Search for $B^0 \to \pi^- \tau^+ \nu_\tau$ with hadronic tagging at Belle

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
"Doctor rerum naturalium"
der Georg-August-Universität Göttingen

-

im Promotionsprogramm ProPhys
der Georg-August University School of Science (GAUSS)

vorgelegt von
Philipp Hamer
aus Paderborn

Göttingen, 2015

Betreuungsausschuss:

Prof. Dr. Ariane Frey
II. Physikalisches Institut, Georg-August Universität Göttingen

Prof. Dr. Arnulf Quadt
II. Physikalisches Institut, Georg-August Universität Göttingen

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Ariane Frey
              II. Physikalisches Institut, Georg-August Universität Göttingen

Korreferent: Prof. Dr. Stan Lai
              II. Physikalisches Institut, Georg-August Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Laura Covi, PhD
Institut für Theoretische Physik, Georg-August Universität Göttingen

Prof. Dr. Wolfram Kollatschny
Institut für Astrophysik, Georg-August Universität Göttingen

Prof. Dr. Arnulf Quadt
II. Physikalisches Institut, Georg-August Universität Göttingen

Prof. Dr. Steffen Schumann
II. Physikalisches Institut, Georg-August Universität Göttingen

Search for $B^0 \to \pi^- \tau^+ \nu_\tau$
with hadronic tagging at Belle

by

Philipp Hamer

Abstract

A search for the decay $B^0 \to \pi^- \tau^+ \nu_\tau$ is presented. The search is performed on the full Belle data set containing 772×10^6 $B\overline{B}$ pairs, collected at the $\Upsilon(4S)$ resonance with the Belle detector at the KEKB asymmetric energy $e^+ e^-$ collider. The τ^+ lepton is reconstructed in the decays $\tau^+ \to e^+ \nu_e \bar{\nu}_\tau$, $\tau^+ \to \mu^+ \nu_\mu \bar{\nu}_\tau$, $\tau^+ \to \pi^+ \bar{\nu}_\tau$, and $\tau^+ \to \rho^+ \bar{\nu}_\tau$. A full reconstruction algorithm based on NeuroBayes is used to reconstruct one B^0 meson, the B_{tag}, in a hadronic decay. The remainder of a signal event contains exactly two charged particles. The separation between signal and background events is performed using boosted decision trees. A fit is performed in the distribution of the extra energy in the electromagnetic calorimeter ECL, which is defined as all energy deposited by neither the B_{tag} nor the reconstructed signal final state particles. No significant signal is observed and an upper limit of $\mathcal{B}\left(B^0 \to \pi^- \tau^+ \nu_\tau\right) < 2.5 \times 10^{-4}$ is obtained at the 90% confidence level. The result is in good agreement with the Standard Model predictions.

Suche nach $B^0 \to \pi^- \tau^+ \nu_\tau$ mit hadronischem Tag bei Belle

von

Philipp Hamer

Zusammenfassung

Eine Suche nach dem Zerfall $B^0 \to \pi^- \tau^+ \nu_\tau$ wird vorgestellt. Die Suche wird auf dem vollständigen Belle Datensatz durchgeführt, der 772×10^6 $B\overline{B}$ Paare beinhaltet, die auf der $\Upsilon(4S)$ Resonanz mit dem Belle Detektor am asymmetrischen e^+e^- KEKB Beschleuniger gesammelt wurden. Das τ^+ Lepton wird in den Zerfallskanälen $\tau^+ \to e^+ \nu_e \bar{\nu}_\tau$, $\tau^+ \to \mu^+ \nu_\mu \bar{\nu}_\tau$, $\tau^+ \to \pi^+ \bar{\nu}_\tau$ und $\tau^+ \to \rho^+ \bar{\nu}_\tau$ rekonstruiert. Eines der beiden B^0 Mesonen, das B_{tag}, wird mittels eines auf NeuroBayes beruhenden Algorithmus vollständig in einem hadronischen Zerfallskanal rekonstruiert. Der Rest des Kollisionsereignisses beinhaltet genau zwei geladene Spuren im Falle eines Signalzerfalls. Die weitere Trennung zwischen Signal und Untergrund wird mithilfe von Boosted Decision Trees durchgeführt. Ein Fit wird in der Verteilung der Extra Energie im elektromagnetischen Kalorimeter ECL, welche definiert ist als alle Energie die weder vom B_{tag} noch von der rekonstruierten Signalseite stammt, durchgeführt. Kein signifikantes Signal wird beobachtet und ein oberes Limit von $\mathcal{B}\left(B^0 \to \pi^- \tau^+ \nu_\tau\right) < 2.5 \times 10^{-4}$ basierend auf einem Vertrauensintervall von 90% wird bestimmt. Das Ergebnis ist in guter Übereinstimmung mit der Vorhersage des Standard Modells.

Contents

1. Introduction

One of the most fundamental questions of mankind is the question of our origin, why do we exist, and where are we? The question is open to interpretation, though, and is probably the basis for many different scientific disciplines, that try to answer different aspects of the question: What is the purpose of life in the universe? How does life work on a biological level? How do molecules and atoms interact with each other? What is the matter part of the universe made of?

Physics is the field of science that studies the properties and interaction of matter in our universe. With more knowledge in this field, the research has evolved in many different, specialized branches. The question of the fundamental constituents of matter led to the discovery of atoms. Further research led to the discovery of electrons and atomic nuclei, then to the discovery of protons and neutrons, followed by the discovery of the up- and down-quarks. This led to the discovery of the currently known six leptons and six quarks, which are the elementary particles that build our universe. The question whether quarks and leptons are really fundamental constituents of matter, or if they are composed by other, smaller particles, is not easily answered, though. Further research will potentially provide an answer.

The elementary particles and their interactions are described by the Standard Model of elementary particle physics (SM). Its formulation is based on two of the big revolutions in theoretical physics of the last century, Einstein's special relativity and quantum mechanics, and was finalized in its current form about 40-50 years ago [1, 2]. The SM is able to describe the interaction of elementary particles and can make predictions on interaction probabilities, decay rates and other properties that can be tested at high energy experiments, such as collider experiments. Many predictions of the SM have been proven correct by these experiments. The last missing piece, the Higgs boson, has been found at the LHC by the ATLAS and CMS experiments in 2012 [3, 4].

There exist, however, limitations of the SM. First of all, we know four fundamental interactions, but only three are described by the SM. The fourth force, gravity, is not part of the SM. Furthermore, we know from cosmological observations, that the SM describes only a small part of matter and energy in the universe. A larger part is filled with what we call *dark matter* and *dark energy*, which are not described in the SM [5].

Physics beyond the Standard Model is actively searched for in all high energy experiments, directly and indirectly. Many direct searches try to find new particles via their direct decays. In order to achieve this goal, new, typically heavy particles have to be produced, so high energy collisions are needed. The ATLAS and CMS experiments at the LHC operate at the energy frontier. The LHC provides collisions with the highest, currently under lab conditions achievable, center-of-mass energy. Experiments at the intensity frontier, such as the Belle experiment, on the other hand, operate at lower energies, but produce a huge amount data of the same processes. These experiments might not able to directly detect a new particle, but may see its contribution to certain processes already known in the Standard Model. Comparing the experimental results

at a high precision with theoretical predictions in the SM, allows one to notice possible deviations which may hint at new particles.

The Belle experiment is the only detector at the KEKB e^+e^- accelerator. The leptons collide at a center-of-mass energy of $\sqrt{s} = 10.58\,\text{GeV}$, where the production of the $\Upsilon(4S)$ meson is enhanced. The $\Upsilon(4S)$ meson in turn decays exclusively into a $B\overline{B}$ pair, which is why KEKB is also called a B-factory. It has been build to study the CP violation in the B meson system, which ultimately led to the Nobel Prize for M. Kobayashi and T. Maskawa in 2008 [6]. The Belle experiment has collected the largest B meson data sample at a lepton collider, containing 772×10^6 $B\overline{B}$ pairs. The large data sample and the clean initial state of the lepton-lepton collision provide not only a perfect environment to understand CP violation, but make the Belle experiment also a perfect experiment to search for rare and experimentally challenging decays involving multiple neutrinos.

(Semi-)leptonic decays of B mesons into τ leptons are of special interest in indirect searches for models involving new, charged particles, like the H^+, whose coupling strength increase with increasing lepton mass. The measured branching fractions of the decay $B \to D^{(*)}\tau\nu$ differ from the SM prediction at a level of 3.9σ [7], which is currently one of the biggest deviations from the SM. A contribution from a charged Higgs boson should result in a similar effect in the decay $B \to \pi\tau\nu$. As the ratio $\mathcal{B}\left(B \to \pi\tau\nu\right)/\mathcal{B}\left(B \to \pi\ell\nu\right)$ is theoretically clean to compute, the decay has been proposed as a cross check [8]. The analysis presented in this thesis is the first search for the decay $B^0 \to \pi^-\tau^+\nu_\tau$ at the Belle experiment. The decay is experimentally challenging due to a small branching fraction and low reconstruction efficiencies because of the involved τ lepton, which is reconstructed in the four one-prong decay channels $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$, $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$, $\tau^+ \to \pi^+\bar{\nu}_\tau$, and $\tau^+ \to \rho^+\bar{\nu}_\tau$. This analysis is the first published result of the decay, no evidence nor upper limit on the branching ratio have been determined before.

This thesis is organized as follows. The theoretical description of the SM in general, and of the decay $B^0 \to \pi\tau\nu$, as well as possible effects from physics beyond the SM are described in Chapter 2, together with a short summary of recent results in flavor physics. The Belle experiment and the KEKB accelerator are described in detail in Chapter 3. Many different analysis techniques have been used in the analysis, among them full hadronic reconstruction of a B meson through neural networks, boosted decision trees and hypothesis tests. They are described in Chapter 4. The reconstruction of the signal decay in an e^+e^- event is then presented in detail in Chapter 5, followed by the results of the analysis in Chapter 6. The analysis is summarized in Chapter 7. Furthermore, an outlook for future experiments is given.

Notation and Conventions

Throughout this thesis, natural units are used, meaning that the speed of light in vacuum c, the reduced Planck constant $\hbar = h/2\pi$ and the Boltzmann constant k_B are set to 1:

$$c = 1 \tag{1.1a}$$

$$\hbar = 1 \tag{1.1b}$$

$$k_B = 1. \tag{1.1c}$$

All physical quantities are then expressed in terms of eV.

Three-vectors are labeled by an arrow above the symbol, like $\vec{p}$, and normal symbols are used for four-vectors, p, with the addition of greek indices, if necessary. The Minkowski metric is used as the space-time metric $g_{\mu\nu}$, contravariant four-momenta are given as $p^\mu = (E, \vec{p})$. The Einstein summation is used, meaning for two four-vectors a and b,

$$a_\mu b^\mu \equiv \sum_{\mu=0}^{3} a_\mu b^\mu. \tag{1.2}$$

In order to improve readability, the following notation is used to denote particle decays. The electric charges, and particle / anti-particle flavor, of the final state particles are usually written in formulas, but not written explicitly in sentences. It should be clear from conservation of electric charge at interaction vertices, though. The charge conjugate decay is not explicitly written but always included in the expression, such that $B^0 \to \pi\tau\nu_\tau$ denotes the decays $B^0 \to \pi^-\tau^+\nu_\tau$ and $\bar{B}^0 \to \pi^+\tau^-\bar{\nu}_\tau$.

If the electric charge of the initial state particle is not explicitly stated, both charged and neutral initial state particle and their corresponding decays are meant, e.g. $B \to D\ell\nu_\ell$ denotes $B^0 \to D^-\ell^+\nu_\ell$, $\bar{B}^0 \to D^+\ell^-\bar{\nu}_\ell$, $B^+ \to D^0\ell^+\nu_\ell$, and $B^- \to D^0\ell^-\nu_\ell$. The letter ℓ denotes the two light leptons, the electron and the muon. Therefore, the expression $B \to D\ell\nu_\ell$ truly labels eight decays, the four written above with $\ell = e$ and $\ell = \mu$.

In Feynman diagrams, time *flows* horizontally.

2. Theoretical Foundations

In this chapter, the underlying theory of elementary particle physics will be described. The Standard Model of particle physics (SM) will be introduced in Section 2.1. It comprises the theoretical description of three of the four fundamental forces in our Universe, the electromagnetic force, the weak force, and the strong force. Gravity is not included in the Standard Model as it is not relevant at small distances and the mass scale of the fundamental particles.

Recent results in flavor physics are presented in Section 2.2. Current highlights and tensions from experimental results are presented in the section. As such, it does not contain much information about the underlying theoretical description, but serves as a motivation for the search for the decay $B^0 \to \pi\tau\nu$.

The theoretical description of the semileptonic decay $B^0 \to \pi\tau\nu$ is given in the Section 2.3. Furthermore, an overview of relevant models of physics beyond the Standard Model, so-called New Physics (NP), and their possible contribution to $B^0 \to \pi\tau\nu$ are shortly discussed.

2.1. The Standard Model of Elementary Particle Physics

The Standard Model of particle physics (SM) describes the interaction between elementary particles via three of the four fundamental forces, the electromagnetic, weak and strong force. It is described well in literature, and this section follows the notation given in Refs. [1, 2, 9]. It is a Lorentz invariant, renormalizable quantum field theory invariant under local gauge transformations of the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. It successfully describes many phenomena on the elementary level.

Figure 2.1 shows all known elementary particles. The particles are grouped into fermions with spin $1/2$, the gauge bosons with spin 1, and the Higgs boson with spin 0. Their properties are listed in Table 2.1 and Table 2.2, respectively. The fermions can be grouped into quarks and leptons, which in turn can be grouped into three families each, displayed as columns in the figure. The SM does not predict a maximum number of families for neither quark nor leptons and the table might need to be expanded in the future. For each fermion, an antifermion exists. An antiparticle is related to the particle by charge conjugation which changes the sign of all internal quantum numbers. Fermions and gauge bosons gain mass through the interaction with the Higgs field, mediated by the Higgs boson.

