problem lies within the insufficient theoretical understanding of the hadronic interactions of the K_L mesons with the detector material, especially for low momentum K_L. This culminates in large differences between simulated and real data and therefore the efficiency to reconstruct a K_L is very different in MC and data. In the study in [43], K_L from $D^0 \rightarrow \phi K_S$, $\phi \rightarrow K_L K_S$ decays are used to calibrate the reconstruction efficiency. As a result of this study, a weighting function for MC samples is implemented and can be used in Belle analyses. In the function, event-by-event weights are calculated. The assigned weight is 1 if no K_L is reconstructed in the event. If one or more K_L candidates are reconstructed, the weight for the MC event will be the probability to also reconstruct such K_L candidates in data. If there is one fake K_L in the K_L candidates, the weight will be zero as the probability to have fake K_L reconstructed in the event is the same for data and MC [43].

In order to use the K_L veto, MC samples are reweighted using the functions implemented in [43]. Data events are rejected if one or more K_L candidates are detected in the KLM.

7.1.2. Variables

E_{ECL} Energy depositions in the ECL clusters are summed up if the following conditions for the energy deposition are met:

- It is not used as a photon candidate in the reconstruction of the $\Upsilon(4S)$ resonance.

- It is not assigned as bremsstrahlung to a nearby track used in the B_{tag} or B_{sig} reconstruction.

- It is greater than 50 MeV, 100 MeV, or 150 MeV if it lies in the barrel, forward cap, or backward cap region of the Belle detector, respectively.

For an event in which B_{tag} and B_{sig} are correctly reconstructed, all energy depositions in the ECL should theoretically be used in the reconstruction. Hence, for a true signal event, E_{ECL} must be either zero or a small value. The observed smearing for the signal component to values larger than zero is due to beam background and secondary interactions in the detector. Another component originates from unreconstructed π^0 candidates in the τ decay, e.g. $\tau^+ \to \rho^+ (\to \pi^+ \pi^0) \nu_\tau$ where only the charged pion is reconstructed.

$M_{\mathrm{bc}}^{\mathrm{tag}}$ The beam-constrained mass of the B_{tag} candidate is defined as

$$M_{\mathrm{bc}}^{\mathrm{tag}} = \frac{1}{c^2} \sqrt{E_{\mathrm{beam}}^2 - \vec{p}_{B_{\mathrm{tag}}}^2 \cdot c^2}, \tag{7.2}$$

with half the CM energy E_{beam} and the reconstructed momentum $\vec{p}_{B_{\mathrm{tag}}}$ of the B_{tag} candidate in the CM system. For events with a correctly reconstructed B_{tag} candidate, $M_{\mathrm{bc}}^{\mathrm{tag}}$ peaks at the mass of the B^0 meson with the world average $M_{B^0} = (5279.58 \pm 0.17)$ MeV/c^2 [17]. As the measured energy of the B_{tag} candidate is replaced with the beam energy, $M_{\mathrm{bc}}^{\mathrm{tag}}$ is almost independent of the mass hypotheses for each particle. There remains a small dependency since the mass hypotheses for each particle have to be assigned before the momentum vector of the B_{tag} candidate can be boosted into the CM frame [19, p. 86]. This variable mainly distinguishes continuum and combinatorial background from correctly reconstructed B_{tag} candidates. It is not a discriminator for different decays that mimic the signature of the signal decay $B^0 \to \tau^+ \tau^-$.

ΔE_{tag} Defined as the difference between the reconstructed energy of the B_{tag} candidate E_{tag}^* and the beam energy E_{beam}, both measured in the CM system and written as

$$\Delta E_{\mathrm{tag}} = E_{\mathrm{tag}}^* - E_{\mathrm{beam}}. \tag{7.3}$$

ΔE_{tag} depends, by construction, on the mass hypotheses used for each particle in the reconstruction of the B_{tag} candidate. Therefore, ΔE_{tag} is helpful for discriminating correctly reconstructed B_{tag} candidates from physics background events involving misidentification [19, p. 86]. ΔE_{tag}, as well as $M_{\mathrm{bc}}^{\mathrm{tag}}$, is a good discriminator between continuum and $B\bar{B}$ events, because ΔE_{tag} peaks at zero for correctly reconstructed B_{tag} candidates.

Table 7.2.: Preselection cuts and corresponding values for the signal and background rejection in percent. The cuts are applied successively.

Cut	Signal rejection in %	Background rejection in %
$E_{\mathrm{ECL}} < 1.2$ GeV	2.65	58.23
$M_{\mathrm{bc}}^{\mathrm{tag}} > 5.27$ GeV/c^2	0.84	76.52
$|\Delta E_{\mathrm{tag}}| < 50$ MeV	14.26	47.29
$\mathcal{N}_{\mathrm{tag}} > 0.05$	34.78	72.08
$M_{\mathrm{miss}}^2 > 0.5$ (GeV/c^2)^2	0.84	2.01

$\mathbf{\mathcal{N}_{tag}}$ The FR algorithm described in Section 4.2 allows for selecting a working point in the analysis. For each B_{tag} candidate a neural network discriminator $\mathcal{N}_{\mathrm{tag}}$ is calculated. This discriminator can be interpreted as the probability of the B_{tag} being a correctly reconstructed B^0 meson. Hence, a cut larger than a threshold on $\mathcal{N}_{\mathrm{tag}}$ lowers the amount of wrongly reconstructed B_{tag} candidates.

$\mathbf{M_{miss}^2}$ The 4-momentum of the missing particles is

$$p_{\mathrm{miss}} = p_{\mathrm{beam}} - (p_{\mathrm{tag}} + p_{\mathrm{sig}}), \tag{7.4}$$

with the reconstructed 4-momenta of the B_{sig} and B_{tag} candidates p_{sig} and p_{tag}, respectively, and the 4-momentum of the beam p_{beam}. The resulting value of the inner product of p_{miss} is the missing mass squared M_{miss}^2:

$$M_{\mathrm{miss}}^2 = E_{\mathrm{miss}}^2 - \vec{p}_{\mathrm{miss}}^2, \tag{7.5}$$

with the missing energy E_{miss} and the missing momentum $\vec{p}_{\mathrm{miss}}$. If everything in an $\Upsilon(4S)$ decay is reconstructed, M_{miss}^2 will be zero. Considering the experimental mass resolution, the same holds if only one neutrino is not reconstructed. In the decay $B^0 \to \tau^+\tau^-$ at least two neutrinos are in the final state. Hence, M_{miss}^2 must have values larger than zero for a signal decay.

7.1.3. Preselection Cuts

Using the vetoes and applying cuts on the variables described above reduces the number of background events. The preselection cuts are not optimized but are motivated by standard procedures of the Belle collaboration. In Table 7.2, the cuts and their signal and background rejection rates are listed. The distributions for signal and background events for the selection variables are shown in Fig. 7.2. The vertical lines in Fig. 7.2 mark the applied cut values. All cuts display a large background rejection of 40% to 77%, except for the one on M_{miss}^2. It only rejects 2% of the background. The signal rejection is high for the cut of $\mathcal{N}_{\mathrm{tag}} > 0.05$. Using this tight cut is motivated as the data–MC agreement is bad for events with a lower value of $\mathcal{N}_{\mathrm{tag}}$ (see validation in Chapter 8).
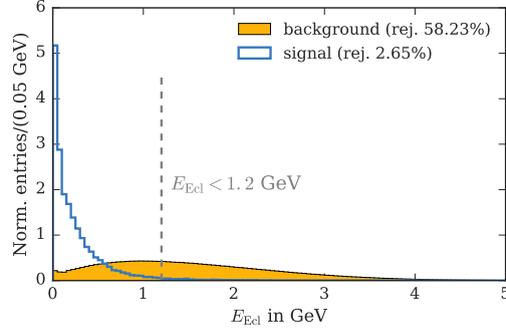
(a) Signal and background E_{ECL} distributions. Events with $E_{\mathrm{Ecl}} < 1.2$ GeV are selected.
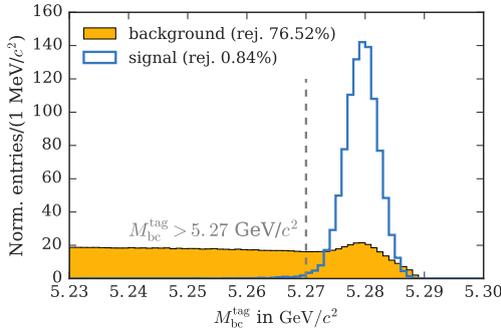
(b) Signal and background $M_{\mathrm{bc}}^{\mathrm{tag}}$ distributions. Events with $M_{\mathrm{bc}}^{\mathrm{tag}} > 5.27$ GeV/c^2 are selected.
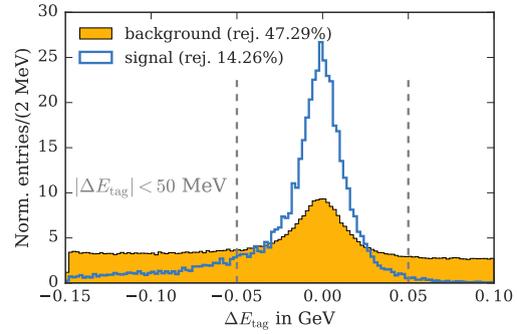
(c) Signal and background ΔE_{tag} distributions. Events with $|\Delta E_{\mathrm{tag}}| < 50$ MeV are selected.
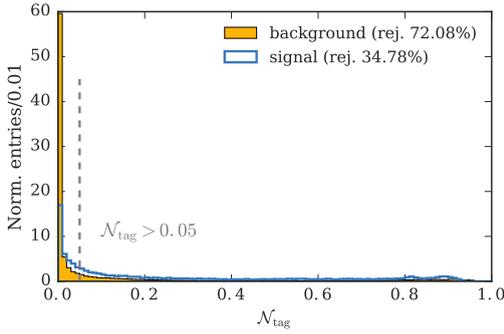
(d) Signal and background $\mathcal{N}_{\mathrm{tag}}$ distributions. Events with $\mathcal{N}_{\mathrm{tag}} > 0.05$ are selected.
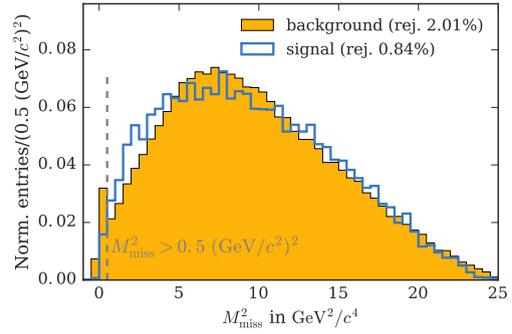
(e) Signal and background M_{miss}^2 distributions. Events with $M_{\mathrm{miss}}^2 > 0.5$ $(\mathrm{GeV}/c^2)^2$ are selected.

Figure 7.2.: Distributions of preselection variables for signal (blue line) and background (yellow) events. All distributions are normalized to one. Each plot is created using events surviving the cut in the previous plots. The same holds for the rejection values. The signal and background rejection is stated in the legend.

Table 7.3.: Possible final states of the B_{sig} candidate. The name of a final state is an abbreviation for the decay modes of the two τ leptons in the decay $B^0 \to \tau^+\tau^-$.

Name	τ decay modes
e^+e^-	$\tau \to e\nu_e\nu_\tau$, $\tau \to e\nu_e\nu_\tau$
$e^\pm\mu^\mp$	$\tau \to e\nu_e\nu_\tau$, $\tau \to \mu\nu_\mu\nu_\tau$
$e^\pm\pi^\mp$	$\tau \to e\nu_e\nu_\tau$, $\tau \to \pi\nu_\tau$
$\mu^+\mu^-$	$\tau \to \mu\nu_\mu\nu_\tau$, $\tau \to \mu\nu_\mu\nu_\tau$
$\mu^\pm\pi^\mp$	$\tau \to \mu\nu_\mu\nu_\tau$, $\tau \to \pi\nu_\tau$
$\pi^+\pi^-$	$\tau \to \pi\nu_\tau$, $\tau \to \pi\nu_\tau$

Table 7.4.: Signal efficiencies for the signal-side separation. The efficiencies are calculated after the preselection was applied.

Final state	Efficiency
e^+e^-	0.741 ± 0.004
$e^\pm\mu^\mp$	0.472 ± 0.005
$e^\pm\pi^\mp$	0.783 ± 0.004
$\mu^+\mu^-$	0.297 ± 0.004
$\mu^\pm\pi^\mp$	0.533 ± 0.005
$\pi^+\pi^-$	0.754 ± 0.004

7.1.4. Signal-side Separation

Reconstructing three different decay channels of the τ lepton (Table 6.2) leads to six different final states, listed in Table 7.3. As the distributions of kinematic variables may differ between all final states, they are separated using the particle identification (PID) variables described in Section 4.4. The following cuts for the different final state particle hypotheses are applied:

μ candidate `muID` > 0.1;

e candidate `muID` < 0.1 and `eID` > 0.1;

π candidate `muID` < 0.1 and `eID` < 0.1 and `kpiID` < 0.4;

With these requirements only one hypothesis is possible for each track. If the `muID` condition is fulfilled the track will be treated as a muon. If the `muID` condition is not fulfilled, it will be checked to determine if it is an electron. If it is not an electron, it can be a pion or a kaon. With the cut on the `kpiID` only pions are selected. After applying these cuts on the PID variables the efficiencies for the six final states are listed in Table 7.4. The efficiencies for final states including a muon are much lower than for the others. One reason for this behavior is the fact that the `muID` is only available for charged tracks with a momentum larger than 600 MeV/c (see Fig. 7.3), as charged particles with a smaller momentum do not reach the KLM system.
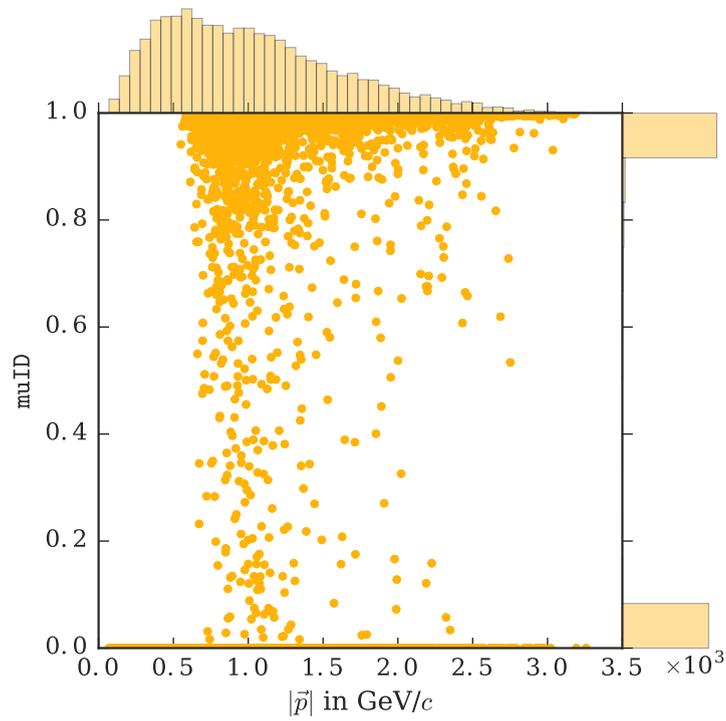
Figure 7.3.: Scatter plot of the `muID` and the momentum of a charged track. On the top axis, the projection of the scatter plot onto the momentum of the charged tracks is shown. For momenta below 600 MeV/c the `muID` is zero. On the right axis, the projection on the `muID` is depicted. The number of events in the distribution of the `muID` in the first and last bin at zero and one, respectively, is much larger than the number of events in the bins between. The plot is created using simulated $B^0 \to \tau^+ \tau^-$ decays, in which both τ leptons decay into $\mu \nu_\mu \nu_\tau$.

7.2. MC Sample Weighting

There are known differences between data and MC samples. In order to incorporate them, MC events are weighted. The discrepancies important for this analysis are the different reconstruction efficiency of the FR, the slightly different efficiency of the lepton ID (`LID`) selection, and differences between simulated and measured branching ratios for dominant background processes.

7.2.1. Tag Efficiency Correction

The tagging efficiency on data and MC of the FR was studied and calibrated in [44, 45]. Here, one B meson in an event is reconstructed in semi leptonic decays of type $B \to X_c \ell \nu$. These decay channels are used as they all have large and well-known branching fractions. The D mesons in the decay are reconstructed in hadronic modes and listed below.

- $B^- \to D^0 \ell^- \bar{\nu}_\ell$
- $B^- \to D^{*0} \ell^- \bar{\nu}_\ell$
- $\bar{B}^0 \to D^+ \ell^- \bar{\nu}_\ell$
- $\bar{B}^0 \to D^{*+} \ell^- \bar{\nu}_\ell$
- $D^0 \to K^- \pi^+$

- $D^0 \to K^- \pi^+ \pi^0$
- $D^0 \to K^- 2\pi^+ \pi^-$
- $D^+ \to K^- 2\pi^+$
- $D^+ \to K^- 2\pi^+ \pi^0$
- $D^+ \to K^- 3\pi^+ \pi^-$

- $D^{*0} \to D^0 \pi^0$
- $D^{*0} \to D^0 \gamma$
- $D^{*+} \to D^0 \pi^+$
- $D^{*+} \to D^+ \pi^0$

The accompanying B is reconstructed using the FR. In order to estimate the data–MC differences, the M^2_{miss} distribution is fitted separately for all tag modes used in the FR. For each tag mode, the average correction factor over all semi leptonic modes (listed above) is calculated. These factors are used to weight events in the MC samples. In Fig. 7.4, ratios of the number of reconstructed tag candidates in data and MC are shown for different tag modes and semi leptonic channels. The authors of [44, 45] provide tables with weight factors depending on $\mathcal{N}_{\mathrm{tag}}$ and the tag mode. These tables are used to assign the tag correction factors to each event with a correctly reconstructed B_{tag} candidate. Fig. 7.5 shows the tag correction factors for the different background samples and the signal sample.

7.2.2. `LID` Efficiency Correction

Another known discrepancy between data and MC lies in the efficiency of a cut on the `LID` variables. The correction factors to account for this systematic data–MC discrepancy as well as their systematic error are studied in [46] using four-fermion events $e^+ e^- \to e^+ e^- \ell^+ \ell^-$ ($\ell = e, \mu$). To validate whether a hadronic environment changes the `LID` efficiency, $B \to X J/\psi(\ell^+ \ell^-)$ decays are compared with the four-fermion events [46, p. 3]. The `LID` efficiency depends on the polar angle θ and the momentum $|\vec{p}|$ of a charged track. Hence, the correction weights are estimated for seven bins in θ and ten bins in $|\vec{p}|$. Similar to the tag correction weights described in Section 7.2.1, the authors of [46] provide tables with the determined `LID` correction factors for different cut values. In each event, the signal
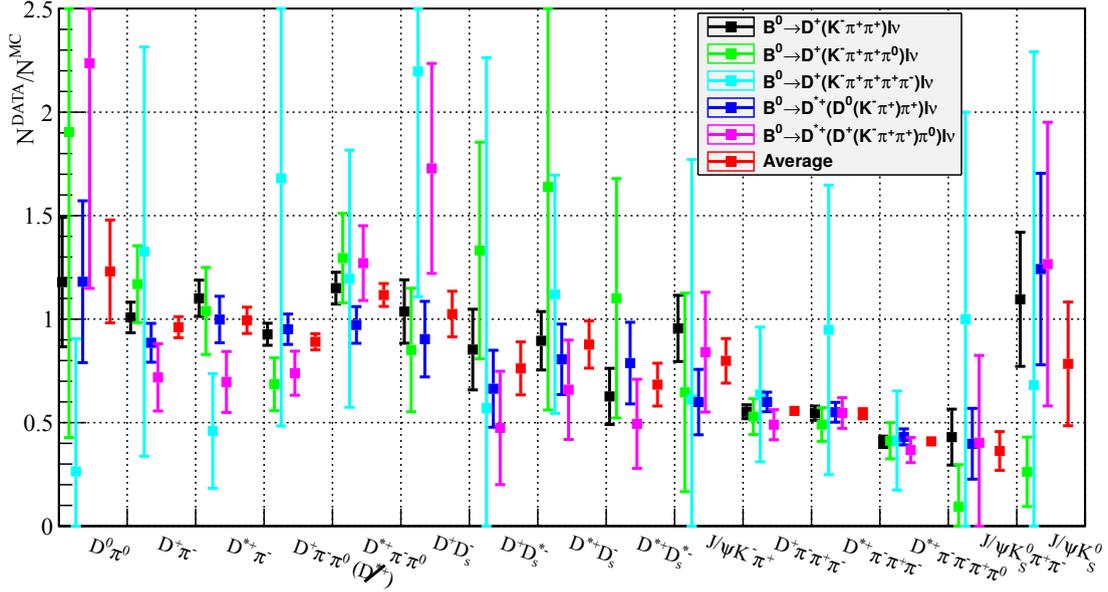
Figure 7.4.: Ratios of reconstructed tag candidates in data and MC for different tag modes (x-axis) from [44]. The ratios are shown for different reconstructed signal decays and their average (red).
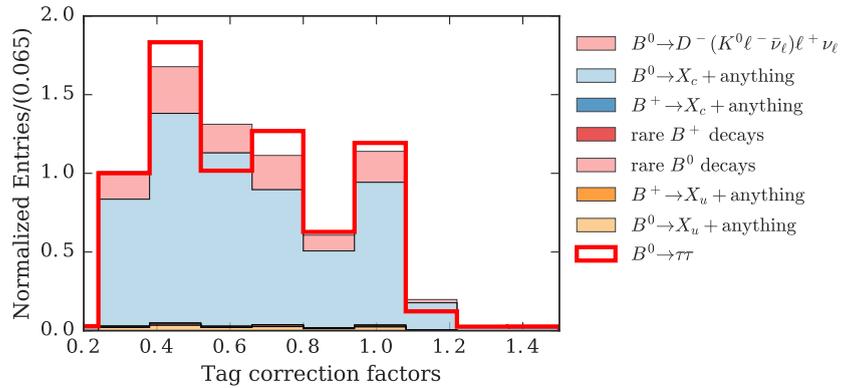
Figure 7.5.: Tag correction factors determined in [45] and used as event-by-event weights. The histograms are normalized.

(a) Final state e^+e^-.

(b) Final state $\mu^+\mu^-$.
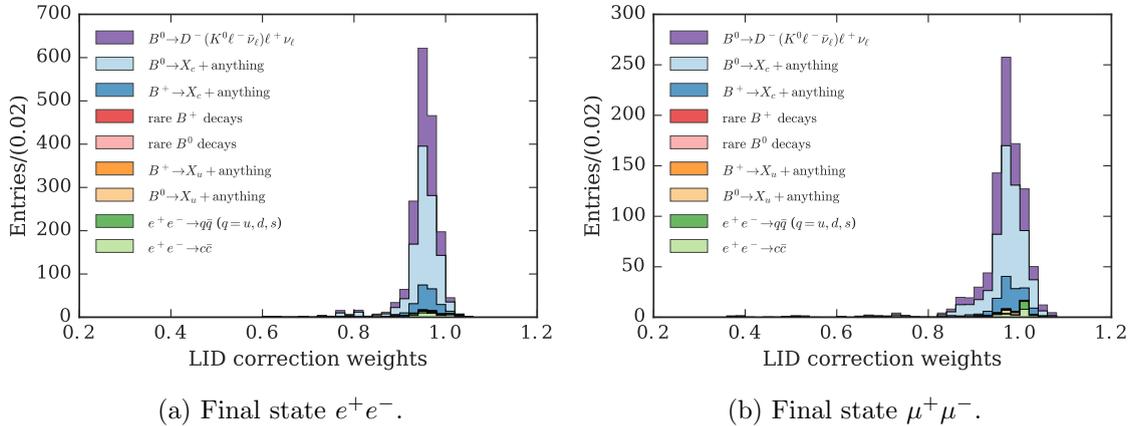
Figure 7.6.: `LID` weights used in the analysis. The weights shown are on event level. This means they are the product of the `LID` weights of the two charged tracks forming a B_{sig} candidate. The different background components are scaled to the integrated luminosity in data. In Fig. 7.6a and Fig. 7.6b, the weights for the final state e^+e^- and $\mu^+\mu^-$, respectively, are shown.

Table 7.5.: Correction weights for events including a $D^- \rightarrow K^0\ell^-\bar{\nu}_\ell$ ($\ell = e, \mu$) decay. Values are taken from [47]. The weights are calculated as the ratio between the measured and the simulated value for the branching ratio.

Process	Simulated $\mathcal{B}$ in %	Measured $\mathcal{B}$ in %	Weight
$D^- \rightarrow K^0 e^- \bar{\nu}_e$	6.8	8.90 ± 0.15	1.31 ± 0.02
$D^- \rightarrow K^0 \mu^- \bar{\nu}_\mu$	6.8	9.3 ± 0.7	1.37 ± 0.10

candidate has two charged tracks. The `LID` weight is calculated for each of them. An event is then weighted using the product of the particular weights. Using these tables, the `LID` weights on event level used in this analysis are shown in Fig. 7.6 for the final states e^+e^- and $\mu^+\mu^-$.

7.2.3. Branching Ratio Corrections

A main background contribution comes from B decays with a charged D^- meson decaying into $K^0\ell^-\bar{\nu}_\ell$ ($\ell = e, \mu$). In addition to the separate treatment of the $BD\ell\nu_\ell$ events, all events in the generic and rare samples including a $D^- \rightarrow K^0\ell^-\bar{\nu}_\ell$ decay are reweighted. The weights are calculated as the ratio of the best known values (taken from [47]) and the simulated branching ratios and are listed in Table 7.5.

7.3. Neural Network Based Selection

A large amount of background events remains after the preselection requirements. To further reduce these backgrounds, a separate NeuroBayes neural network is trained to distinguish between correctly reconstructed B_{sig} candidates and background for each of the six final states. The optimization of the cut on the resulting network discriminator $\mathcal{N}_{\text{sig}}$ is described in Section 7.3.2.

7.3.1. Training

Besides $M_{\text{bc}}^{\text{tag}}$, $\mathcal{N}_{\text{tag}}$, and M_{miss}^2 (described in Section 7.1.2) more variables are used as inputs for the neural network. The variables are grouped into those which are measured or calculated in the laboratory system and those estimated in the rest frame of the B_{sig} candidate. Table 7.6 shows a complete list of the input variables.

Variables in the Laboratory System As described in Section 6.2, the only final state particles (FSP) of the signal decay are two stable charged particles with opposite electrical charge. The momentum and energy of such a charged particle is estimated during the track fitting. As input for the neural networks the transverse momenta p_{T} of the daughters of B_{sig}, and the corresponding energies are used. Both are measured in the laboratory frame. The transverse momentum p_{T} denotes the component of the momentum perpendicular to the beam axis and is defined as $p_{\text{T}} = \sqrt{p_x^2 + p_y^2}$.

Variables also estimated in the laboratory and used as input are the angle between the two charged particles $\theta_{0 \triangleleft 1}$, and the momentum asymmetry defined as

$$A_{01} = \frac{|\vec{p}_0| - |\vec{p}_1|}{|\vec{p}_0| + |\vec{p}_1|}, \tag{7.6}$$

with the three-momenta $\vec{p}_{0,1}$ of the charged FSP.

Furthermore, the visible, reconstructed invariant mass of the B_{sig} candidate $M(B_{\text{sig}})$, is fed into the networks. It is calculated as $M(B_{\text{sig}}) = \sqrt{(p_0 + p_1)^2}$, with the four-momenta of both charged particles $p_{0,1}$.

The missing momentum $\vec{p}_{\text{miss}}$ can be calculated, because the complete event is reconstructed. As input, the absolute value $|\vec{p}|_{\text{miss}}$, the transverse momentum $p_{T,\text{miss}}$, and the polar angle of the missing momentum $\cos\theta_{\text{miss}}$ are used, as in the main background components the massive K_L meson is not reconstructed and carries away momentum and energy.

The charged particles forming a B_{sig} candidate are fitted to a common vertex[a]. Hence, the distance in the x-y-plane between the fitted vertex and the IP, and the significance of the distance, are also fed into the nets.

---

[a]Details about vertex fitting can be found in [19].
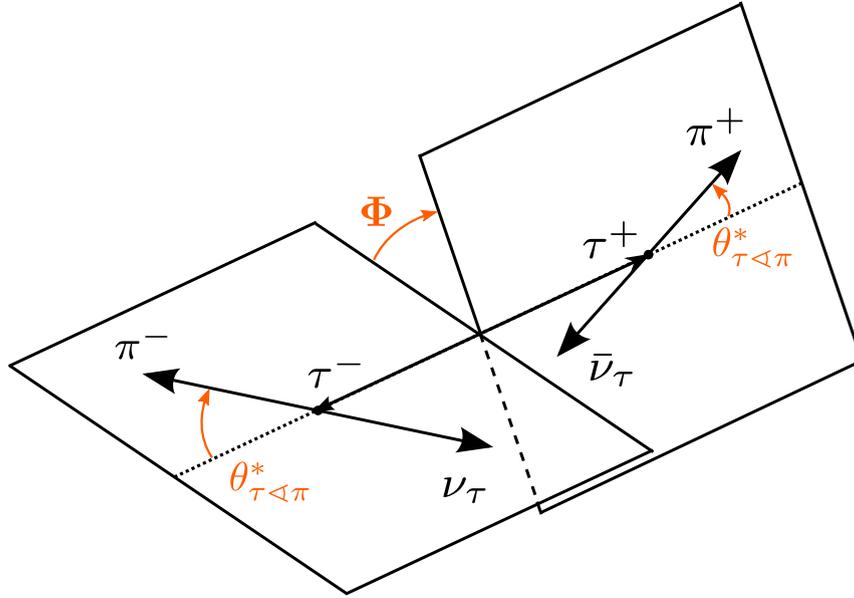
Figure 7.7.: $B^0 \to \tau^+\tau^-$ decay chain with both τ leptons decaying into a charged pion and a neutrino in the rest frame of the B_{sig} candidate.

Variables in the B_{sig} Rest Frame In order to estimate the kinematic variables in the rest frame of the B_{sig} candidate, the four-momentum p_{sig} must be known. Since at least two neutrinos are present in the final state, p_{sig} cannot directly be measured and boosted into the B_{sig} rest frame. However, the B_{tag} candidate is reconstructed in a hadronic final state. Hence, its four-momentum p_{tag} is entirely known. Using

$$p_{\mathrm{sig}} + p_{\mathrm{tag}} = 0, \tag{7.7}$$

which holds in the CM system, boosting into the B_{sig} rest frame is possible using the negative 4-momentum of the B_{tag} after boosting into the CM system[b].

Besides the momentum variables of the FSP in the laboratory frame, the absolute value of the momentum $|\vec{p}^*|$ of both charged FSP measured in the B_{sig} rest frame as well as the angle between the two charged FSP particles $\cos\theta^*_{0\vartriangleleft 1}$ are used as input.

For charged particles with a π hypothesis (see Section 7.1.4) the variable $\cos\theta^*_{\tau\vartriangleleft\pi}$ is calculated. It denotes the cosine of the angle between the charged pion and the τ lepton in the rest frame of the B_{sig} candidate. In Fig. 7.7, the decay chain $B^0 \to \tau^+\tau^-$, where both τ leptons decay into a charged pion and a neutrino, is visualized. The angle is calculated using momentum and energy conservation, as both $B^0 \to \tau^+\tau^-$

---

[b]Variables marked with * are boosted into the B_{sig} rest frame.

and $\tau^+ \to \pi^+ \bar{\nu}_\tau$ are two-body decays:

$$\boldsymbol{p}_\tau = \boldsymbol{p}_\pi + \boldsymbol{p}_\nu$$
$$\Rightarrow \quad \boldsymbol{p}_\nu = \boldsymbol{p}_\tau - \boldsymbol{p}_\pi$$
$$\Rightarrow \quad 0 = (\boldsymbol{p}_\tau - \boldsymbol{p}_\pi)^2$$
$$\Rightarrow \quad 0 = m_\tau^2 c^2 + m_\pi^2 c^2 - 2\boldsymbol{p}_\tau \boldsymbol{p}_\pi$$
$$\Rightarrow \quad 2(E_\tau E_\pi - \vec{p}_\tau \vec{p}_\pi) = m_\tau^2 c^2 + m_\pi^2 c^2$$
$$\Rightarrow \quad \cos(\theta_{\tau \vartriangleleft \pi})|\vec{p}_\tau||\vec{p}_\pi| = \frac{E_\tau E_\pi}{c^2} - \frac{m_\tau^2 c^2 + m_\pi^2 c^2}{2}$$
$$\Rightarrow \cos(\theta_{\tau \vartriangleleft \pi}) = \frac{2E_\tau E_\pi - m_\tau^2 c^4 - m_\pi^2 c^4}{2c^2|\vec{p}_\tau||\vec{p}_\pi|}, \tag{7.8}$$

where $\boldsymbol{p}_\tau$, $\boldsymbol{p}_\pi$ and $\boldsymbol{p}_\nu$ denote the 4-momenta of τ, π and ν_τ, respectively. The energy is denoted as E_i, $\vec{p}_i$ is the momentum and m_i the rest mass of particle $i = \tau, \pi$. It is possible to measure the momentum $\vec{p}_\pi$ of the charged π, but since neutrinos escape the detector unseen, we cannot reconstruct the momentum of the τ candidates. To calculate $\cos(\theta_{\tau \vartriangleleft \pi})$ the energy and momentum of the τ candidates are required.

In the rest frame of the B_{sig}, the energy and absolute value of the momentum of both τ leptons are given by the energy and momentum of the B_{sig} candidate. $B^0 \to \tau^+ \tau^-$ is a two-body decay. Hence, in the rest frame of B_{sig}, τ^+ and τ^- fly back-to-back with the same momentum and both have the energy $E_\tau = m_B c^2/2$. The absolute value of their momentum can be computed as follows:

$$E_\tau^2 = m_\tau^2 c^4 + |\vec{p}_\tau|^2 c^2$$
$$\Rightarrow \quad |\vec{p}_\tau| = \sqrt{\frac{E_\tau^2}{c^2} - m_\tau^2 c^2}$$
$$\Rightarrow \quad |\vec{p}_\tau| \overset{E_\tau = m_B c^2/2}{=} c \cdot \sqrt{\frac{m_B^2}{4} - m_\tau^2} \tag{7.9}$$

Thus with Equation 7.9, $\cos(\theta_{\tau \vartriangleleft \pi})$ in Equation 7.8 can be expressed in the rest frame of the B_{sig} as

$$\cos(\theta^*_{\tau \vartriangleleft \pi}) \overset{7.9}{=} \frac{m_B E_\pi^* - m_\tau^2 c^2 - m_\pi^2 c^2}{2c|\vec{p}_\pi^*|} \left(\sqrt{\frac{m_B^2}{4} - m_\tau^2} \right)^{-1}, \tag{7.10}$$

where E_π^* and $\vec{p}_\pi^*$ are the energy and momentum of the pion in the rest frame of B_{sig}, respectively.

Before the trainings of the networks an additional cut in the final states $e^\pm \pi^\mp$, $\mu^\pm \pi^\mp$, and $\pi^+ \pi^-$ is applied on the variable $\cos \theta^*_{\tau \vartriangleleft \pi}$ defined as

$$-1 < \cos \theta^*_{\tau \vartriangleleft \pi} < 1. \tag{7.11}$$

The cut is only applied to the tracks with the π hypothesis.
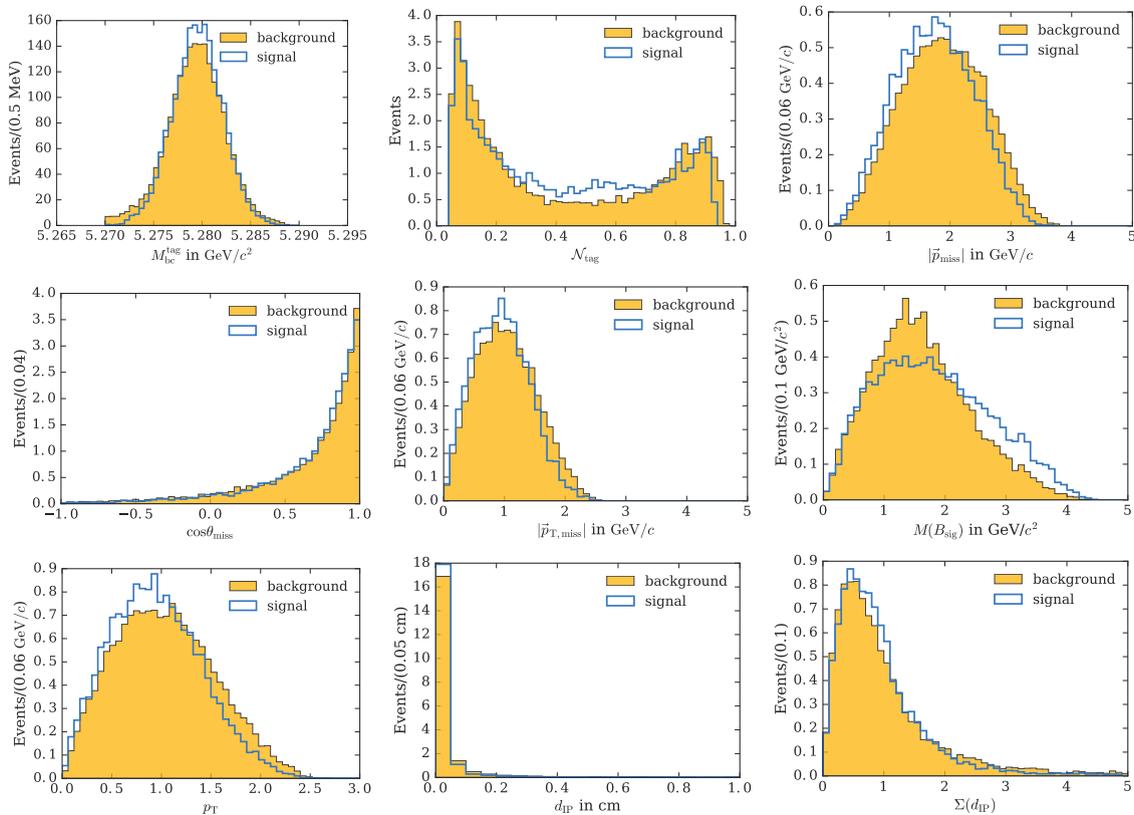
Figure 7.8.: Distributions of signal (blue line) and background (solid yellow) events for training variables used in the $e^{\pm}\pi^{\mp}$ training.

Angle between Daughter and rec. B_{sig} Momentum Another variable used as input for the networks is the angle between the momentum of one daughter and the measured momentum of the B_{sig} candidate, with the latter in the CMS and the daughter momentum boosted with the reconstructed B_{sig} momentum. This variable would be called pseudo helicity angle, if all decay products of the B meson were reconstructed and hence, the measured momentum of B_{sig} was the real momentum of it. For simplicity, the variable is denoted as $\cos\theta_{\text{hel},0}$.

The distributions for signal and background of the variables used in the training are shown in Figs. 7.8 and 7.9 exemplarily for the final state $e^{\pm}\pi^{\mp}$.

The differences between the signal (blue line) and the background (yellow) distributions in Figs. 7.8 and 7.9 vary from variable to variable. For most of the input variables the difference is very small. The largest difference for signal and background can be found in the variable $\cos\theta^{*}_{\tau\lhd\pi}$ for both daughters (see first and fourth row in Fig. 7.9).

In the trainings of the six neural networks, the background sample consists of three streams of generic MC including the decays of type $BD\ell\nu_{\ell}$, and the full rare and $u\ell\nu_{\ell}$ samples. Signal events are taken from simulated samples, where $B^{0} \rightarrow \tau^{+}\tau^{-}$ events decay into the
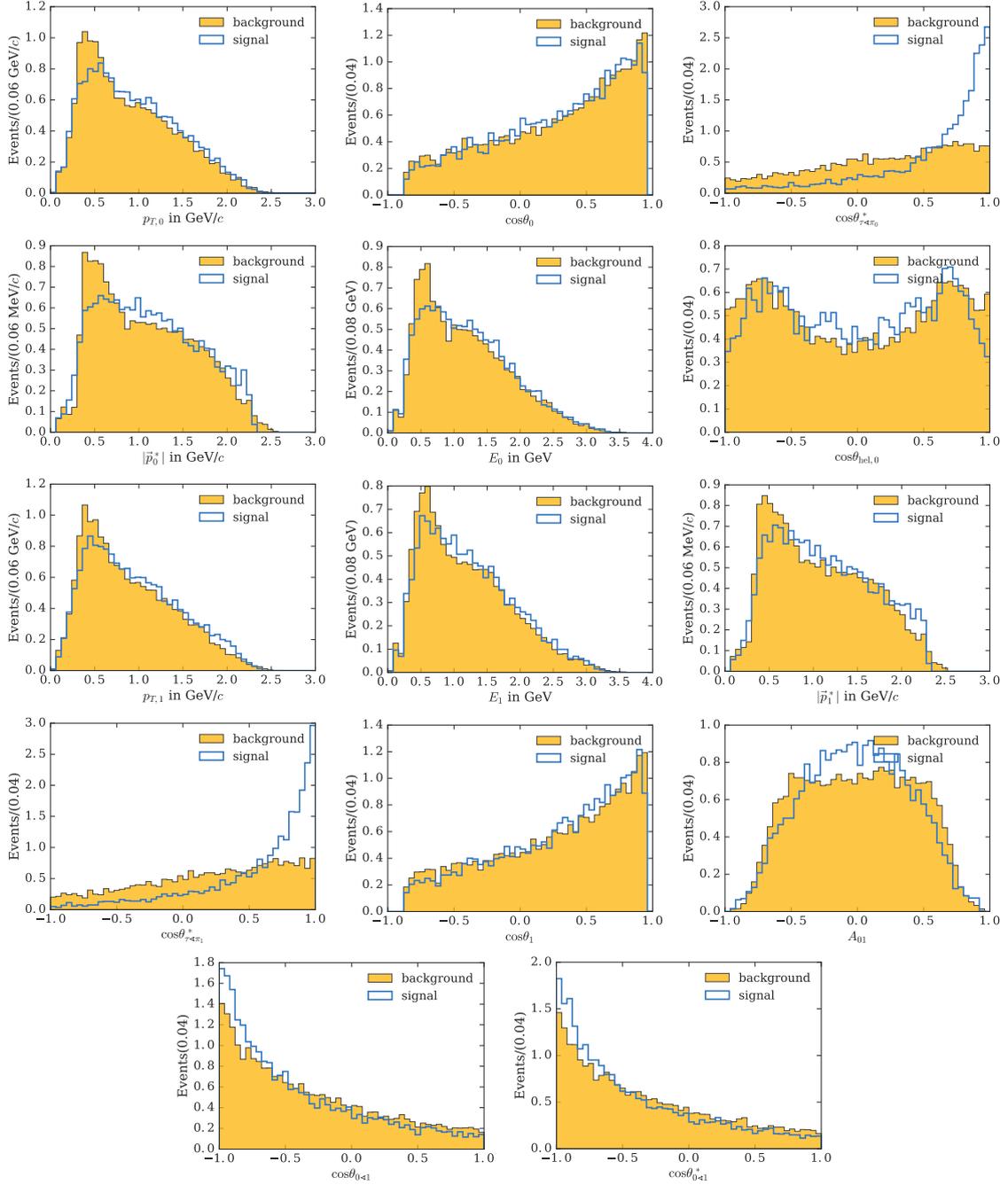
Figure 7.9.: Distributions for signal (blue line) and background (yellow) events of the input variables related to the daughters of the B_{sig} candidate used in the $e^{\pm}\pi^{\mp}$ training.

Table 7.6.: Input variables of the neural nets.

	Variable	Short description
Lab. frame	$p_{\mathbf{T},i}$	Transverse momentum of B_{sig} daughters
	E_i	Energy of B_{sig} daughters
	$\cos\theta_i$	Polar angle of B_{sig} daughters
	$\cos\theta_{0\triangleleft 1}$	Angle between B_{sig} daughters
	A_{01}	Momentum asymmetry of B_{sig} daughters
	$M(B_{\mathbf{sig}})$	Reconstructed mass of B_{sig}
	$p_{\mathbf{T}}$	Reconstructed transverse momentum of B_{sig}
	$M^2_{\mathbf{miss}}$	Missing mass squared of the event
	$\lvert\vec{p}_{\mathbf{miss}}\rvert$	Absolute value of the missing momentum in the event
	$\cos\theta_{\mathbf{miss}}$	Polar angle of the missing momentum
	$\lvert\vec{p}_{\mathbf{T,miss}}\rvert$	Absolute value of the transverse component of the missing momentum in the event
	$d_{\mathbf{IP}}$	Distance of B_{sig} vertex and IP
	$\Sigma(d_{\mathbf{IP}})$	Significance of d_{IP}
B_{sig} rest frame	$\lvert\vec{p}^{\,*}_i\rvert$	Absolute value of the momentum of B_{sig} daughters
	$\cos\theta^*_{0\triangleleft 1}$	Angle between B_{sig} daughters
	$\cos\theta^*_{\tau\triangleleft\pi}$	Angle between τ and B_{sig} daughter with π hypothesis
	$\cos\theta_{\mathbf{hel,0}}$	Angle between daughter 0 and the reconstructed momentum of B_{sig}

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.
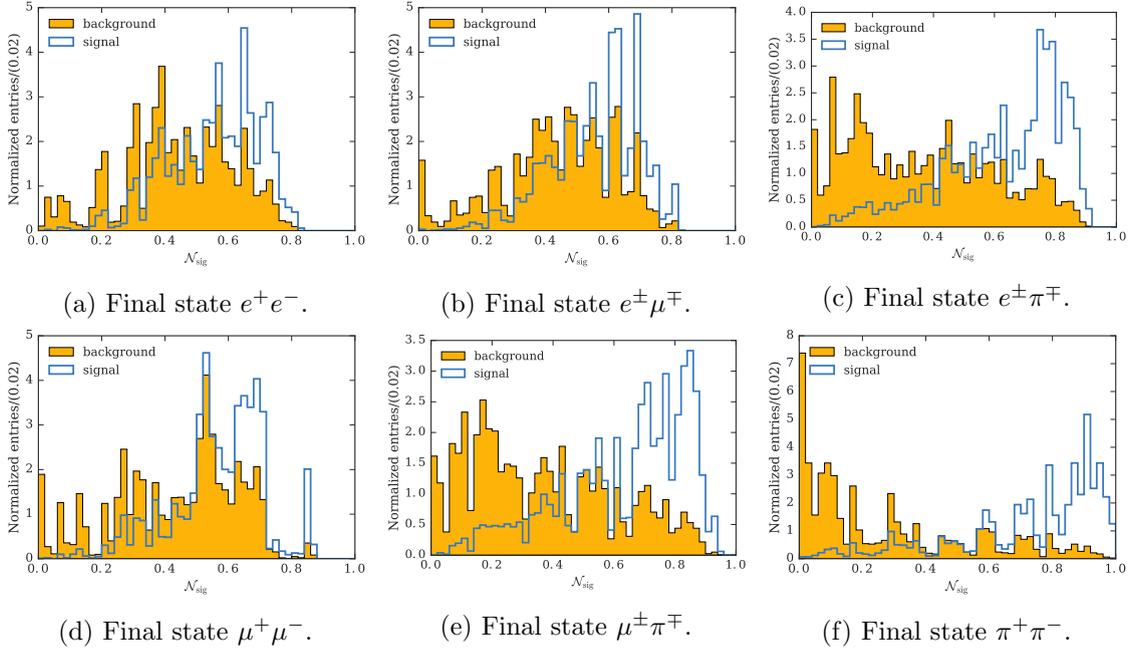
(f) Final state $\pi^+\pi^-$.

Figure 7.10.: NeuroBayes neural net output variable $\mathcal{N}_{\text{sig}}$ plotted for signal (blue line) and background (solid yellow) on the training samples.

specific final state. The signal to background ratio is fixed to 1:1, since no prior knowledge of the signal probability is assumed. As NeuroBayes calculates the significance of each input variable and only uses the most significant ones, the final set of used input variables differs for each final state.

The final discriminating variable $\mathcal{N}_{\text{sig}}$ is shown in Fig. 7.10 for the six neural networks. For final states including a π, the discrimination between signal and background events is slightly larger. One reason is the usage of the variable $\cos\theta^*_{\tau\triangleleft\pi}$, which is only defined for these final states, as input. For the pure letponic final states the variables are not defined as there are two neutrinos in the τ decay. During the training phase NeuroBayes calculates the importance of each input variable. In the final state e^+e^- the most important variables are $\mathcal{N}_{\text{tag}}$, p_{T}, and the pseudo helicity angle $\cos\theta^*_{\text{hel},0}$. $\mathcal{N}_{\text{tag}}$, $|\vec{p}_{\text{miss}}|$, and $\cos\theta_{\text{miss}}$ are most important in the $e^\pm\mu^\mp$ final state. The three most important variables in the networks of the final states $e^\pm\pi^\mp$ and $\mu^\pm\pi^\mp$ are the angles $\cos\theta^*_{\tau,\pi_0,1}$, and $\mathcal{N}_{\text{tag}}$. For $\mu^+\mu^-$ the angle between the two muon candidates $\cos\theta_{0\triangleleft1}$, $\mathcal{N}_{\text{tag}}$, and $\Sigma(d_{\text{IP}})$ are the most relevant variables. In the $\pi^+\pi^-$ final state the missing mass squared M^2_{miss}, the transverse missing momentum $|\vec{p}_{\text{T,miss}}|$, and $\mathcal{N}_{\text{tag}}$ are most significant in the trainings. All variables used in the networks are listed in the tables in Chapter A together with their estimated rank.

7.3.2. Optimization

The final decision of whether an event is selected or not will be made using a cut on $\mathcal{N}_{\text{sig}}$. In order to find the optimal value for the cut on $\mathcal{N}_{\text{sig}}$, Punzi's figure of merit (FOM) for searches for new effects [48], given by

$$f(\epsilon, B) = \frac{\epsilon}{\sigma/2 + \sqrt{B}}, \tag{7.12}$$

with the reconstruction efficiency ϵ, the number of background events B, and the desired one-tailed significance of an observation $\sigma = 3$, is used. Both, the efficiency and the number of background events are estimated in the signal window

$$E_{\text{Ecl}} < 0.2 \text{ GeV}. \tag{7.13}$$

In the signal samples used for the optimization both τ leptons decay only into the specific final state. Hence, no cross feed is included. For 10000 cut values in the interval $[0, 1.0]$, the reconstruction efficiency and the number of surviving background events are estimated and the corresponding value of the FOM is calculated. The optimal cut value is the one where the FOM reaches its maximum. Independent samples, not used in the training of the nets, are used for both signal and background in order to optimize the cut on $\mathcal{N}_{\text{sig}}$. In Fig. 7.11, the process of the optimization for each neural network is visualized additionally to the reconstruction efficiency for each cut value.

7.4. Efficiency

The signal efficiency is determined using a sample of 50×10^6 $B\bar{B}$ events with B_{sig} decaying into $B^0 \to \tau^+\tau^-$ with all possible τ final states, while the B_{tag} decays via a $b \to c$ transition. The uncertainty on the efficiency is calculated as

$$\sigma_\epsilon = \sqrt{\frac{N_{\text{sel}} \cdot (N_{\text{gen}} - N_{\text{sel}})}{N_{\text{gen}}^3}}, \tag{7.14}$$

with the number of selected and generated signal events N_{sel} and N_{gen}, respectively. The reconstruction efficiencies in the six final states for correctly reconstructed events and cross feed are listed in Table 7.7. Cross feed in a final state are $B^0 \to \tau^+\tau^-$ events, which are reconstructed in this final state but originally are simulated in another final state, e.g. a simulated $e^\pm\mu^\mp$ decay is reconstructed in the e^+e^- final state since the muon is reconstructed as an electron. For the final states with a pion ($e^\pm\pi^\mp$, $\mu^\pm\pi^\mp$, $\pi^+\pi^-$), the amount of cross feed is comparable with the amount of correctly reconstructed signal events, or, in the case of the final states $\mu^\pm\pi^\mp$ and $\pi^+\pi^-$, even larger. Hence, the cross feed component is split up into different simulated channels, e.g. $\rho^\pm\rho^\mp$, where both τ leptons decay into $\rho(770)^-\bar{\nu}_\tau$. In Table 7.8, the individual cross feed components are listed for the different final states. For the reconstructed final state $e^\pm\pi^\mp$ the largest component of the cross feed originates from $e^\pm\rho^\mp$ decays, in which the π^0 from the ρ^+ decay is not reconstructed. Another large component are events for which the μ is misidentified as π. In the case of the final state $\mu^\pm\pi^\mp$ ($\pi^+\pi^-$) it is analogous: the main cross feed stems from $\mu^\pm\rho^\mp$ ($\pi^\pm\rho^\mp$) and $\mu^+\mu^-$ ($\mu^\pm\pi^\mp$) decays.

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state e^+e^-.

(e) Final state $e^\pm\mu^\mp$.

(f) Final state $e^\pm\pi^\mp$.

(g) Final state $\mu^+\mu^-$.

(h) Final state $\mu^\pm\pi^\mp$.

(i) Final state $\pi^+\pi^-$.

(j) Final state $\mu^+\mu^-$.

(k) Final state $\mu^\pm\pi^\mp$.
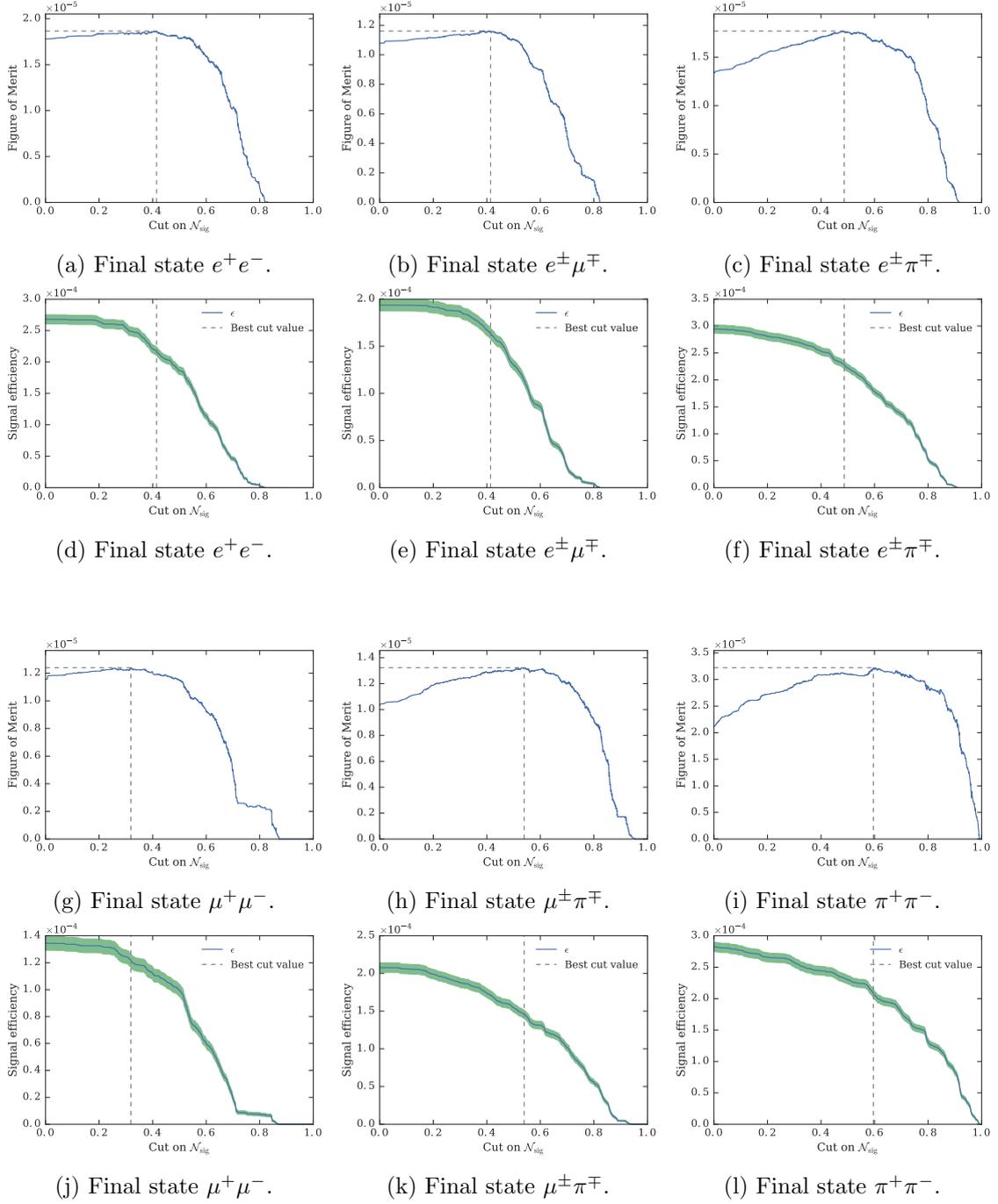
(l) Final state $\pi^+\pi^-$.

Figure 7.11.: In the first and third row, the FOM for the cuts on $\mathcal{N}_{\text{sig}}$ for the different final states are shown. Below (second and fourth row) are the corresponding reconstruction efficiencies versus the cuts on $\mathcal{N}_{\text{sig}}$.

Table 7.7.: Reconstruction efficiencies for all final states (given in 10^{-5}) split up into correctly reconstructed signal events and cross feed. The efficiencies include the tag efficiency and the τ branching ratios of the reconstructed final state.

Final state	Corr. rec. (in 10^{-5})	Cross feed (in 10^{-5})	Total (in 10^{-5})
e^+e^-	1.26 ± 0.05	0.048 ± 0.010	1.30 ± 0.05
$e^\pm\mu^\mp$	1.56 ± 0.06	0.204 ± 0.020	1.77 ± 0.06
$e^\pm\pi^\mp$	1.54 ± 0.06	1.52 ± 0.06	3.06 ± 0.08
$\mu^+\mu^-$	0.549 ± 0.033	0.130 ± 0.016	0.68 ± 0.04
$\mu^\pm\pi^\mp$	0.82 ± 0.04	0.95 ± 0.04	1.78 ± 0.06
$\pi^+\pi^-$	0.438 ± 0.030	0.84 ± 0.04	1.28 ± 0.05

Table 7.8.: Reconstruction efficiencies (in 10^{-5}) for all final states. The columns are the six reconstructed final states. The rows are the different simulated final states of the $B^0 \to \tau^+\tau^-$ decays. The elements in bold are the efficiencies where the simulated and reconstructed final state are the same.

	e^+e^-	$e^\pm\mu^\mp$	$e^\pm\pi^\mp$	$\mu^+\mu^-$	$\mu^\pm\pi^\mp$	$\pi^+\pi^-$
e^+e^-	$\mathbf{1.26 \pm 0.05}$	0.01 ± 0.00	0.09 ± 0.01	-	-	-
$e^\pm\mu^\mp$	0.01 ± 0.01	$\mathbf{1.56 \pm 0.06}$	0.40 ± 0.03	0.01 ± 0.00	0.03 ± 0.01	0.00 ± 0.00
$e^\pm\pi^\mp$	0.01 ± 0.00	0.06 ± 0.01	$\mathbf{1.54 \pm 0.06}$	0.00 ± 0.00	0.01 ± 0.00	0.03 ± 0.01
$\mu^+\mu^-$	-	0.01 ± 0.01	0.00 ± 0.00	$\mathbf{0.55 \pm 0.03}$	0.18 ± 0.02	0.01 ± 0.00
$\mu^\pm\pi^\mp$	-	0.01 ± 0.00	0.01 ± 0.01	0.05 ± 0.01	$\mathbf{0.82 \pm 0.04}$	0.18 ± 0.02
$\pi^+\pi^-$	-	-	0.01 ± 0.00	-	0.04 ± 0.01	$\mathbf{0.44 \pm 0.03}$
$e^\pm\rho^\mp$	0.02 ± 0.01	0.07 ± 0.01	0.76 ± 0.04	0.00 ± 0	0.01 ± 0.00	0.01 ± 0.01
$\mu^\pm\rho^\mp$	-	0.01 ± 0.00	0.01 ± 0.00	0.04 ± 0.01	0.50 ± 0.03	0.06 ± 0.01
$\pi^\pm\rho^\mp$	-	-	0.02 ± 0.01	0.00 ± 0.00	0.05 ± 0.01	0.36 ± 0.03
$\rho^\pm\rho^\mp$	-	-	0.00 ± 0.00	-	0.01 ± 0.01	0.08 ± 0.01
Other	0.01 ± 0.00	0.03 ± 0.01	0.21 ± 0.02	0.03 ± 0.01	0.13 ± 0.02	0.10 ± 0.01
Sum	1.30 ± 0.05	1.77 ± 0.06	3.06 ± 0.08	0.68 ± 0.04	1.78 ± 0.06	1.28 ± 0.05

7.5. Background Composition

Using information from the generator level, it is possible to extract the simulated decay on the signal side for each event. The main background in all final states are misreconstructed B^0 decays. In Table 7.9 the decays contributing with a fraction larger than two percent after the cut on $\mathcal{N}_{\text{sig}}$ are listed for each final state. The main background contribution in all final states is due to missing massive particles like K_L and π^0 mesons, or the slow pion from D^* decays. In the final state e^+e^- the decay chain $B^0 \to D^-e^+\nu_e$, with the D^- decaying further into $K_Le^-\bar{\nu}_e$, is the most dominant background process. It is very likely that the K_L is not detected. Therefore, the two electrons in the final state, together with the missing energy from the neutrinos and the K_L, exactly mimic the signature of the $B^0 \to \tau^+\tau^-$ decay. Analogous effects occur for the other leptonic final states $e^\pm\mu^\mp$ and $\mu^+\mu^-$. In decays including ρ, the π^0 of its decay is not reconstructed. Another background originates in the additional misreconstruction of a charged track, e.g. in the final state $e^\pm\pi^\mp$ in which a μ is reconstructed as a charged π.

The background is decomposed and decays are grouped according to the B decays

- $B^0 \to D^-\ell^+\nu_\ell$,

- $B^0 \to D^-\tau^+\nu_\tau$,

- $B^0 \to D^*(2010)^-\ell^+\nu_\ell$, and

- $B^0 \to D_0^*(2400)^-\ell^+\nu_\ell$.

Decays of the type $B^0 \to D^-\ell^+\nu_\ell$ are further subdivided according to their D^- decays

- $D^- \to K^0\ell^-\bar{\nu}_\ell$,

- $D^- \to K^0\pi^-$, and

- $D^- \to K^0\pi^-\pi^0$.

Here, ℓ can either be a e or a μ. The E_{ECL} distributions for the different background components are plotted in Fig. 7.12 for all six final states. In the pure leptonic final state e^+e^-, the dominant component is the decay $B^0 \to D^-\ell^+\nu_\ell$ where the D^- decays into a $K_L\ell^-\bar{\nu}_\ell$. This is the only peaking component. In the case of the $\mu^\pm\pi^\mp$ final state, another contribution has a peaking E_{ECL} distribution, namely $B^0 \to D^-\ell^+\nu_\ell$, but here the D^- decays into a K_L and a π^-. In both cases, the K_L is not detected which results in missing momentum in the event. The two charged tracks in the final state are then taken from the B and D decay, respectively.

Table 7.9.: Dominant background processes after the final selection.

channel	Decay	events	fraction
e^+e^-	$B^0(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)e^+\nu_e)$	151.2	0.32
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)\pi^0)e^+\nu_e)$	40.2	0.09
	$B^0(\to D^-(\to K^0(\to K_S^0)e^-\bar{\nu}_e)e^+\nu_e)$	19.9	0.04
	$B^0(\to D^*(2010)^-(\to \bar{D}^0(\to K^+e^-\bar{\nu}_e)\pi^-)e^+\nu_e)$	15.0	0.03
	$B^0(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)e^+\nu_e\gamma)$	14.2	0.03
	$B^0(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e\gamma)e^+\nu_e)$	12.3	0.03
	$B^0(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)\tau^+(\to e^+\bar{\nu}_\tau\nu_e)\nu_\tau)$	10.3	0.02
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)\pi^0)e^+\nu_e)$	9.5	0.02
$e^\pm\mu^\mp$	$B^0(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)\mu^+\nu_\mu)$	130.7	0.19
	$B^0(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)e^+\nu_e)$	103.3	0.15
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)\pi^0)\mu^+\nu_\mu)$	35.8	0.05
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)\pi^0)e^+\nu_e)$	24.4	0.04
	$B^0(\to D^-(\to K^0(\to K_S^0)e^-\bar{\nu}_e)\mu^+\nu_\mu)$	19.0	0.03
	$B^0(\to D^*(2010)^-(\to \bar{D}^0(\to K^+e^-\bar{\nu}_e)\pi^-)\mu^+\nu_\mu)$	16.2	0.02
	$B^0(\to D^-(\to K^0(\to K_S^0)\mu^-\bar{\nu}_\mu)e^+\nu_e)$	13.6	0.02
$e^\pm\pi^\mp$	$B^0(\to D^-(\to K^0(\to K_L^0)\pi^-)e^+\nu_e)$	47.9	0.06
	$B^0(\to D^-(\to \rho(770)^-(\to \pi^-\pi^0)K^0(\to K_L^0))e^+\nu_e)$	27.7	0.04
	$B^0(\to D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e)\mu^+\nu_\mu)$	25.8	0.03
	$B^0(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)e^+\nu_e)$	18.1	0.02
	$B^0(\to \rho(770)^+(\to \pi^+\pi^0)D^-(\to K^0(\to K_L^0)e^-\bar{\nu}_e))$	14.9	0.02
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)\pi^-)\pi^0)e^+\nu_e)$	17.3	0.02
	$B^0(\to D^-(\to K^0(\to K_L^0)\pi^-\pi^0)e^+\nu_e)$	13.3	0.02
$\mu^+\mu^-$	$B^0(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu)$	102.8	0.34
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)\pi^0)\mu^+\nu_\mu)$	27.5	0.09
	$B^0(\to D^-(\to K^0(\to K_S^0)\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu)$	13.6	0.04
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)\pi^0)\mu^+\nu_\mu)$	7.9	0.03
	$B^0(\to D^*(2010)^-(\to \bar{D}^0(\to K^+\mu^-\bar{\nu}_\mu)\pi^-)\mu^+\nu_\mu)$	7.4	0.02
	$B^0(\to D^-(\to K^*(892)^0(\to K^0(\to K_L^0)\pi^0)\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu)$	4.8	0.02
$\mu^\pm\pi^\mp$	$B^0(\to D^-(\to K^0(\to K_L^0)\pi^-)\mu^+\nu_\mu)$	36.1	0.07
	$B^0(\to D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu)\mu^+\nu_\mu)$	26.5	0.05
	$B^0(\to D^-(\to \rho(770)^-(\to \pi^-\pi^0)K^0(\to K_L^0))\mu^+\nu_\mu)$	19.6	0.04
	$B^0(\to D^-(\to K^0(\to K_L^0)\pi^-\pi^0)\mu^+\nu_\mu)$	12.8	0.03
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)\pi^-)\pi^0)\mu^+\nu_\mu)$	12.5	0.02
	$B^0(\to \rho(770)^+(\to \pi^+\pi^0)D^-(\to K^0(\to K_L^0)\mu^-\bar{\nu}_\mu))$	9.4	0.02
	$B^0(\to D^*(2010)^-(\to \bar{D}^0(\to K^+\pi^-)\pi^-)\mu^+\nu_\mu)$	9.0	0.02
$\pi^+\pi^-$	$B^0(\to \rho(770)^+(\to \pi^+\pi^0)D^-(\to K^0(\to K_L^0)\pi^-))$	5.1	0.03
	$B^0(\to D^-(\to K^0(\to K_L^0)\pi^-)\mu^+\nu_\mu)$	4.9	0.03
	$B^0(\to D^*(2010)^-(\to D^-(\to K^0(\to K_L^0)\pi^-)\pi^0)\mu^+\nu_\mu)$	4.0	0.02
	$B^0(\to D^-(\to \rho(770)^-(\to \pi^-\pi^0)K^0(\to K_L^0))\mu^+\nu_\mu)$	3.1	0.02
	$B^0(\to D^*(2010)^-(\to \bar{D}^0(\to K^+\pi^-)\pi^-)\mu^+\nu_\mu)$	2.8	0.02
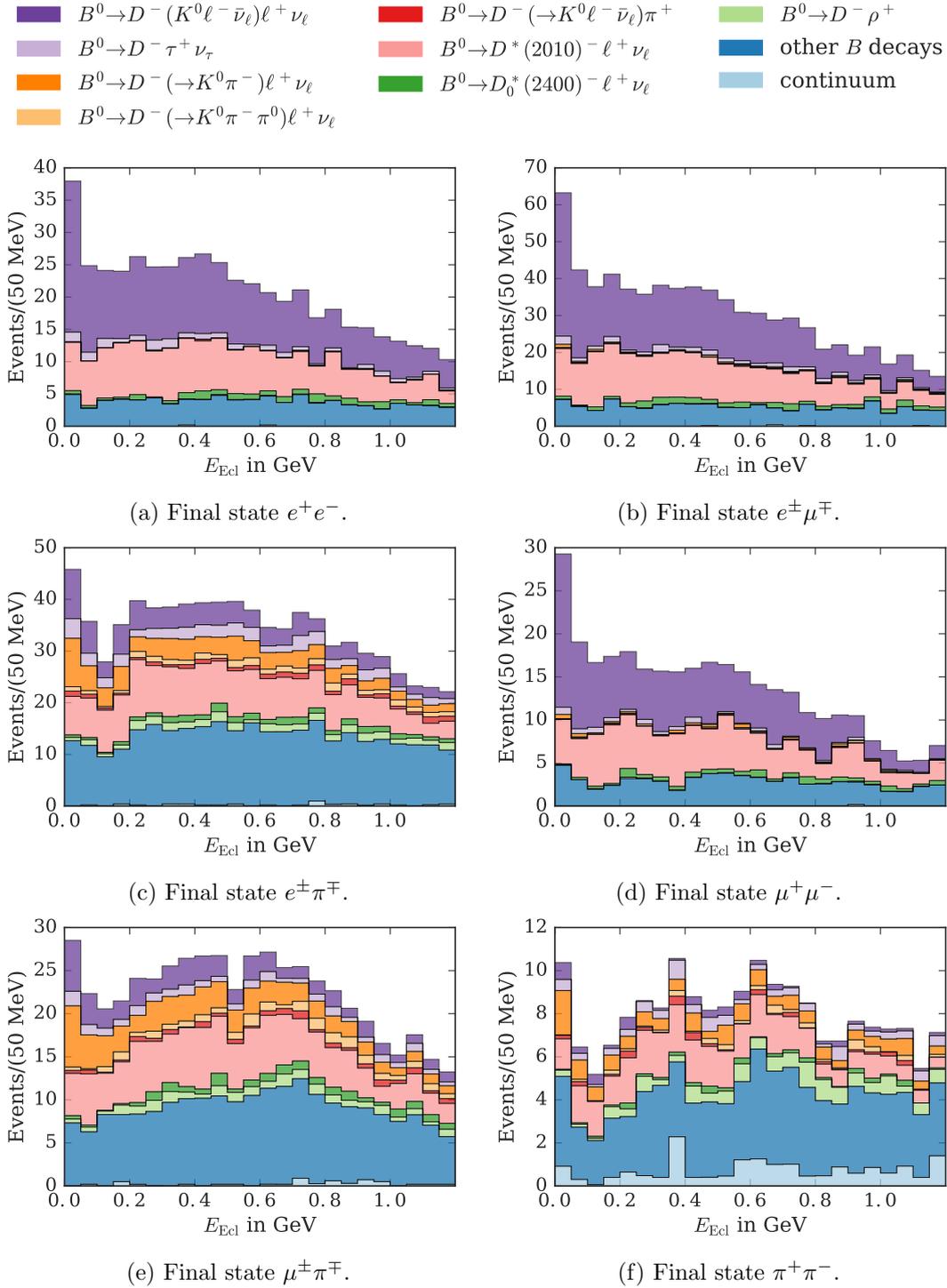
Figure 7.12.: E_{ECL} distributions of simulated background events for the six final states. The background events are grouped by their B and D decays, respectively.

8. Validation

In a blind analysis, the model is developed and optimized on MC. It is essential to verify that the measured data is correctly described by the MC. Background processes with the same signatures as signal simulated with a wrong branching ratio can have a big influence on the resulting branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$. As the branching ratio is extracted using a fit on the variable E_{ECL}, data and MC distributions of this variable are investigated in the following tests.

8.1. Off-Resonance Data

As described in Section 3.1 the cross section for a continuum process at the $\Upsilon(4S)$ energy is about three times larger than the one for producing a $\Upsilon(4S)$ itself. Therefore, deviations between data and MC can also have their origin in wrongly modeled continuum background samples. In order to test for possible differences in the continuum background, so-called off-resonance data and MC are processed and the resulting outcomes are compared to each other. Off-resonance data (MC) denotes a recorded (simulated) sample with a CM energy 60 MeV below the $\Upsilon(4S)$ mass. The Belle experiment recorded an off-resonance data sample with an integrated luminosity of $\mathcal{L}_{\mathrm{off}} = 79~\mathrm{fb}^{-1}$. The off-resonance samples are processed and the preselection cuts described in Section 7.1 are applied with only one change: the cut on $\mathcal{N}_{\mathrm{tag}}$ is released to $\mathcal{N}_{\mathrm{tag}} > 0.01$ in order to enhance the number of events in the resulting samples. Both, the tag correction and the LID correction weights are applied. In Fig. 8.1, the E_{ECL} distributions of data and MC are compared to each other. The two-sample Kolmogorov–Smirnov (KS) test [49] results in a test statistic of 0.06 and the corresponding p-value of 0.97. The null hypothesis H_0 that both distributions originate from a common distribution cannot be rejected at the 95% confidence level (CL) by this test. But the simulation overestimates the number of selected events in eleven of the twelve bins. The crucial point for the analysis is that after the preselection the off-resonance data shows no peaking component as expected. After the final selection, in the six final states, five events are selected from off-resonance data while 1.5 events are expected from MC.
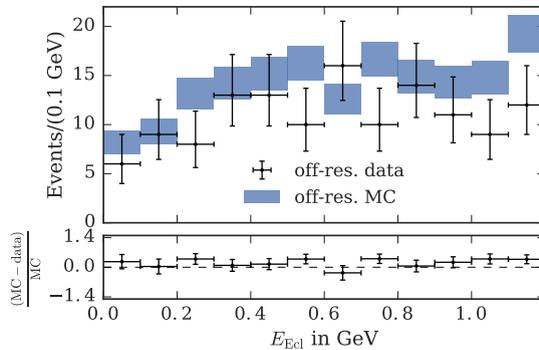
Figure 8.1.: E_{ECL} distribution for off-resonance data and MC samples after applying the preselection cuts (Table 7.2). The cut on $\mathcal{N}_{\mathrm{tag}}$ is released to 0.01 as otherwise the number of events is almost zero. The blue boxes are the 1σ band of the number of expected events in each bin.

Also in the final selection, no peaking background components originating from continuum events are expected. The simulated continuum events consistently describes the shape of data distributions. Thus the shape of the continuum background is fixed to the shape obtained from MC, while the normalization is floated in the fit for the branching ratio extraction (see Chapter 9).

8.2. Data–Monte Carlo Comparison in Sideband Samples

The dominant background component in the final selected samples does not originate from continuum events but from misreconstructed B decays. Hence, different sideband selections are applied on the reconstructed samples in order to validate other types of background than continuum. The following three different sideband selections are used, as for each sideband the focus lies on a different background type:

ΔE_{tag} sideband Test background from wrongly reconstructed B_{tag} candidates (Section 8.2.1)

K_S sideband Test background from B meson decays with a K_L meson in the final state (Section 8.2.2)

E_{ECL} sideband Test combined background distributions (Section 8.2.3)

The MC samples used in all sideband validation tests consist of generic, $BD\ell\nu_\ell$, continuum, rare, and $u\ell\nu_\ell$ events scaled to the integrated luminosity of data. Event-by-event weights are applied to all MC samples to correct for the known data–MC discrepancies described in Section 7.2.

8.2.1. ΔE_{tag} Sideband

B_{tag} candidates with a large value of ΔE_{tag} are wrongly reconstructed. In order to test whether the E_{ECL} distribution for events with a misreconstructed B_{tag} candidate is well

simulated in the MC samples, the E_{ECL} distribution is compared for data and MC events in the ΔE_{tag} sideband region defined as

$$0.05 < |\Delta E_{\text{tag}}| < 0.1 \text{ GeV}. \tag{8.1}$$

Besides the change of the cut on ΔE_{tag}, the cut on the network output of the B_{tag} candidate is released to $\mathcal{N}_{\text{tag}} > 0.025$ in order to enhance the statistical significance of the comparison. The other cuts on the variables used in the preselection are applied without modification.

In Fig. 8.2, the distributions after the final selection of data and MC are statistically in agreement. Solely in the final state $e^{\pm}\mu^{\mp}$ (Fig. 8.2b) a larger deviation between data and MC histograms can be seen. In the other five final states the distributions for data and MC are in good agreement. No further MC correction procedure is derived from this test.