Quarks are the only fermions that interact via the strong force. They obtain an additional degree of freedom called the color charge which is conserved in strong interactions. Three color charges exist, usually labeled r, g, and b, of which quarks carry one and gluons carry one color and one anticolor. As gluons couple to color and carry color themselves, gluon-gluon interaction is possible through 3-gluon and 4-gluon vertices.

All fermions interact via the weak force which is mediated by the charged vector bosons $W^\pm$ and
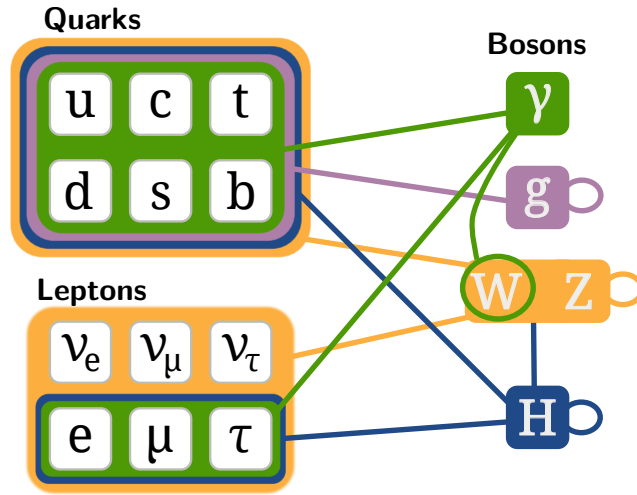
Figure 2.1.: The fundamental particles described in the Standard Model and their interaction.

the neutral Z^0. The weak force couples to the weak isospin I, which is $I = 1/2$ for left-handed fermions, $I = 0$ for right-handed fermions. The third component of the isospin I_3 is conserved in the charged weak interactions and only left-handed fermions interact with the $W^{\pm}$ meson. The quark and lepton families are formed by the weak isospin doublets where up-type quarks (u, c, t) and neutrinos are assigned $I_3 = +1/2$ while down-type quarks (d, s, b) and electrically charged leptons are assigned $I_3 = -1/2$.

Photons mediate the electromagnetic force and as such couple to electric charge. The only particles that do not interact via the electromagnetic force therefore are the photons themselves, the gluons, the Z^0, the neutrinos, and the Higgs boson.

2.1.1. Strong Interaction

The strong interaction is described by the underlying $SU(3)$ gauge symmetry. It couples to the color charge which means that quarks are the only fermions that are affected by the strong force. The eight generators of the $SU(3)$ group are the gluons, the mediators of the strong interaction. Color is conserved at strong interaction vertices. The interacting quarks do not need to carry the same color though, as gluons always carry one *unit* of color and one *unit* of anti-color. Only color singlet states are allowed in nature. Particles that carry a net charge of color can therefore not occur as free particles, which is in agreement with the fact that no free quarks or gluons are observed. Mesons always contain a quark and anti-quark with the same (anti-)color. The combination of all three colors results in a *colorless* or *white* state, as present in all baryons, for example.

The coupling strength of the strong interaction is given by $g_s = \sqrt{4\pi\alpha_s}$, with α_s being the strong coupling constant. The term *constant* is not correct as the coupling constant depends on the energy at which the interaction strength is probed, a result of the vacuum polarization. With

Fermion	Q	I_3	Color	c_A	c_V	Mass [MeV]
u: up Quark	$2/3$	$1/2$	✓	$1/2$	$1/2 - 4/3 \sin^2 \theta_W$	2.3
d: down Quark	$-1/3$	$-1/2$	✓	$-1/2$	$-1/2 - 2/3 \sin^2 \theta_W$	4.8
c: charm Quark	$2/3$	$1/2$	✓	$1/2$	$1/2 - 4/3 \sin^2 \theta_W$	1,275
s: strange Quark	$-1/3$	$-1/2$	✓	$-1/2$	$-1/2 - 2/3 \sin^2 \theta_W$	95
t: top Quark	$2/3$	$1/2$	✓	$1/2$	$1/2 - 4/3 \sin^2 \theta_W$	4,180
b: bottom Quark	$-1/3$	$-1/2$	✓	$-1/2$	$-1/2 - 2/3 \sin^2 \theta_W$	173,210
ν_e: Electron Neutrino	0	$1/2$	–	$1/2$	$1/2$	< 0.002
e: Electron	-1	$-1/2$	–	$-1/2$	$-1/2 + 2 \sin^2 \theta_W$	0.511
ν_μ: Muon Neutrino	0	$1/2$	–	$1/2$	$1/2$	< 0.19
μ: Muon	-1	$-1/2$	–	$-1/2$	$-1/2 + 2 \sin^2 \theta_W$	105.7
ν_τ: Tau Neutrino	0	$1/2$	–	$1/2$	$1/2$	< 18.2
τ: Tau	-1	$-1/2$	–	$-1/2$	$-1/2 + 2 \sin^2 \theta_W$	1,776.8

Table 2.1.: Properties of the fermions in the Standard Model. While the electric charge is a clear concept and given in units of the absolute charge of the electron, the weak isospin I, color and weak vertex factors $c_{A,V}$ are explained in the text. The masses are taken from [10].

Interaction	Boson	Gauge Coupling	Q	I_3	Color	Mass $[GeV/c^2]$
Weak	W^+	$\sqrt{4\pi\alpha}/\sin\theta_W$	$+1$	$+1/2$	-	80.835
	W^-	$\sqrt{4\pi\alpha}/\sin\theta_W$	-1	$-1/2$	-	80.835
	Z^0	$\sqrt{4\pi\alpha}/\left(\sin\theta_W \cos\theta_W\right)$	0	0	-	91.188
Electromagnetic	Photon (γ)	$\sqrt{4\pi\alpha}$	0	0	-	0
Strong	Gluon (g)	$\sqrt{4\pi\alpha_s}$	0	0	✓	0

Table 2.2.: Properties of the gauge bosons in the Standard Model. grouped by the interaction they mediate. The parameters α, α_s and θ_W are described below in Sections 2.1.1 and 2.1.2. The electric charge is denoted by Q while the third component of the weak isospin is I_3. The masses are taken from [10].

Q being the momentum transfer at the interaction vertex, $\alpha_s(Q^2)$ is given by

$$\alpha_s\left(Q^2\right) = \frac{12\pi}{(33 - 2n_f)\log\left(Q^2/\Lambda^2\right)}, \tag{2.1}$$

with n_f being the number of quark flavors and Λ being the QCD cut-off parameter. Depending on n_f, the value of Λ is in the range of $\approx 200\,\text{MeV}$[10] and QCD interactions can be calculated perturbatively only for $Q^2 \gg \Lambda^2$. Virtual quark anti-quark production and annihilation processes result in an increase of the coupling strength at high energies, while gluon-gluon loop corrections result in a decrease. The latter effect dominates the corrections at high energies which results in a property of the strong force called asymptotic freedom at high energies. In this regime, $\lim_{Q^2 \to \infty} \alpha_s\left(Q^2\right) \to 0$, and a pertubative approach can be used to mathematically describe strong interaction. At low energies, the value of α_s increases. If the distance between two quarks gets too large, enough energy is released to create new quark anti-quark pairs, with the result that no free quarks can be observed. Instead, bound quark states, the mesons, are

created. This effect is called confinement.

The Z^0 mass is often chosen as a reference point for $|Q|$ and the world average of α_s at this momentum transfer is [10]

$$\alpha_s \left(m_Z^2 \right) = 0.1185 \pm 0.0006. \tag{2.2}$$

2.1.2. Electroweak Interaction

The electromagnetic and weak interaction have been unified in the electroweak interaction by Glashow, Weinberg and Salam which is why the unified description is often referred to as the GWS model [11–13]. The underlying gauge symmetry is the $SU(2)_L \times U(1)_Y$ group, where the subscript L indicates that the $SU(2)$ interactions couple only to left-handed particles, while Y denotes the weak hypercharge $Y = 2\left(Q - I_3\right)$ with the electric charge Q. The three generators of the non-Abelian $SU(2)$ provide the fields $W_1^\mu, W_2^\mu, W_3^\mu$ and the generator of the Abelian $U(1)$ group provides the field B^μ.

The Lagrangian of the weak interaction may not contain explicit mass terms for the fermion and boson fields in order to maintain local gauge invariance. The mediator bosons of the weak interaction, $W^\pm$ and Z^0, are massive though. The electroweak symmetry breaking through the Higgs mechanism, described further below, solves this problem. As a result of the electroweak symmetry breaking, the four fields of the electroweak interaction mix to yield the four physically observable gauge bosons $W^\pm, Z^0$ and the photon, denoted A here:

$$\begin{pmatrix} W_\mu^+ \\ W_\mu^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} W_\mu^1 \\ W_\mu^2 \end{pmatrix} \tag{2.3a}$$

$$\begin{pmatrix} A_\mu \\ Z_\mu^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}. \tag{2.3b}$$

The angle θ_W is the weak mixing angle, also called Weinberg angle, defined in terms of the coupling constants of the $U(1)_Y$ and $SU(2)_L$ interaction, g' and g, respectively,

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}. \tag{2.4}$$

Its value depends on the renormalization scheme and is $\sin^2\theta_W(M_Z) = 0.23126(5)$ [10] at the Z^0 mass.

Electromagnetic Interaction

The electromagnetic interaction in the GWS model is described by *Quantum Electrodynamics*, QED. The underlying symmetry is the Abelian $U(1)$ group. The physical result of the Abelian property is that photons do not couple to each other. As in QCD, the loop corrections lead to a scale dependency of the electromagnetic coupling constant. Only fermion-boson interactions, namely pair production and annihilation processes, are possible contributors to the loop corrections which means that only *shielding* effects are observed. At the vertex four-momentum transfer Q and a value at the reference momentum μ by

$$\alpha_{\mathrm{em}} \left(Q^2, \mu^2 \right) = \frac{\alpha(\mu^2)}{1 - \frac{\alpha(\mu^2)}{3\pi} \log\left(\frac{Q^2}{\mu^2}\right)}, \tag{2.5}$$

which is valid for $Q^2 \gg \mu^2$. At very low $Q^2 \to 0$, $\alpha_{\text{em}} \approx 1/137$ is known as the fine-structure constant. Its value increases with higher Q^2 such that $\alpha_{\text{em}}(m_W^2) \approx 1/128$ [10].

Weak Interaction

The weak interaction is mediated by the massive vector bosons $W^\pm$ and Z^0. Based on the charge of the mediating particles, the interactions are also called charged and neutral weak interactions, respectively. The weak isospin I is the corresponding *charge* of the weak interaction which is described mathematically by the non-Abelian $SU(2)$ gauge symmetry. An important property of the weak interaction is that it violates parity, and the chirality of the fermions play an important role in the interaction. The chirality, also called handedness, is defined as the eigenvalue of the chirality operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, where γ^μ are the Dirac matrices. A fermion can be decomposed into the left- and right-handed state and written as

$$\psi = \psi_L + \psi_R \tag{2.6}$$

by applying the projection operators

$$P_L = \frac{1}{2}\left(1 - \gamma^5\right), \quad P_L\psi = \psi_L \tag{2.7a}$$

$$P_R = \frac{1}{2}\left(1 + \gamma^5\right), \quad P_R\psi = \psi_R. \tag{2.7b}$$

The fermions of the Standard Model are then be grouped into weak isospin doublets and singlets

$$
\begin{array}{ccccccc}
I_3 & = +\frac{1}{2} & \begin{pmatrix}\nu_e\\e^-\end{pmatrix}_L & \begin{pmatrix}\nu_\mu\\\mu^-\end{pmatrix}_L & \begin{pmatrix}\nu_\tau\\\tau^-\end{pmatrix}_L & \begin{pmatrix}u\\d'\end{pmatrix}_L & \begin{pmatrix}c\\s'\end{pmatrix}_L & \begin{pmatrix}t\\b'\end{pmatrix}_L \\
I_3 & = -\frac{1}{2} & & & & & & \\
& I_3 = 0 & e_R^- & \mu_R^- & \tau_R^- & u_R, d_R & c_R, s_R & t_R, b_R.
\end{array}
$$

Only left-handed fermions take part in the charged weak interaction, while the electromagnetic and strong interaction do not differ between the chirality states. The inverse is true for anti-particles. Only right-handed anti-fermions interact weakly if a charged $W^\pm$ is involved. This property of the charged weak current is described by the vector - axial vector $(V - A)$ structure. Therefore, not only parity P is violated by the weak interaction, charge-conjugation C is violated, too, as $C\psi_L = \overline{\psi}_L$. The combination of both transformations is conserved in most weak interactions, though, as $CP\psi_L = \overline{\psi}_R$. Neutral weak interactions mediated by the Z^0 do not possess a strict $V - A$ structure in the quark sector, but still prefer to couple to left-handed fermions. The mathematical description of the charged and neutral weak interaction vertex is

$$\frac{-ig_w}{2\sqrt{2}}\gamma^\mu\left(1 - \gamma^5\right) \qquad \text{and} \tag{2.8a}$$

$$\frac{-ig_z}{2}\gamma^\mu\left(c_V^f - c_A^f\gamma^5\right), \tag{2.8b}$$

respectively, where $g_{w,z}$ are the coupling constants and $c_{A,V}^f$ depend on the fermion flavor f. The *vector* part of the interaction is described by γ^μ, while $\gamma^\mu\gamma^5$ describes the axial vector part. The values of $c_{A,V}^f$ are listed in Table 2.1. The coupling constants are related to the electromagnetic coupling constant g_{em} via the Weinberg angle by $g_w = g_{\text{em}}/\sin\theta_W$ and $g_z = g_{\text{em}}/\cos\theta_W\sin\theta_W$.