8.2.2. K_S Sideband

As the main background comes from events where a K_L is not detected, events with an additional K_S candidate are a good way to validate this background and check whether some discrepancies between data and MC occur. Events where a K_S candidate is reconstructed (in addition to the B_{tag} and B_{sig} candidates) are selected for the K_S sideband. The cut on the charged track veto in the preselection is modified and events where additional tracks are used in K_S candidates are not discarded. Only K_S candidates which fulfill the goodKS selection criteria described in Section 7.1.1 are taken into account. The final selection is applied on the K_S sideband samples. In Fig. 8.3, the E_{ECL} distributions in the K_S sideband are shown. The distributions in all final states exhibit no obvious hint that the data is not described by MC. In the region of $E_{\text{Ecl}} \approx 300$ MeV data is underestimated in the final states e^+e^- and $\mu^+\mu^-$. The reason for this deviation can be, that the momentum distribution of the K_S in data is softer than the simulated distribution. However, no large discrepancy is found.

8.2.3. E_{ECL} Sideband

In addition to the tests in the ΔE_{tag} and K_S sideband, distributions for data and MC events in the E_{ECL} sideband, defined as

$$E_{\text{Ecl}} > 0.2 \text{ GeV}, \tag{8.2}$$

are compared to each other. The final selection is applied to data and MC samples without further modifications.

E_{ECL} Distribution

In Fig. 8.4, the E_{ECL} distributions are depicted. In each final state a two-sample KS test is performed to check the compatibility of the E_{ECL} distribution in data and MC. For all six final states the hypothesis that the data and the MC distributions are comparable is not rejected at the 95% CL. Hence, within the statistical uncertainties the E_{ECL} distributions of data and MC agree with each other. In Table 8.1, the results of the KS tests are tabulated. The E_{ECL} distributions in Fig. 8.4 show a good agreement between data and MC.

(a) Final state e^+e^-.

(b) Final state $e^{\pm}\mu^{\mp}$.

(c) Final state $e^{\pm}\pi^{\mp}$

(d) Final state $\mu^+\mu^-$

(e) Final state $\mu^{\pm}\pi^{\mp}$

(f) Final state $\pi^+\pi^-$
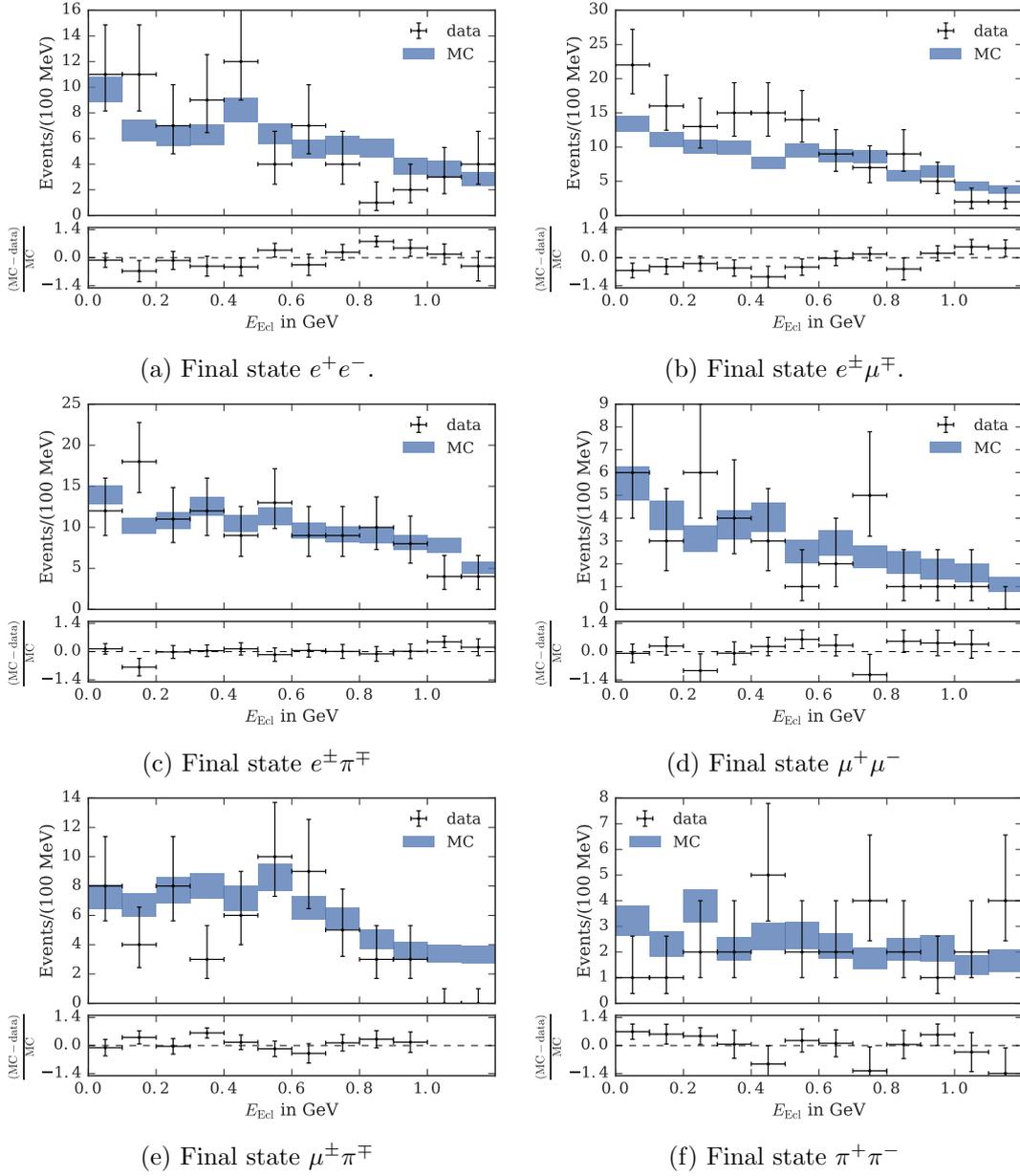
Figure 8.2.: E_{ECL} distributions for ΔE_{tag} sideband events after the final selection. The blue bars represent simulated MC events, the points data.

(a) Final state e^+e^-.

(b) Final state $e^{\pm}\mu^{\mp}$.

(c) Final state $e^{\pm}\pi^{\mp}$

(d) Final state $\mu^+\mu^-$

(e) Final state $\mu^{\pm}\pi^{\mp}$

(f) Final state $\pi^+\pi^-$
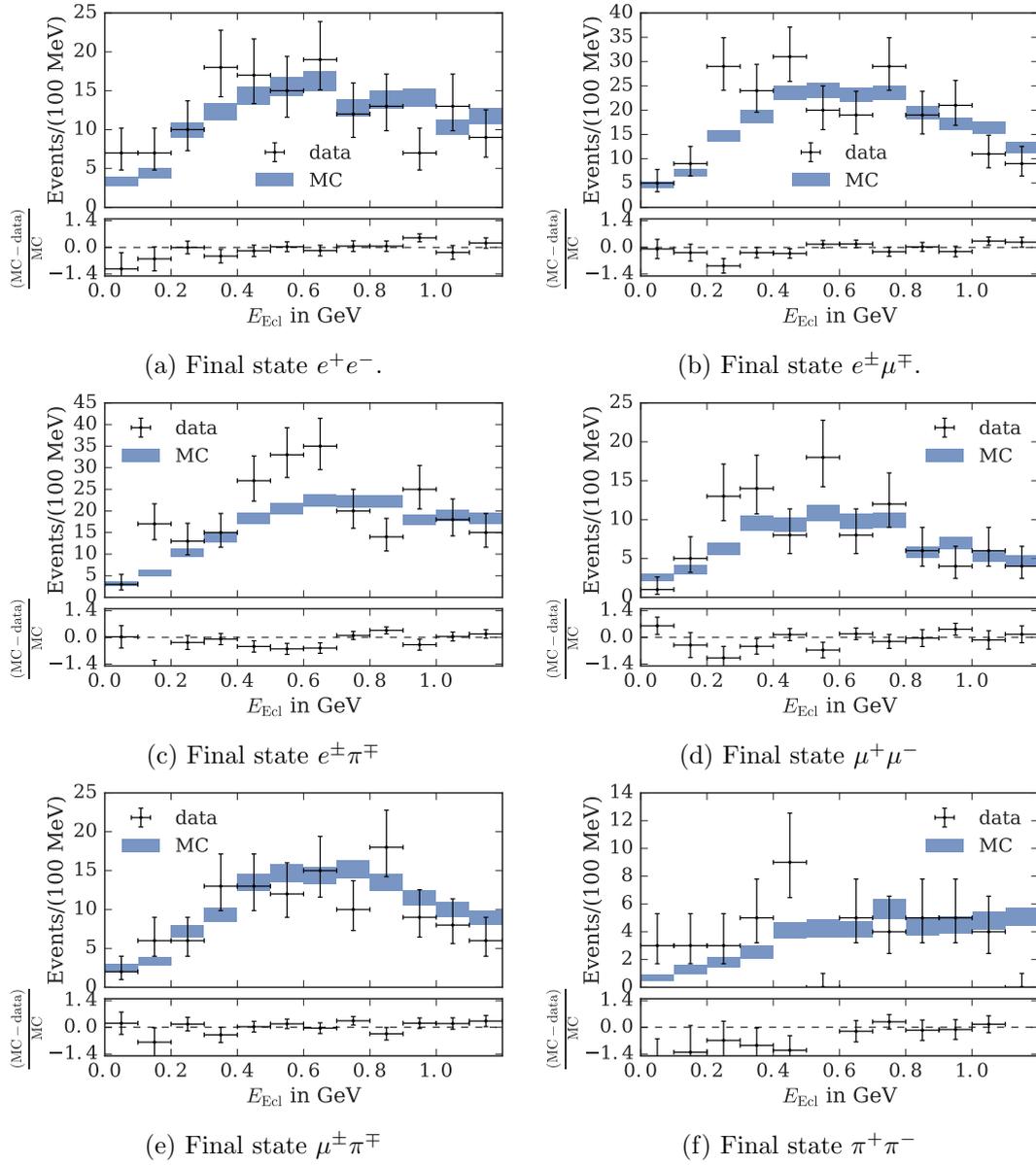
Figure 8.3.: E_{ECL} distributions for K_S sideband events. The blue bars represent simulated MC events, the points data.

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$

(d) Final state $\mu^+\mu^-$

(e) Final state $\mu^\pm\pi^\mp$

(f) Final state $\pi^+\pi^-$
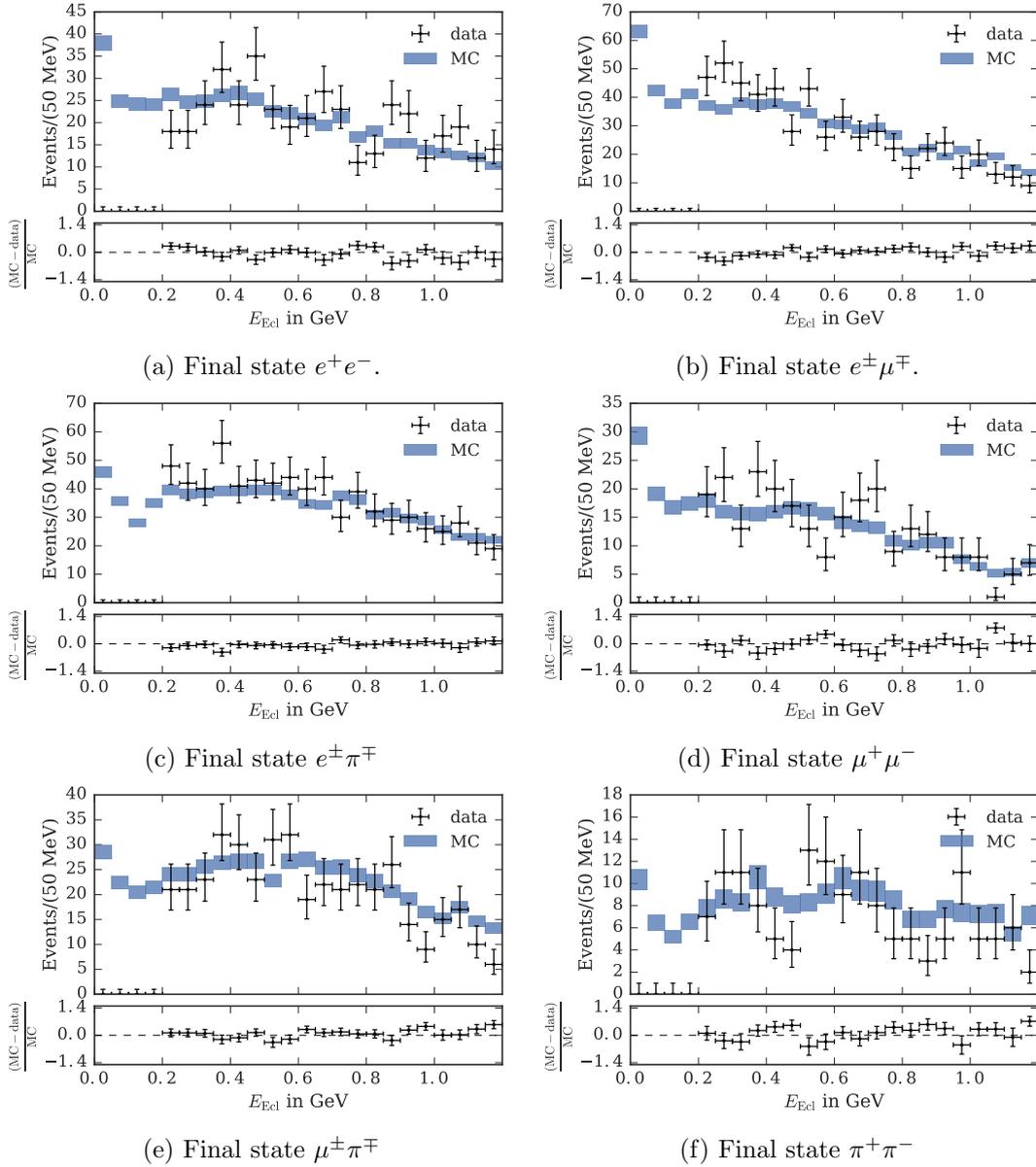
Figure 8.4.: E_{ECL} distributions for E_{ECL} sideband events. The blue bars represent the 1σ uncertainty band of the number of MC events. Data is plotted as black points. In the signal region $E_{\mathrm{Ecl}} < 0.2$ GeV only MC can be plotted since at this point the analysis has not been unblinded yet.

Table 8.1.: Results of the KS test for the E_{ECL} distributions in the E_{ECL} sideband region. H_0 is the null hypothesis that both distributions originate from a common distribution.

Final state	t	p	reject H_0
e^+e^-	0.06	0.52	no
$e^\pm\mu^\mp$	0.07	0.10	no
$e^\pm\pi^\mp$	0.05	0.29	no
$\mu^+\mu^-$	0.05	0.89	no
$\mu^\pm\pi^\mp$	0.06	0.42	no
$\pi^+\pi^-$	0.09	0.59	no

Tag Modes

In order to validate the performance of the FR algorithm the frequencies of the reconstructed decay modes of the B_{tag} candidates in data and MC are compared to each other after the final selection. The frequencies for data and MC are depicted in Fig. 8.5. The distributions for data and MC are in good agreement within the statistical uncertainty.
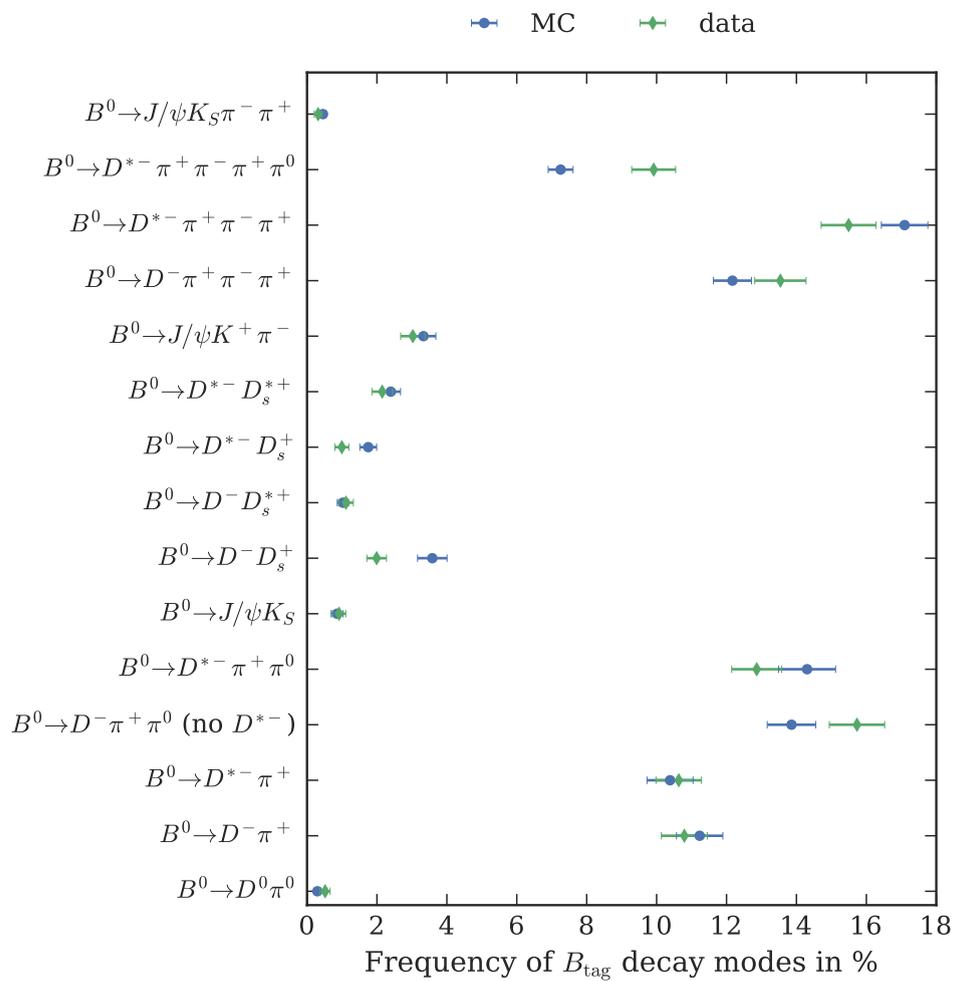
Figure 8.5.: Frequency of the decay modes in which B_{tag} is reconstructed. The blue dots represent the MC expectation. The values corresponding to data are drawn as green diamonds.

9. Branching Ratio Extraction

The signal yields are obtained using a simultaneous extended unbinned maximum likelihood fit to E_{ECL} in all six final states. The fit is implemented using the `RooFit` package [50]. The fitting procedure is outlined in Section 9.1. The details about the determination of the probability density function (PDF) of the E_{ECL} variable are described in Section 9.2. The fit is tested on MC samples (Section 9.3) and toy studies are performed to test the stability of the fit (Section 9.4).

9.1. Fitting Procedure

The fitting procedure is as follows:

1. Creation of the PDF template for signal and background components.

2. Estimation of the normalization of the $BD\ell\nu_\ell$ background component.

3. Estimation of the relative fractions of all remaining background components.

4. Fit to each final state separately to obtain the individual signal and background normalizations.

5. Simultaneous fit on all final states. The normalizations estimated in Step 4 are used as starting parameters and are allowed to vary.

The shape of the E_{ECL} distribution of the signal and all background components is determined from MC and is fixed.

Due to its large yield, the $BD\ell\nu_\ell$ component is removed from the generic background sample and is generated as an individual component. Its normalization is scaled to the number of expected $BD\ell\nu_\ell$ events in the Belle data sample using the current world averages for the branching ratios $\mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0)$, $\mathcal{B}(B^0 \to D^-\ell^+\nu)$, and $\mathcal{B}(D^- \to K^0\ell^-\bar{\nu}_\ell)$, with $\ell = e, \mu$ (see Table 6.1). The normalization of the $BD\ell\nu_\ell$ component is fixed in the fit.

The remaining generic, continuum, rare, and $u\ell\nu_\ell$ backgrounds are treated as individual components in the fit. The relative fractions of these components are estimated from MC and are fixed in the fit, while the overal normalization is allowed to vary.

The branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is determined in each final state as

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = \frac{N_{\text{sig},i}}{2 \cdot N_{B\bar{B}} \cdot \mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0) \cdot \epsilon_i}, \qquad (9.1)$$

where $N_{\text{sig},i}$ and ϵ_i are the number of fitted signal events and reconstruction efficiency, respectively, for final state i. Here, $N_{B\bar{B}}$ denotes the number of $B\bar{B}$ events in the Belle data sample, and $\mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0)$ is the probability that a $\Upsilon(4S)$ decays into a pair of neutral B mesons. The factor 2 in the denominator is required as the number of neutral B mesons and not the number of pairs of neutral B mesons must be used in the calculation of the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$. The signal yields $N_{\text{sig},i}$ in each final state are related to $\mathcal{B}(B^0 \to \tau^+\tau^-)$ via the individual reconstruction efficiency. Therefore, $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is estimated simultaneously in all final states. After all, the fit has seven floating parameters: the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$ and the normalization of the background components (excluding $BD\ell\nu_\ell$) in each of the six final states.

9.2. Description of the PDF

The combined PDF is defined as

$$P(E_{\text{Ecl}}) = \sum_i \left[N_{\text{sig},i}P_{\text{sig},i}(E_{\text{Ecl}}) + N_{BD\ell\nu_\ell,i}P_{BD\ell\nu_\ell,i}(E_{\text{Ecl}}) + N_{\text{bkg},i}\sum_j f_{i,j}P_{\text{bkg},i,j}(E_{\text{Ecl}}) \right], \quad (9.2)$$

where the index i denotes the final state. The normalization of the signal, background, and $BD\ell\nu_\ell$ components are denoted by N_{sig}, N_{bkg}, and $N_{BD\ell\nu_\ell}$, respectively. The index j marks the different background components and f_j is the relative fraction. The fractions are constrained by $\sum_j f_j = 1$. The PDF of the signal is denoted as P_{sig}, that of the $BD\ell\nu_\ell$ component as $P_{BD\ell\nu_\ell}$, and P_{bkg} denotes the PDFs for the remaining background components. The signal PDFs include cross feed. As the number of signal events in each final state is constrained by its reconstruction efficiency, in principle not the six signal yields but the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is fitted.

9.2.1. Histogram Template PDFs

As there is no analytical description of the E_{ECL} distribution for the different components entering the fit, histogram PDFs are used. All shapes are determined from simulated events. For the background components with neutral and charged $b \to c$ decays, the shape is determined using a sample of simulated events which corresponds to nine times the recorded luminosity in data. The continuum, rare, and $u\ell\nu_\ell$ components are determined using 5, 50, and 20 times the luminosity of data, respectively. The template for the $BD\ell\nu_\ell$ component is determined using a sample of 50 million generated decays. To account for the known data–MC discrepancies, the simulated samples are reweighted before generating the templates

using the corrections described in Section 7.2, namely the tag correction, `LID` efficiency correction, and the weights for correcting the branching ratio of $D^0 \to K^0 \ell^- \bar{\nu}_\ell$ decays. The templates for the final states $e^+ e^-$ and $\pi^+ \pi^-$ are shown in Fig. 9.1, respectively Figs. 9.2 and 9.3. The templates for the other four final states can be found in Chapter B.

(a) Signal component

(b) $BD\ell\nu_\ell$ decays.

(c) $B^0 \to X_c +$ anything decays.

(d) $B^\pm \to X_c +$ anything decays.

(e) Neutral rare decays.

(f) $B^0 \to X_u +$ anything decays
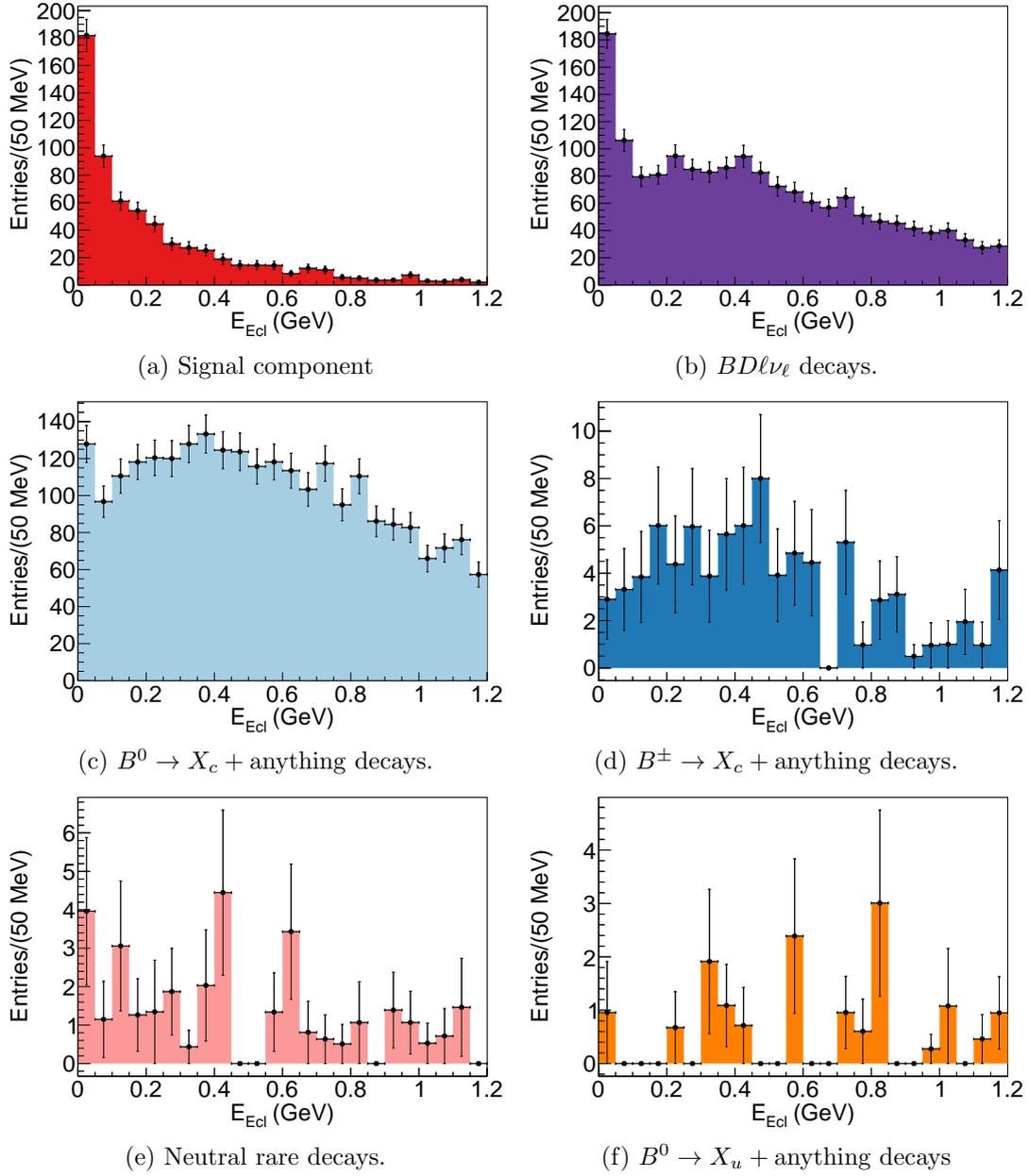
Figure 9.1.: Templates used in the fit for the final state e^+e^-.
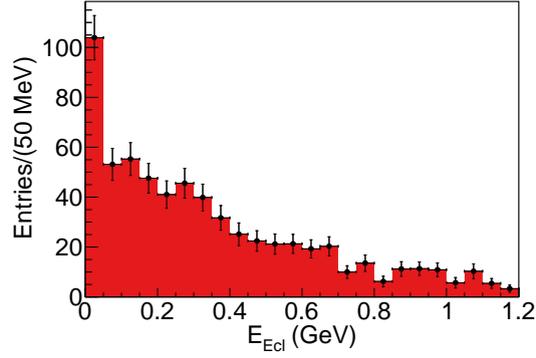
(a) Signal component
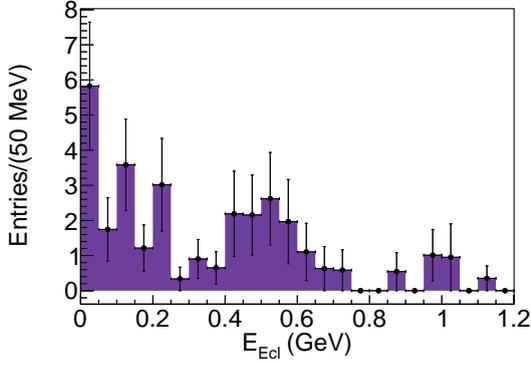
(b) $BD\ell\nu_\ell$ decays.
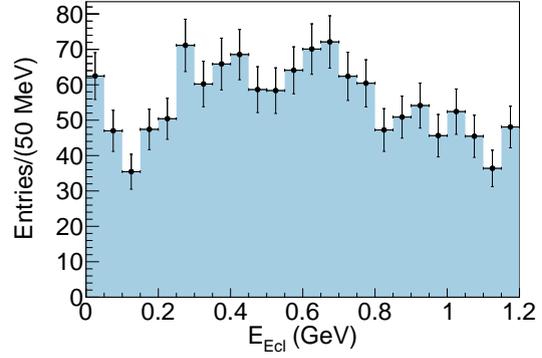
(c) $B^0 \to X_c$ + anything decays.
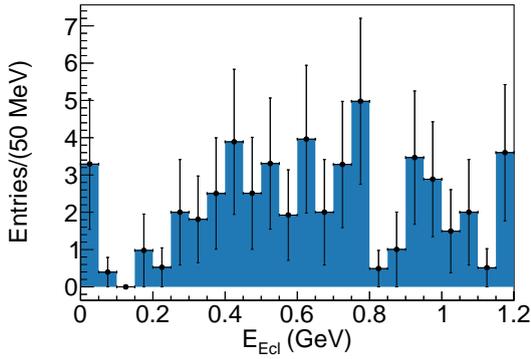
(d) $B^\pm \to X_c$ + anything decays
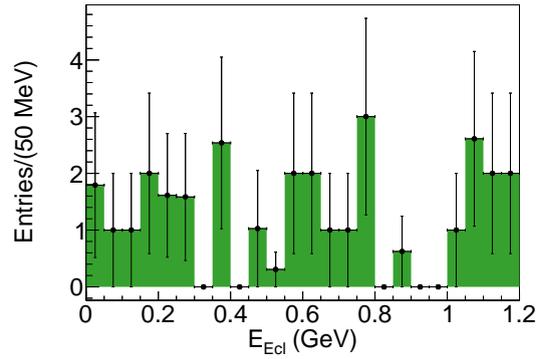
(e) $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$ decays.

Figure 9.2.: Templates used in the fit for the final state $\pi^+\pi^-$.

(a) $e^+e^- \to c\bar{c}$ decays.

(b) Neutral rare decays.

(c) Charged rare decays.

(d) $B^0 \to X_u + $ anything decays
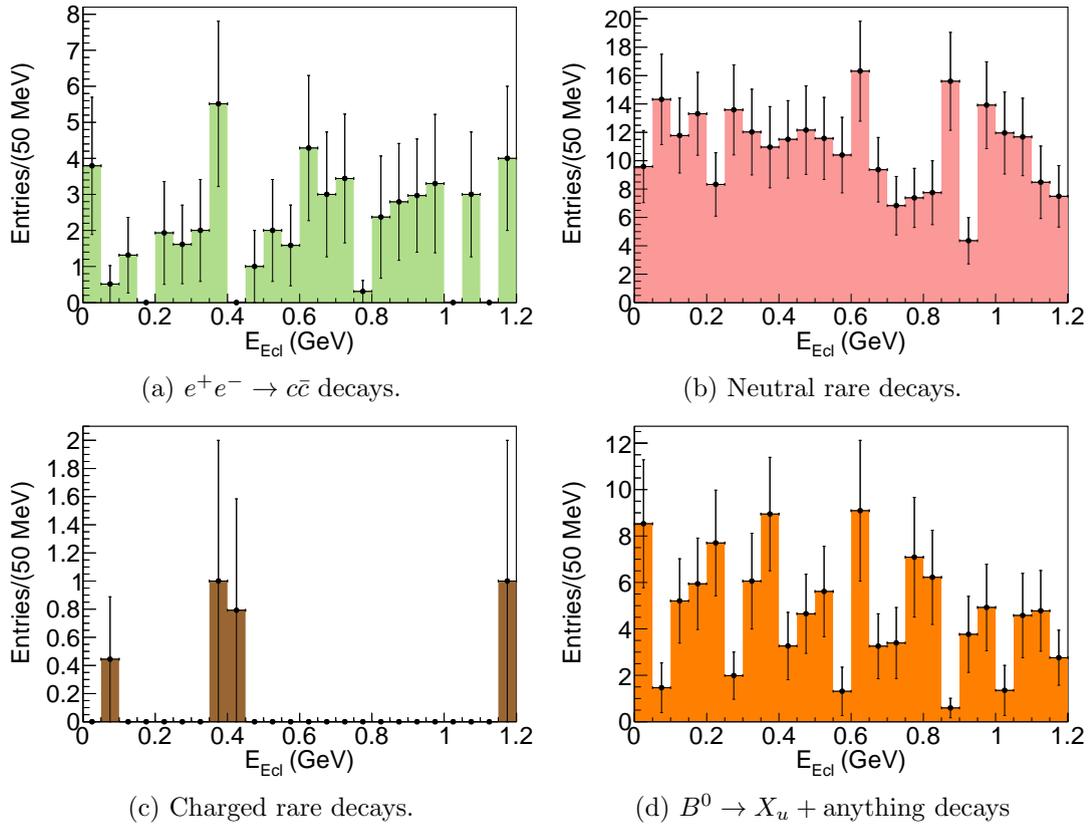
Figure 9.3.: Templates used in the fit for the final state $\pi^+\pi^-$.

9.3. Fit on Simulated Data

The MC samples are separated into sub-samples which are used to both: estimate the PDF templates for the fit, and apply the fit itself. As the rare and $u\ell\nu_\ell$ samples cannot be divided, both are used for fitting and template generation. For generic decays, ten samples are available. Hence, fits on ten different samples can be performed. But since there are only six streams of simulated continuum events, the fit is performed on only six different samples. The resulting branching ratios for the different samples are plotted in Fig. 9.4. Statistically, all six fits, with an average of $\mathcal{B}(B^0 \to \tau^+\tau^-) = (0.16 \pm 0.30) \cdot 10^{-3}$, are consistent with the input branching ratio of $3 \cdot 10^{-8}$. The fit results on Sample 1 are illustrated in Fig. 9.5.
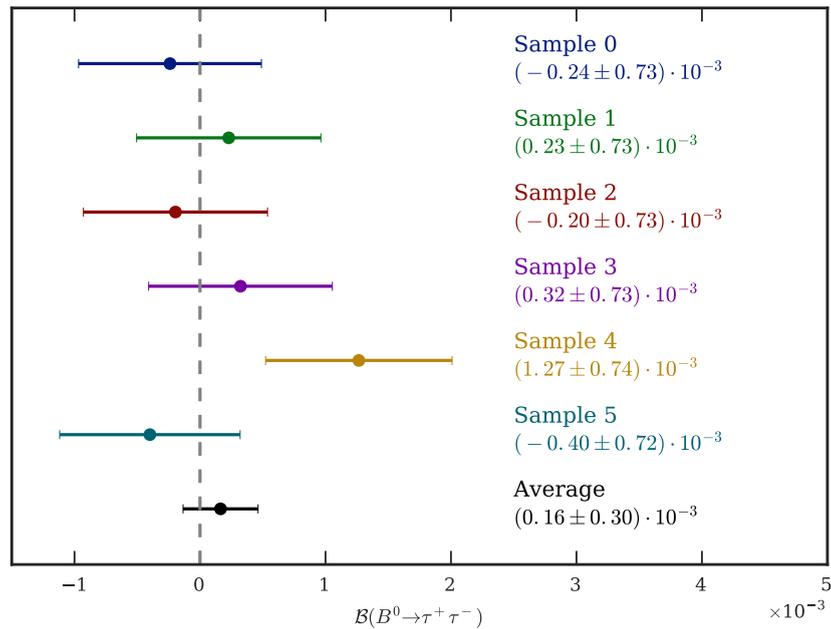
Figure 9.4.: Fitted branching ratios for the different MC samples. The dashed line shows the input value. Only the statistical uncertainties are plotted.

(a) Final state e^+e^-

(b) Final state $e^\pm\mu^\mp$

(c) Final state $e^\pm\pi^\mp$

(d) Final state $\mu^+\mu^-$

(e) Final state $\mu^\pm\pi^\mp$

(f) Final state $\pi^+\pi^-$
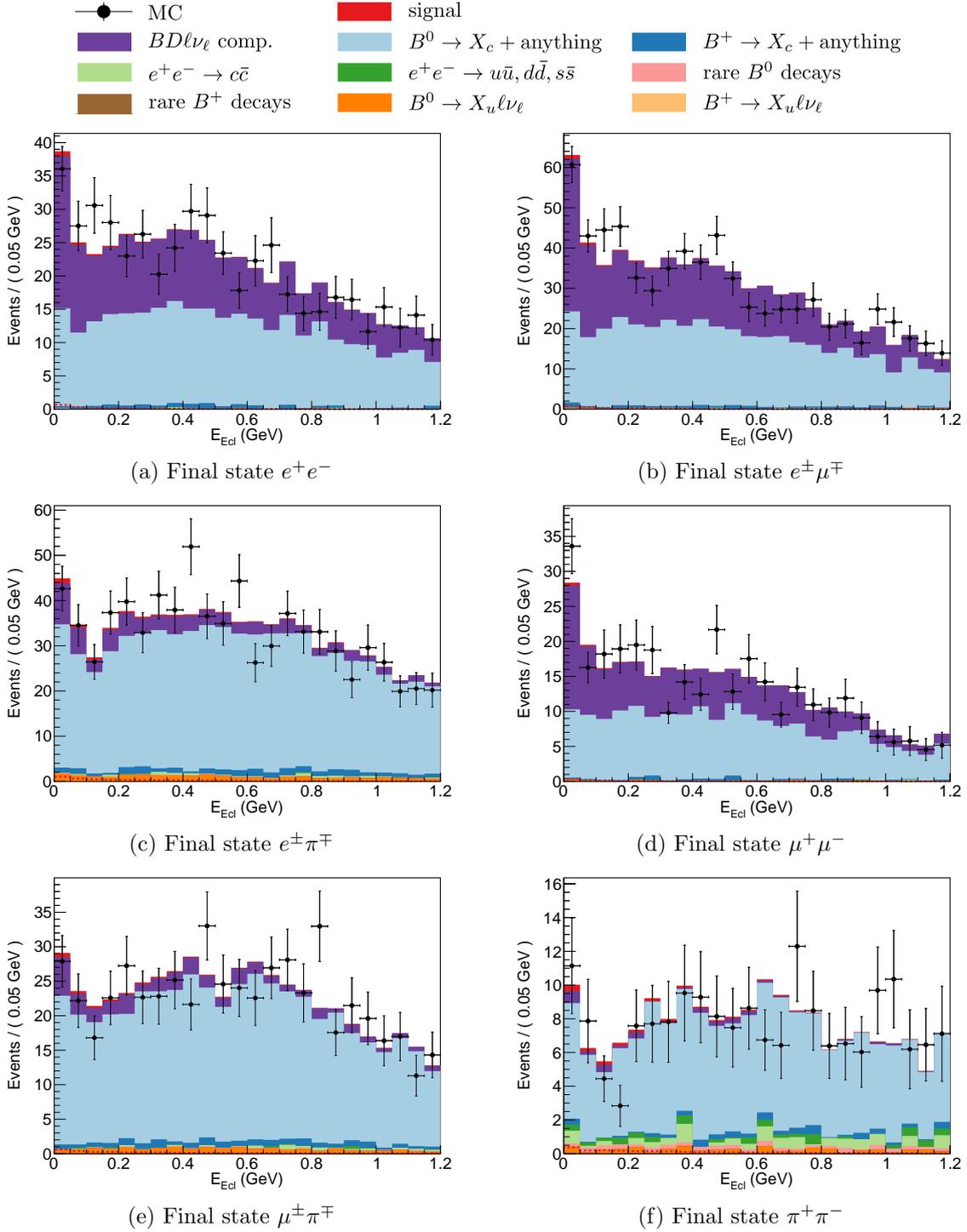
Figure 9.5.: Results of the fit on Sample 1.