In order to describe weak decays like $B^0 \to \pi\tau\nu$, the propagator term of the $W^\pm$-boson is needed. It is given by

$$-i\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{m_W^2}\right)\frac{1}{q^2 - m_W^2}, \tag{2.9}$$

where $g_{\mu\nu}$ is the metric tensor, q is the transfered 4-momentum and m_W is the mass of the W-boson. In processes with $|q^2| \ll m_W^2$, as in the case of $B^0 \to \pi\tau\nu$, the propagator term can be approximated by

$$i\frac{g_{\mu\nu}}{m_W^2}. \tag{2.10}$$

In this limit, the weak interaction can be described as a point like four-fermion interaction, originally described by Fermi [14]. In this description, the Fermi constant $G_F = 1.1663787(6) \times 10^{-5}\,\mathrm{GeV}^{-2}$ [10] can be used to describe the matrix element of the weak interaction,
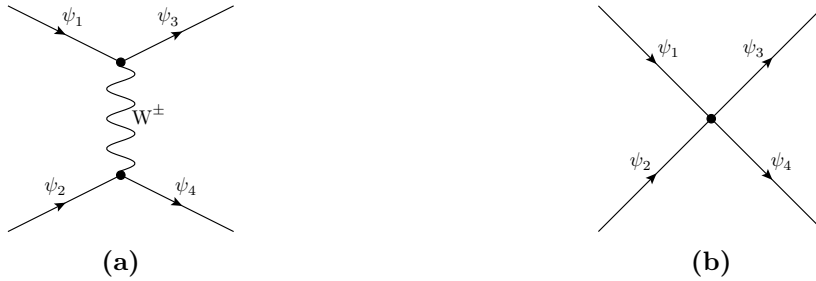
$$\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8m_W^2}. \tag{2.11}$$

The reduction of a charged weak process to a four-fermion interaction is shown as two diagrams in Figure 2.2. With the initial state i, final state f, and Dirac spinors ψ_j, the matrix element of the $i \to f$ transition in both diagrams is given by

$$\mathcal{M}_{fi} = -i\left[\frac{g_W}{\sqrt{2}}\overline{\psi}_3\frac{1}{2}\gamma^\mu(1-\gamma^5)\psi_1\right]\left[\frac{g_{\mu\nu} - q_\mu q_\nu/m_W^2}{q^2 - m_W^2}\right]\left[\frac{g_W}{\sqrt{2}}\overline{\psi}_4\frac{1}{2}\gamma^\nu(1-\gamma^5)\psi_2\right] \tag{2.12a}$$

$$\xrightarrow{q^2 \ll m_W^2} \mathcal{M}_{fi} = i\frac{G_F}{\sqrt{2}}g_{\mu\nu}\left[\overline{\psi}_3\gamma^\mu(1-\gamma^5)\psi_1\right]\left[\overline{\psi}_4\gamma^\nu(1-\gamma^5)\psi_2\right], \tag{2.12b}$$

respectively.

**(a)**   **(b)**

Figure 2.2.: (a) A charged weak scattering process $\psi_1 + \psi_2 \to \psi_3 + \psi_4$ with the exchange of a $W^\pm$-boson. In the limit of low momentum transfer $|q^2| \ll m_W^2$, the process can be described by an effective point like four-fermion interaction, as shown in (b).

The weak isospin eigenstates d', s', b' are superpositions of quark mass eigenstates. As the charged weak force interacts with the weak isospin doublets, a quark flavor change is possible in this type of interaction. The weak isospin eigenstates are described in terms of the mass eigenstates by the Cabibbo-Kobayashi-Maskawa (CKM) Matrix V_{CKM} [15, 16] as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}\begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.13}$$

The CKM matrix is unitary in the SM, which reduces the number of free parameters to three mixing angles and a complex phase. The complex phase introduces a description of CP violation in the SM. Matrix elements for processes involving quark q_i to quark q_j transitions include the CKM matrix element V_{ij}. As a result, $|V_{ij}|^2$ can be interpreted as the probability for a transition $i \to j$ to happen. Transitions in the same quark family are preferred with respect to cross-family transitions. The absolute values $|V_{ij}|$ [10] are shown below in Equation (2.14) and shown graphically in Figure 2.3.

$$|V_{ij}| = \begin{pmatrix} 0.97525 \pm 0.00022 & 0.2254 \pm 0.008 & 0.00413 \pm 0.00049 \\ 0.225 \pm 0.008 & 0.986 \pm 0.016 & 0.0411 \pm 0.0013 \\ 0.0084 \pm 0.0006 & 0.0400 \pm 0.0027 & 1.021 \pm 0.032 \end{pmatrix} \quad (2.14)$$
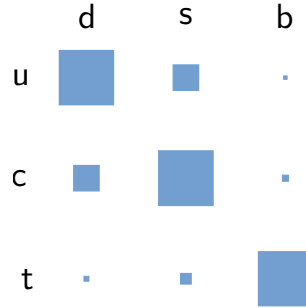
Figure 2.3.: Graphical representation of the absolute values of the CKM Matrix elements. The area of each square is equal to $|V_{ij}|$, values taken from [10].

In the analysis presented here, a search for the decay $B^0 \to \pi\tau\nu$ is performed. The B^0 decays via the weak $b \to u$ transition. A $b \to c$ transition is roughly 100 times more likely to happen, as can be seen in Equation (2.14).

The CKM matrix only covers the charged currents in the quark sector. Flavor changing neutral currents (FCNC) have not been observed and are not part of the Standard Model. In the lepton sector, mixing has been observed between the neutrinos, described by the PMNS matrix [17–19].

### 2.1.3. Electroweak symmetry breaking

As stated above, the electroweak $SU(2)_L \times U(1)_Y$ symmetry is broken which results in the massive gauge bosons $W^{\pm}$ and Z^0. The symmetry breaking is implemented through a mechanism developed by Brout, Englert, Higgs, Guralnik, Hagen and Kibble at around the same time [20–22]. Four real scalar fields $\phi_{1..4}$ are introduced in a $SU(2)_L \times U(1)_Y$ gauge invariant Lagrangian. They are arranged in a weak isospin doublet ϕ with weak hypercharge $Y = 1$, proposed by Weinberg [12], as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (2.15)$$

with the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi \right)^2, \quad (2.16)$$

where $\mu^2 < 0$ and $\lambda > 0$. The form of the Higgs potential is shown in Figure 2.4.

(a) **(b)**
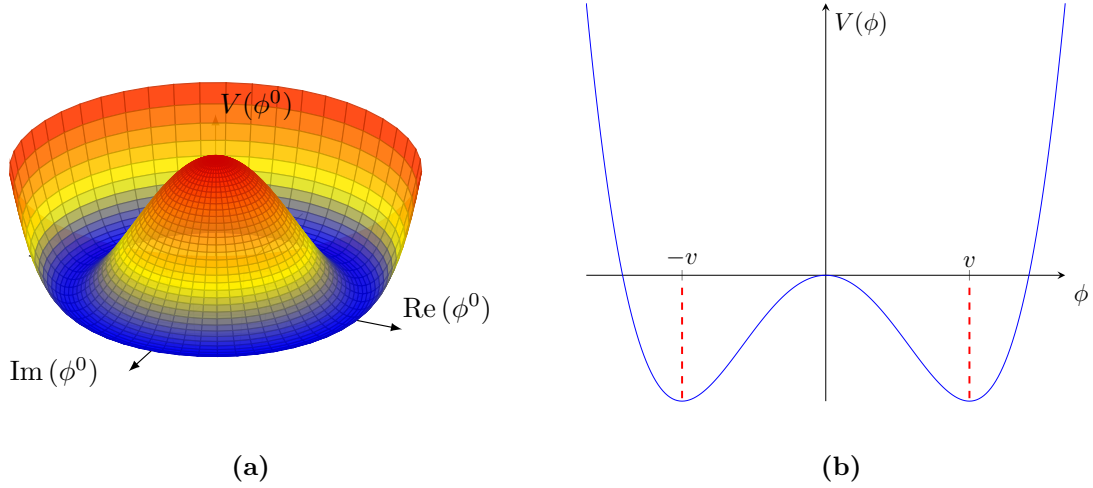
Figure 2.4.: Sketches of the Higgs potential given in Equation (2.16), as a three dimensional sketch in (a) and in one dimension in (b). The sketches are included to better illustrate the shift of the vacuum expectation value v, but do not represent the actual potential.

As can be seen, the minimum is not at $|\phi| = 0$ but instead at

$$|\phi|_{\min} = \sqrt{\frac{-\mu^2}{2\lambda}} = v, \tag{2.17}$$

with v being the vacuum expectation value. $\phi(x)$ is expanded around the minimum chosen at

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.18}$$

The expansion introduces a massive scalar particle, the Higgs boson, with mass $m_h^2 = 2v^2\lambda$, and three massless Goldstone bosons. Through an additional gauge, the Higgs mechanism, the three Goldstone bosons get absorbed as the longitudinal polarization of the already known gauge fields W^1, W^2 and W^3. The vacuum expectation value v of the Higgs field is related to the Fermi constant $G_F = 1.1663787(6) \times 10^{-5}\,\text{GeV}^{-2}$ [10] and thus can be calculated to

$$v^2 = \frac{1}{\sqrt{2}G_F} \qquad \Rightarrow v = 246.22\,\text{GeV}. \tag{2.19}$$

The Higgs mechanism directly produces mass terms for the $W^\pm$ and Z^0 with the help of the weak couplings g and g' introduced in Equation (2.4),

$$M_W = \frac{1}{2}vg \tag{2.20a}$$

$$M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \tag{2.20b}$$

$$\Rightarrow \frac{M_W}{M_Z} = \cos\theta_W. \tag{2.20c}$$

In the Standard Model, the fermions, except for neutrinos, obtain mass through a Yukawa coupling λ_f to the Higgs field, where f denotes the fermion flavor. Neutrinos are massless in the SM, though. As the fermions gain mass through the coupling, the coupling constant λ_f is related to the fermion mass by

$$\lambda_f = \sqrt{2}\frac{m_f}{v} \tag{2.21a}$$

$$\Leftrightarrow m_f = \frac{\lambda_f v}{\sqrt{2}}. \tag{2.21b}$$

The higher the mass of the fermion, the stronger the coupling to the Higgs field. With the world-average of m_{top} and G_F [10], the strongest coupling λ_f is the coupling to the top quark with $\lambda_{\text{top}} = 0.995$.

The Higgs boson was discovered in 2012 by the ATLAS [3] and CMS [4] experiments, located at the Large Hadron Collider (LHC) in Geneva. The latest combined measurements of the Higgs boson mass from ATLAS and CMS in the decay channels $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ results in $m_H = 125.09 \pm 0.21 \pm 0.11\,\text{GeV}$ [23]. Further studies [24, 25] show that spin 0 and positive parity of the discovered particle are the most likely hypotheses, which is in agreement with the nature of the SM Higgs boson.

2.1.4. Limitations of the Standard Model

The Standard Model of particle physics is a very successful model of the most fundamental properties of our universe. There are, however, limitations in it, some of which will be listed here. It might be the case that the electroweak and strong force belong to the same underlying, unifying force which is broken similarly to how the electroweak force is broken into the electromagnetic and weak force. Furthermore, the gravitational force is not included in the theory. These are mostly aesthetic arguments though.

Hints for neutrino oscillations have been seen in 2001 by the Super-Kamiokande experiment [26] and the Sudbury Neutrino Observatory (SNO) [27]. One year later, the SNO confirmed the existence of oscillations between neutrino flavor states [28]. Neutrinos need to have a non-vanishing mass for oscillations to occur, and they have been found to be very small, see Table 2.1, such that the massless approximation works well in many SM calculations. While there are possibilities to include neutrino masses in the SM, neutrinos would not obtain mass from interactions with the Higgs boson; not without any modification to the SM neutrinos, see for example the review *Neutrino Mass, Mixing, and Oscillations* in [10].

It is assumed that the same amount of matter and antimatter has been created in the Big Bang. Obviously, not all particles annihilated themselves with their anti-partner. Furthermore, astrophysical observations show no hint of local preference of antimatter or matter in some regions. While CP violation in the SM through the complex phase in the CKM matrix is able to explain different behavior between matter and antimatter, the effect is too small to explain the difference observed in the universe [29].

Observations in astrophysics show hints for dark matter and dark energy. The description *dark* describes the fact that they both do not interact via the electromagnetic force. Dark matter is hypothesized in order to explain the orbital velocities of galaxies in clusters [30]. Dark energy on the other hand is hypothesized in order to explain the accelerated expansion of the universe [31].

Recent measurements from Planck [5] estimate that 84.5% of all matter in the universe is dark matter, while dark matter plus dark energy make up for 95.1% of the total mass-energy content of the universe. Neither dark matter nor dark energy are included in the SM.

2.2. Recent Results in Flavor Physics

The search for New Physics can be done directly and indirectly. New, massive particles can be created at high energy experiments like the LHC and future experiments at the energy frontier, and thus allow for a direct search of NP particles. An alternative are low energy, precision measurements of certain decay processes. Optimal processes are decays that are theoretically predictable and experimentally measureable with a small uncertainty. Discrepancies between the theoretical prediction and the experimental result may occur and provide indirect hints for NP. The field of flavor physics investigates, among others, the decays of Kaons and B mesons. The purpose of this section is to present some of the recent discrepancies around the 3σ level between theory predictions and experimental results and thereby provide a motivation for the search for the decay $B^0 \to \pi\tau\nu$.

The CKM matrix elements $|V_{cb}|$ and $|V_{ub}|$ can be determined by measuring exclusive decays like $B \to \pi\ell\nu_\ell$, or inclusively by measuring all $B \to X_u\ell\nu_\ell$ decays. The results of both methods differ by roughly 3σ though, shown in Table 2.3. Preliminary results of exclusive measurement of $|V_{cb}|$ by Belle via $\mathcal{B}(B \to D\ell\nu_\ell)$, use improved models and lattice QCD results and are in better agreement with the value obtained from inclusive determination [32].

Recently, measurements of $b \to s\ell^+\ell^-$ processes, have produced intriguing results. These decays are forbidden at tree level in the SM, but can proceed via box or penguin diagrams, shown in Figure 2.5. New, heavy particles may contribute in these diagrams and affect the decay process. The decay rates for the processes shown below are quite low, in the order of $\mathcal{O}(10^{-7})$. Furthermore, the theoretical computation is more difficult than tree-level decays. In order to better understand the deviations from the SM, both theory calculations have to be improved, and more data has to be taken.

In the SM, the gauge bosons couple to all leptons with the same coupling strength, a phenomenon called lepton universality. Recent measurements of the ratio R_K of the branching fractions $\mathcal{B}(B^+ \to K^+\mu^+\mu^-)$ and $\mathcal{B}(B^+ \to K^+e^+e^-)$ by LHCb [33], show a deviation of 2.6σ. Earlier measurements by Belle [34] and BaBar [35] are in agreement with the SM prediction, but have a higher statistical uncertainty. The results of all experiments are shown in Figure 2.6a. The result of $\mathcal{B}(B^+ \to K^+e^+e^-)$ alone is compatible with the SM prediction, according to LHCb, though, so the discrepancy is likely to originate from the $b \to s\mu^+\mu^-$ transition.

The angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ shows deviations from the SM predictions, too. The

	$|V_{cb}|$	$|V_{ub}|$
Inclusive	$(42.2 \pm 0.7) \times 10^{-3}$	$(4.41 \pm 0.15 \, ^{+0.15}_{-0.17}) \times 10^{-3}$
Exclusive	$(39.5 \pm 0.8) \times 10^{-3}$	$(3.28 \pm 0.29) \times 10^{-3}$

Table 2.3.: World averages of $|V_{cb}|$ and $|V_{ub}|$, obtained from inclusive and exclusive determinations [10]. The new, preliminary result from Belle is not yet included in the exclusive value of $|V_{cb}|$.

theoretical calculation of the angular distribution of the observable P_5' does not depend heavily on a good understanding of the involved hadronic form factors and is therefore a useful observable to test the SM. It is a composite variable constructed from observables of the angular analysis of the branching fraction and is defined in [36, 37]. The LHCb experiment measured P_5' in bins of the mass squared of the muon pair, q^2, and compared the results against recent theory calculations [38]. They obtain agreement with the SM calculation in the low q^2 region, but observe a deviation at the 3.7σ level in the region $4.0 \, \text{GeV}^2 < q^2 < 8.0 \, \text{GeV}^2$. The results are shown in Figure 2.6b.

A third deviation in a $b \to s\ell^+\ell^-$ process has been observed in the differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$ by LHCb. The difference between theory predictions [39, 40] and experiment in the region $1.0 \, \text{GeV}^2 < q^2 < 6 \, \text{GeV}^2$, where precise theoretical calculations are available, is found to be at the 3.5σ level. The results are shown in Figure 2.6c.
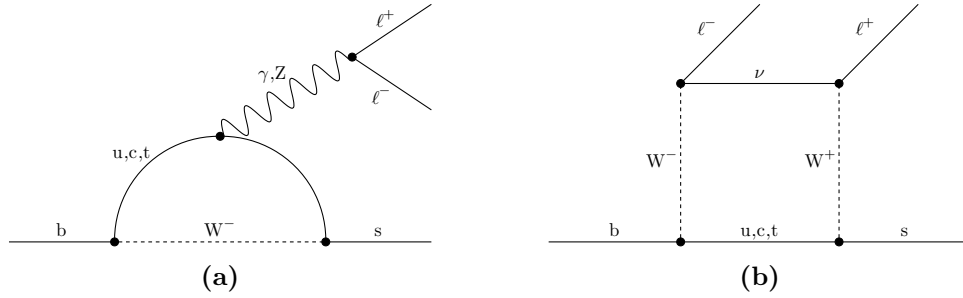

Figure 2.5.: (a) Penguin and (b) box Feynman diagram for the $b \to s\ell^+\ell^-$ process in the SM.

As mentioned above, the $b \to s\ell^+\ell^-$ transitions are higher order processes described by penguin or box diagrams, and therefore challenging both theoretically and experimentally. There are, however, also tree-level decays that are sensitive to NP scenarios. Models that include a charged Higgs boson can influence (semi-)leptonic B decays into a $\tau\nu_\tau$ pair, as for example $B^0 \to \pi\tau\nu$. At the tree-level, these decays are theoretically clean. On the other hand, the decays are experimentally challenging. Due to the short lifetime of the τ lepton, it decays inside of the detector, and has to be reconstructed from its decay products. The final state therefore contains 2-3 neutrinos, which are not directly detectable, but result in missing momentum. Advanced reconstruction techniques have to be applied, as will be described in more detail in Chapters 4 and 5. While the branching fraction of the decays is in the order of $\mathcal{O}(10^{-4})$ or higher, the reconstruction efficiency is usually quite low.

Two decays involving a τ lepton have been studied by Belle and BaBar, before, and will briefly be presented here. First measurements of the branching fraction $\mathcal{B}\left(B^+ \to \tau^+\nu_\tau\right)$ showed a deviation from the SM prediction in the order of 2σ [43–45]. However, more recent measurements by Belle [46, 47] using improved analysis methods are in good agreement with the SM.

The measurement of the ratio $R(D^{(*)}) = \mathcal{B}\left(B \to D^{(*)}\tau\nu_\tau\right)/\mathcal{B}\left(B \to D^{(*)}\ell\nu_\ell\right)$ with ℓ being a light lepton, $\ell = e$ or μ, by BaBar [48] showed a deviation of 2.4σ from the theory prediction. While new results from Belle [49] are in better agreement with the SM, LHCb [50] sees a similar disagreement in $R(D^*)$, but did not measure $R(D)$. Belle and BaBar reconstruct the τ lepton into $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$ and $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$, while LHCb uses $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$, only. The Heavy Flavor Averaging Group (HFAG) prepared a combination of the three results for the EPS-HEP conference 2015 [7]. $R(D)$ and $R(D^*)$ exceed the SM predictions by 1.7σ and 3.0σ, respectively. Combining both measurements shows a deviation from the SM prediction at the 3.9σ level.
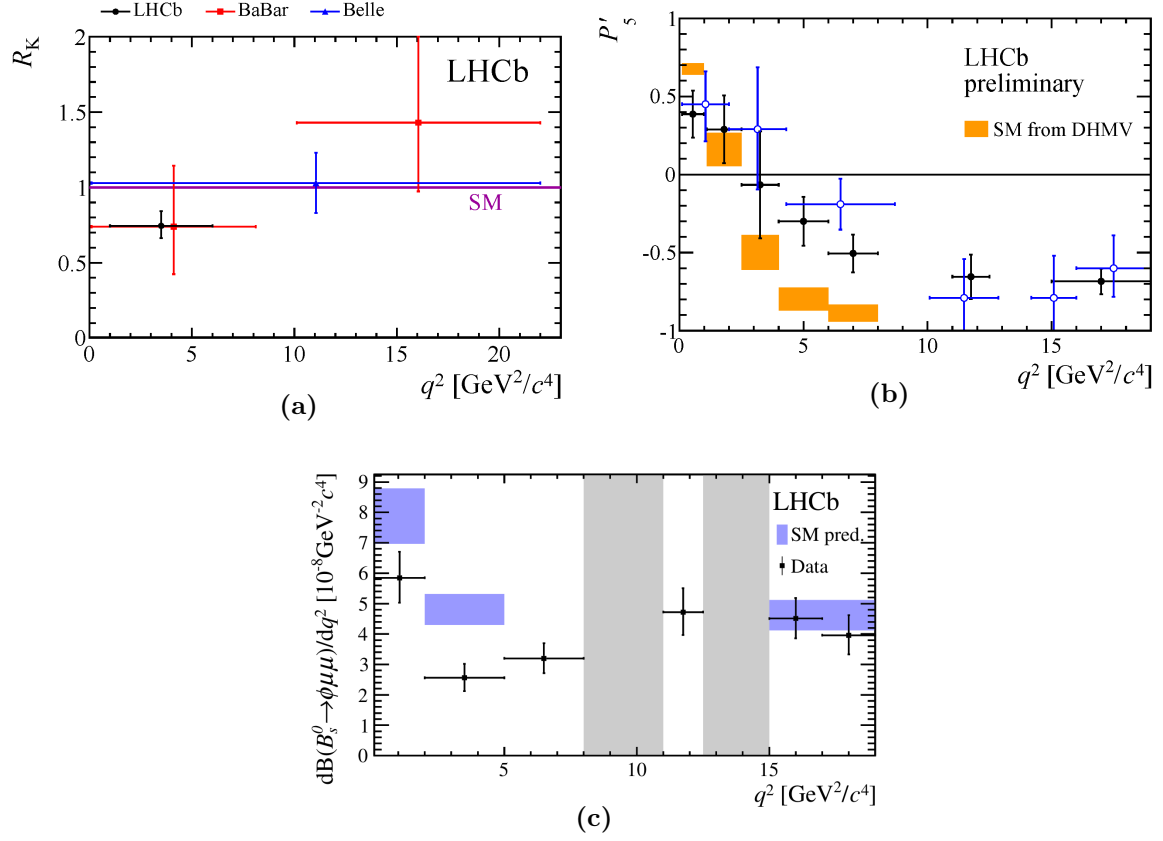
Figure 2.6.: (a) Results of R_K by BaBar, Belle, and LHCb. Plot taken from Ref. [41]. (b) Results on P'_5 by LHCb. The figure includes the results on the 2011 data (blue) and the results on the full LHC run 1 dataset (black). Plot taken from Ref. [42]. (c) Measurement of the differential branching fraction $d\mathcal{B}\left(B_s^0 \to \phi\mu^+\mu^-\right)/dq^2$ by LHCb. The grey areas indicate vetoes to exclude charmonium resonances.

The results on both observables, as well as the combination of both observables is shown in Figure 2.7.

The analysis presented in this thesis is the search for $B^0 \to \pi^- \tau^+ \nu_\tau$ at the Belle experiment and will be described in more detail below.
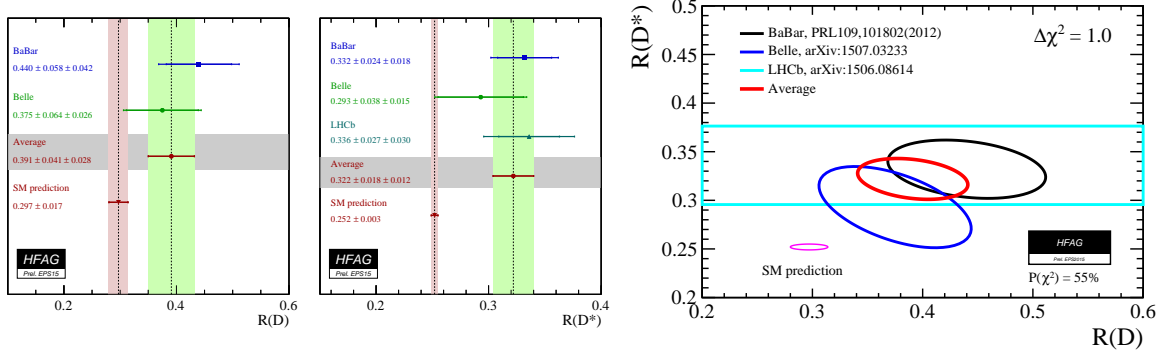
Figure 2.7.: Average of $R(D)$ and $R(D^*)$ and of the combination for the EPS-HEP 2015 by the HFAG [7]. See the text for an explanation of how the combination is obtained.
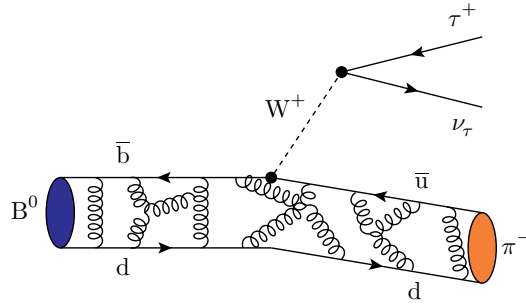
2.3. The decay $B \to \pi\tau\nu$

The section gives a brief overview of the theoretical description of the decay $B^0 \to \pi\tau\nu$ in the Standard Model. Possible effects of physics beyond the Standard Model will be listed in Section 2.3.2.

2.3.1. Standard Model

The notation in this section follows the review article published in [51].

The decay of $B^0 \to \pi\tau\nu$ is a weak process, mediated by the exchange of a $W^{\pm}$-boson. Figure 2.8 shows the Feynman graph of the decay. The b-quark decays into an u-quark by emitting a W-boson with the appropriate charge, given by the initial B^0 or $\bar{B}^0$, followed by the decay of the W-boson into a τ-ν_τ pair.

Figure 2.8.: Feynman graph of the decay $B^0 \to \pi\tau\nu$.

The decay rate for the multi-body decay $B \to d_1 + d_2 + d_3 + \dots$ can be computed using Fermi's Golden Rule [2]

$$\Gamma(B \to \sum_i d_i) = \frac{1}{2m_B} \left| \mathcal{M}(B \to \sum_i d_i) \right|^2 d\text{LIPS}, \tag{2.22}$$

where m_B is the mass of the B-meson, $\mathcal{M}$ is the matrix element of the transition and dLIPS is

the Lorentz-invariant phase space given by

$$d\text{LIPS} = (2\pi)^4 \delta^{(4)}(p_B - \sum_i p_i) \prod_i \frac{d^3\vec{p_i}}{(2\pi)^3 2E_i}. \tag{2.23}$$

Here, p_B is the four-momentum of the decaying B-meson and $p_i = (E_i, \vec{p_i})$ are the four-momenta of the final state particles.