9.4. Stability Test

Besides fitting different MC samples, the stability of the fit is tested using pseudo experiments. The latter is a data sample whose E_{ECL} distribution follows the PDF used in the fit. The normalization of the background PDFs are set to the expected number of events in MC. The branching ratio of the signal process $B^0 \to \tau^+\tau^-$ can be set to arbitrary values. After the generation of the sample, it is fitted using the same PDF. Hence, the validity of the model in terms of the correct description of the real data is not tested. This test can only reveal numerical instabilities or a bias.

For each of 20 branching ratios between 0 and 0.01, a thousand pseudo experiments are simulated and fitted. In Fig. 9.6, the mean values of the fitted branching ratios are compared to the simulated ones. A linear function is fitted to the data points in Fig. 9.6, in order to estimate the size of a possible bias. The slope of the linear function is with 1.003 ± 0.002 in very good agreement with one. No offset has been found, since the axis intercept $(-6.8 \pm 1.1) \times 10^{-6}$ is two orders of magnitude smaller than the statistical uncertainty of the fit.
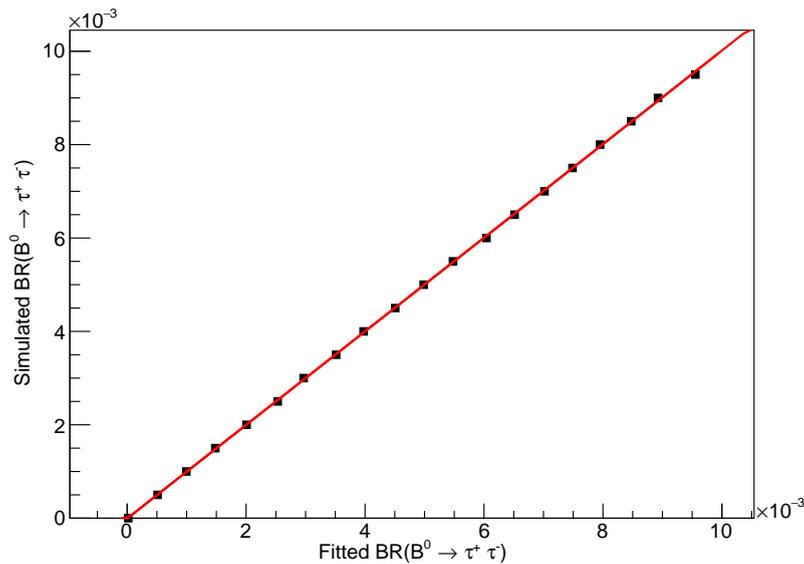
Figure 9.6.: Linearity test result of the fit procedure. The points show the mean values (for 1000 toys) of the fitted branching ratios. The errors are included, but so small that they are not visible here. A linear function formally $f(x) = c_0 + c_1 \cdot x$ is fitted to the points and shown in red. The estimated parameters are: $c_0 = (-6.8 \pm 1.1) \times 10^{-6}$ and $c_1 = 1.003 \pm 0.002$. Hence, the fit perfectly reproduces the generated branching ratios.

9.5. Limit Estimation

The fitted values of $\mathcal{B}(B^0 \to \tau^+\tau^-)$ on the fits to MC samples, for which the decay $B^0 \to \tau^+\tau^-$ is simulated with a branching ratio of 3×10^{-8}, are not significant. Therefore, no signal is expected to be found. An expected upper limit on $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is estimated using the profile likelihood method. For that the fit is repeated for 6000 values for $\mathcal{B}(B^0 \to \tau^+\tau^-)$ in the interval $[-0.001, 0.005]$. The branching ratio is fixed in the fit and the likelihood function is maximized with respect to the remaining parameters. With this, the profile likelihood $\mathcal{P}(\mathcal{B})$ function is estimated. The upper limit at the 90% confidence level (CL) on the branching ratio $\mathcal{B}_{\mathrm{ul}}$ is determined from

$$0.9 = \int_0^{\mathcal{B}_{\mathrm{ul}}} d\mathcal{B} \; \mathcal{P}(\mathcal{B}). \tag{9.3}$$

The lower integration boundary is set to zero in order to exclude the non-physical parameter region. Fig. 9.7 shows the profile likelihood function and the determined upper limit for Sample 1. The calculated limits for the six different MC samples are listed in Table 9.1.
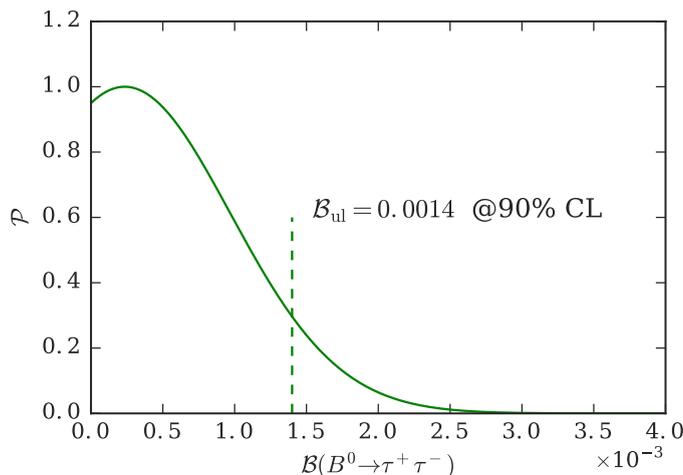
Figure 9.7.: Profile likelihood function $\mathcal{P}$ and expected upper limit of $\mathcal{B}(B^0 \to \tau^+\tau^-)$ determined on Sample 1.

Table 9.1.: Expected upper limits on $\mathcal{B}(B^0 \to \tau^+\tau^-)$ for the six MC samples.

Sample	0	1	2	3	4	5
$\mathcal{B}_{\mathrm{ul}}$ @90% CL (10^{-3})	1.1	1.4	1.1	1.5	2.3	1.0

10. Fit on Data

After the development and validation of the event selection and the fitting procedure on MC, the fit is performed on experimental data. The results of the final fit are summarized in Section 10.1. Afterwards, the estimation of the systematic uncertainties (Section 10.2) and the determination of the significance (Section 10.3) of the fitted result are described.

10.1. Results

The resulting E_{ECL} distributions for the fit on data are shown in Fig. 10.1. A large signal component (red) is observed in all six final states. This is unexpected, as in the studies on MC (Sections 9.3 to 9.5), using the SM expectation for $\mathcal{B}(B^0 \to \tau^+\tau^-)$, no significant signal would have been awaited.

If it is assumed that all background components as in the SM, the excess in data could be interpreted as $B^0 \to \tau^+\tau^-$ events and the resulting value for the branching ratio is

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = (4.39^{+0.80}_{-0.83}) \times 10^{-3}, \tag{10.1}$$

with the statistical uncertainty only. The fitted number of signal events and the branching ratios for the separate final states are listed in Table 10.1. A χ^2 test is performed in order to check the consistency of the fitted branching ratios in the individual final states. The null hypotheses that the individual fit results are consistent with each other cannot be rejected at the 95% confidence level. The results of the tests are shown in Table 10.2. The goodness of the combined fit is measured by a χ^2 test. For that, the χ^2 between the fitted PDF and the data histogram in each final state are cumulated. The number of degrees of freedom is the number of bins (144) minus the number of floating parameters in the fit (7). The resulting reduced χ^2 value is 0.88. The probability for getting this value or a larger one is 35%.

The expected and fitted numbers of background events in each final state are listed in Table 10.3. Solely in the final state $\pi^+\pi^-$, the fitted and expected number of background

(a) Final state e^+e^-.

(b) Final state $e^{\pm}\mu^{\mp}$.

(c) Final state $e^{\pm}\pi^{\mp}$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^{\pm}\pi^{\mp}$.

(f) Final state $\pi^+\pi^-$.
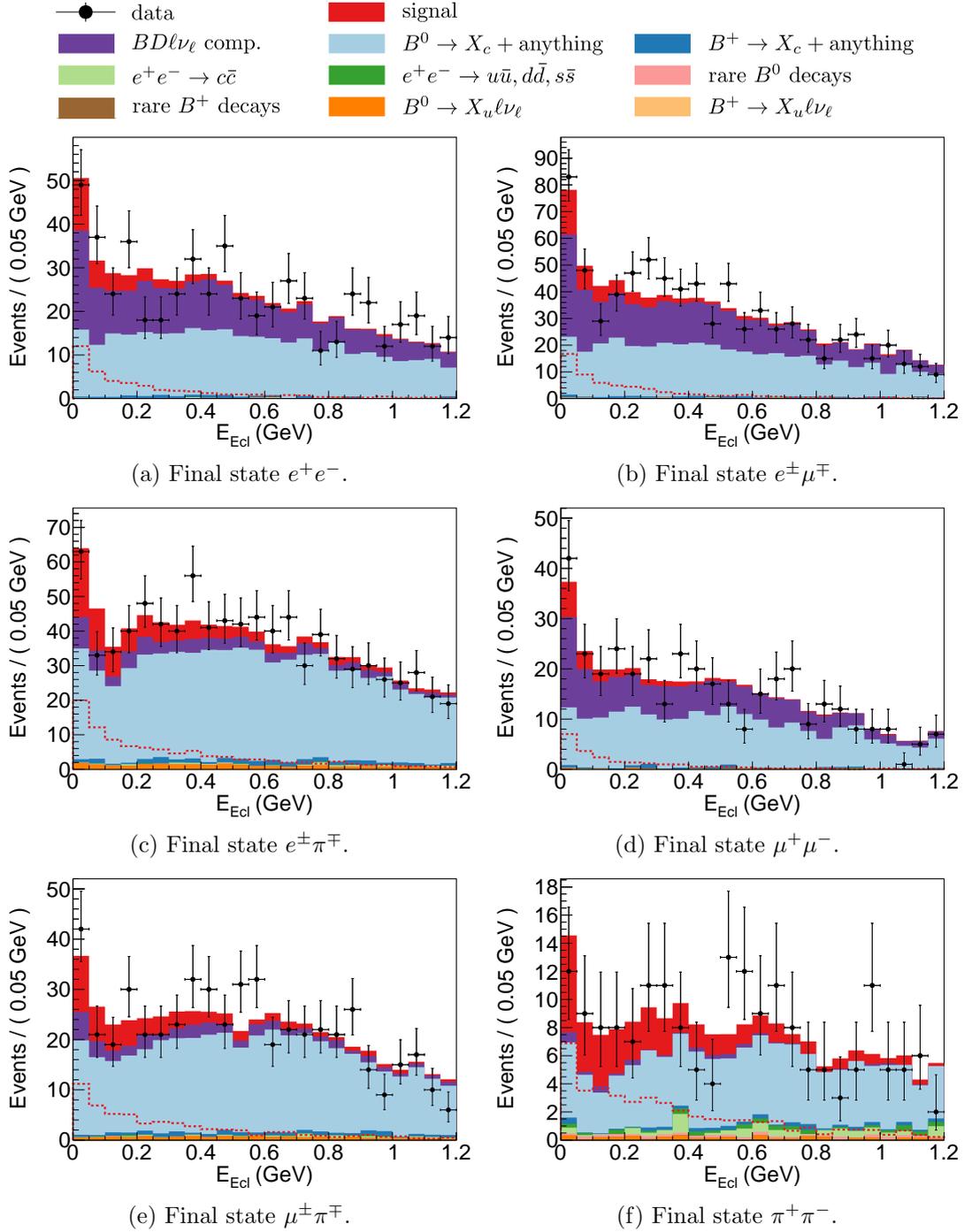
Figure 10.1.: Resulting E_{ECL} distributions from the fit to data. The different components are shown as stacked histogram (solid). Additionally, the signal distribution is separately shown as dashed line.

events differ significantly. In the other five final states, the ratio is in good agreement with one, which indicates a good understanding of the background components.

Table 10.1.: Signal yields and branching ratios for the separate final states and the simultaneous fit on all final states.

Final state	N_{sig}	$\mathcal{B}(B^0 \to \tau^+\tau^-)$ (in 10^{-3})
e^+e^-	33 ± 21	$3.33^{+2.23}_{-2.08}$
$e^\pm\mu^\mp$	73 ± 27	$5.52^{+2.09}_{-1.97}$
$e^\pm\pi^\mp$	70 ± 34	$3.05^{+1.53}_{-1.47}$
$\mu^+\mu^-$	40 ± 18	$7.87^{+3.68}_{-3.40}$
$\mu^\pm\pi^\mp$	63 ± 26	$4.76^{+2.03}_{-1.88}$
$\pi^+\pi^-$	44 ± 18	$4.56^{+1.96}_{-1.84}$
Combined	325 ± 60	$4.39^{+0.80}_{-0.83}$

Table 10.2.: Result of the χ^2 whether the branching ratios fitted in the individual states are consistent with each other.

	Test statistic	p-value	Reject @95% CL
χ^2	2.78	0.733	no

Table 10.3.: Expected and fitted number of background events for the individual final states.

Final state	Expected	Fitted	Ratio
e^+e^-	292 ± 5	309 ± 24	1.06 ± 0.08
$e^\pm \mu^\mp$	431 ± 6	397 ± 28	0.92 ± 0.07
$e^\pm \pi^\mp$	716 ± 8	719 ± 34	1.00 ± 0.05
$\mu^+\mu^-$	200 ± 4	214 ± 19	1.07 ± 0.10
$\mu^\pm \pi^\mp$	481 ± 6	427 ± 25	0.89 ± 0.05
$\pi^+\pi^-$	182 ± 4	137 ± 15	0.75 ± 0.08

10.2. Systematic Uncertainties

10.2.1. Track Reconstruction Efficiency

In [51] the track reconstruction efficiency was studied using the decay chain $D^* \to D^0\pi$, $D^0 \to \pi\pi K_S$, and $K_S \to \pi^+\pi^-$. In order to have a partially reconstructed D^*, one of the pions from the K_S is allowed to be not explicitly reconstructed. Kinematic constraints on the D^*, D^0, and K_S masses are used to recover the un-reconstructed track. By calculating the ratio between the number of events with such a constrained track and the number of events in which the constrained track is actually reconstructed, the tracking efficiency is estimated [51, p. 5-6]. The estimated tracking efficiencies in data and MC are compared and found to be consistent within 0.35% for one track. In the $B^0 \to \tau^+\tau^-$ final state two tracks are reconstructed. Hence, a systematic error of $\pm 0.7\%$ is assigned.

10.2.2. PID Selection

The systematic uncertainties due to the selection efficiency of the lepton identification are estimated according to the studies in [46]. A correction factor together with its uncertainty is determined. The latter is taken as systematic error and results in 2.6%. Similarly the systematic uncertainty due to the efficiency of the particle identification of pions was investigated in [52]. For this, the decay $D^* \to D^0\pi^+$ with the D^0 meson further decaying to $K^-\pi^+$ was reconstructed. Since the kaon and pion from the D^0 decay can be identified by their charge, the selection efficiency of the PID for pions can be estimated in data and MC. The uncertainty is found to be negligible compared to the identification of leptons.

10.2.3. Number of $B\bar{B}$ Pairs

The number of produced $B\bar{B}$ pairs in the Belle data sample is $(771.581 \pm 10.566) \times 10^6$. Its uncertainty contributes as a systematic error of $\pm 1.4\%$ on the branching ratio.

10.2.4. Branching Ratio $\mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0)$

In the calculation of $\mathcal{B}(B^0 \to \tau^+\tau^-)$ the branching ratio of the decay $\Upsilon(4S) \to B^0\bar{B}^0$ enters. Its uncertainty contributes with a systematic error of $\pm 1.2\%$.

10.2.5. Signal Efficiency

The signal reconstruction efficiency is estimated on MC samples with limited statistics. Hence, it is only known with an uncertainty which is stated in Table 7.7. The fit is repeated with the signal efficiencies varied by $\pm 1\sigma$. The difference between the result of the nominal analysis and the ones with the varied efficiencies is taken as systematic uncertainty. It is evaluated to be $\pm 1.5\%$.

10.2.6. τ Branching Ratios

The branching ratios of the decays of the τ lepton have uncertainties themselves. In order to validate a possible influence on $\mathcal{B}(B^0 \to \tau^+\tau^-)$ the branching ratios of $\tau \to e\nu_e\nu_\tau$, $\tau \to \mu\nu_\mu\nu_\tau$, $\tau \to \pi\nu_\tau$, and $\tau \to \rho\nu_\tau$ are successively varied by reweighting the events in the signal MC. Afterwards the fit is repeated using the reweighted signal sample. The differences between the nominal fit and those with the reweighted signal samples are summed in quadrature resulting in a systematic error of $\pm0.3\%$. The same method was used in the search for the decay $B^+ \to \tau^+\nu_\tau$ in [53, 3].

10.2.7. Tag Efficiency Correction

In [44] the tag correction weights are determined. With the application of the weights, an additional source of a systematic uncertainty occurs. The systematic error was estimated in [44] and is $\pm4.5\%$.

10.2.8. Scaling of the $BD\ell\nu_\ell$ Component

Another source of systematic uncertainty has its origin in the normalization of the $BD\ell\nu_\ell$ component, as it is fixed in the fit. To account for it, the fit is repeated with the normalization varied by $\pm1\sigma$. In Table 11.2 the normalizations and their uncertainties are listed. The difference between the nominal and the modified fits is taken as systematic error. It is estimated to be $\pm5.9\%$.

10.2.9. Correction of Background Branching Ratios

Events with a B decay including a $D^- \to K^0\ell^-\bar{\nu}_\ell$ decay are weighted to correct for the discrepancy between simulated and world average branching ratios. The systematic uncertainty due to this reweighting is estimated by varying the weights by $\pm1\sigma$ before the fit templates are generated. The variation is done separately for $\ell = e$ and $\ell = \mu$. Afterwards the data sample is refitted. The difference between the results of the modified and the nominal fit are quadratically summed and taken as a systematic error. The systematic uncertainty is estimated to be $\pm0.3\%$

10.2.10. Histogram PDF Shapes

The shapes of the histogram PDFs for signal and background used in the fit are estimated on MC samples. In order to validate the influence of statistical fluctuations of the underlying samples on $\mathcal{B}(B^0 \to \tau^+\tau^-)$, the content of the histograms are varied in each bin. The variation follows a Poisson distribution with the original bin content as the mean value. Afterwards, the nominal fit is performed and the resulting value for $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is stored. This procedure is repeated 1000 times and the width of the distribution of the fitted $\mathcal{B}(B^0 \to \tau^+\tau^-)$ values is taken as systematic error. In Fig. 10.2 the fit results for the 1000 fits and the fitted Gaussian distribution are shown. The width of the Gaussian function is 0.0002 which corresponds to a systematic error of 4.5%.
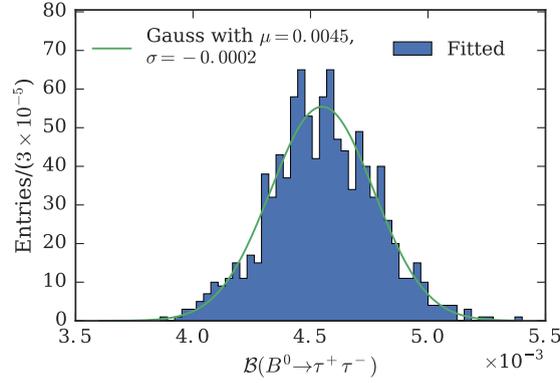
Figure 10.2.: Distribution of the branching ratios (blue) for the different fits with shuffled PDF templates. The green line represents the Gaussian distribution fitted to the histogram.

10.2.11. K_L Veto Weights

The K_L veto weights are fraught with uncertainties. They depend on the momentum of the simulated K_L mesons and are determined in [43]. The systematic error caused by the weights' uncertainties is estimated by varying the weights by $\pm 1\sigma$ before generating the templates and afterwards repeating the fit. The assigned systematic error on $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is the difference between the results of the nominal fit and the modified ones and reaches a total of $\pm 4.3\%$.

Final Result

The final result of $\mathcal{B}(B^0 \to \tau^+\tau^-)$ including all systematic uncertainties is given by

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = (4.39^{+0.80}_{-0.83} \pm 0.45) \times 10^{-3}. \tag{10.2}$$

The first error is statistical and the second systematic.

Table 10.4.: Systematic errors which enter the calculation of the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$. The total systematic error is the square root of the quadratic sum of the individual errors.

Source	Relative uncertainty in %
Track reconstruction efficiency	±0.7
PID selection	±2.6
Number of $B\bar{B}$ pairs	±1.4
$\mathcal{B}(\Upsilon(4S) \to B^0\bar{B}^0)$	±1.2
Signal efficiency	±1.5
τ branching ratios	±0.3
Tag efficiency correction	±4.5
$BD\ell\nu_\ell$ scaling	±5.9
Background branching ratios	±0.3
PDF shape	±4.5
K_L veto weights	±4.3
Total	±10.3

10.3. Significance of the Result

"The significance of the result is the probability of the background-only hypothesis to result in the observed signal-strength, or larger" also known as p-value [19, p. 130]. To calculate it the ratio of the maximum likelihood of the background-only hypothesis $\mathcal{L}_0$ and the maximum likelihood of the signal-plus-background hypothesis $\mathcal{L}_{\max}$ is used as test statistic. It is written as

$$\lambda_0 = \log \frac{\mathcal{L}_0}{\mathcal{L}_{\max}}. \tag{10.3}$$

Wilks' theorem [54] implies that λ_0 follows a χ^2 distribution with one degree-of-freedom (dof) for each parameter of interest. In this particular case, the only parameter of interest is the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$, hence λ_0 follows a χ^2 distribution with one dof. Commonly, the significance Σ is expressed in Gaussian standard deviations and can be calculated as

$$\Sigma = \sqrt{-2\log \frac{\mathcal{L}_0}{\mathcal{L}_{\max}}} = \sqrt{2\Delta\mathcal{L}}. \tag{10.4}$$

In order to account for the systematic uncertainties the profile likelihood function is convolved with a Gaussian function with the width equal to the systematic error. The fit is repeated with the branching ratio fixed to 10000 values in the interval $[-0.1 \times 10^{-3}, 5 \times 10^{-3}]$ in order to determine the profile likelihood curve. In order to get the significance of discarding the hypothesis $B^0 \to \tau^+\tau^-$ events to be zero, only the systematic uncertainties related to the signal yield are taken into account - namely those of the $BD\ell\nu_\ell$ scaling, the background branching ratios, the PDF shape, and the K_L veto weights. The curves before and after the convolution are shown in Fig. 10.3. The significance of the branching ratio including all systematic uncertainties is 5.0σ.
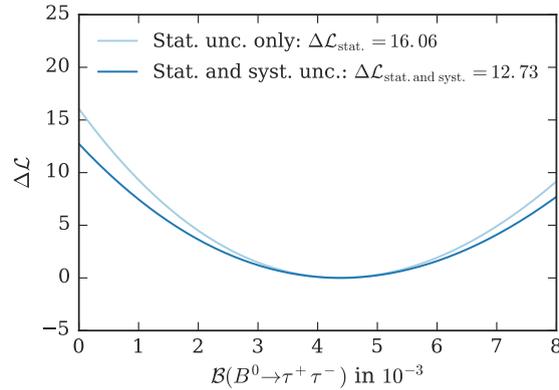
Figure 10.3.: Profile likelihood ratios $\Delta\mathcal{L}$ for $\mathcal{B}(B^0 \to \tau^+\tau^-)$. The minimum is at the nomninal fit result of $\mathcal{B}(B^0 \to \tau^+\tau^-) = 4.39 \times 10^{-3}$.

11. Cross-Checks

This analysis has been performed as a blind analysis, i.e. the selection, optimization, and fit procedure have been developed on MC events. Assuming the SM value for the branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$, an expected upper limit of $\mathcal{B}_{\mathrm{ul}} = 2.3 \times 10^{-3}$ is estimated on fits on MC. However, the fit to data yields a significant signal component and a resulting branching ratio of $\mathcal{B}(B^0 \to \tau^+\tau^-) = (4.39^{+0.80}_{-0.83}) \times 10^{-3}$. The following additional cross-checks are performed to help understand the observed excess.

11.1. Influence of $B^0 \to K^0\tau^+\tau^-$ Decays

A possible background is the decay $B^0 \to K^0\tau^+\tau^-$, in which the K^0 can be either in the K_S or K_L state. If the kaon in the decay is not reconstructed, the final state particles are the same as for the signal decay $B^0 \to \tau^+\tau^-$. After the final selection has been applied on the background MC samples, no $B^0 \to K^0\tau^+\tau^-$ event is selected. However, the branching ratio of $B^0 \to K^0\tau^+\tau^-$ is not measured yet. In the rare MC sample it is simulated using the branching ratio $\mathcal{B}(B^0 \to K^0\tau^+\tau^-)_{\mathrm{sim}} = 1.3 \times 10^{-7}$, which was predicted in [55]. Hence, the number of simulated $B^0 \to K^0\tau^+\tau^-$ events is about 5000. To have a statistically significant conclusion, an additional sample of 5 million $\Upsilon(4S) \to B^0\bar{B}^0$ events, where one B^0 meson decays via a $b \to c$ transition and the other via $B^0 \to K^0\tau^+\tau^-$, is generated. $B^0 \to K^0\tau^+\tau^-$ is reconstructed and the same selection as for $B^0 \to \tau^+\tau^-$ is applied. The resulting distributions of the E_{ECL} variable are shown in Fig. 11.1. The number of selected events in each final state are listed in Table 11.1. A larger branching ratio of the decay $B^0 \to K^0\tau^+\tau^-$ can have an influence on the signal component, but the enhancement has to be on the order of 50000 or larger.

11.2. Constraining the Background Normalization

In the nominal fit, the fitted background normalizations are close to the expected ones. Only in the $\pi^+\pi^-$ final state the fitted value exhibits a larger deviation from the expectation.

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.

(f) Final state $\pi^+\pi^-$.
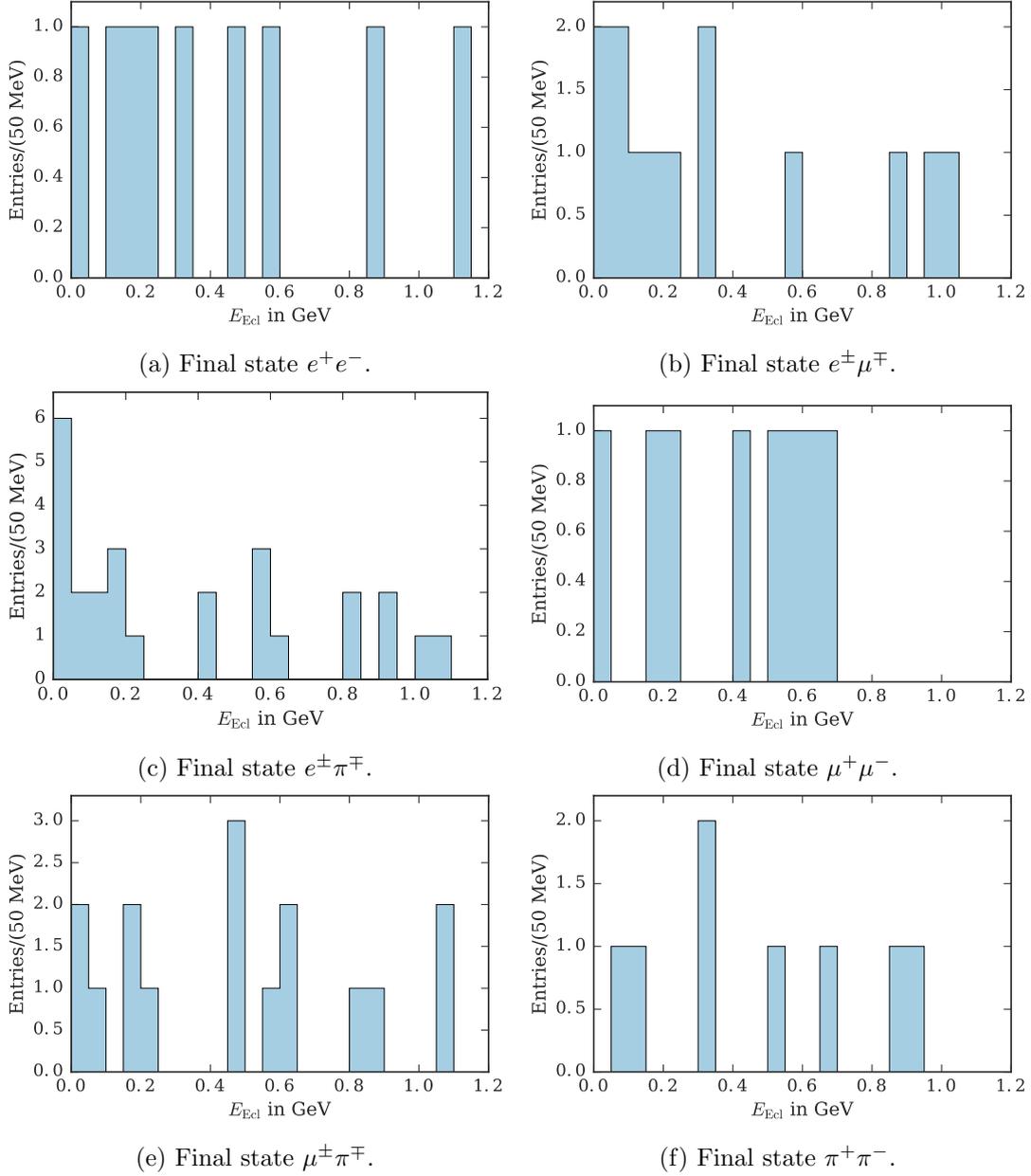
Figure 11.1.: E_{ECL} distributions of the reconstructed $B^0 \to K^0\tau^+\tau^-$ sample after the final selection. The number of simulated events corresponds to a branching ratio of 50000 times the expectation in [55].

Table 11.1.: Number of selected $B^0 \to K^0\tau^+\tau^-$ events in the different final states.

Final State	Selected events	Scaled to SM
e^+e^-	9	0.00018
$e^\pm\mu^\mp$	13	0.00026
$e^\pm\pi^\mp$	26	0.00052
$\mu^+\mu^-$	8	0.00016
$\mu^\pm\pi^\mp$	16	0.00032
$\pi^+\pi^-$	8	0.00016

For testing the influence of the background normalization on the signal component, the fitting procedure is slightly modified. The background normalizations of the individual final states are constrained with a common factor f_{bkg}. This factor scales the expected normalizations in the fit and is defined in

$$N_{\mathrm{bkg},i}^{\mathrm{fit}} = f_{\mathrm{bkg}} \cdot N_{\mathrm{bkg},i}^{\mathrm{exp}}. \tag{11.1}$$

Here, the fitted and expected background normalizations are denoted with $N_{\mathrm{bkg},i}^{\mathrm{fit}}$ and $N_{\mathrm{bkg},i}^{\mathrm{exp}}$, respectively. The normalization of the $BD\ell\nu_\ell$ component remains fixed in the fit. Using the modified fit procedure, the fitted branching ratio is

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = (3.98 \pm 0.80) \times 10^{-3}. \tag{11.2}$$

The background scaling factor estimated in the fit is

$$f_{\mathrm{bkg}} = 0.962 \pm 0.033. \tag{11.3}$$

The branching ratio has a significance, including only the statistical uncertainties, of 5.2σ. Constraining the background normalization yields a significant signal comparable with the one of the nominal fit. The fitted value of the scaling factor f_{bkg}, which is in good agreement with one, confirms the good understanding of the background components.

11.3. Investigation of the $BD\ell\nu_\ell$ Component

As previously stated (Section 7.5), the $B^0 \to D^-(K_L\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$ decays are not only the dominant background component but they also peak at zero in the E_{ECL} distribution; thus, a large $BD\ell\nu_\ell$ sample is used to determine the PDF shape. In order to validate the influence of the $BD\ell\nu_\ell$ component on the fit result, several checks are performed.

11.3.1. Variation of the Scaling

The normalizations of the $BD\ell\nu_\ell$ component used in the fit are defined in Eq. (6.2) and listed in Table 11.2 with their uncertainty for the six final states. In Table 11.3, the uncertainty of the normalization is split into the different sources. The largest uncertainty

Table 11.2.: Normalization of the $BD\ell\nu_\ell$ component for the different final states. The uncertainty of the normalization comes from the uncertainties on the branching ratios and $N_{B\bar{B}}$ used to calculate $f_{BD\ell\nu_\ell}$ and the tagging efficiency correction.

Final state	$N_{BD\ell\nu_\ell}$	$\Delta N_{BD\ell\nu_\ell}$
e^+e^-	205.2	17.9
$e^\pm\mu^\mp$	302.3	26.3
$e^\pm\pi^\mp$	72.7	6.3
$\mu^+\mu^-$	124.3	10.8
$\mu^\pm\pi^\mp$	41.0	3.6
$\pi^+\pi^-$	3.9	0.3

Table 11.3.: Relative uncertainty of the $BD\ell\nu_\ell$ normalization split into different sources.

Source	Rel. uncertainty in %
$N_{B\bar{B}}$	1.4
$(\Upsilon(4S) \to B^0\bar{B}^0)$	1.2
$\mathcal{B}(B^0 \to D^-\ell^+\nu_\ell)$	5.5
$\mathcal{B}(D^- \to K^0 e^-\bar{\nu}_e)$	1.2
$\mathcal{B}(D^- \to K^0 \mu^-\bar{\nu}_e)$	3.9
Tag correction	4.5
PID Selection	2.6
Track reconstruction	0.7
Total	8.7

Table 11.4.: Results of the fits with varied normalization of the $BD\ell\nu_\ell$ component.

	$\mathcal{B}(B^0 \to \tau^+\tau^-)$ in 10^{-3}
$+\Delta N_{BD\ell\nu_\ell}$	$4.14^{+0.80}_{-0.82}$
$-\Delta N_{BD\ell\nu_\ell}$	$4.64^{+0.81}_{-0.83}$

Table 11.5.: Expected and fitted normalizations of the $BD\ell\nu_\ell$ component as a result of floating the $BD\ell\nu_\ell$ normalization in the fit.

Final state	Expected $N_{BD\ell\nu_\ell}$	Fitted $N_{BD\ell\nu_\ell}$	Ratio
e^+e^-	205 ± 14	$(2.1 \pm 1.0) \times 10^2$	1.0 ± 0.5
$e^\pm\mu^\mp$	302 ± 21	$(5.2 \pm 1.1) \times 10^2$	1.7 ± 0.4
$e^\pm\pi^\mp$	73 ± 5	$(9 \pm 9) \times 10^1$	1.2 ± 1.2
$\mu^+\mu^-$	124 ± 9	$(2.3 \pm 0.5) \times 10^2$	1.8 ± 0.4
$\mu^\pm\pi^\mp$	41.0 ± 2.9	$(1.0 \pm 0.5) \times 10^2$	2.3 ± 1.3
$\pi^+\pi^-$	3.91 ± 0.28	20 ± 17	5 ± 4

comes from the branching ratio of the decay $B^0 \to D^-\ell^+\nu_\ell$ and the tag correction. For each final state, the relative uncertainty of the normalization is the same, as it has its origin solely in the scaling factor.

The fit is repeated with the $BD\ell\nu_\ell$ normalizations varied once with $+\Delta N_{BD\ell\nu_\ell}$ in all final states and once with $-\Delta N_{BD\ell\nu_\ell}$. This is also a systematic uncertainty and is described in Section 10.2.8. The fit results are listed in Table 11.4 and the resulting E_{ECL} distributions are shown in Fig. 11.2 (Fig. 11.3) for the upwards (downwards) variation of the scaling. The upscaling of the $BD\ell\nu_\ell$ background by one standard deviation still yields a significant signal component and deviates from the nominal fit result by 5.9%.

11.3.2. Floating Normalization

In another cross-check, the normalizations of the $BD\ell\nu_\ell$ component are floated in the fit in each final state separately, analogous to N_{bkg}. Its result is

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = (2.6 \pm 1.6) \times 10^{-3}. \tag{11.4}$$

The shape of the template of the $BD\ell\nu_\ell$ component is very signal-like. Hence, floating the normalization in the fit gives a much smaller signal component, and, in addition, the statistical uncertainty increases. In Fig. 11.4, the E_{ECL} distributions after the fit for the individual final states are shown. In Table 11.5, the expected and fitted normalizations of the $BD\ell\nu_\ell$ component and their ratios are shown. The fitted values of the $BD\ell\nu_\ell$ normalization have large uncertainties and vary strongly in the six final states. Not only the signal component gets eaten by the $BD\ell\nu_\ell$ background, but the normalization of the other backgrounds is also smaller for some of the final states, e.g. $e^\pm\mu^\mp$. The expected and

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.