The four-momentum transfer q is defined as $q = p_B - p_\pi$, leading to the Lorentz-invariant term $q^2 = (p_B - p_\pi)^2$. The minimal value of q^2 is obtained when the lepton pair is at rest, that is $q^2_{min} = m_\tau^2$, while the maximum value is reached when the pion gains no kinetic energy, $q^2_{max} = (m_B - m_\pi)^2$. In the allowed range $3.16\,\text{GeV}^2 < q^2 < 26.43\,\text{GeV}^2$, the relation $q^2 \ll m_W^2$ is valid, and using Equation (2.12b), the matrix element can be written as the product of the weak currents

$$\mathcal{M}\left(B^0 \to \pi\tau\nu\right) = -i\frac{G_F}{\sqrt{2}}V_{ub}L^\mu H_\mu. \tag{2.24}$$

The leptonic and hadronic current are given by

$$L^\mu = \overline{u}_\tau \gamma^\mu (1 - \gamma^5) v_\nu \tag{2.25a}$$

$$H_\mu = \langle \pi^+ | \overline{u}\gamma_\mu (1 - \gamma_5) b | \overline{B} \rangle, \tag{2.25b}$$

where u_τ and v_ν are Dirac spinors. The leptonic current can be computed easily, while the hadronic current carries QCD contributions, as is also shown in Figure 2.8. Both the B-meson and the π-meson are pseudoscalar mesons. From Lorentz invariance and parity considerations one can show [52, 53], that $P \to P$ transitions have no axial-vector contributions. With p being the momentum of the pion, the hadronic current is usually written in terms of the vector and scalar form factors f^+ and f^0,

$$\langle \pi(p) | \overline{u}\gamma_\mu b | \overline{B}(p+q) \rangle = f^+(q^2) \left[2p_\mu + \left(1 - \frac{m_B^2 - m_\pi^2}{q^2} \right) q_\mu \right] + f^0(q^2) \frac{m_B^2 - m_\pi^2}{q^2} q_\mu, \tag{2.26}$$

with $f^+(0) = f^0(0)$. Both form factors can be computed using experimental input. The distance of the interaction in mesons is around 1 fm, a scale at which α_s is too large to perform pertubative QCD calculations. Lattice QCD [54–56] enables calculations in the non-pertubative regime by performing numerical calculations on a discrete four dimensional Euclidean space-time lattice with distance a. The fermion fields are represented on the lattice points, while the links between the lattice points represent the gauge fields. The computation time on the lattice is one of the major limitations, constraining the lattice spacing a. The form factors can only be computed in the upper $q^2 \geq 16\,\text{GeV}^2$ range using lattice QCD. A more detailed description of lattice QCD techniques is beyond the scope of this work, the interested reader may find more information in [57, 58]. In the lower region, $0 \leq q^2 \leq \tilde{q}^2_{max}$ with $\tilde{q}^2_{max}$ ranging from $12 - 16\,\text{GeV}^2$, the form factors are computed using QCD light-cone sum rules (LCSR) [8, 59, 60]. LCSR allow a continous calculation in contrast to the discretized calculation on the lattice. The original QCD sum rules by Shifman, Vainshtein and Zakharov [61] (SFV sum rules) are applied by performing an operator product expansion (OPE) of the $B \to \pi$ correlation function near the light-cone $x^2 \approx 0$. Reviews and more detailed descriptions of LCSR calculations for the $B \to \pi$ form factors can be found at [62–64] and the references therein.

The form factors can be described analytically in the whole valid q^2 region by making use of the so-called z expansion [65]. The variable q^2 is mapped to a new variable z by

$$z(q^2, t_0) = \frac{\sqrt{(m_B + m_\pi)^2 - q^2} - \sqrt{(m_B + m_\pi)^2 - t_0}}{\sqrt{(m_B + m_\pi)^2 - q^2} + \sqrt{(m_B + m_\pi)^2 - t_0}} \tag{2.27}$$

with m_B and m_π being the B- and π-meson mass. The free parameter t_0 is used to limit the maximum value of z and set to $t_0 = (m_B + m_\pi)^2 - 2\sqrt{m_B m_\pi}\sqrt{(m_B + m_\pi)^2 - q_{\min}^2}$ in [8, 56]. This limits the value to $|z| \lesssim 0.3$ in the case of $B \to \pi\ell\nu$, depending on the choice of $q_{\min}^2$. The two form factors are then expressed in terms of $z = z(q^2, t_0)$, known as the Bourrely-Caprini-Lellouch (BCL) expansion [66],

$$f^+(q^2) = \frac{1}{1 - q^2/m_{B^*}^2} \sum_{k=0}^{N-1} b_k^+ \left[z^k - (-1)^{k-N} \frac{k}{N} z^N \right] \tag{2.28a}$$

$$f^0(q^2) = \sum_{k=0}^{N} b_k^0 z^k. \tag{2.28b}$$

Plots of $f^+(q^2)$ and $f^0(q^2)$ in the valid q^2 range are shown in Figure 2.9.
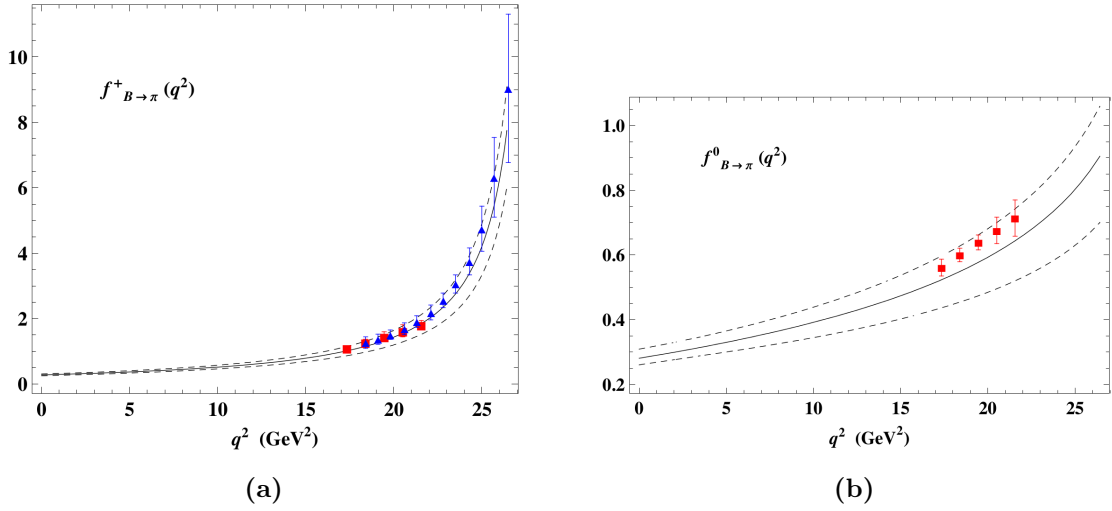
Figure 2.9.: The vector (a) and scalar (b) form factors, calculated from LCSR and fitted to the BCL parametrization (solid) with uncertainties (dashed). The results are compared with results from HPQCD [54] (red squares) and FNAL/MILC [55] (blue triangles). Plot taken from Khodjamirian, *et al* [8].

With the knowledge of the form factors, the branching fraction of $B^0 \to \pi\tau\nu$ can be computed. Calculations based on the form factors calculated form LCSR [8] have been performed by Dutta, *et al* [67], resulting in a SM prediction of

$$\mathcal{B}(B^0 \to \pi\tau\nu) = (8.91^{+6.49}_{-3.98}) \times 10^{-5}. \tag{2.29}$$

The uncertainty on the result is relatively large. While not exactly stated in the publication, part of the uncertainty might be due to the large uncertainty of $|V_{ub}|$ and the discrepancy of the inclusive and exclusive determination, being $|V_{ub}| = (4.41 \pm 0.15^{+0.15}_{-0.17}) \times 10^{-3}$ and $|V_{ub}| = (3.28 \pm 0.29) \times 10^{-3}$, respectively [10].

$|V_{ub}|$ cancels out when computing the ratio

$$\frac{d\Gamma(B \to \pi\tau\nu_\tau)/dq^2}{d\Gamma(B \to \pi\ell\nu_\ell)/dq^2} = \frac{(q^2 - m_\tau^2)^2}{(q^2)^2} \left(1 + \frac{m_\tau^2}{2q^2} \right) \left\{ 1 + \frac{3m_\tau^2(m_B^2 - m_\pi^2)^2}{4(m_\tau^2 + 2q^2)m_B^2 p_\pi^2} \frac{|f^0(q^2)|^2}{|f^+(q^2)|^2} \right\} \quad , \tag{2.30}$$

or integrated over the allowed q^2 range,

$$R(\pi) = \frac{\Gamma(B \to \pi\tau\nu_\tau)}{\Gamma(B \to \pi\ell\nu_\ell)}. \tag{2.31}$$

$R(\pi)$ thus depends only on the ratio of the two form factors as the summand including the scalar form factor f^0 tends to zero for small lepton masses, $\ell = e$ or μ, which makes this ratio theoretically clean and highly sensitive to new physics effects [8, 62, 68]. Dutta, *et al.* [67], compute the ratio (2.31) based on LCSR form factor calculations from [8] to $R(\pi) = 0.698^{+0.035}_{-0.044}$. With $\mathcal{B}\left(B^0 \to \pi\ell\nu\right) = (1.45 \pm 0.05) \times 10^{-4}$ [10], the expected branching fraction is

$$\mathcal{B}(B^0 \to \pi\tau\nu) = (10.1^{+0.7}_{-0.8}) \times 10^{-5}. \tag{2.32}$$

Latest lattice QCD calculations of the form factors, with the z expansion fitted to low q^2 data from $B^0 \to \pi\ell\nu$ by BaBar and Belle [56], allow to calculate the branching fraction in the SM [69] to

$$\mathcal{B}(B^0 \to \pi\tau\nu) = (9.35 \pm 38) \times 10^{-5}. \tag{2.33}$$

2.3.2. Possible Effects from New Physics

While many different possible extensions to the Standard Model exist, a certain set of so-called New Physics (NP) has been in the focus of searches in Flavor Physics. The set introduces a minimal extension in the scalar sector of the SM by introducing a second weak isospin doublet in the Higgs sector, leading to a set called the Two-Higgs-Doublet Model (2HDM) [70]. This section will give a very short description of the general idea, based on the review by Branco, *et al.* [71]. Possible consequences will be described in the second half of this section.

The electroweak symmetry breaking in the SM is achieved by introducing a single $SU(2) \times U(1)$ doublet, see Equation (2.15), in the Lagrangian. As the name suggests, in the 2HDM, eight fields are introduced in two such doublets,

$$\phi_a = \begin{pmatrix} \phi_a^+ \\ (v_a + \rho_a + i\eta_a)/\sqrt{2} \end{pmatrix}, \qquad a = 1, 2, \tag{2.34}$$

with vacuum expectation values of

$$\langle\phi_1\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \qquad \langle\phi_2\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2. \end{pmatrix} \tag{2.35}$$

The ratio of the two vacuum expectation values is called $\tan\beta = v_2/v_1$. Of the eight new fields in Equation (2.34), three get absorbed to give mass to the $W^\pm$ and Z^0 gauge bosons, as described in Section 2.1.3, leaving five new physical scalar fields. In this way, five Higgs particles are created, two neutral Higgs particles h^0 and H^0, two charged Higgs particles $H^\pm$ and a pseudosalar A^0.

While there are many motivations for 2HDMs, three of the strongest are supersymmetry, the strong CP problem and baryogenesis. In supersymmetric models [72–75], the scalar fields are arranged in chiral multiplets while their complex conjugates carry the opposite chirality. A second Higgs doublet is then needed in order to give mass of both up- and down-type quarks. The minimal supersymmetric standard model (MSSM) has been one of the favorite expansions on the SM and is actively searched for [76–78]. Supersymmetry (SUSY) introduces a new set

of particles, doubling the number of particles. Each fermion in the SM is assigned a bosonic partner in the SUSY, and each boson in the SM has a fermionic supersymmetric partner. The supersymmetric partners of the leptons and quarks are called sleptons and squarks, respectively. R-parity is a new symmetry introduced in the MSSM to avoid proton decay, where all SM particles have R-parity of $+1$, while all sypersymmetric particles have R-parity of -1. If R-parity is conserved, the lightest supersymmetric particle (LSP) is stable and would be a candidate for dark matter.

The strong CP problem refers to the fact that the QCD Lagrangian may contain a CP violating term. Phenomenologially, the CP violation in the strong interaction has to be very small. 2HDMs allow the CP violating term to be rotated away by allowing the construction of a global $U(1)$ symmetry and thus solve the strong CP problem [79].

Furthermore, additional sources for CP violation may arise in 2HDMs, which could provide an explanation for the baryon asymmetry observed in the Universe [80–82].

A major problem in all multi-Higgs-Doublet models is the possibility of tree level flavor-changing neutral currents (FCNC). The diagonalization of the mass matrix, in case of the 2HDM,

$$M_{ij} = y_{ij}^1 \frac{v_1}{\sqrt{2}} = y_{ij}^2 \frac{v_2}{\sqrt{2}}, \tag{2.36}$$

does not simultaneously diagonalize the Yukawa couplings y^1 and y^2, which means that the Yukawa interaction will not be flavor diagonal. Neutral Higgs particles will mediate FCNC in this way, which have not been observed yet, though. There are different ways to suppress FCNC in the 2HDM, one of which is that each class of charged fermions couples to only one Higgs doublet ϕ_i. Depending on the coupling, the 2HDM is separated into different types, shown in Table 2.4. As the 2HDM type II is a subset of the MSSM, it has been in the focus of new

Type	up-type quarks	down-type quarks	charged leptons
Type I	ϕ_2	ϕ_2	ϕ_2
Type II	ϕ_2	ϕ_1	ϕ_1
Lepton-specific	ϕ_2	ϕ_2	ϕ_1
Flipped	ϕ_2	ϕ_1	ϕ_2

Table 2.4.: Types of the 2HDM which suppress FCNC. The up- and down-type quarks and charged leptons couple to only one Higgs doublet. By convention, ϕ_2 couples to the up-type quarks.

physics effects in flavor physics. Furthermore, in 2HDM other than type II, scalar contributions to decays of type $B \to X\tau\nu$ are assumed to be less relevant [83].

Models like the 2HDM may modify the $b \to u\tau\nu$ transition by mediating the process through a charged Higgs boson instead of a $W^{\pm}$-boson, as shown in Figure 2.10. As the decay $B^0 \to \pi\tau\nu$ has long been believed unaccessible by current generation experiments, there are only few theoretical calculations concerning new physics effects. The ratio Equation (2.30) has been examined in the presence of a charged Higgs contribution by Khodjamirian, *et al.* [8], in a
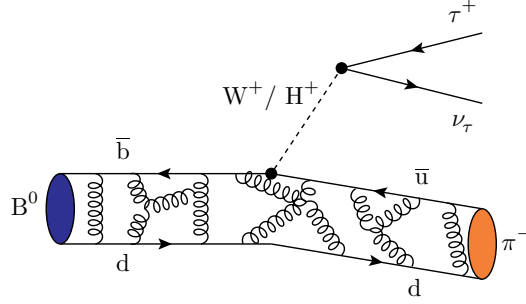
Figure 2.10.: Feynman graph of the decay $B^0 \to \pi\tau\nu$. In NP scenarios like the 2HDM, the decay can be mediated by a $W^{\pm}$ boson or a $H^{\pm}$ boson.

generic approach defining an effective Hamiltonian

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} V_{ub} \bigg\{ \bar{u}\gamma_\mu(1-\gamma_5)b\bar{\tau}\gamma^\mu(1-\gamma_5)\nu_\tau$$
$$- \frac{\bar{m}_b m_\tau}{m_B^2} \bar{u}\left(g_S + g_P\gamma_5\right)b\bar{\tau}(1-\gamma_5)\nu_\tau \bigg\} + h.c., \tag{2.37}$$

where g_S and g_P are the effective scalar and pseudoscalar coupling constants of the new interaction, with $g_S = g_P$ in the MSSM. The ratio Equation (2.30) changes by replacing

$$f^0(q^2) \to \left(1 - \frac{g_S q^2}{m_B^2}\right) f^0(q^2). \tag{2.38}$$

Assuming two values of $g_S = -0.4$ and $g_S = 2.4$, motivated by measurements of $\mathcal{B}(B \to \pi\ell\nu_\ell)$ and $\mathcal{B}(B^+ \to \tau^+\nu_\tau)$, the ratio is plotted in Figure 2.11a. The integrated ratio depends on g_S, too, and is shown in Figure 2.11b.


(a)   (b)
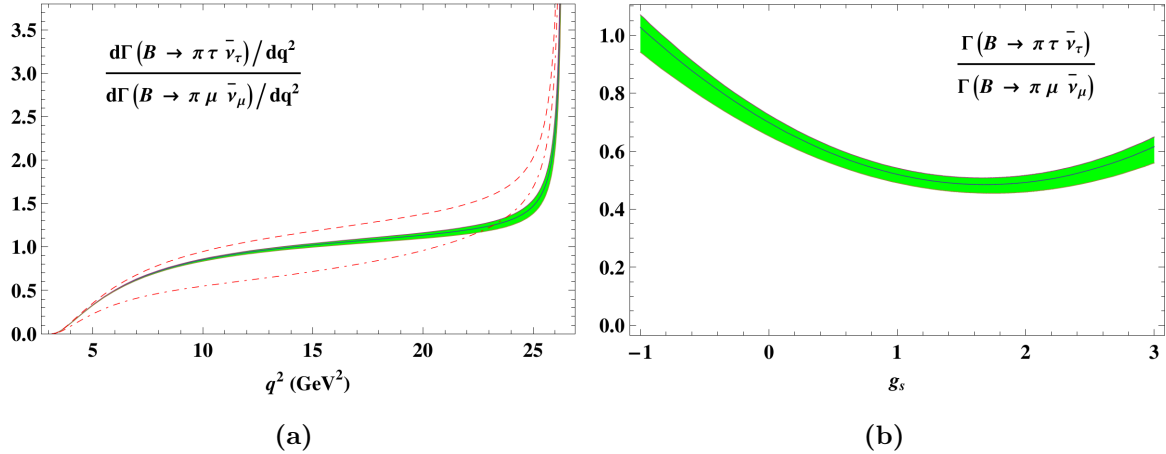
Figure 2.11.: Possible effects of a charged Higgs contribution on the ratio Equation (2.30) for $g_S = -0.4$ (*dashed, red*) and $g_S = 2.4$ (*dash-dotted, red*) in comparison to the SM prediction based on LCSR (*sold, error band in green*) in (a) and the integrated ratio over $g_S \in [-1, 1]$ in (b). Plots and calculations are taken from [8].

The branching fraction in the MSSM has been computed by Kim and Wang [84]. While their result of the SM branching fraction is higher than the latest results shown above, the effect of

new physics relative to the SM prediction should still hold true. In their calculations, $\tan\beta/m_H$ is constrained by the measurement of $\mathcal{B}\left(B^+ \to \tau\nu\right)$ from BaBar [85] and an old Belle result from 2006 [86]. The effects of a fixed $\tan\beta = 50$ and floating m_H has been discussed in [87] but will not be discussed here. They show, that in the case of R-parity conserving (RPC) MSSM, $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$ is not very sensitive to a change of $\tan\beta/m_H$. Furthermore, the forward-backward (FB) asymmetry of the charged lepton

$$\overline{\mathcal{A}}_{FB} = \frac{\int_0^{+1} \frac{d^2\mathcal{B}\left(B^0\to\pi\tau\nu\right)}{dq^2 d\cos\theta} d\cos\theta - \int_{-1}^0 \frac{d^2\mathcal{B}\left(B^0\to\pi\tau\nu\right)}{dq^2 d\cos\theta} d\cos\theta}{\int_0^{+1} \frac{d^2\mathcal{B}\left(B^0\to\pi\tau\nu\right)}{dq^2 d\cos\theta} d\cos\theta + \int_{-1}^0 \frac{d^2\mathcal{B}\left(B^0\to\pi\tau\nu\right)}{dq^2 d\cos\theta} d\cos\theta} \tag{2.39}$$

is very sensitive to charged Higgs effects in the RPC MSSM. R-parity violating (RPV) MSSM includes two additional Feynman diagrams which contribute to $B^0 \to \pi\tau\nu$, shown in Figure 2.12, where slepton and squark currents contribute to the $b \to u\tau\nu$ transition. The transition can
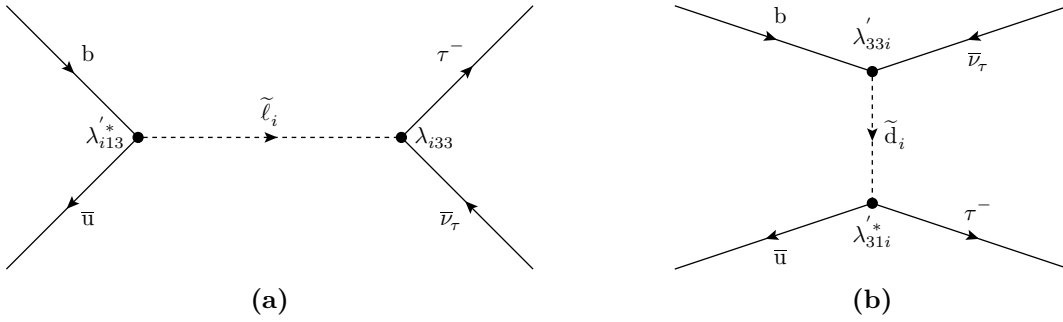
Figure 2.12.: Feynman diagrams of (b) slepton and (b) squark currents contributing to $b \to u\tau\nu$ in the R-parity violating MSSM [84].

then be mediated by squarks and sleptons, the superpartners of the SM quarks and leptons, respectively [73, 75]. The RPV MSSM affects the branching fraction and the $\overline{\mathcal{A}}_{FB}$ in a different way, enhancing the branching fraction in the case of squark mediation and slightly broadening the range in case of slepton contributions. The effects on $d\mathcal{B}(B^0 \to \pi\tau\nu)/ds$, with s being the momentum transfer, and $\overline{\mathcal{A}}_{FB}$, depending on s, are shown in Figure 2.13. Each dot in the plots is one calculation at a random point of the 1σ range of the input parameters. Their results on the branching fraction is listed in Table 2.5.

Effects of different, more general configurations of New Physics couplings on $B^0 \to \pi\tau\nu$ has been examined by Dutta, *et al.* [67]. Their results depend on the coupling in question, and it can be concluded, that NP can have both increasing and decreasing effects on $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$. Within the 1σ range, the ratio $R(\pi)$ (2.31), takes a minimal value of 0.36 and a maximum value of 7.06.
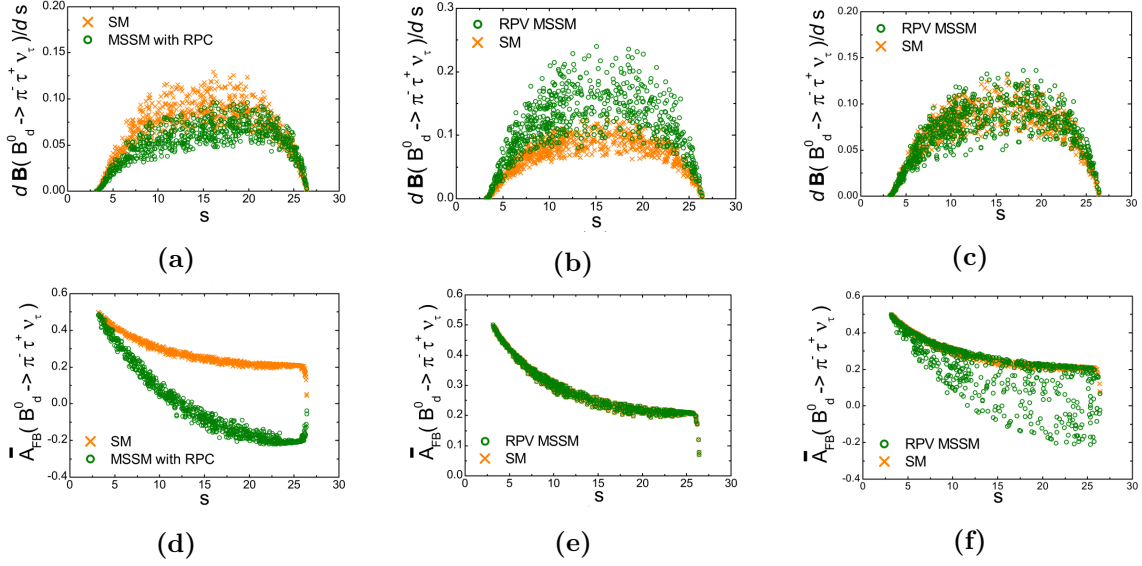
Figure 2.13.: Effects of charged Higgs contributions on $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$ and $\overline{\mathcal{A}}_{FB}$ in the RPC MSSM in (a, d). Effects in the RPV MSSM depending on $|\lambda_{33i}^{'*}\lambda_{31i}^{'}|$ and $|\lambda_{i33}^{*}\lambda_{i13}^{'}|$ are shown in (b, e) and (c, f), respectively. $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$ are shown in units of 10^{-4}. s denotes the momentum transfer q^2. Plots are taken from [84].

Model	$\mathcal{B}\left(B^0 \to \pi\tau\nu\right)\ [10^{-4}]$
SM	$[1.12, 2.28]$
RPC MSSM	$[0.80, 1.79]$
RPV MSSM with $\lambda_{33i}^{'*}\lambda_{31i}^{'}$	$[1.45, 4.59]$
RPV MSSM with $\lambda_{i33}^{*}\lambda_{i13}^{'}$	$[0.91, 2.41]$

Table 2.5.: $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$ in the MSSM predicted by Kim and Wang [84]. The ranges are obtained by calculating the branching fractions at random points in the 1σ range of the input parameters.

.

3. Experimental Setup

This chapter gives an overview of the Belle experiment. The experiment includes the KEKB accelerator and storage complex and the Belle detector. While it was originally built to study CP violation in the B-meson system, it has also been successfully used to perform high precision studies on CKM elements and angles, as well as rare B decays [88]. It produced the largest dataset of $e^+e^- \rightarrow B\bar{B}$ pairs, operating at a center-of-mass energy of $10.58\,\mathrm{GeV}$, which is slightly above the $\Upsilon(4S)$ resonance. The Belle experiment therefore is also called a B factory.

Its astounding success in confirming the CP violation in the B system led to the Nobel Prize for the two Japanese physicists Kobayashi and Maskawa in 2008 [6].

The KEKB accelerator started operation in December 1998 and was turned off June 30th, 2010. Both the accelerator and detector are currently being upgraded into SuperKEKB and Belle II, respectively. The upgrade is designed to deliver 40 times the instantaneous luminosity of KEKB, with improved tracking especially for low momentum particles on the detector side.

Section 3.1 describes the KEKB accelerator. In Section 3.2, the Belle detector is presented followed by the particle identification methods in Section 3.3. While not part of the hardware, but still part of the experiment, the used and generated datasets are described in Section 3.4.

3.1. The KEKB Accelerator

KEKB [89] is an asymmetric electron-positron collider with a circumference of $3\,\mathrm{km}$. Electrons are accelerated in a linac-complex. A part of the electron bunches are directed onto a tungsten target in order to produce positrons. In 2006, the tungsten plate was replaced by a tungsten crystal, which increased the positron efficiency [90]. The electrons are accelerated to an energy of $8.0\,\mathrm{GeV}$ and stored in the high energy ring, HER, whereas the positrons are accelerated to $3.5\,\mathrm{GeV}$ and stored in the low energy ring, LER. This results in a center-of-mass energy of $10.58\,\mathrm{GeV}$, boosted with $\beta\gamma = 0.425$ with respect to the detector reference frame. The energy is slightly above the mass of the $\Upsilon(4S)$, which decays into a $B\overline{B}$ pair with a decay width $\Gamma/\Gamma_{\mathrm{tot}} > 96\%$ [10]. Figure 3.1 shows a schematic view of the accelerator complex. The design luminosity of the KEKB collider is $1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ with a bunch crossing rate of $509\,\mathrm{MHz}$. The bunches cross at the interaction point with an collision angle of $22\,\mathrm{mrad}$. In 2009, the design instantaneous luminosity was exceeded by a factor of 2, with a peak luminosity of $\mathcal{L} = 2.11 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [90, 91]. This World Record was achieved with the help of crab cavities, which help to increase the interaction volume of two crossing bunches. In order to achieve this, the crab cavities rotate the bunches to collide head-on. The basic principle is shown in Figure 3.2.