(f) Final state $\pi^+\pi^-$.
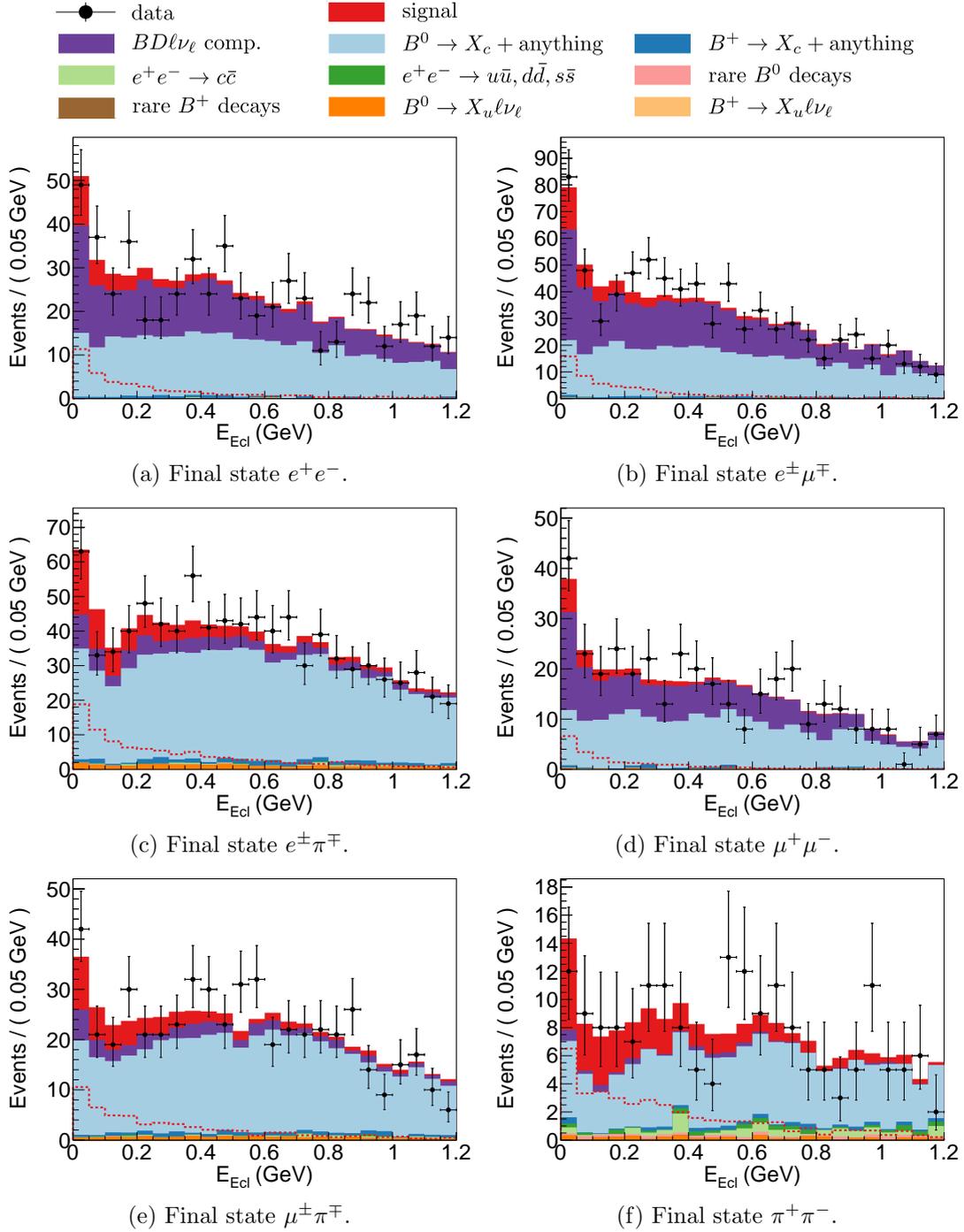
Figure 11.2.: Resulting E_{ECL} distributions of the fit with the $BD\ell\nu_\ell$ normalization plus one sigma.

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.

(f) Final state $\pi^+\pi^-$.
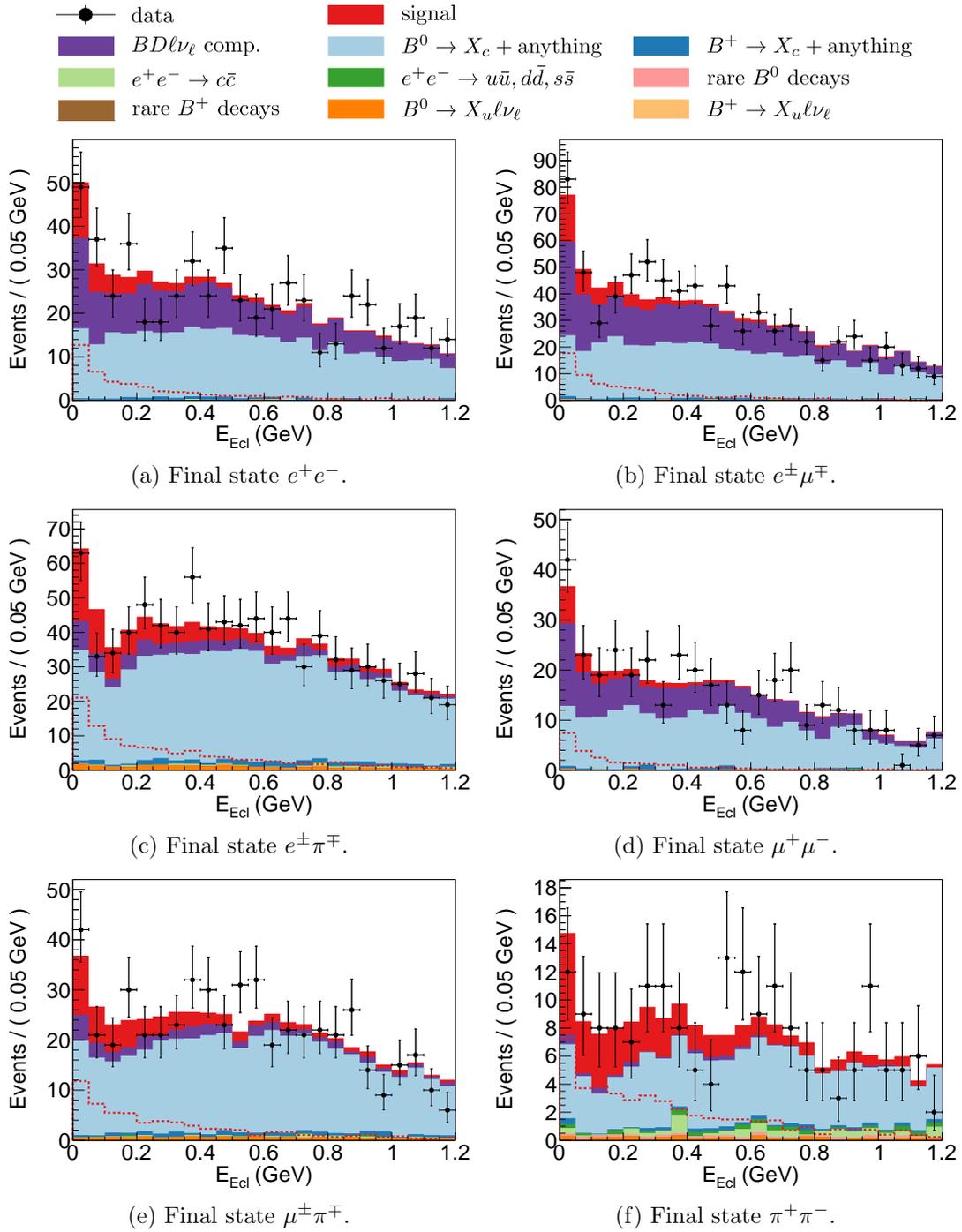
Figure 11.3.: Resulting E_{ECL} distributions of the fit with the $BD\ell\nu_\ell$ normalization minus one sigma.

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.

(f) Final state $\pi^+\pi^-$.
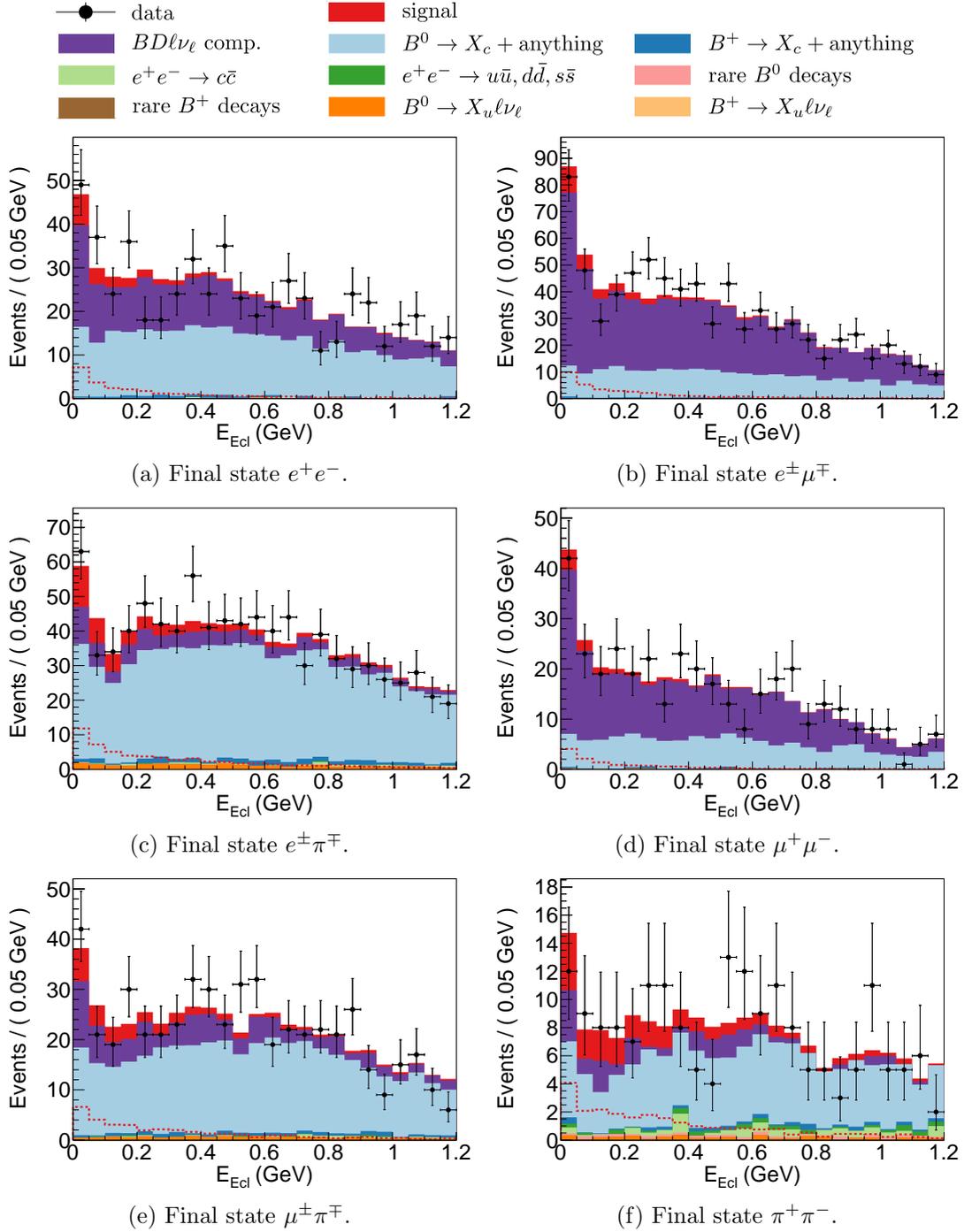
Figure 11.4.: Resulting E_{ECL} distributions of the fit with the $BD\ell\nu_\ell$ normalization floating in the fit.

Table 11.6.: Expected and fitted normalizations of the background components (other than $BD\ell\nu_\ell$) as a result of floating the $BD\ell\nu_\ell$ normalization in the fit.

Final state	Expected N_{bkg}	Fitted N_{bkg}	Ratio
e^+e^-	292 ± 5	$(3.2 \pm 0.9) \times 10^2$	1.10 ± 0.31
$e^\pm\mu^\mp$	431 ± 6	$(2.1 \pm 1.0) \times 10^2$	0.50 ± 0.22
$e^\pm\pi^\mp$	716 ± 8	$(7.4 \pm 0.7) \times 10^2$	1.04 ± 0.10
$\mu^+\mu^-$	200 ± 4	$(1.2 \pm 0.5) \times 10^2$	0.62 ± 0.24
$\mu^\pm\pi^\mp$	481 ± 6	$(4.0 \pm 0.5) \times 10^2$	0.82 ± 0.10
$\pi^+\pi^-$	182 ± 4	138 ± 19	0.76 ± 0.11

fitted numbers of background events are listed in Table 11.6. Due to the large variations of the ratios fitted to expected normalization in the six final states, no clear hint exists that the $BD\ell\nu_\ell$ background is a reasonable explanation for the signal component.

11.3.3. Constraining the $BD\ell\nu_\ell$ Normalization

Additionally, the fit procedure is modified in such a way that the normalization of the $BD\ell\nu_\ell$ component in each final state is constrained by a common factor $f^{\text{fit}}_{BD\ell\nu_\ell}$. The normalization used in the fit is defined as

$$N^{\text{fit,i}}_{BD\ell\nu_\ell} = f^{\text{fit}}_{BD\ell\nu_\ell} \cdot N^i_{BD\ell\nu_\ell}, \tag{11.5}$$

with expected normalization $N^i_{BD\ell\nu_\ell}$ in the final state i. The factor $f^{\text{fit}}_{BD\ell\nu_\ell}$ is floated in the fit. Using this modified fit procedure $\mathcal{B}(B^0 \to \tau^+\tau^-)$ results in

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = (2.8 \pm 1.1) \times 10^{-3}, \tag{11.6}$$

and $f^{\text{fit}}_{BD\ell\nu_\ell}$ is

$$f^{\text{fit}}_{BD\ell\nu_\ell} = 1.56 \pm 0.26. \tag{11.7}$$

The fitted E_{ECL} distributions are depicted in Fig. 11.5 and the expected and fitted numbers of background events are tabulated in Table 11.7. The uncertainty on the calculated scaling is 8.7% (Table 11.3). Hence, this uncertainty is too small to justify a factor of 1.56.

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.
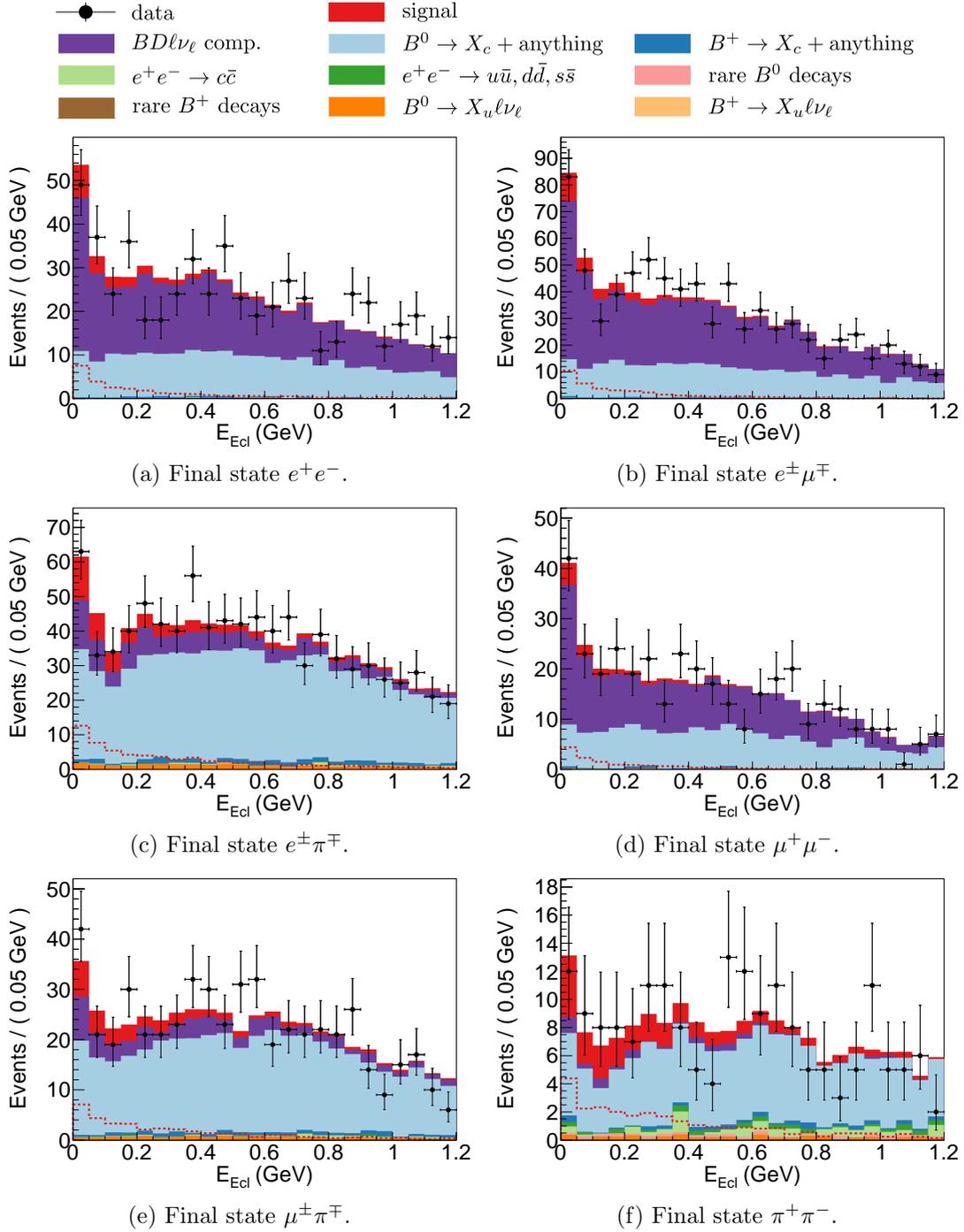
(f) Final state $\pi^+\pi^-$.

Figure 11.5.: Resulting E_{ECL} distributions of the fit with the common scaling factor $f^{\mathrm{fit}}_{BD\ell\nu_\ell}$ for the $BD\ell\nu_\ell$ component in the different final states.

Table 11.7.: Expected and fitted normalizations of the background components (other than $BD\ell\nu_\ell$) as a result of using a common scaling factor for the $BD\ell\nu_\ell$ normalization in each final state.

Final state	Expected N_{bkg}	Fitted N_{bkg}	Ratio
e^+e^-	292 ± 5	$(2.1 \pm 0.5) \times 10^2$	0.74 ± 0.18
$e^\pm\mu^\mp$	431 ± 6	$(2.5 \pm 0.7) \times 10^2$	0.59 ± 0.17
$e^\pm\pi^\mp$	716 ± 8	714 ± 34	1.00 ± 0.05
$\mu^+\mu^-$	200 ± 4	156 ± 33	0.78 ± 0.17
$\mu^\pm\pi^\mp$	481 ± 6	425 ± 25	0.88 ± 0.05
$\pi^+\pi^-$	182 ± 4	149 ± 16	0.82 ± 0.09

11.4. $BD\ell\nu_\ell$-Enriched Sample

In the three-body decay $B^0 \to D^-\ell^+\nu_\ell$ the charged lepton on average has a higher momentum than a charged lepton in the final state of a $B^0 \to \tau^+\tau^-$ decay. To obtain a sample which is enriched with $BD\ell\nu_\ell$ decays the sample is split into subsamples depending on the product of the charge of the daughter with the higher momentum of the B_{sig} and the flavor of the B_{tag} candidate. This product is denoted as `tag_times_charge`. The flavor of the B_{tag} candidate is $+1$ or -1 if it is reconstructed as a B^0 or a $\bar{B}^0$, respectively. Assuming no CP violation, the time-integrated mixing probability for the B^0 meson is $\chi_d = 0.1874 \pm 0.0018$ [17]. Hence, with knowledge of the flavor of the B_{tag} candidate the flavor of the B_{sig} candidate is mostly known. If the B_{tag} candidate is reconstructed as a B^0 ($\bar{B}^0$), the B_{sig} candidate is a $\bar{B}^0$ (B^0). Therefore, the charge of the lepton is -1 ($+1$) and with it `tag_times_charge` is -1 (-1). The fractions of the events in each subsample for the dominant MC types are shown in Fig. 11.6. Signal MC is represented with an equal amount of events in each subsample. The fit is performed on each subsample and the results are tabulated in Table 11.8. The fitted branching ratios in the two samples are statistically compatible with each other and with the result of the nominal fit.
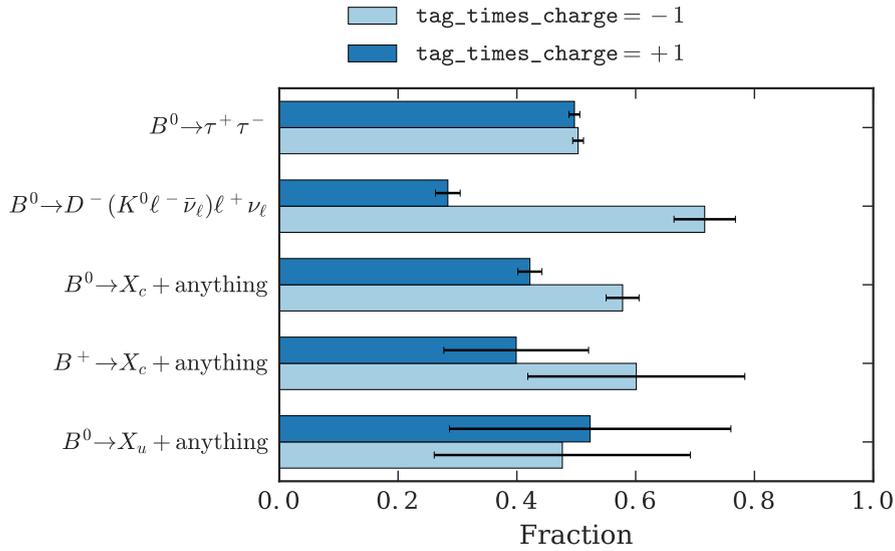
Figure 11.6.: Fractions of the events of the different MC types in the subsamples for `tag_times_charge` ± 1.

Table 11.8.: Fitted $\mathcal{B}(B^0 \to \tau^+\tau^-)$ for the subsamples `tag_times_charge` $= \pm 1$.

`tag_times_charge`	$\mathcal{B}(B^0 \to \tau^+\tau^-)$ in 10^{-3}	Significance
$+1$	$3.9^{+1.1}_{-1.0}$	4.0σ
-1	5.1 ± 1.2	4.5σ

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.

(f) Final state $\pi^+\pi^-$.
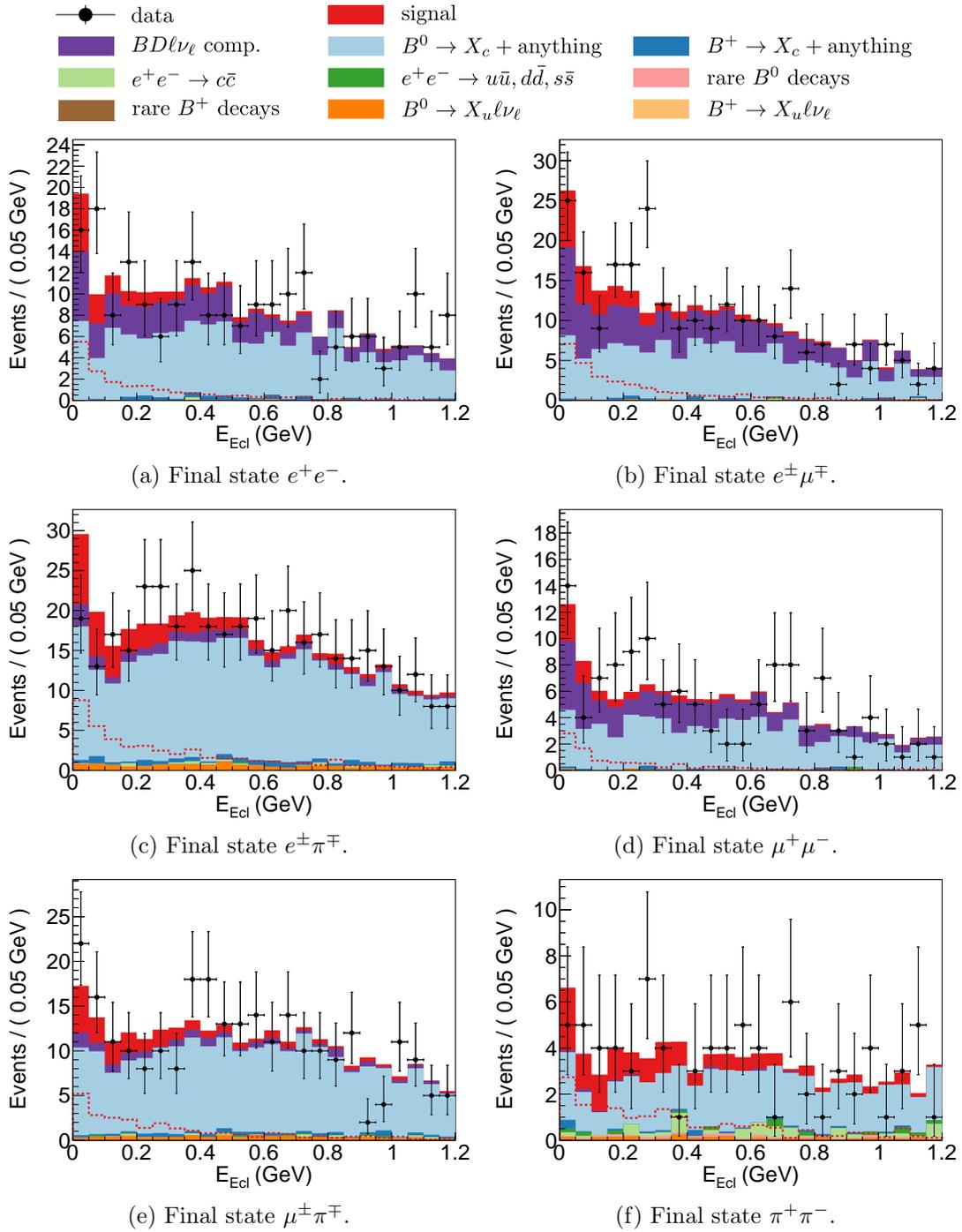
Figure 11.7.: Resulting E_{ECL} distributions of the fit on the sample with a decreased $BD\ell\nu_\ell$ component (`tag_times_charge` $= +1$).

(a) Final state e^+e^-.

(b) Final state $e^\pm\mu^\mp$.

(c) Final state $e^\pm\pi^\mp$.

(d) Final state $\mu^+\mu^-$.

(e) Final state $\mu^\pm\pi^\mp$.

(f) Final state $\pi^+\pi^-$.
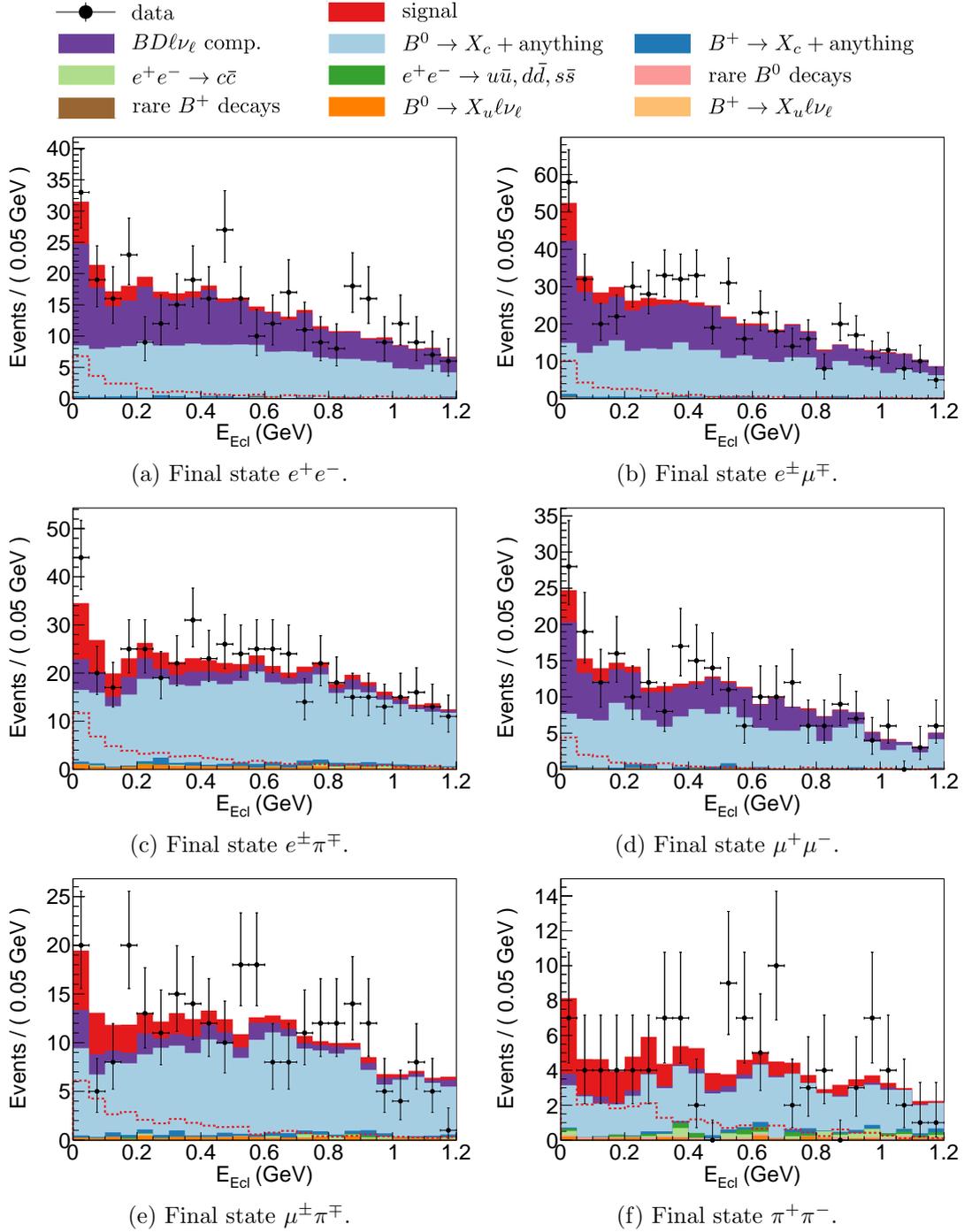
Figure 11.8.: Resulting E_{ECL} distributions of the fit on the sample enriched with $BD\ell\nu_\ell$ events (`tag_times_charge` $= -1$).

11.5. Data-MC Comparison

In Section 8.2, data and MC sideband samples are compared to each other and found to be consistent within the statistical uncertainty. In order to test whether the $B^0 \to \tau^+\tau^-$ model describes the excess in data, data and MC distributions of the variables used in the networks are shown after the final selection and without additional sideband selections. For MC, the simulated background sample plus the signal sample using the normalization from the nominal fit are used.

In Figs. 11.9 to 11.14, the distributions for the three most important features for the individual final states are depicted. The distributions for the remaining input variables can be found in Section C.1. All variables used in the trainings and E_{ECL} are depicted with only the preselection applied in Section C.2. The distributions of data and MC show a good agreement after both, the preselection and the final selection. Hence, the $B^0 \to \tau^+\tau^-$ decays are a possible candidate for the excess in data.

(a) $\mathcal{N}_{\mathrm{tag}}$ distribution in the final state e^+e^-.

(b) Transverse momentum distribution in the final state e^+e^-.

(c) Pseudo helicity angle in the final state e^+e^-.
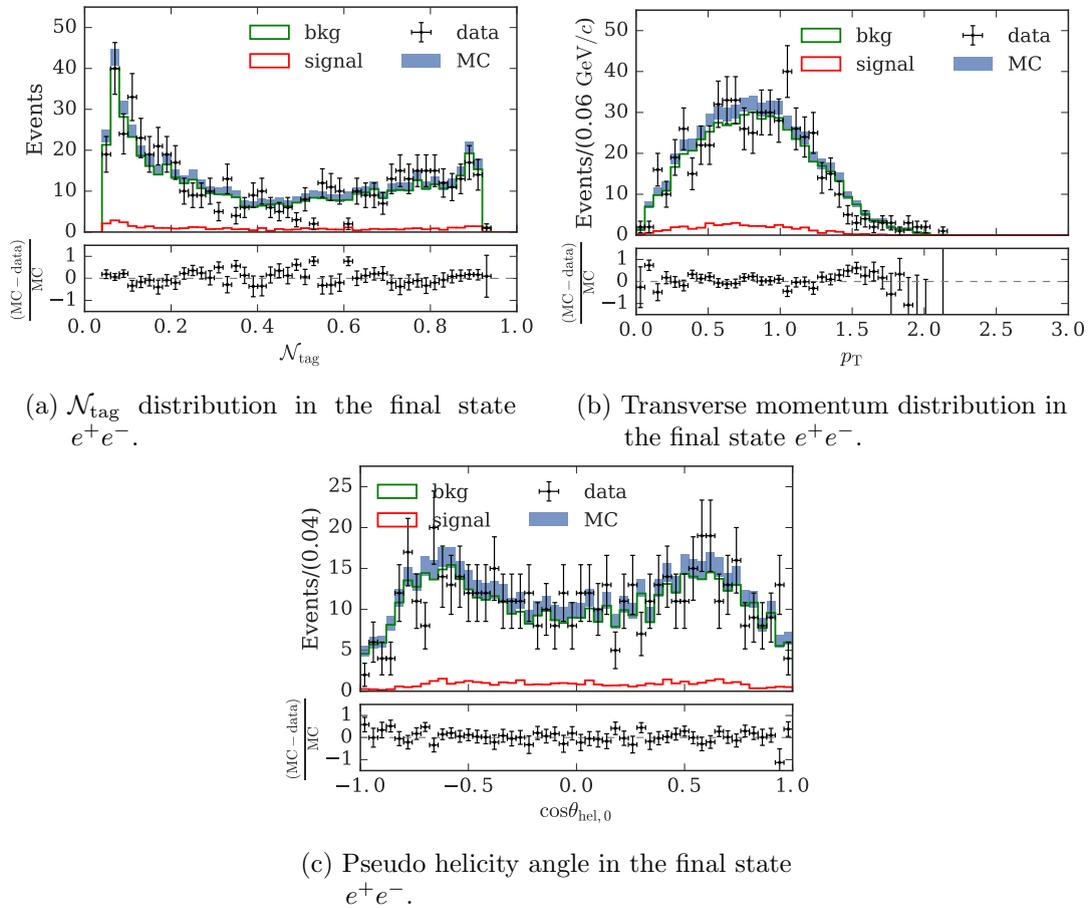
Figure 11.9.: Distributions of the three most important input variables used in the neural network in the final state e^+e^-.
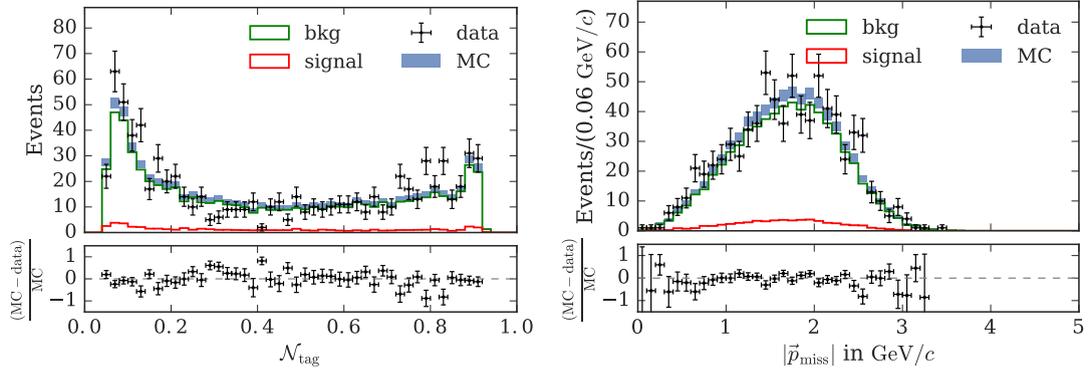
(a) $\mathcal{N}_{\text{tag}}$ distribution in the final state $e^{\pm}\mu^{\mp}$.

(b) Missing momentum distribution in the final state $e^{\pm}\mu^{\mp}$.
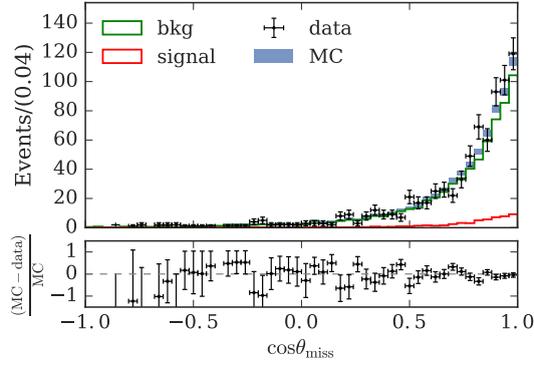
(c) Polar angle of the missing momentum vector in the final state $e^{\pm}\mu^{\mp}$.

Figure 11.10.: Distributions of the three most important input variables used in the neural network in the final state $e^{\pm}\mu^{\mp}$.

(a) Angle between daughter 0, with hypothesis pion, and the tau in the B_sig rest frame.

(b) Angle between daughter 1, with hypothesis pion, and the tau in the B_sig rest frame.

(c) $\mathcal{N}_\mathrm{tag}$ distribution in the final state $e^\pm\pi^\mp$.
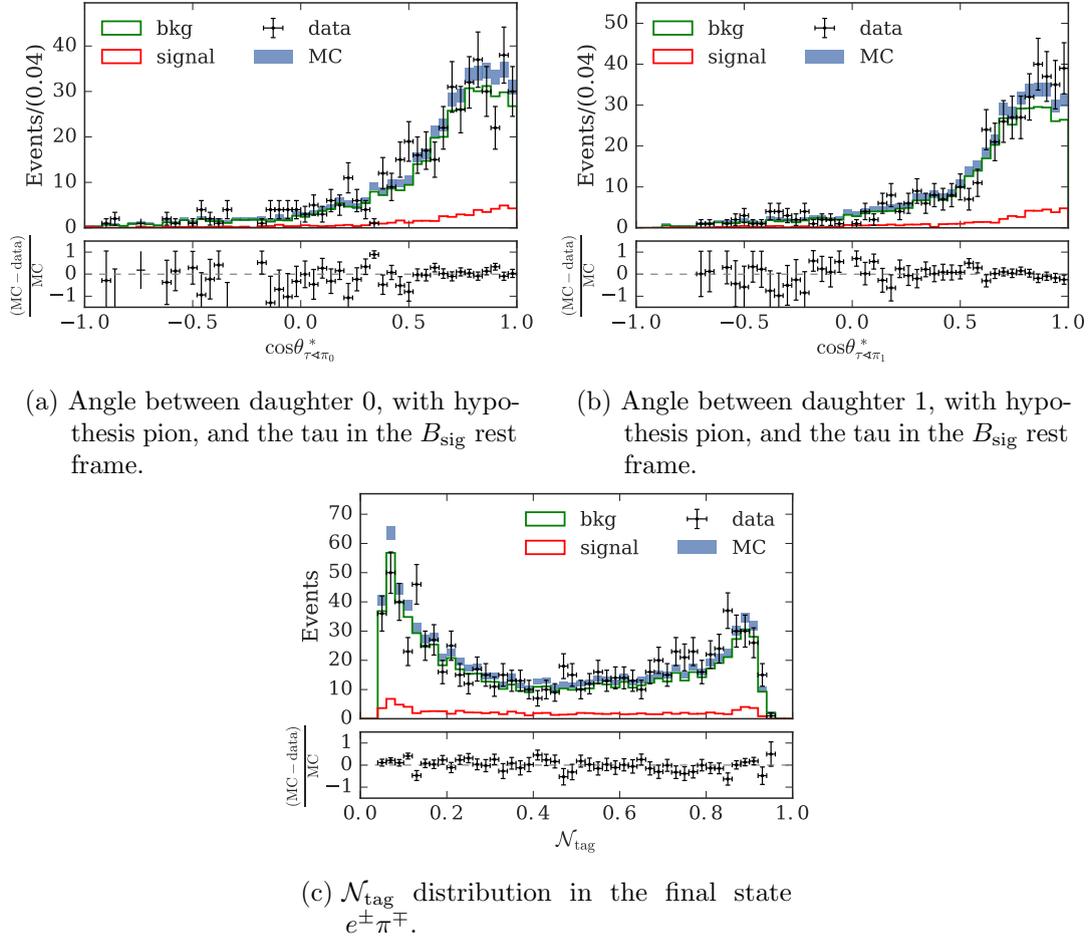
Figure 11.11.: Distributions of the three most important input variables used in the neural network in the final state $e^\pm\pi^\mp$.
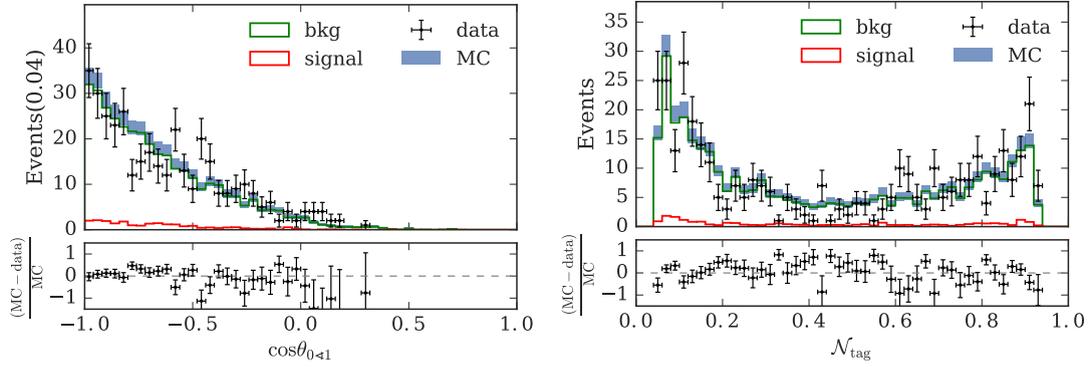
(a) Angle between both muons in the final state $\mu^+\mu^-$.