By the end of June 2010, the total integrated luminosity was $\int \mathcal{L}dt = 1040.863\,\mathrm{fb}^{-1}$ [92]. The integrated luminosity over time is shown in Figure 3.3.
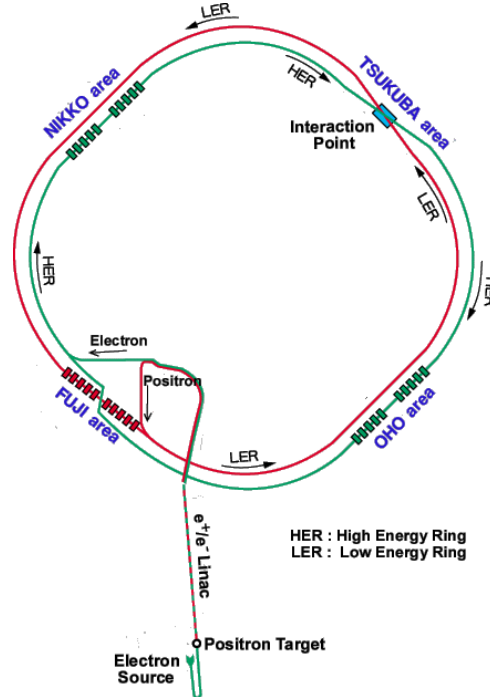
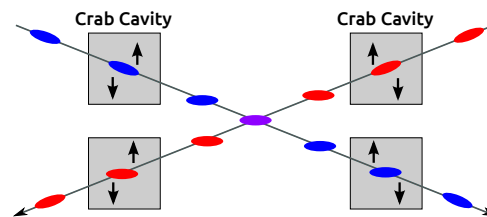**Figure 3.1.:** Schematic view of the KEKB accelerator.

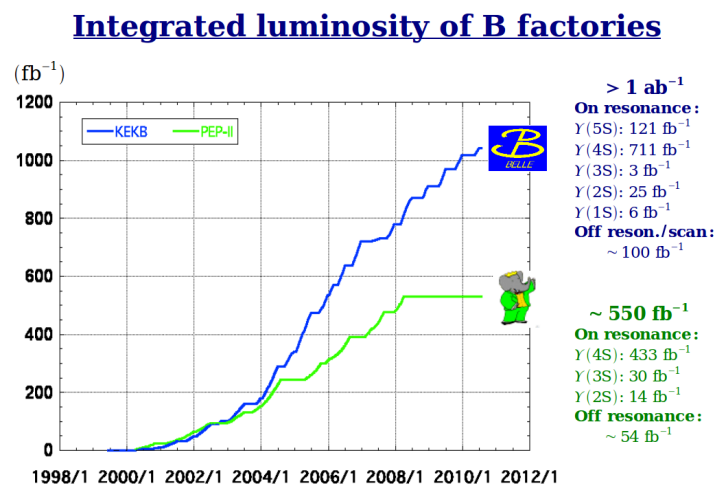**Figure 3.2.:** Crab cavities rotate the particle bunches to collide head-on.

Figure 3.3.: Integrated luminosity over time at Belle (blue) and BaBar (green).

3.2. The Belle Detector

The Belle detector is a general purpose 4π detector built around the only interaction point of the KEKB storage ring. The x-axis is horizontal away from the center of the KEKB ring, the y-axis is vertical and the z-axis is in the opposite direction of the positron beam. It covers a polar angle range of $17° < \Theta < 150°$. As the interaction products are boosted in direction of the electron momentum direction, the detector is slightly asymmetric. The detector and its sub components are described in detail in Ref. [93], which is the basis for the overview given here. Figure 3.4 shows a perspective scheme of the detector and its components. The following sections give a short description of the hardware of each subdetector as well as their purpose in particle identification. The methods of particle identification are explained in Section 3.3.
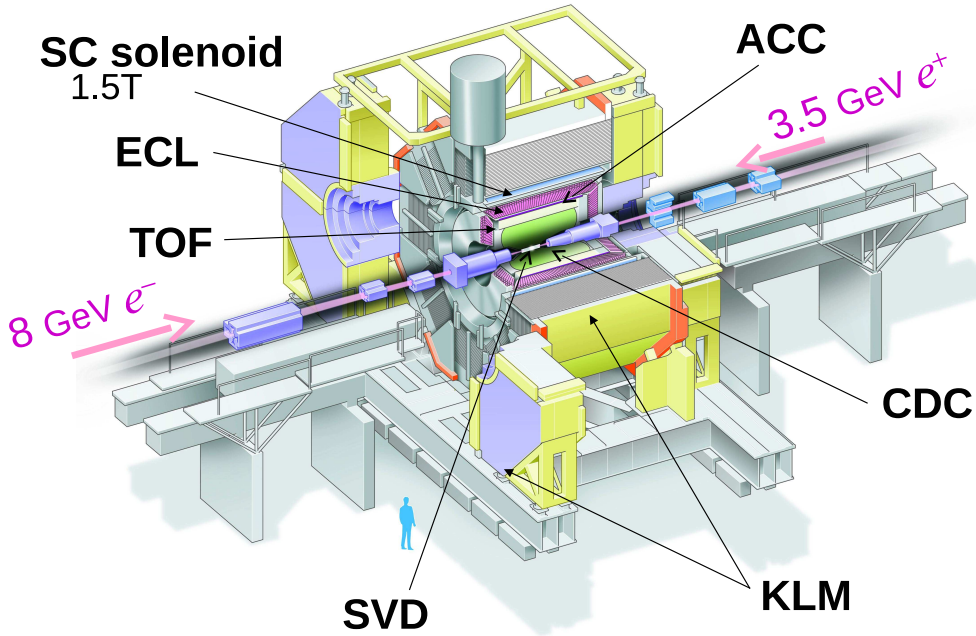
**Figure 3.4.:** Perspective view of the Belle detector.

Silicon Vertex Detector - SVD

The Silicon Vertex Detector is the innermost subdetector in Belle. It is a silicon strip detector and important for the measurement of decay vertices as well as track reconstruction. The SVD is made of double sided silicon strip sensors, of which a scheme is shown in Figure 3.5. Charged particles passing through the sensitive area create electron/hole pairs which produce a measurable signal. By having two layers, crossed at an angle, a position measurement of the passing track in two dimensions is possible. Two versions of the SVD have been used during the Belle operational time.

The first version, SVD1, was in use until 2003, when it got replaced by its successor SVD2. SVD1 consisted of three layers, with the innermost layer located at a radius of $3.0\,\mathrm{cm}$. It covered a polar angle range of $23° < \Theta < 140°$.

In order to improve the vertex resolution, a fourth layer was added in SVD2 [94]. Furthermore, the radius of the beam pipe was reduced from $2.0\,\mathrm{cm}$ to $1.5\,\mathrm{cm}$, allowing the innermost SVD layer at a radius of $2.0\,\mathrm{cm}$, now covering also the whole polar angle range of $17° < \Theta < 150°$. The overall radius of the SVD increased, with the outermost layer then at a farther radius than in SVD1. The impact parameter resolution, determined using cosmic ray studies, has improved slightly in ρ direction, from $\sigma_\rho = (19.2 \pm 0.8) \oplus (54 \pm 0.8)/(p\beta(\sin\theta)^{3/2})$ in SVD1 to $\sigma_\rho = (17.4 \pm 0.3) \oplus (34.3 \pm 0.7)/(p\beta(\sin\theta)^{3/2})$ in SVD2. The improvement is better in the z-direction, being $\sigma_z = (42.2 \pm 1) \oplus (44.3 \pm 1)/(p\beta(\sin\theta)^{5/2})$ in SVD1 and $\sigma_z = (26.3 \pm 0.4) \oplus (32.9 \pm 0.8)/(p\beta(\sin\theta)^{5/2})$ in SVD2 [94].

Both SVD versions are shown for comparison in Figure 3.6.
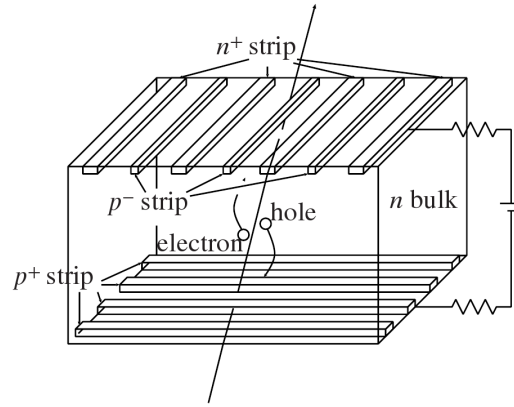
Figure 3.5.: Double-sided silicon vertex detector, DSSD, as used in the SVD.
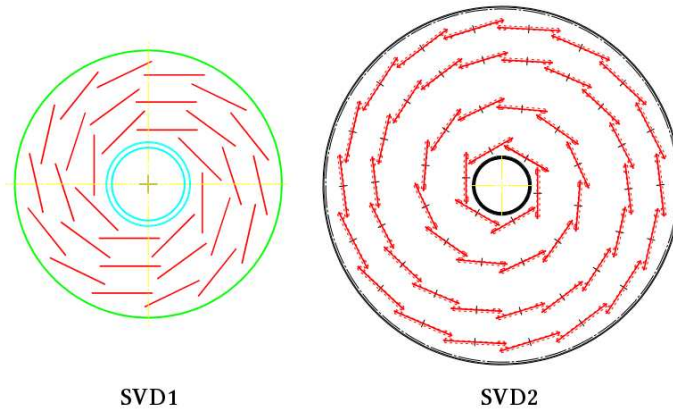
Figure 3.6.: Comparison between SVD1 and SVD2 layout. Both illustrations use the same scale, with the radius of the beam pipe being $2.0\,\mathrm{cm}$ in SVD1 and $1.5\,\mathrm{cm}$ in SVD2.

Central Drift Chamber - CDC

The Central Drift Chamber is the main tracking subdetector of the Belle detector. It is asymmetric in the z-direction and covers the angular region of $17° < \Theta < 150°$. The CDC consists of 8400 drift cells, each of which is made of one sense wire and 8 field wires. The electric field strength is lower than $20\,\mathrm{kV/cm}$ at the surface of the electric field wires in order to avoid radiation damage. The cells are filled with a 1:1 gas mixture of He and C_2H_6 which provides low multiple scattering for particles with a momentum lower than $1\,\mathrm{GeV/c}$ due to the low atomic number. The gas has a radiation length of $640\,\mathrm{m}$ and the drift velocity saturates at $4\,\mathrm{cm}/\mu s$ at a field strength of rougly 2-4 kV/cm. Charged particles passing through the CDC ionize the gas, with the produced charged ions drifting towards the sense wire due to the applied electric field. The CDC is also important in the particle identification methods, which make use of the dE/dx measurements. Although the atomic number of the gas mixture is relatively low, the large ethane component still provides a good dE/dx resolution of 7% for minimum-ionizing particles [88]. The measurement of the track momentum is enabled by the $1.5\,\mathrm{T}$ magnetic field provided by the superconducting solenoid coil. The momentum resolution is given by $\sigma_{p_T}/p_T = 0.0019 p_T \oplus 0.0030/\beta$ with $[p_T : GeV/c]$ [88].
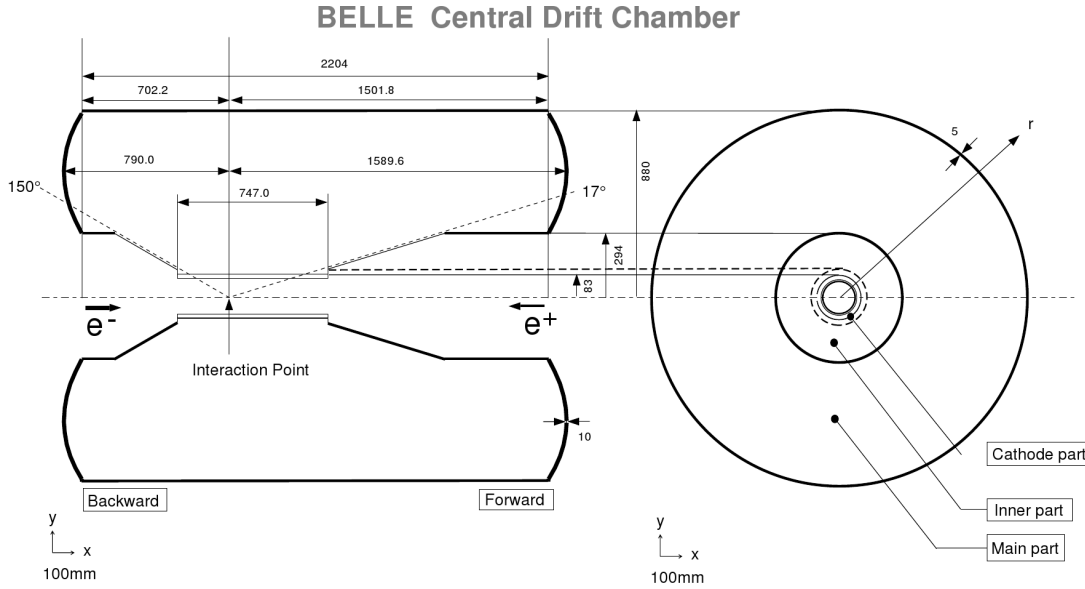
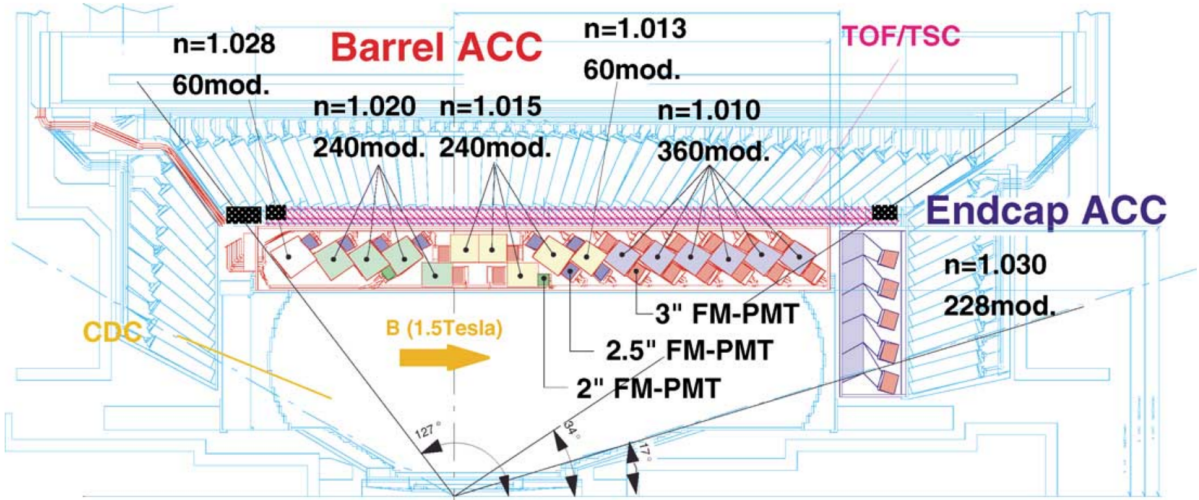

Figure 3.7.: Sideview of the CDC along the z-axis.