(b) $\mathcal{N}_{\text{tag}}$ distribution in the final state $\mu^+\mu^-$.
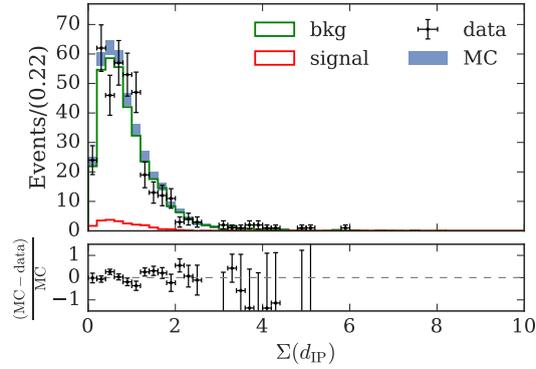
(c) Significance of the distance of the vertex to the IP in the final state $\mu^+\mu^-$.

Figure 11.12.: Distributions of the three most important input variables used in the neural network in the final state $\mu^+\mu^-$.
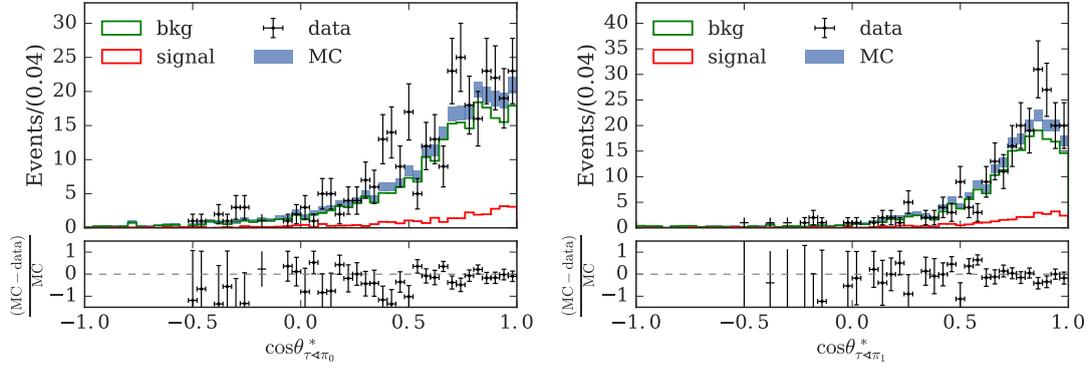
(a) Angle between daughter 0, with hypothesis pion, and the tau in the B_{sig} rest frame.

(b) Angle between daughter 1, with hypothesis pion, and the tau in the B_{sig} rest frame.
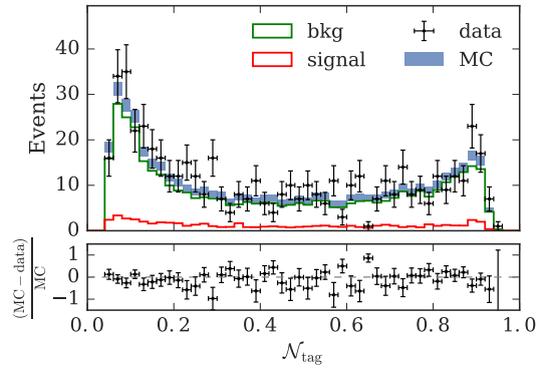
(c) $\mathcal{N}_{\text{tag}}$ distribution in the final state $\mu^{\pm}\pi^{\mp}$.

Figure 11.13.: Distributions of the three most important input variables used in the neural network in the final state $\mu^{\pm}\pi^{\mp}$.
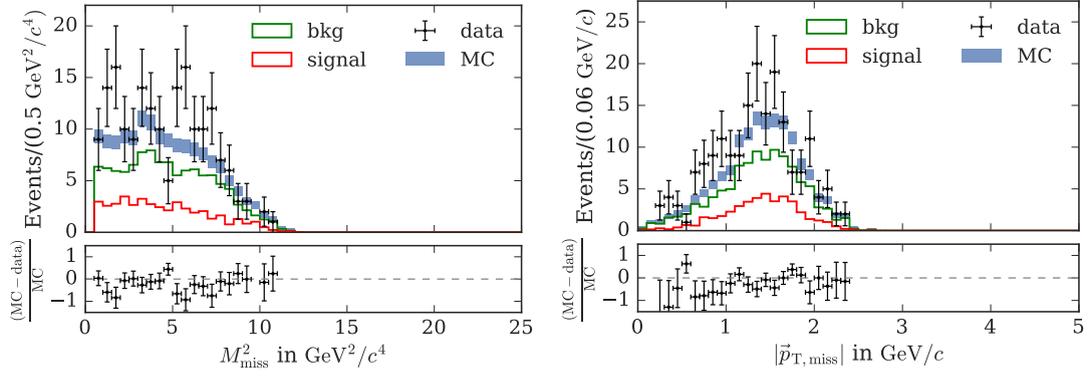
(a) Missing mass squared in the final state $\pi^+\pi^-$.

(b) Missing transverse momentum distribution in the final state $\pi^+\pi^-$.
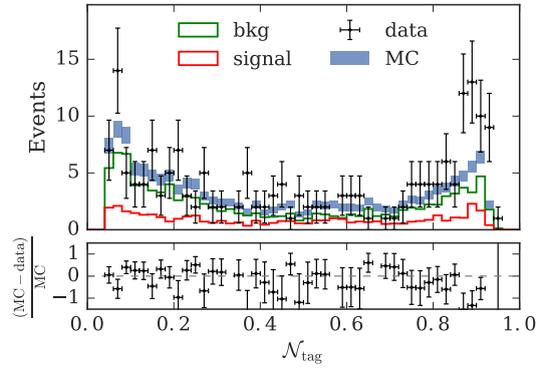
(c) $\mathcal{N}_{\text{tag}}$ distribution in the final state $\pi^+\pi^-$.

Figure 11.14.: Distributions of the three most important input variables used in the neural network in the final state $\pi^+\pi^-$.

11.6. Conclusion of the Cross-Checks

Depending on the treatment of the $BD\ell\nu_\ell$ component in the fit, the size of the signal component changes immensely. With the $BD\ell\nu_\ell$ normalization as a free parameter in the fit, a 40% smaller branching ratio is fitted. Even with the constraint of a common scaling factor on the normalization in each final state, the change in the branching ratio is 36%. The different ratios between expected and fitted $BD\ell\nu_\ell$ normalizations in the six final states and the large common scaling factor cannot be explained by the uncertainties of the contributing branching ratios. Additionally, the normalization factors of the other background components deviate from their expectation.

In both the $BD\ell\nu_\ell$ enriched and $BD\ell\nu_\ell$ deprived sample, a significant signal is measured. The combined significance is even larger as the one of the nominal result. These results show that the $BD\ell\nu_\ell$ component alone cannot be a consistent explanation for the excess in data without relinquishing the trust in the MC itself.

Comparing the distributions of the variables used as input for the NeuroBayes neural networks for data and MC does not give an indication for a process which causes the large signal component.

12. Summary and Conclusion

In this thesis, the search for the decay $B^0 \to \tau^+\tau^-$ is presented. For the first time, the search for the decay $B^0 \to \tau^+\tau^-$ is performed on the complete data set recorded with the Belle detector at the $\Upsilon(4S)$ resonance at the KEKB collider containing 772 millions $B\bar{B}$ pairs.

The Full Reconstruction algorithm is used to reconstruct the B_{tag} candidate in a fully hadronic mode. With the remaining tracks and calorimeter entries not used for the B_{tag} recombination, the signal candidate B_{sig} is reconstructed in the $B^0 \to \tau^+\tau^-$ mode. The B_{sig} candidates are divided into six subsamples according to their final state. The large amount of background is suppressed using a NeuroBayes neural network for each final state separately. The dominant background process $B^0 \to D^-\ell^+\nu_\ell$ with the D^- decaying into $K^0\ell^-\bar{\nu}_\ell$ (with $\ell = e, \mu$) is simulated in a separate sample and the normalization is determined using the best known values of the corresponding branching ratios. The search is performed as a blind analysis and is validated on sideband selection samples in which no signal is expected. The data distributions in the sideband samples are well-described by the background Monte Carlo samples. The branching ratio $\mathcal{B}(B^0 \to \tau^+\tau^-)$ is extracted with a simultaneous extended maximum likelihood fit in all final states. The expected upper limit for $\mathcal{B}(B^0 \to \tau^+\tau^-)$ at the 90% confidence level is estimated to be $\mathcal{B}_{\mathrm{ul}} < 2.3 \cdot 10^{-3}$. The expected limit is five orders of magnitude larger than the Standard Model expectation of $(2.22 \pm 0.19) \times 10^{-8}$. The measurement performed on experimental data yields a signal excess. This excess can only be interpreted as $B^0 \to \tau^+\tau^-$ decays, if all background components are exactly as in the Standard Model. Under this assumption the size of the signal component can be expressed as the branching ratio for the decay $B^0 \to \tau^+\tau^-$ of

$$\mathcal{B}(B^0 \to \tau^+\tau^-) = (4.39^{+0.80}_{-0.83} \pm 0.45) \times 10^{-3},$$

where the first and second error denote the statistical and the systematic error, respectively.

The observation of a significant signal five orders of magnitude larger than the Standard Model prediction was not expected. Although the selection procedure was carefully validated

before the unblinding, due to the unexpected and significant excess in data, additional cross-checks were performed in order to understand the origin of the large excess in data.

The fitting procedure was modified to test the influence of the dominant background component. For events with the decay chain $B^0 \to D^-(\to K^0 \ell'^- \bar{\nu}_{\ell'})\ell^+ \nu_\ell$ the normalization constant was used as a free parameter in the fit. As a result, the fitted branching ratio was smaller and no longer significant. However, at the same time the fitted normalizations of the other background processes in the different final states moved away from their expectation. Additionally, the ratio between the fitted and expected normalization of the $B^0 \to D^-(\to K^0 \ell'^- \bar{\nu}_{\ell'})\ell^+ \nu_\ell$ component in the modified fit was different in the final states. If constrained to a common factor, it was more than 5σ larger than the expected normalization. Besides the examination of the background component the influence of B decays with the same topology, namely two oppositely charged tracks and missing energy, like $B^0 \to K_L \tau^+ \tau^-$ was examined. The enhancement of its predicted branching ratio must be of the same order of magnitude as the enhancement of $\mathcal{B}(B^0 \to \tau^+ \tau^-)$, in order to explain the excess. In two separate fits to a $B^0 \to D^-(\to K^0 \ell'^- \bar{\nu}_{\ell'})\ell^+ \nu_\ell$ enriched and deprived sample, two significant branching ratios were fitted with a combined significance larger than the one of the nominal result. All cross-checks did not show a clear indication of the origin of the excess.

In summary, the analysis presented within the scope of this work yields a significant excess in B^0 decays with the topology of two oppositely charged tracks and missing energy. However, the analysis does not conclude that the origin of the excess in data is caused by $B^0 \to \tau^+ \tau^-$ decays.

In order to clarify the origin of the excess and the underlying process other independent searches for the decay $B^0 \to \tau^+ \tau^-$ must be performed at future experiments like Belle II at the SuperKEKB collider. The design luminosity of the SuperKEKB is 40 times larger than the luminosity of the KEKB collider. Therefore, Belle II will record a data sample as large as the complete Belle data set in one to two years. Then the results of the search for $B^0 \to \tau^+ \tau^-$ at Belle II may shed light on the situation.

For now, the current analysis procedure can be extended by reconstructing τ leptons in additional decay channels like three-prong decays in order to exclude background decays with two charged tracks and missing energy as the detector signature. A signal in these final states would strengthen the hypothesis that the excess observed in this thesis is caused by $B^0 \to \tau^+ \tau^-$ decays.

A. Network Information

Table A.1.: Variables used in the network for the final state e^+e^- sorted by their rank.

Variable	Added	Without	Only	Global correlation	Rank
$\mathcal{N}_{\text{tag}}$	17.43	14.35	17.43	0.14	1
p_{T}	14.42	3.38	15.34	0.89	2
$\cos\theta_{\text{hel},0}$	12.51	7.42	15.52	0.73	3
$M_{\text{bc}}^{\text{tag}}$	10.21	9.37	10.86	0.08	4
$|\vec{p}_{\text{miss}}|$	7.99	5.26	10.77	0.65	5
$\Sigma(d_{\text{IP}})$	6.62	3.07	8.33	0.77	6
M_{miss}^2	5.21	1.71	12.78	0.86	7
$\cos\theta_{0\triangleleft1}$	4.81	3.37	6.79	0.93	8
E_1	3.34	2.04	10.40	0.88	9
E_0	3.58	2.40	10.86	0.88	10

Table A.3.: Variables used in the network for the final state $e^{\pm}\pi^{\mp}$ sorted by their rank.

Variable	Added	Without	Only	Global correlation	Rank
$\cos\theta^*_{\tau\lhd\pi_1}$	33.50	30.48	33.50	0.74	1
$\cos\theta^*_{\tau\lhd\pi_0}$	33.18	31.35	33.18	0.74	2
$\mathcal{N}_{\text{tag}}$	20.66	16.93	23.55	0.14	3
$\cos\theta_{\text{hel},0}$	14.55	8.90	8.24	0.68	4
$M^{\text{tag}}_{\text{bc}}$	13.44	12.66	17.91	0.10	5
E_1	12.18	9.95	11.17	0.87	6
E_0	11.57	8.13	10.65	0.84	7
$|\vec{p}_{\text{miss}}|$	12.31	7.77	14.90	0.71	8
$\Sigma(d_{\text{IP}})$	10.13	6.45	14.88	0.79	9
p_{T}	5.45	3.10	8.31	0.91	10
$\cos\theta^*_{0\lhd1}$	6.23	2.89	12.83	1.00	11
$\cos\theta_1$	2.81	3.31	5.38	0.58	12
$p_{T,0}$	2.49	2.57	10.99	0.86	13
$\cos\theta_0$	2.45	2.51	4.40	0.45	14
$M(B_{\text{sig}})$	2.20	1.48	23.32	0.90	15

Table A.2.: Variables used in the network for the final state $e^{\pm}\mu^{\mp}$ sorted by their rank.

Variable	Added	Without	Only	Global correlation	Rank
$\mathcal{N}_{\text{tag}}$	17.81	15.10	17.81	0.14	1
$|\vec{p}_{\text{miss}}|$	10.65	4.33	11.70	0.80	2
$\cos\theta_{\text{miss}}$	11.22	3.56	7.08	0.83	3
$\Sigma(d_{\text{IP}})$	11.41	5.21	12.23	0.78	4
$\cos\theta_{\text{hel},0}$	10.35	8.08	13.21	0.67	5
$M^{\text{tag}}_{\text{bc}}$	10.08	9.79	11.28	0.08	6
M^2_{miss}	5.29	6.56	9.60	0.47	7
$\cos\theta_{0\lhd1}$	6.14	0.40	10.54	0.99	8
p_{T}	2.63	2.50	14.21	0.88	9
$M(B_{\text{sig}})$	2.21	2.64	9.79	0.80	10

Table A.4.: Variables used in the network for the final state $\mu^+\mu^-$ sorted by their rank.

Variable	Added	Without	Only	Global correlation	Rank
$\cos\theta_{0\triangleleft 1}$	12.48	2.00	12.48	0.98	1
$\mathcal{N}_{\mathrm{tag}}$	11.38	10.36	12.20	0.12	2
$\Sigma(d_{\mathrm{IP}})$	8.27	3.85	8.47	0.80	3
M^2_{miss}	7.70	4.50	6.97	0.55	4
$M^{\mathrm{tag}}_{\mathrm{bc}}$	7.35	7.05	8.20	0.10	5
$\cos\theta_{\mathrm{hel},0}$	5.86	4.45	5.08	0.59	6
$|\vec{p}_{\mathrm{miss}}|$	3.99	4.01	7.66	0.77	7
$\cos\theta_{\mathrm{miss}}$	3.72	2.41	5.19	0.81	8
E_1	2.07	1.83	3.48	0.48	9

Table A.5.: Variables used in the network for the final state $\mu^\pm\pi^\mp$ sorted by their rank.

Variable	Added	Without	Only	Global correlation	Rank
$\cos\theta^*_{\tau\triangleleft\pi_1}$	29.83	20.76	29.83	0.85	1
$\cos\theta^*_{\tau\triangleleft\pi_0}$	27.96	19.83	28.02	0.83	2
$\mathcal{N}_{\mathrm{tag}}$	17.10	14.05	19.33	0.13	3
$\cos\theta_{\mathrm{hel},0}$	16.17	8.28	18.49	0.74	4
$M^{\mathrm{tag}}_{\mathrm{bc}}$	10.65	10.28	14.59	0.10	5
$|\vec{p}_{\mathrm{miss}}|$	7.80	7.35	22.10	0.88	6
$\cos\theta_{\mathrm{miss}}$	10.37	6.58	0.20	0.88	7
$\Sigma(d_{\mathrm{IP}})$	8.85	5.40	13.95	0.79	8
$\cos\theta^*_{0\triangleleft 1}$	7.46	1.49	18.07	0.99	9
E_1	4.44	4.58	19.60	0.91	10
E_0	5.21	4.85	17.54	0.88	11
M^2_{miss}	4.68	2.97	16.35	0.77	12
$\cos\theta_1$	3.58	4.94	7.97	0.70	13
p_{T}	3.65	3.47	18.80	0.90	14
$\cos\theta_0$	3.46	3.77	6.59	0.64	15
$|\vec{p}^*_1|$	1.91	3.12	20.84	0.91	16
$p_{T,1}$	3.46	2.95	19.90	0.92	17
$|\vec{p}^*_0|$	2.90	2.74	20.19	0.88	18

Table A.6.: Variables used in the network for the final state $\pi^+\pi^-$ sorted by their rank.

Variable	Added	Without	Only	Global correlation	Rank
M^2_{miss}	57.27	5.03	57.27	0.98	1
$|\vec{p}_{\mathrm{T,miss}}|$	18.30	2.94	46.85	0.95	2
$\mathcal{N}_{\mathrm{tag}}$	15.00	12.61	22.03	0.20	3
$M^{\mathrm{tag}}_{\mathrm{bc}}$	11.64	11.21	21.86	0.20	4
$\Sigma(d_{\mathrm{IP}})$	7.25	5.00	17.11	0.75	5
$\cos\theta^*_{0\triangleleft 1}$	4.79	2.36	13.80	0.98	6
$p_{T,1}$	3.14	2.67	38.32	0.97	7
$p_{T,0}$	5.73	2.34	36.52	0.97	8
$\cos\theta_{\mathrm{hel},0}$	5.12	5.53	22.15	0.53	9
$M(B_{\mathrm{sig}})$	3.19	2.79	35.66	0.90	10
A_{01}	2.77	2.68	18.17	0.72	11
$|\vec{p}_{\mathrm{miss}}|$	1.33	2.79	9.49	0.73	12
$\cos\theta_{\mathrm{miss}}$	2.62	1.39	41.06	0.93	13
p_{T}	2.30	2.41	48.08	0.94	14

B. Template Histograms

(a) Signal component.

(b) Background component: $B^0 \to D^-(K^0\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$.

(c) Background component: $B^0 \to X_c +$ anything.

(d) Background component: $B^+ \to X_c +$ anything.

(e) Background component: rare B^0 decays.

(f) Background component: $B^0 \to X_u +$ anything.

(g) Background component: $B^+ \to X_u +$ anything.

(h) Background component: $e^+e^- \to c\bar{c}$.
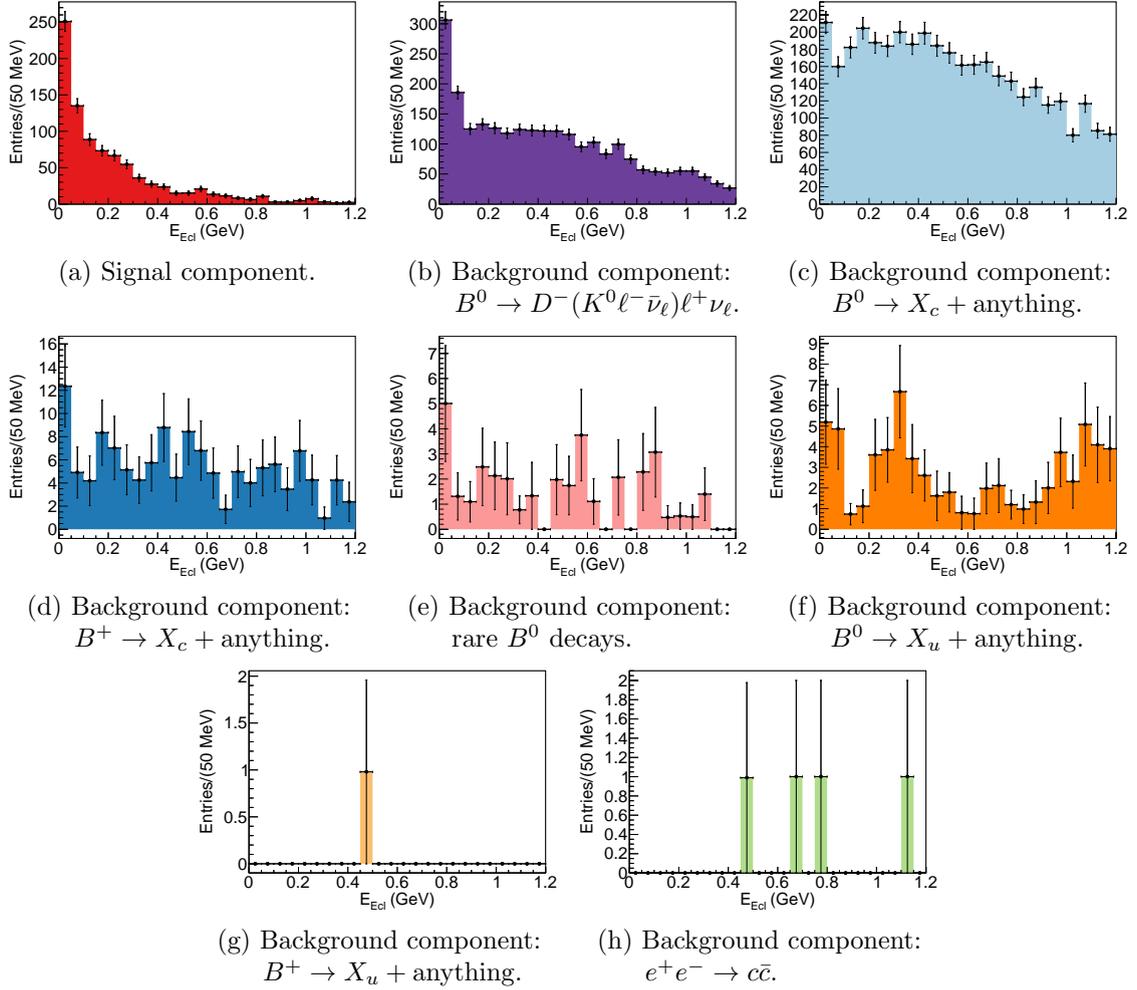
Figure B.1.: Histogram templates used in the fit for the final state $e^\pm\mu^\mp$.

(a) Signal component.

(b) Background component: $B^0 \to D^-(K^0\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$.

(c) Background component: $B^0 \to X_c +$ anything.

(d) Background component: $B^+ \to X_c +$ anything.

(e) Background component: rare B^0 decays.

(f) Background component: rare B^+ decays.

(g) Background component: $B^0 \to X_u +$ anything.

(h) Background component: $B^+ \to X_u +$ anything.

(i) Background component: $e^+e^- \to c\bar{c}$.

(j) Background component: $e^+e^- \to q\bar{q}$ $(q = u, d, s)$.
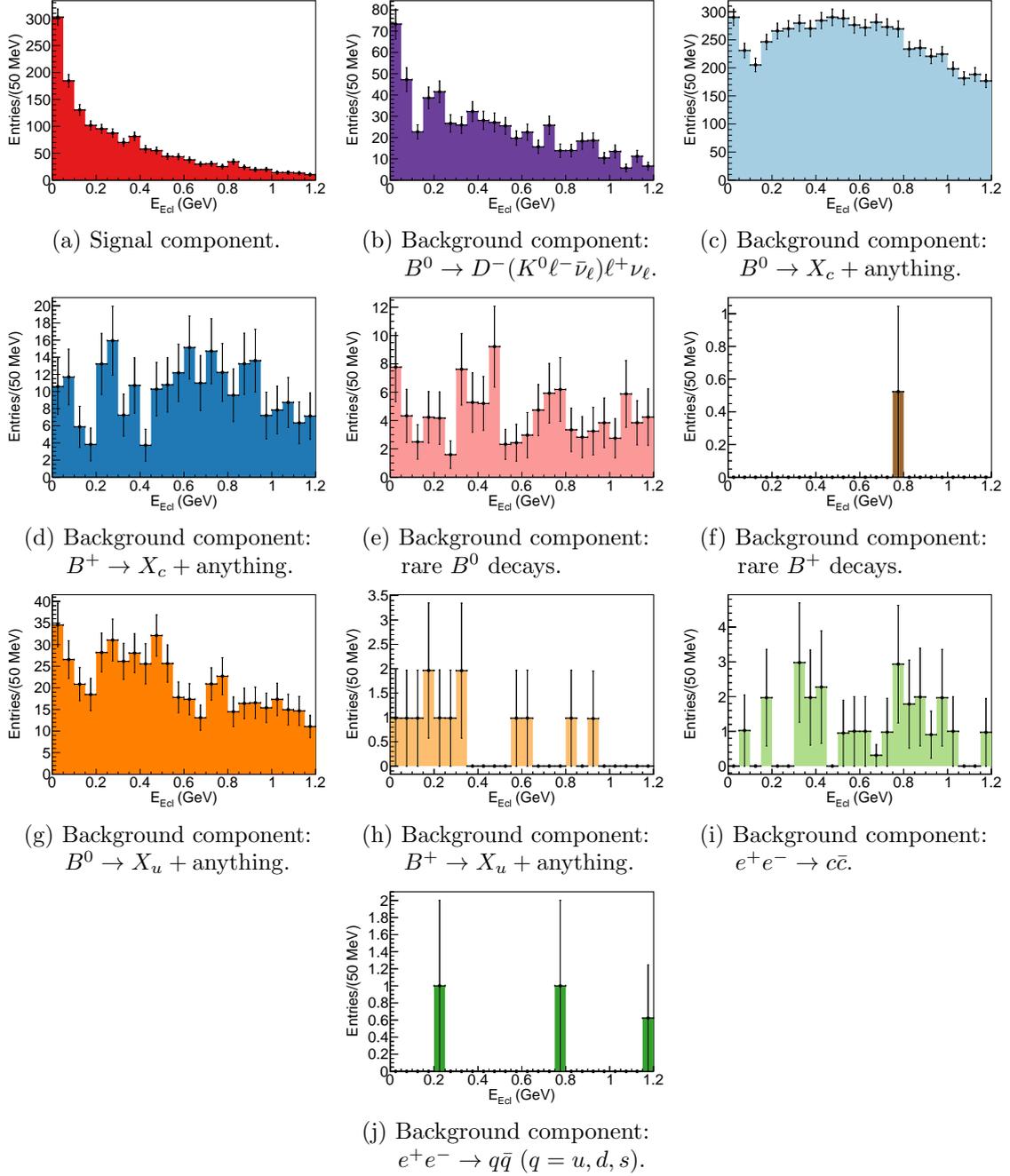
Figure B.2.: Histogram templates used in the fit for the final state $e^\pm\pi^\mp$.

(a) Signal component.

(b) Background component: $B^0 \to D^-(K^0\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$.

(c) Background component: $B^0 \to X_c$ + anything.

(d) Background component: $B^+ \to X_c$ + anything.

(e) Background component: rare B^0 decays.

(f) Background component: $B^0 \to X_u$ + anything.

(g) Background component: $e^+e^- \to c\bar{c}$.

(h) Background component: $e^+e^- \to q\bar{q}\,(q = u, d, s)$.
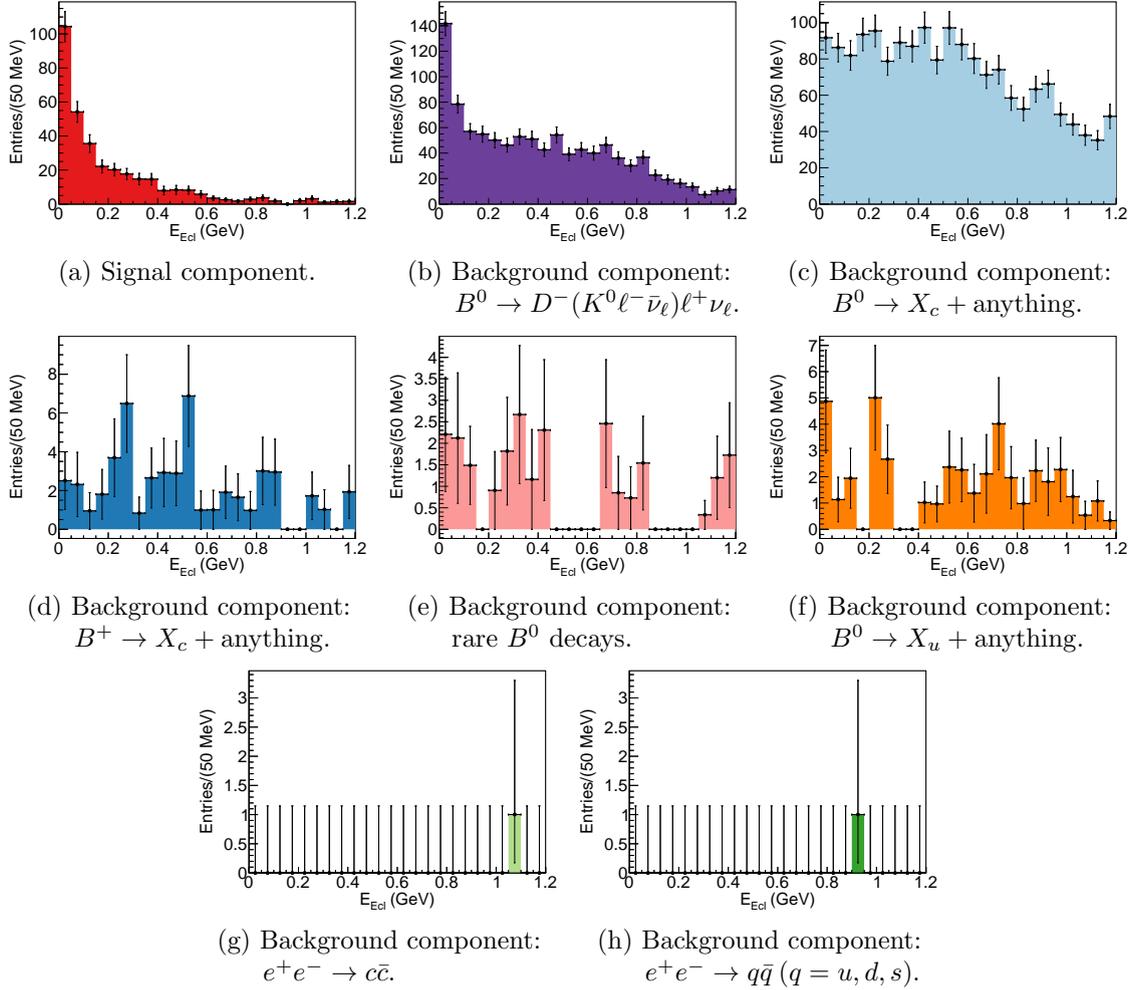
Figure B.3.: Histogram templates used in the fit for the final state $\mu^+\mu^-$.

(a) Signal component.

(b) Background component:
$B^0 \to D^-(K^0\ell^-\bar{\nu}_\ell)\ell^+\nu_\ell$.

(c) Background component:
$B^0 \to X_c$ + anything.

(d) Background component:
$B^+ \to X_c$ + anything.

(e) Background component:
rare B^0 decays.

(f) Background component:
rare B^+ decays.

(g) Background component:
$B^0 \to X_u$ + anything.

(h) Background component:
$B^+ \to X_u$ + anything.

(i) Background component:
$e^+e^- \to c\bar{c}$.

(j) Background component:
$e^+e^- \to q\bar{q}$ $(q = u, d, s)$.
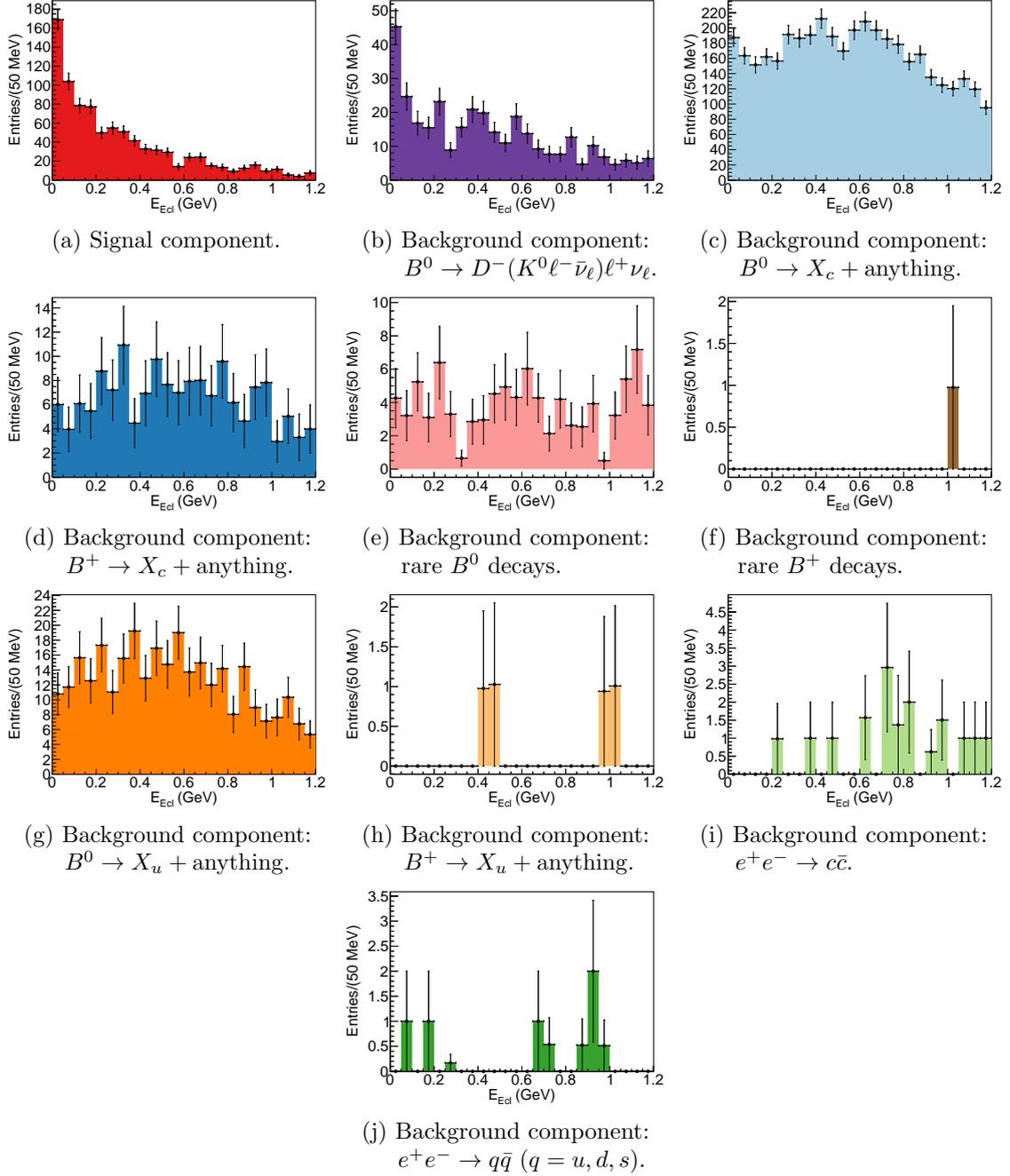
Figure B.4.: Histogram templates used in the fit for the final state $\mu^\pm\pi^\mp$.

C. Data-MC Comparison - Including Signal

C.1. Variables after the Final Selection
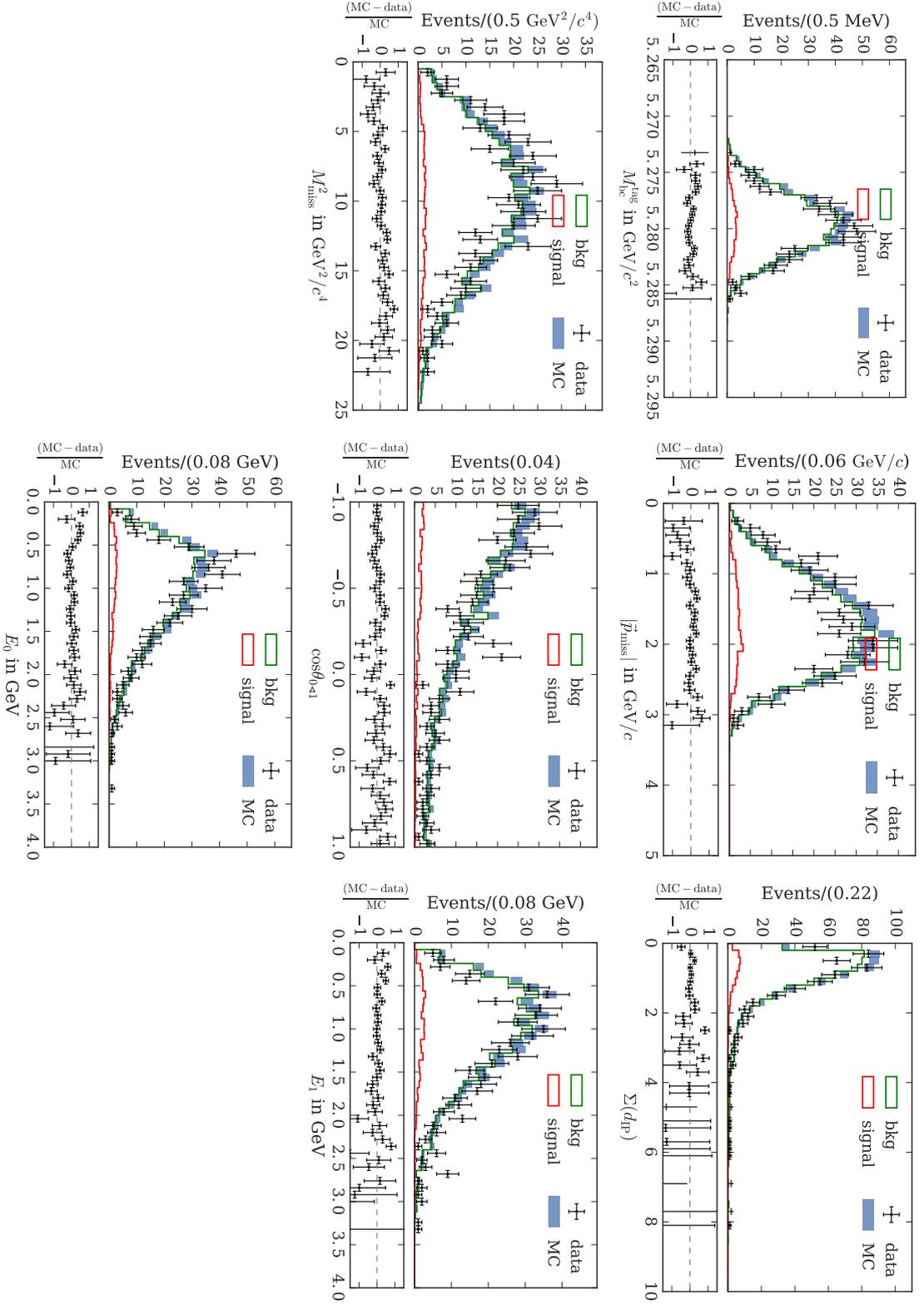
Figure C.1.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state e^+e^-. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.
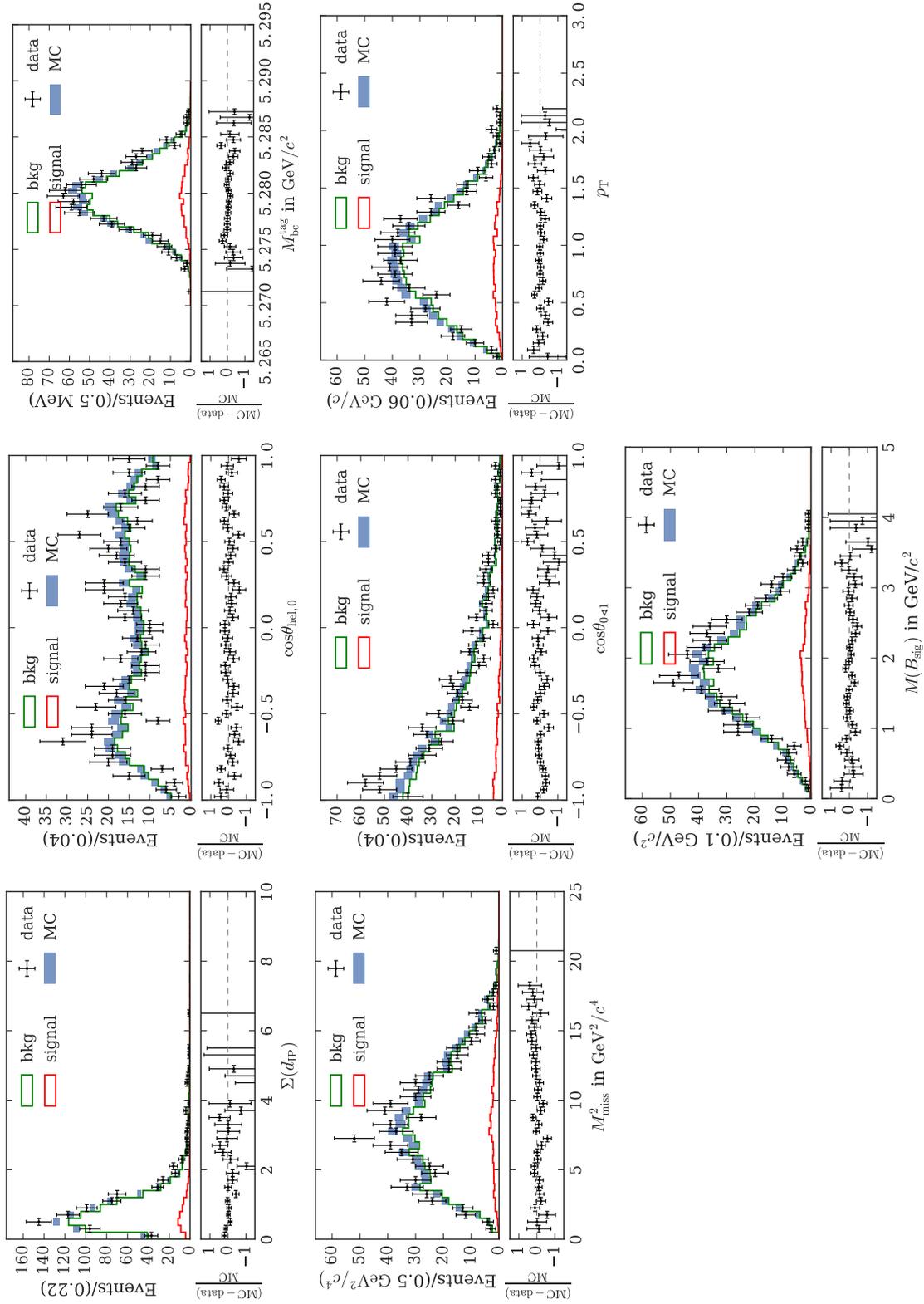
Figure C.2.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^{\pm}\mu^{\mp}$. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.
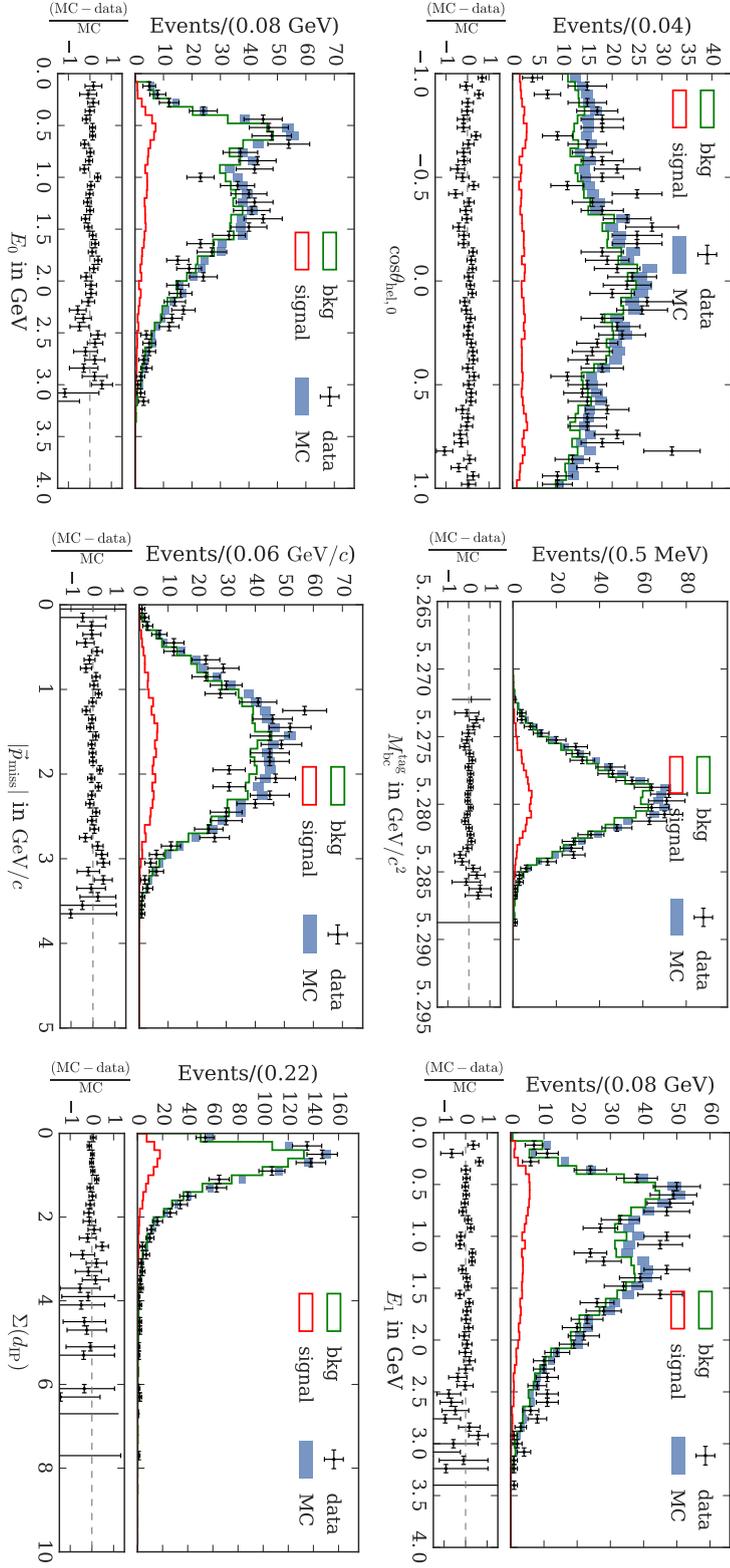
Figure C.3.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^{\pm}\pi^{\mp}$. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.
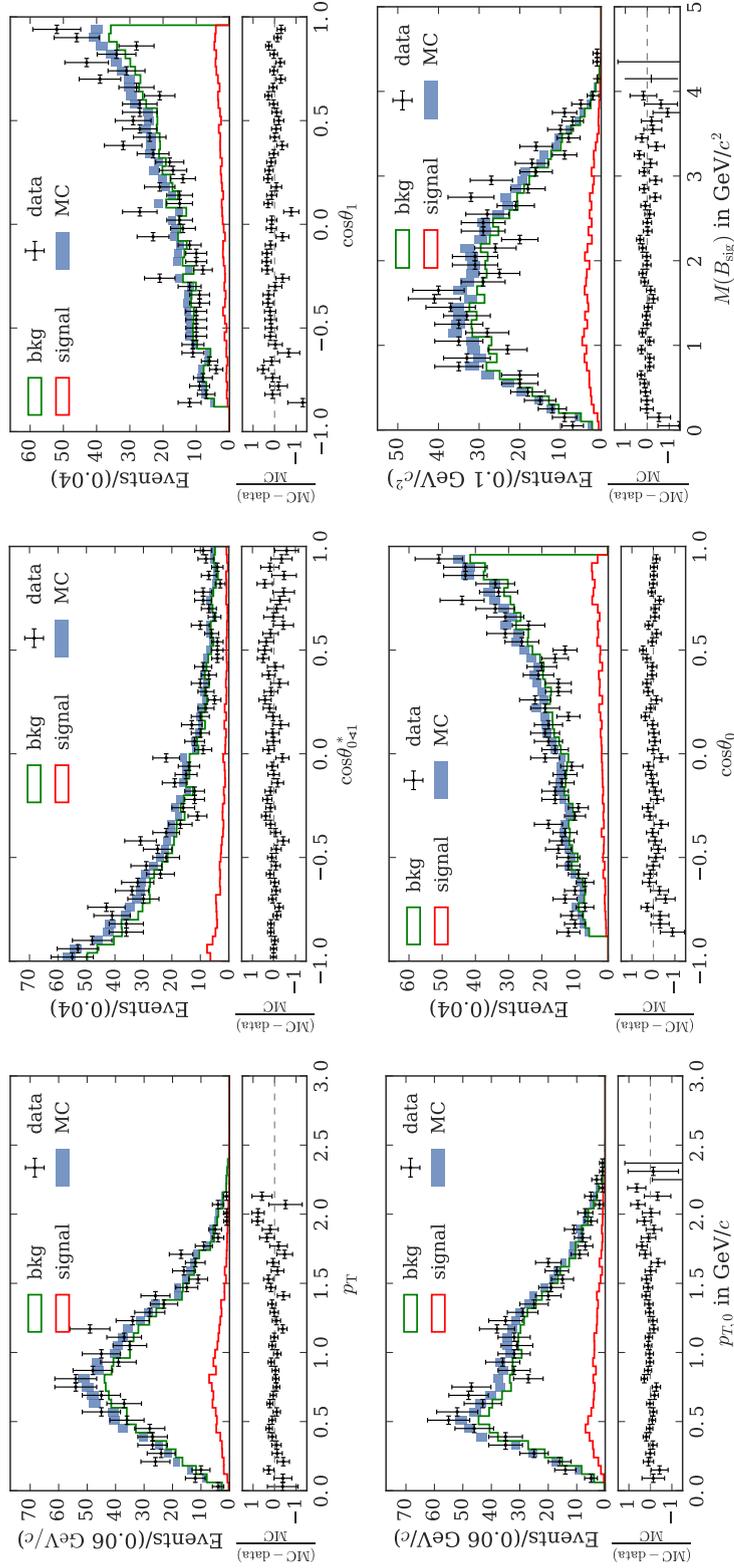
Figure C.4.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^{\pm}\pi^{\mp}$. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.
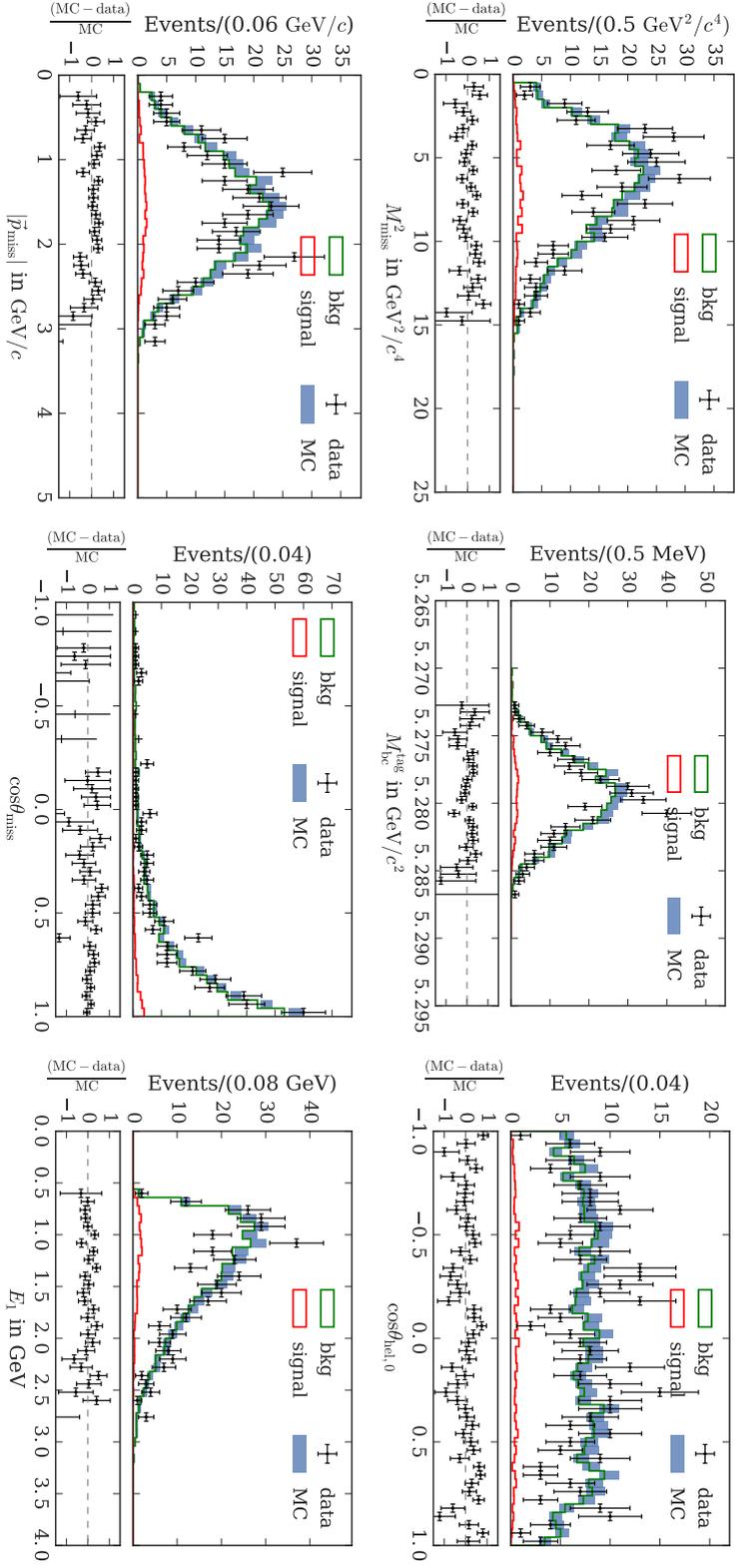
Figure C.5.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^+\mu^-$. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.

116

Figure C.6.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^\pm \pi^\mp$. The signal and background normalization are taken from the fit to data. The final selection is applied on the events.
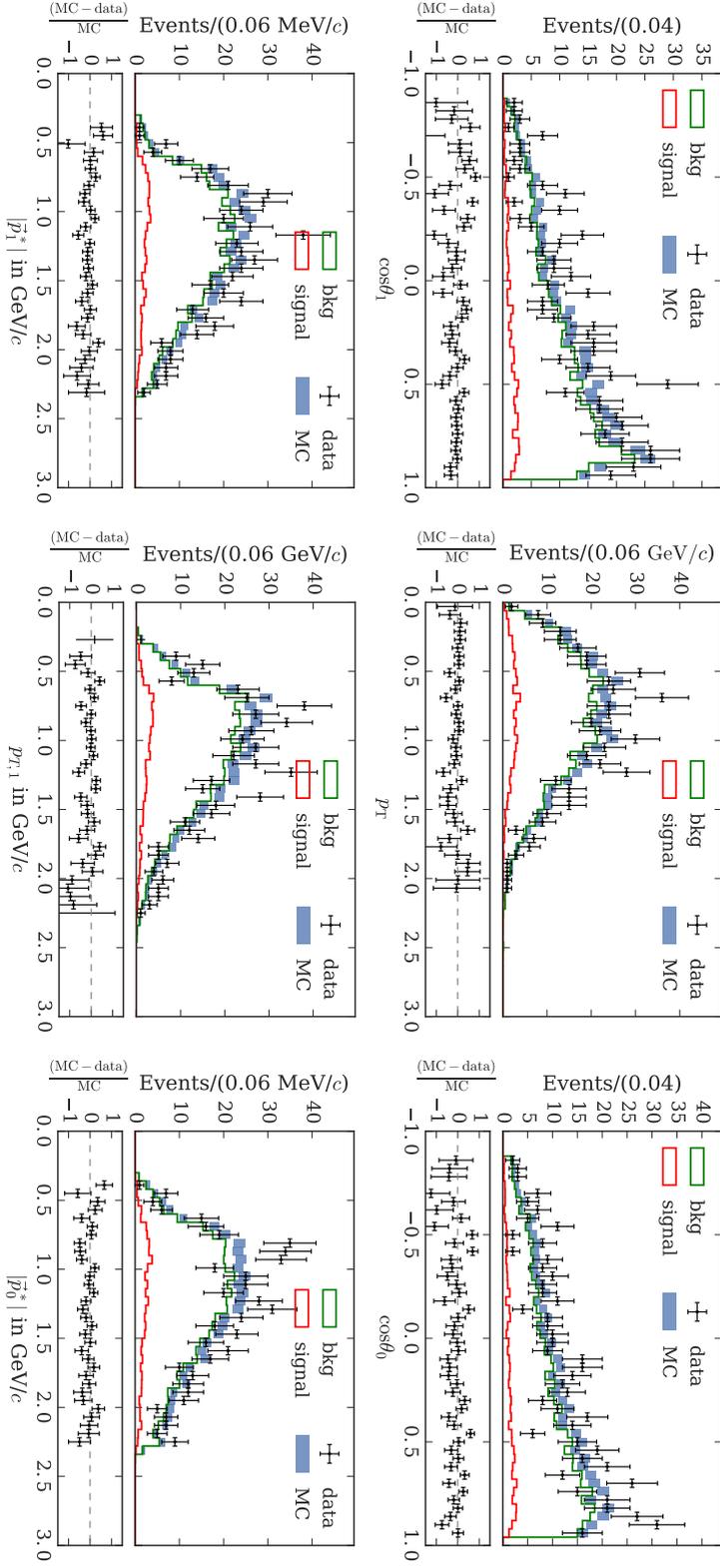
Figure C.7.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^{\pm}\pi^{\mp}$. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.
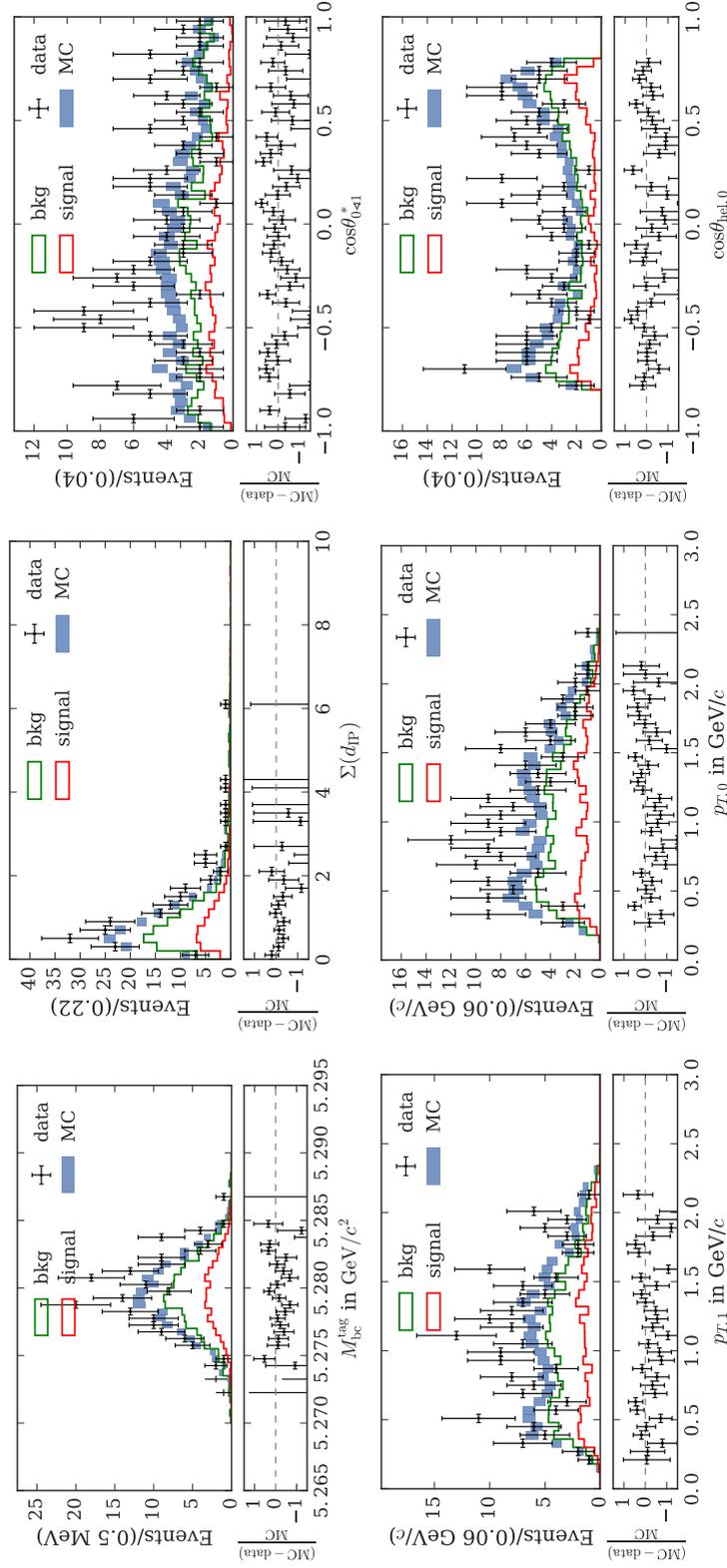
Figure C.8.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\pi^+\pi^-$. The signal and background normalization are taken from the fit to data. The final selection is applied on the events.
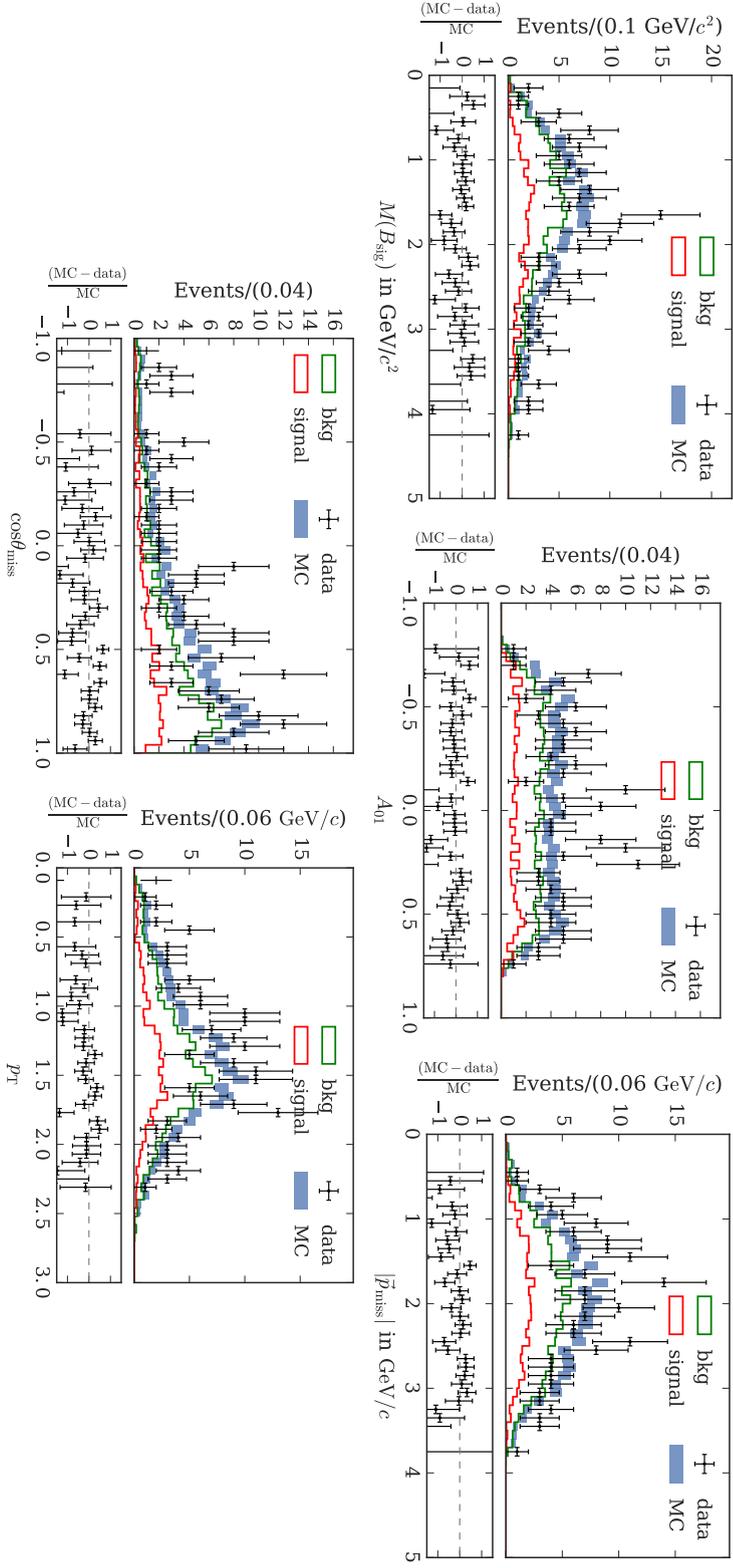
Figure C.9.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\pi^+\pi^-$. The final selection is applied on the events. The signal and background normalization are taken from the fit to data.

C.2. Variables after the Preselection
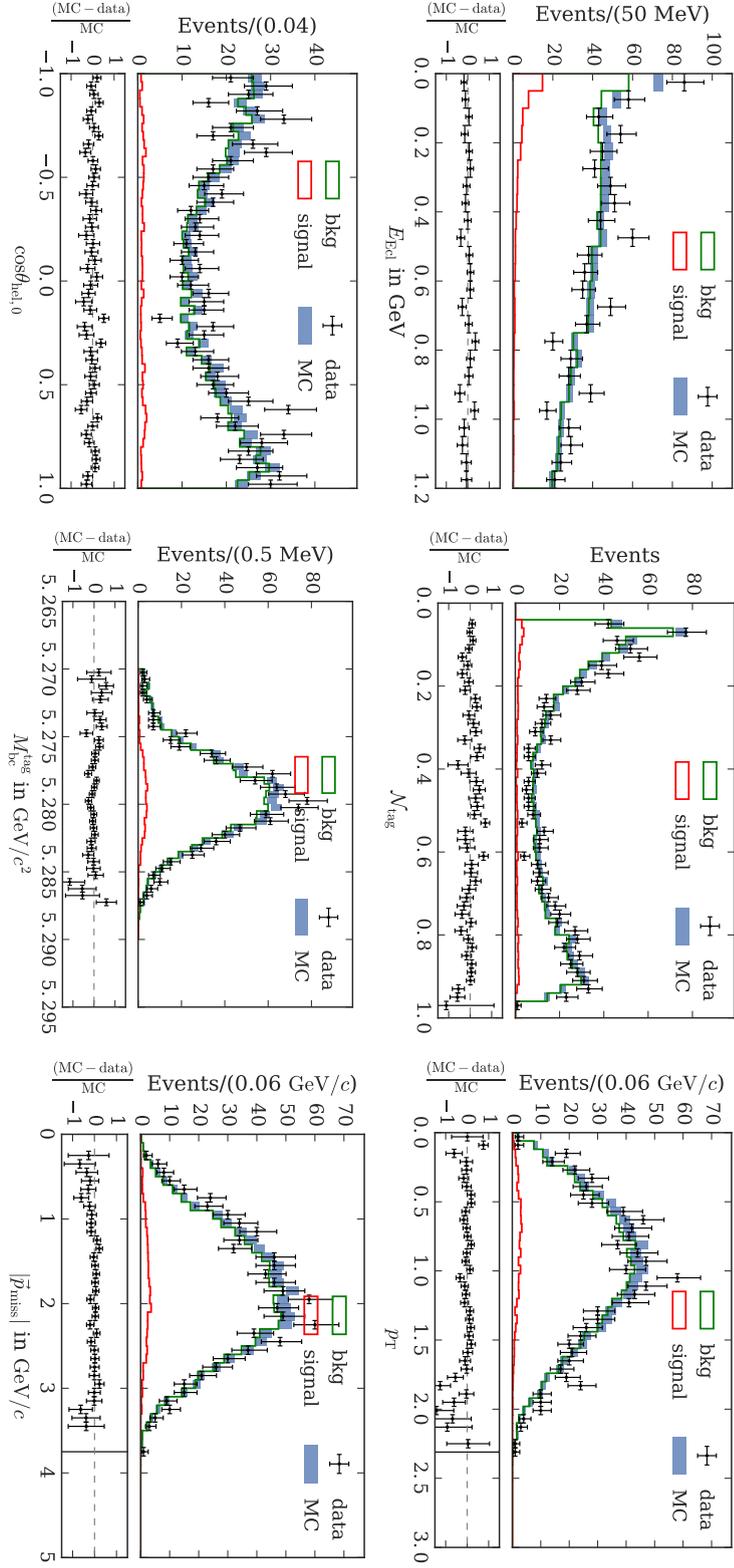
Figure C.10.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state e^+e^-. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
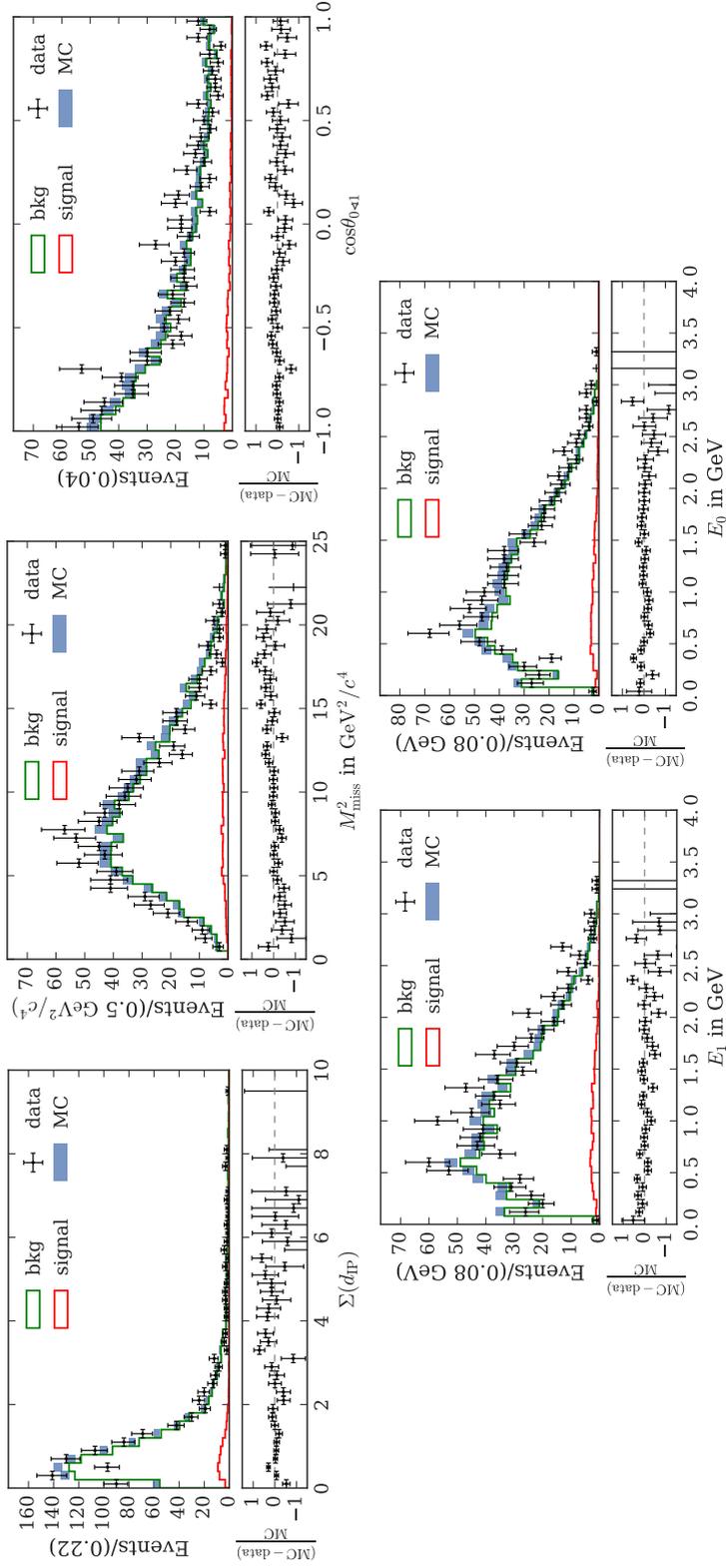
Figure C.11.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state e^+e^-. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
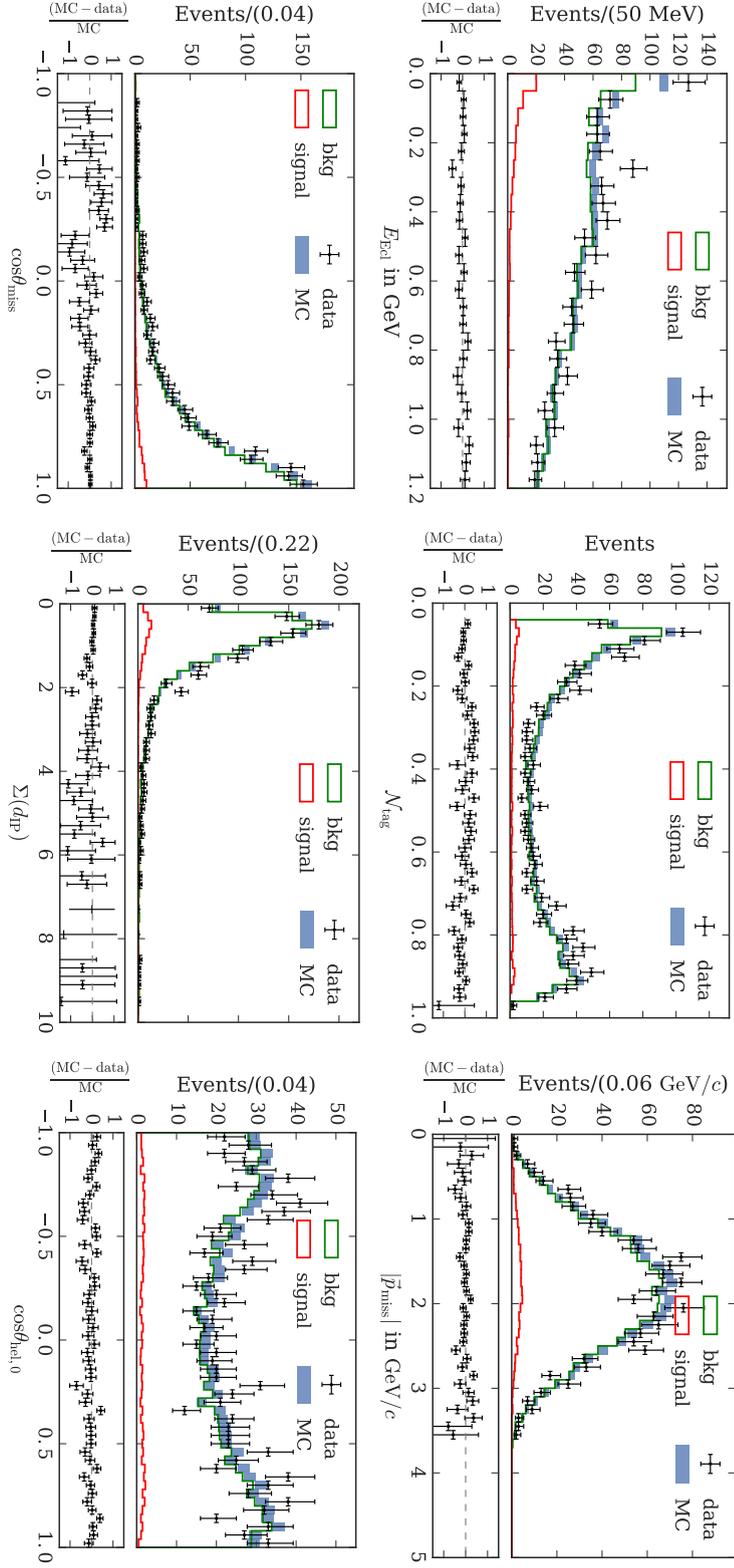
Figure C.12.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^{\pm}\mu^{\mp}$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
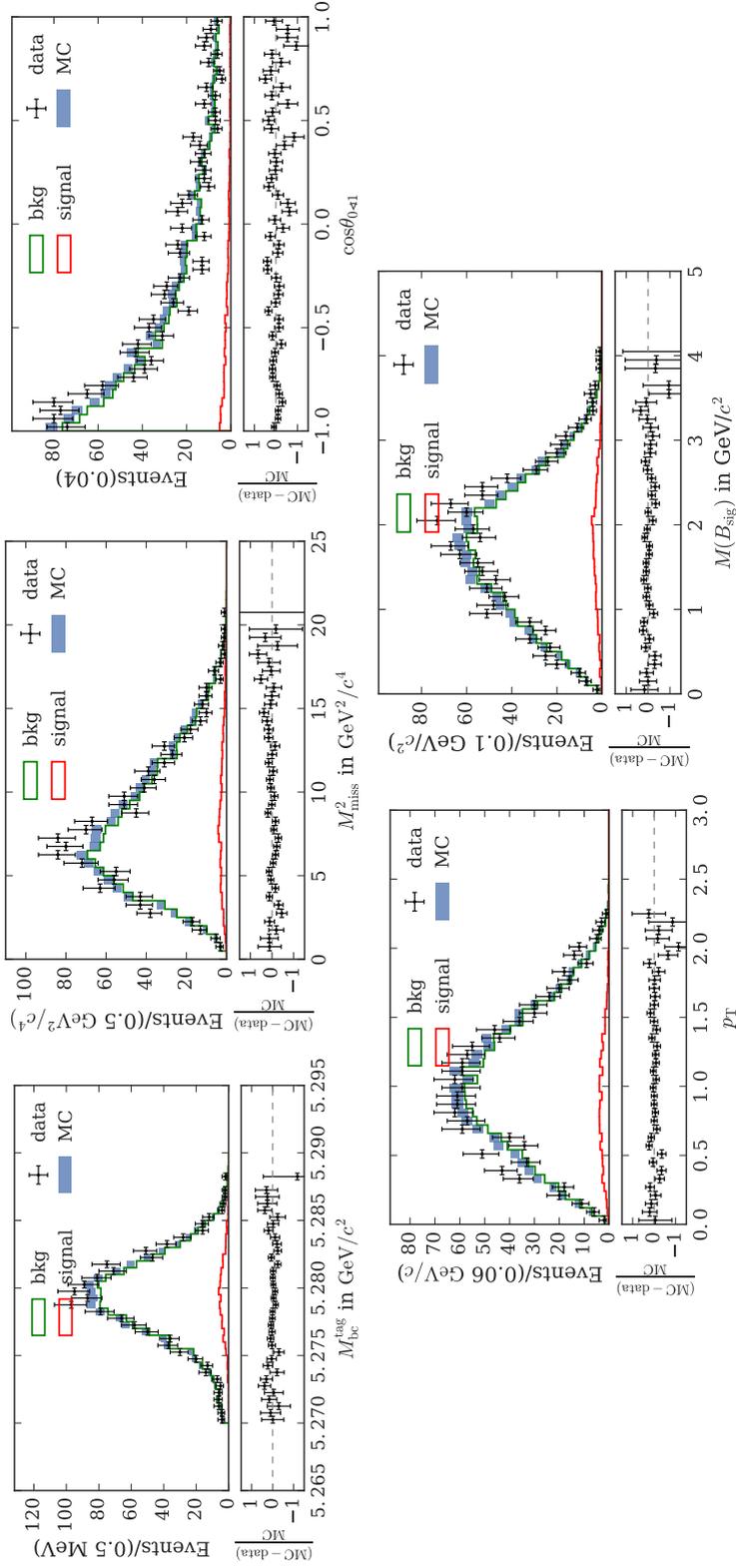
Figure C.13.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^\pm\mu^\mp$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
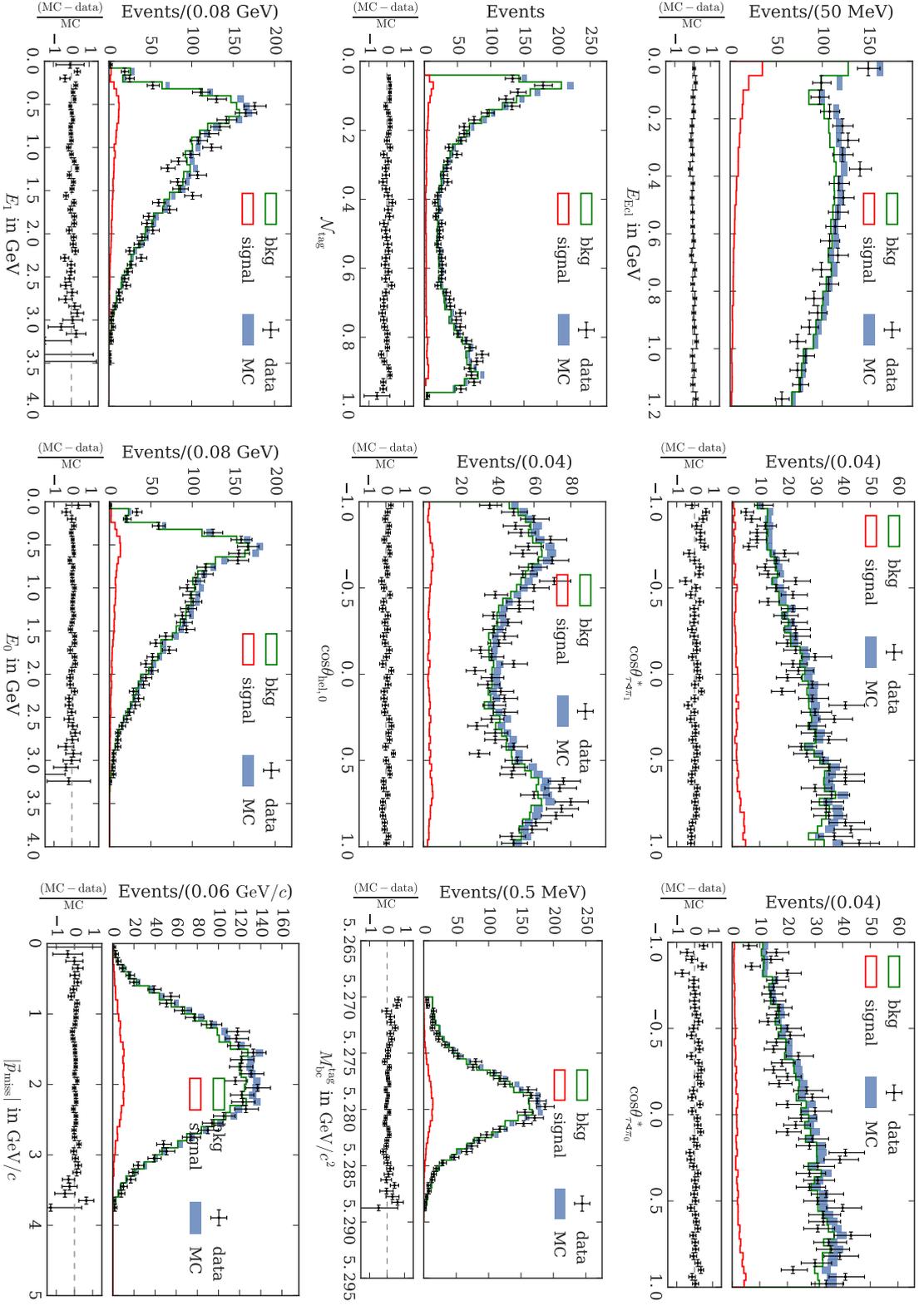
Figure C.14.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^{\pm}\pi^{\mp}$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
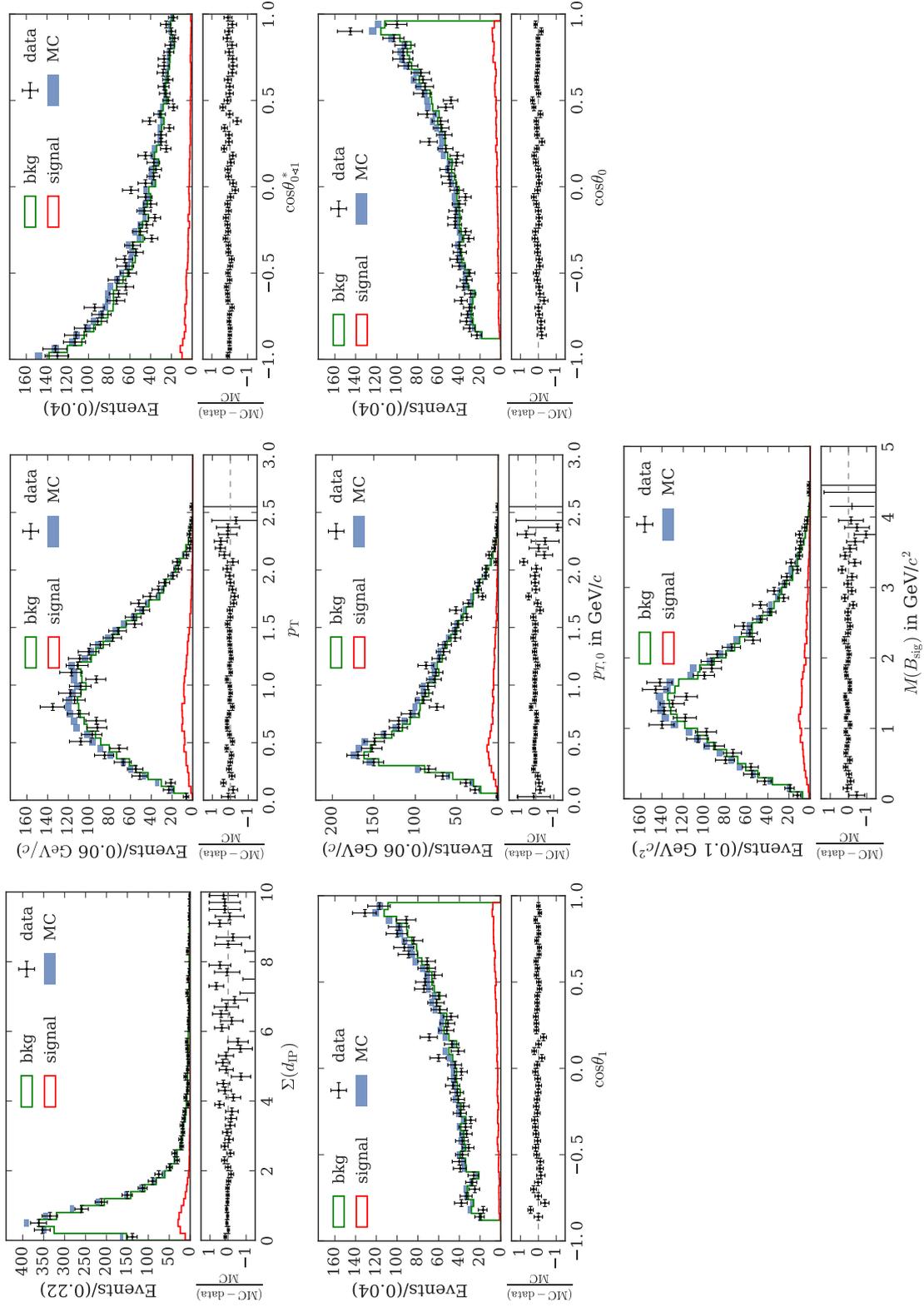
Figure C.15.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $e^{\pm}\pi^{\mp}$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.

Figure C.16.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^+\mu^-$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
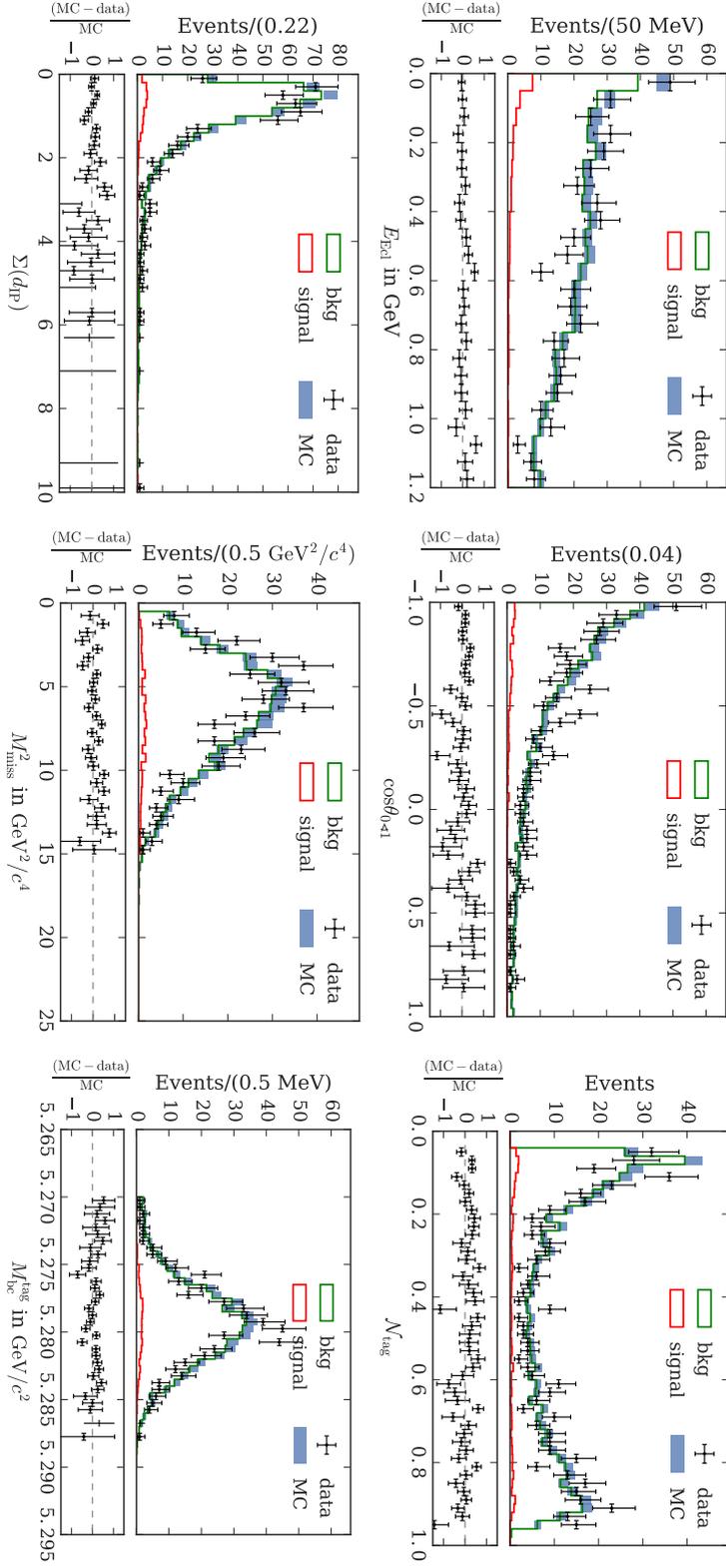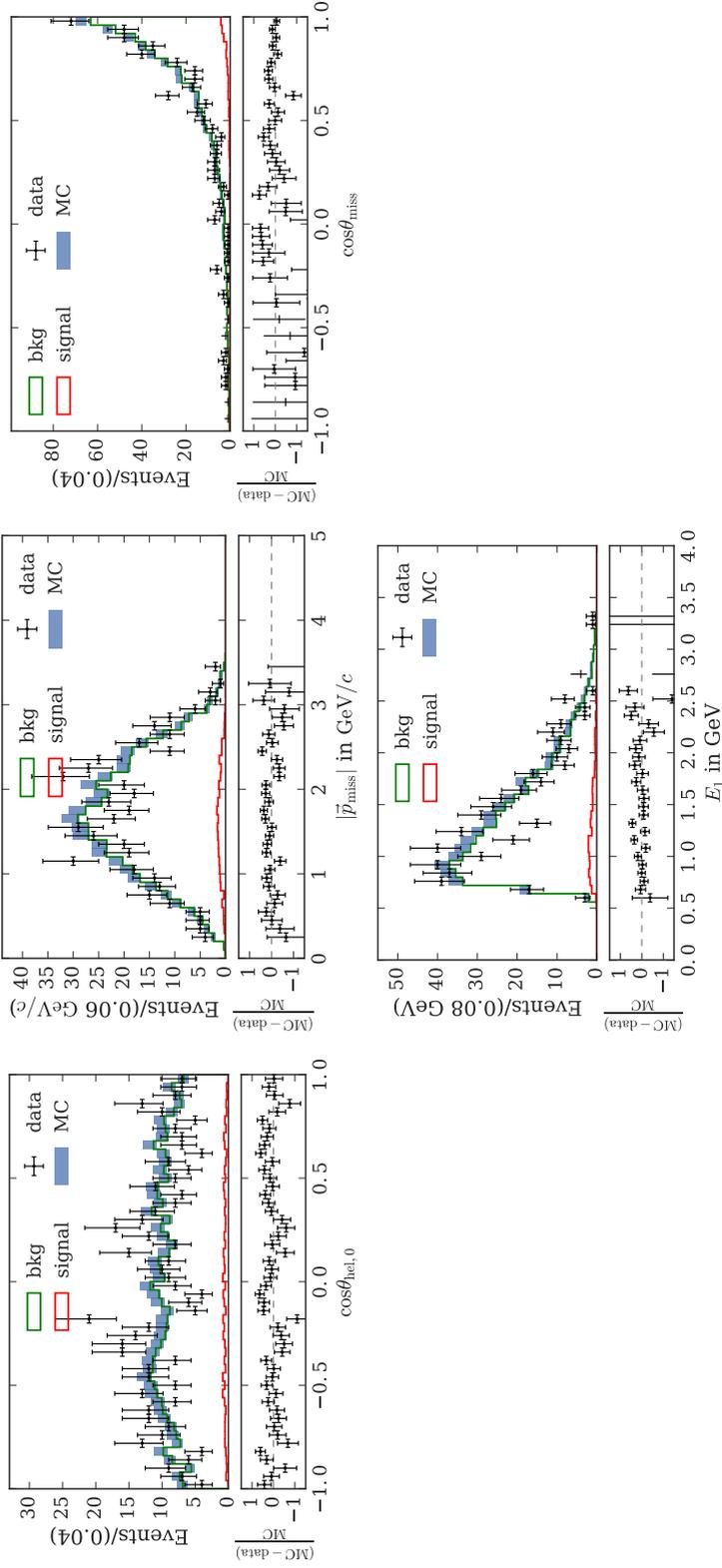
Figure C.17.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^+\mu^-$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
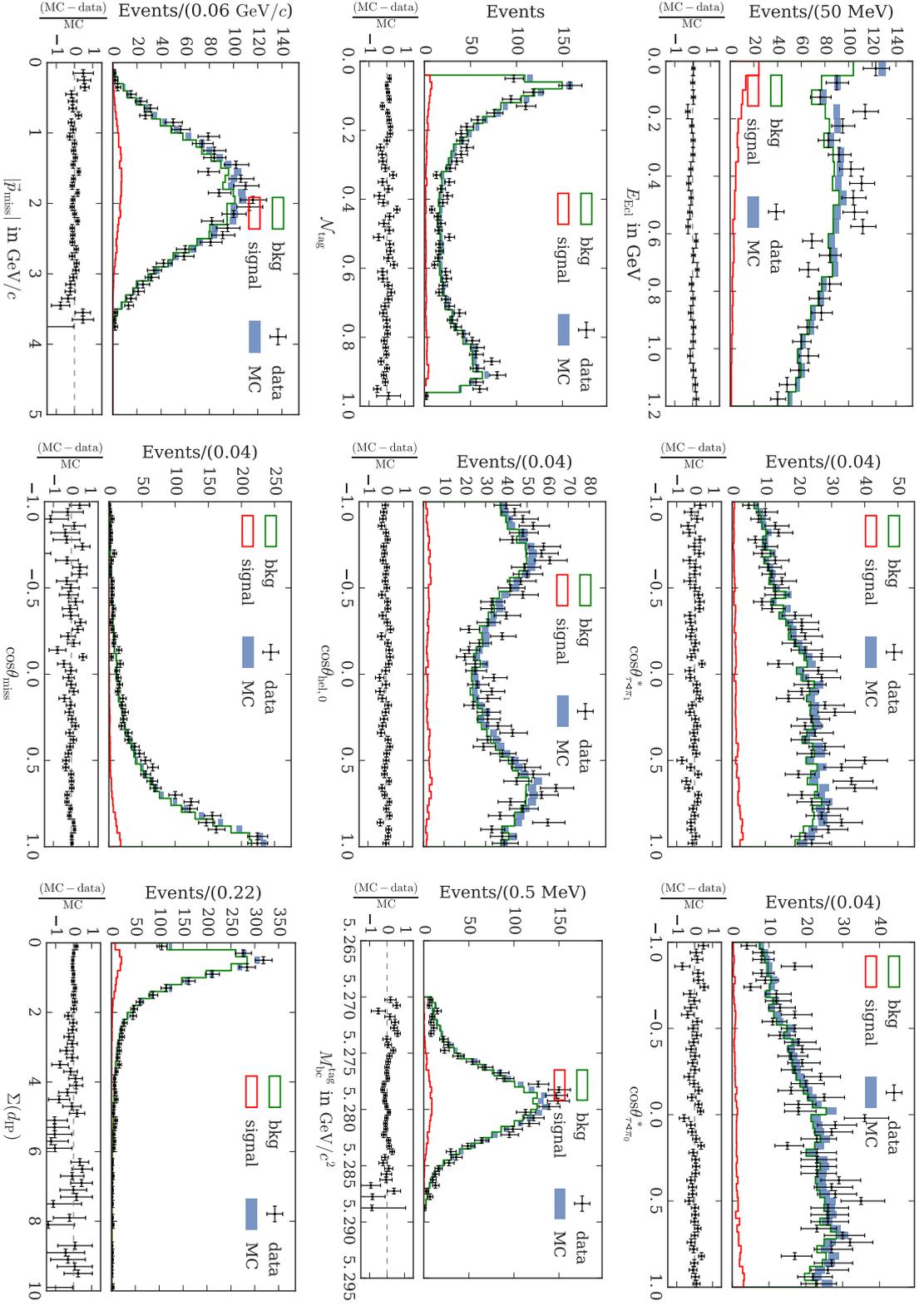
Figure C.18.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^{\pm}\pi^{\mp}$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.

Figure C.19.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^{\pm}\pi^{\mp}$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.

Figure C.20.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\mu^{\pm}\pi^{\mp}$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.
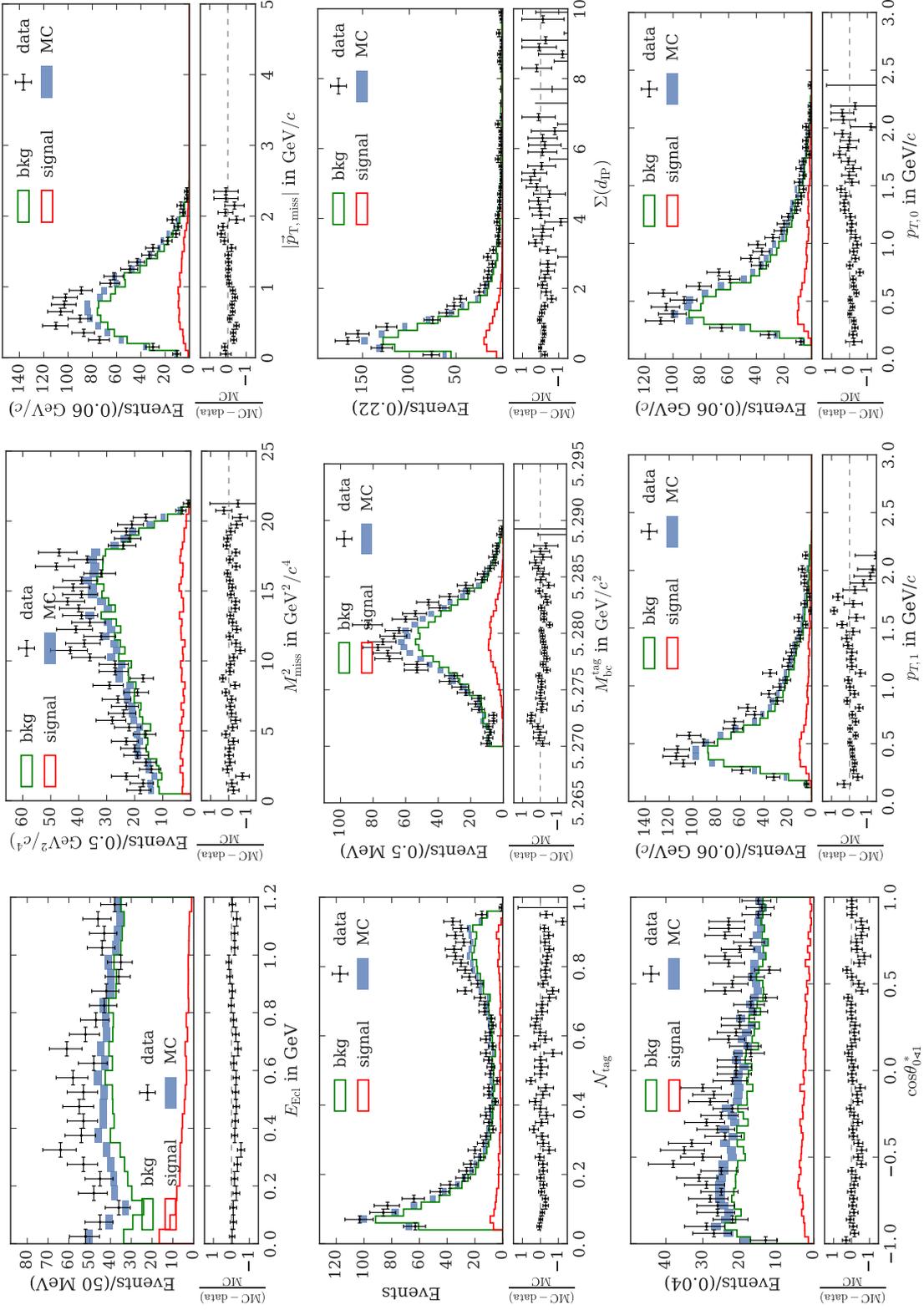
Figure C.21.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\pi^+\pi^-$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.

Figure C.22.: Each plot shows the distributions of an input variable (see x-axis label) used in the network of the final state $\pi^+\pi^-$. Only the preselection is applied on the events. The signal and background normalization are taken from the fit to data.

Bibliography

[1] D. Griffiths, *Introduction to Elementary Particles*, Physics textbook. Wiley, 2008.

[2] K. Abe, K. Abe *et al.*, "Observation of large *CP* violation in the neutral *B* meson system," *Phys. Rev. Lett.*, vol. 87, p. 091802, Aug 2001. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.87.091802

[3] B. Kronenbitter, M. Heck *et al.*, "Measurement of the branching fraction of $B^+ \to \tau^+ \nu_\tau$ decays with the semileptonic tagging method," *Phys. Rev. D*, vol. 92, p. 051102, Sep 2015. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.92.051102

[4] A. Heller, P. Goldenzweig *et al.*, "Search for $B^+ \to \ell^+ \nu_\ell \gamma$ decays with hadronic tagging using the full belle data sample," *Phys. Rev. D*, vol. 91, p. 112009, Jun 2015. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.91.112009

[5] S.-H. Kyeong, Y.-J. Kwon *et al.*, "Measurements of charmless hadronic $b \to s$ penguin decays in the $\pi^+ \pi^- K^+ \pi^-$ final state and first observation of $B^0 \to \rho^0 K^+ \pi^-$," *Phys. Rev. D*, vol. 80, p. 051103, Sep 2009. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.80.051103

[6] G. Isidori, "Flavor physics and CP violation," in *Proceedings, 2012 European School of High-Energy Physics (ESHEP 2012): La Pommeraye, Anjou, France, June 06-19, 2012*, 2014, pp. 69–105. [Online]. Available: http://inspirehep.net/record/1217584

[7] H. E. Logan and U. Nierste, "$B_{s,d} \to \ell^+ \ell^-$ in a two-Higgs-doublet model," *Nuclear Physics B*, vol. 586, no. 1-2, pp. 39–55, Oct. 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S055032130000417X

[8] A. Buras, "Weak Hamiltonian, CP violation and rare decays," *Arxiv preprint hep-ph/9806471*, no. June, p. 250, Jun. 1998. [Online]. Available: http://arxiv.org/abs/hep-ph/9806471

[9] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics*, vol. 49, no. 2, pp. 652–657, Feb. 1973. [Online]. Available: http://ptp.oxfordjournals.org/content/49/2/652

[10] C. Bobeth, M. Gorbahn *et al.*, "$B_{s,d} \to \ell^+ \ell^-$ in the Standard Model with Reduced Theoretical Uncertainties," *Physical Review Letters*, vol. 112, no. 10, p. 101801, 2014. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.112.101801

[11] B. Grzadkowski and P. Krawczyk, "Higgs particle effects in flavour changing transitions," *Zeitschrift für Physik C Particles and Fields*, vol. 18, no. 1, pp. 43–45, Mar. 1983. [Online]. Available: http://link.springer.com/10.1007/BF01571704

[12] C. Davies, "Standard Model Heavy Flavor physics on the Lattice," p. 21, Mar. 2012. [Online]. Available: http://arxiv.org/abs/1203.3862

[13] B. Gripaios, "Composite leptoquarks at the LHC," *Journal of High Energy Physics*, vol. 2010, no. 2, p. 45, Feb. 2010. [Online]. Available: http://link.springer.com/10.1007/JHEP02(2010)045

[14] D. B. Kaplan, "Flavor at ssc energies: A new mechanism for dynamically generated fermion masses," *Nuclear Physics B*, vol. 365, no. 2, pp. 259–278, Nov. 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0550321305800215

[15] S. Sahoo and R. Mohanta, "Scalar leptoquarks and the rare B meson decays," *Physical Review D*, vol. 91, no. 9, p. 094019, May 2015. [Online]. Available: http://journals.aps.org/prd/abstract/10.1103/PhysRevD.91.094019

[16] CMS Collaboration and LHCb Collaboration, "Observation of the rare $B_s \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data," *Nature*, vol. 522, pp. 68–72, 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature14474

[17] K. Olive *et al.* (Particle Data Group), "Review of Particle Physics," *Chinese Physics C*, vol. 38, no. 9, p. 090001, 2014. [Online]. Available: http://stacks.iop.org/1674-1137/38/i=9/a=090001?key=crossref.47735154e79ac9c858085df9b5a1f93a

[18] B. Aubert, R. Barate *et al.*, "Search for the rare decay $B^0 \to \tau^+\tau^-$ at BABAR." *Physical review letters*, vol. 96, no. 24, p. 241802, Jun. 2006. [Online]. Available: http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.96.241802

[19] A. J. Bevan, B. Golob *et al.*, "The Physics of the B Factories," *The European Physical Journal C*, vol. 74, no. 11, p. 3026, Nov. 2014. [Online]. Available: http://link.springer.com/10.1140/epjc/s10052-014-3026-9

[20] D. Besson and T. Skwarnicki, "Upsilon Spectroscopy: Transitions in the Bottomonium System," *Annual Review of Nuclear and Particle Science*, vol. 43, no. 1, pp. 333–378, Dec. 1993. [Online]. Available: http://www.annualreviews.org/doi/abs/10.1146/annurev.ns.43.120193.002001

[21] S. Okubo, "Phi meson and unitary symmetry model," *Phys. Lett.*, vol. 5, pp. 165–168, 1963.

[22] G. Zweig, "An SU$_3$ model for strong interaction symmetry and its breaking; Version 2," no. CERN-TH-412, p. 80 p, Feb 1964, version 1 is CERN preprint 8182/TH.401, Jan. 17, 1964. [Online]. Available: http://cds.cern.ch/record/570209

[23] J. Iizuka, "Systematics and phenomenology of meson family," *Prog. Theor. Phys. Suppl.*, vol. 37, pp. 21–34, 1966.

[24] M. Röhrken, *Time-Dependent CP Violation Measurements*, Springer Theses. Cham: Springer International Publishing, 2014. [Online]. Available: http://link.springer.com/10.1007/978-3-319-00726-7

[25] S. Kurokawa and E. Kikutani, "Overview of the KEKB accelerators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, no. 1, pp. 1–7, Feb. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900202017710

[26] K. Hosoyama, K. Hara *et al.*, "Development of the KEKB Superconducting Crab Cavity," *Proceedings of EPAC08, Genoa, Italy*, pp. 2927–2931, 2008.

[27] T. Abe, I. Adachi *et al.*, "Belle II Technical Design Report," *ArXiv e-prints*, Nov. 2010. [Online]. Available: https://arxiv.org/abs/1011.0352

[28] "KEKB B Factory Design Report," *KEK Report 95-7, High Energy Accelerator Resarch Organization, Japan*, 1995. [Online]. Available: http://www-acc.kek.jp/KEKB/publication/KEKB_design_report/KEKB%20Design%20Report.html

[29] A. Abashian, K. Gotow *et al.*, "The Belle detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 479, no. 1, pp. 117–232, Feb. 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900201020137

[30] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006. [Online]. Available: http://research.microsoft.com/en-us/um/people/cmbishop/prml/

[31] M. Feindt, "A Neural Bayesian Estimator for Conditional Probability Densities," p. 26, Feb. 2004. [Online]. Available: http://arxiv.org/abs/physics/0402093

[32] M. Feindt and U. Kerzel, "The NeuroBayes neural network package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 559, no. 1, pp. 190–194, 2006.

[33] M. Feindt, F. Keller *et al.*, "A Hierarchical NeuroBayes-based Algorithm for Full Reconstruction of B Mesons at B Factories," *Nucl. Instrum. Meth.*, vol. A654, pp. 432–440, Feb. 2011. [Online]. Available: http://arxiv.org/abs/1102.3876

[34] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes," *Physical Review Letters*, vol. 41, no. 23, pp. 1581–1585, Dec. 1978. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.41.1581

[35] E. Nakano, "Belle PID," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 494, no. 1, pp. 402–408, 2002.

[36] K. Hanagaki, H. Kakuno *et al.*, "Electron identification in Belle," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 485, no. 3, pp. 490–503, 2002.

[37] A. Abashian, K. Abe *et al.*, "Muon identification in the Belle experiment at KEKB," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 491, no. 1, pp. 69–82, 2002.

[38] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 462, no. 1, pp. 152–155, 2001.

[39] T. Sjöstrand, P. Edén *et al.*, "High-energy-physics event generation with pythia 6.1," *Comput. Phys. Commun.*, vol. 135, no. 2, pp. 238–259, Apr. 2001. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0010465500002368

[40] R. Brun, F. Bruyant *et al.*, *GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version*. Geneva: CERN, 1987. [Online]. Available: https://cds.cern.ch/record/1119728

[41] K.-F. Chen, F. Fang *et al.*, "Time-dependent CP-violating asymmetries in $b \to s\bar{q}q$ transitions," *Phys. Rev. D*, vol. 72, no. 1, p. 012004, Jul. 2005. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.72.012004

[42] F. Fang, "Study of $K_S \to \pi^+\pi^-$ Selection," *Belle Note*, no. 323, 2000.

[43] K. Hara, "Calibration of low momentum K_L^0 efficiency for veto usage in missing E analyses," *Belle Note*, no. 1228, 2012.

[44] A. Sibidanov and K. Varvell, "Exclusive $B \to X_u \ell \nu$ Decays using New Full Reconstruction Tagging," *Belle Note*, no. 1206, 2012.

[45] A. Sibidanov, K. E. Varvell *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," *Physical Review D*, vol. 88, no. 3, p. 032005, Aug. 2013. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.88.032005

[46] L. Hinz, "Lepton ID efficiency correction and systematic error," *Belle Note*, no. 954, 2006.

[47] C. Patrignani *et al.* (Particle Data Group), "Review of Particle Physics," *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.

[48] G. Punzi, "Sensitivity of searches for new signals and its optimization," in *Conf. Stat. Probl. Part. Physics, Astrophys. Cosmol.*, Aug. 2003, p. 5. [Online]. Available: http://arxiv.org/abs/physics/0308063

[49] Encyclopedia of Mathematics. Kolmogorov-Smirnov test. [Online]. Available: http://www.encyclopediaofmath.org/index.php?title=Kolmogorov-Smirnov_test&oldid=22659

[50] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling," *eConf*, vol. C0303241, p. MOLT007, 2003, [,186(2003)].

[51] B. Bhuyan, "High P_T Tracking Efficiency Using Partially Reconstructed D^* Decays," *Belle Note*, no. 1165, 2010.

[52] S. Nishida, "Study of Kaon and Pion Identification Using Inclusive D^* Sample," *Belle Note*, no. 779, 2005.

[53] B. Kronenbitter, "Measurement of the branching fraction of $B^+ \to \tau^+ \nu_\tau$ decays at the Belle experiment," Ph.D. dissertation, KIT, 2014. [Online]. Available: https://ekp-invenio.physik.uni-karlsruhe.de/record/48604

[54] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938. [Online]. Available: http://www.jstor.org/stable/2957648

[55] A. Ali, P. Ball *et al.*, "Comparative study of the decays $B \to (K, K^*)\ell^+\ell^-$ in the standard model and supersymmetric theories," *Phys. Rev. D*, vol. 61, p. 074024, Mar 2000. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.61.074024

List of Figures

List of Tables

Danksagung