Aerogel Cherenkov Counter - ACC

Particles traveling through a medium at a speed higher than the speed of light in that medium emit light cones in flight direction, called Cherenkov radiation [95]. This fact can be used to separate heavy and light particles, as light particles travel at higher velocities when having the same momentum. Specifically in case of the Belle detector, the ACC is used for separation between kaons and pions. The counter modules are shown in Figure 3.8. The ACC consists of 960 modules in the barrel region, segmented into 60 cells in the ϕ direction, and 228 modules in 5 concentric layers in the forward end-cap. In order to provide best separation, the refractive

indices of the aerogels are varied between 1.01 and 1.03, depending on the polar angle, such that the average momentum of pions will be high enough to emit light when passing through the module, while on average, kaons will be slower and will not emit Cherenkov light. The regions and the module configuration in the ACC is shown in Figure 3.9. The emitted photons are collected by photomultipliers tubes.

(a)   **(b)**
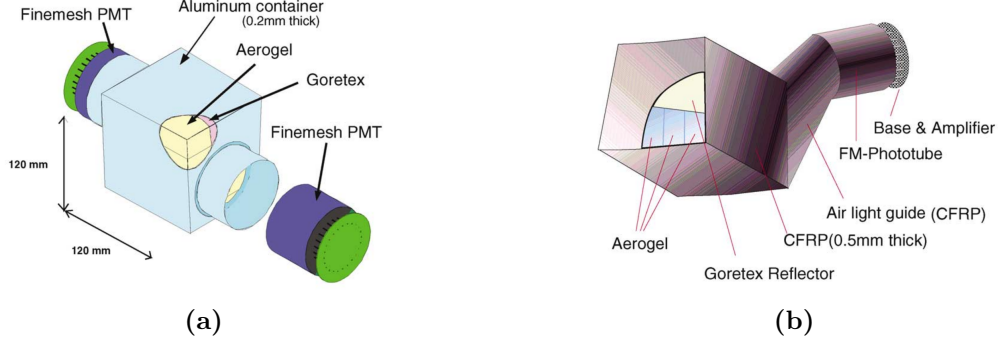
Figure 3.8.: Schematic drawing of the ACC modules in the Belle detector, in the (a) barrel and (b) end-cap region. The figure uses the abbrevations PMT for photomultiplier tube and CFRP for carbon-fiber reinforced plastic.

Figure 3.9.: Sideview of the ACC along the z-axis.

Time of Flight System - TOF

The Time of Flight system consists of a barrel of 128 plastic scintillator counters along ϕ, with a width of approximately $6\,\mathrm{cm}$. Each counter covers $-27.5\,\mathrm{cm} < z < 182.5\,\mathrm{cm}$ at a radius of $r = 122.0\,\mathrm{cm}$. It measures the time between interaction and arrival of charged particles with a time resolution of $100\,\mathrm{ps}$. Combined with the CDC information, this system can be used for separation between pions and kaons with a momentum of $p < 1.2\,\mathrm{GeV/c}$.

Electromagnetic Calorimeter - ECL

The Electromagnetic Calorimeter provides an angular coverage of $17° < \Theta < 150°$ at an inner radius of 1.25 m. The barrel region is 3.0 m long, the endcaps are installed at $z = +2.0$ m and $z = -1.0$ m. It is a homogeneous calorimeter containing 8736 CsI(Tl) counters, which point towards the interaction point. A typical dimension of the front facing crystal side is 55 mm×55 mm, and 65 mm × 65 mm at the rear face in the barrel area, while the actual size varies depending on the crystal position. The crystals provide a total interaction length of $16.2 \, X_0$ and the size is determined by the requirement that approximately 80% of the total energy of a photon injected in the center of the crystal is contained in the crystal. This choice is a compromise between good position resolution and two-photon separation on the one hand, and the required number of channels and energy resolution on the other hand. Increasing the number of crystals would increase the number of gaps and inactive material between the crystals, thus reducing the energy resolution. The material in front of the ECL ranges from 0.3 to $0.8 \, X_0$. A sideview of the ECL is shown in figure 3.10.

The ECL is used to measure the energy of charged mesons, electrons and photons. Particles reaching the detector produce electromagnetic showers in the ECL crystals, which are registered by scintillators. Electrons and photons are generally stopped completely in the ECL, depositing their full remaining energy. Muons and mesons may pass through the ECL, depending on their momentum, while still depositing a certain amount of energy in the ECL.

The energy resolution varies from 4% at 100 MeV to 1.6% at 8 GeV in the barrel region, and 2.85% in the endcaps. The angular resolution at these respective energies are 13 and 3 mrad. This allows a π^0 mass resolution of 4.5 MeV/c^2.

The average position resolution is described by $\sigma_X \, (\text{mm}) = 0.27 + \frac{3.4}{\sqrt{E}} + \frac{1.8}{\sqrt[4]{E}}$.

Additionally, extreme forward and backward calorimeters (EFC) are installed, covering a range of $6.4° < \Theta < 11.5°$ and $163.3° < \Theta < 171.2°$. They are made of radiation hard Bismuth Germanate Oxide crystals.

K_L and Muon Detector - KLM

The KLM covers a polar angle range of $45° < \Theta < 125°$ in the barrel region and up to 20° and 150° in the forward and backward end cap region. It consists of alternating layers of double-gap resistive plate counters and 4.7 cm thick iron plates. The iron plates provide a total of 3.9 radiation lengths for hadrons. Muons and hadrons that reach the KLM will produce showers in the iron plates that are then detected in the RPCs. Since muons are not completely stopped in the KLM, it is not possible to measure their energy, only the direction. Together with informations from the SVD and CDC, the main purpose of the KLM is to discriminate muons from K_L, both producing clusters in the KLM. Not all K_L are completely stopped in the KLM. The minimum momentum threshold for muon to reach the KLM is 0.6 MeV.

Superconducting Solenoid Magnet

The 1.5 T magnetic field in the Belle detector is provided by the superconducting solenoid with a cylindrical volume of 3.4 m in diameter and 4.4 m in length. The iron flux-return yoke consists
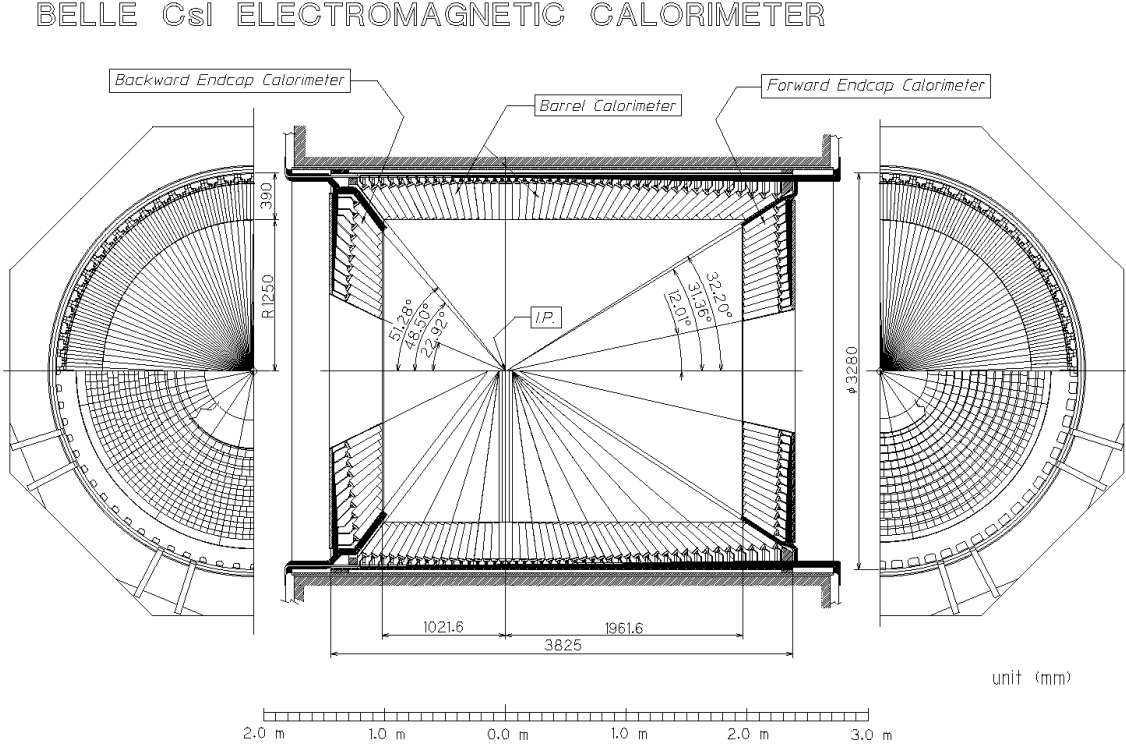
Figure 3.10.: Sideview of the ECL along the z-axis. The lower half of the barrel area shows the front side in the $x - y$ direction.

of 8 of the 14 iron layers of the KLM and 20 cm thick iron plates that surround the outermost KLM layers.

Trigger System

The data acquisition system limits the data rate to 500 Hz, while the collision rate is 509 MHz. However, only a small fraction of the collisions produces physically interesting events. The trigger system is used to identify these events, while ignoring the rest. The events interesting for physics analysis are shown in Table 3.1. The Bhabha scattering and $e^+e^- \rightarrow \gamma\gamma$ events are used for the calibration and luminosity monitoring. They occur at a very high rate, and are scaled down by the trigger system at a fixed rate of 1/100. At the design luminosity of $1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the total rate of interesting events is about 100 Hz, with additionally about 120 Hz of beam background events. As stated above, the design luminosity was exceeded and the detector was operated at double the design intensity, which doubled the rate of physically interesting events. Beam background events increased with a higher factor.

The trigger system has several levels. The SVD requires a faster decision than the main Level-1 trigger can provide. A fast Level-0 trigger is therefore provided by the TOF to the SVD, only, which blocks the readout of the SVD in case of a possibly interesting event. The TOF provides the signal with a delay of approximately $0.85\,\mu$s. The Level-1 trigger operates at a fixed timing of $2.2\,\mu$s. It evaluates information from all subdetectors, except the SVD, and scans the event for properties of relevant physics processes. The Level-1 trigger is implemented in programmable

Physics process	Rate [Hz]
$\Upsilon(4S) \to B\overline{B}$	12
$e^+e^- \to q\bar{q}$, $q \in \{u, d, s, c\}$	28
$e^+e^- \to \ell\bar{\ell}$, $\ell \in \{\mu, \tau\}$	16
Bhabha ($\theta \leq 17°$)	4.4 (scaled 1/100)
$e^+e^- \to \gamma\gamma$ ($\theta \leq 17°$)	0.24 (scaled 1/100)
2γ processes with $p_t \geq 0.3\,\text{GeV}$, $\theta \leq 17°$)	35
Total	~ 96

Table 3.1.: Rate of physics processes at Belle with $\mathcal{L} = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ at the $\Upsilon(4S)$ resonance.

hardware to provide the fast decision time. The next stage is the Level-3 trigger, which is implemented as a software on the online computing farm. A fast track finding algorithm is implemented in order to accept only events with tracks originating from the interaction point. It accepts hadronic events with an efficiency of nearly 100% and reduces the overall data size by a factor of roughly 1/2. Events passing the Level-3 trigger are written to permanent storage on the offline computing farm, where they are further processed by the Level-4 trigger. The Level-4 trigger applies requirements on the deposited energy and requires at least one track originating from the IP region with an improved track reconstruction algorithm. The overall data size is reduced to 26.7%, while all processes of interest are kept with an efficiency of roughly 100%.

3.3. Particle Identification

This section describes the particle identification (PID) algorithms for the most relevant final state particles of this analysis.

Most of the charged final state particles in B-meson decays are pions. The charged particles stable enough to be detected in the detector are electrons, muons, pions, kaons and protons, where the relevant charged particles for the search for $B^0 \to \pi\tau\nu$ are electrons, muons and pions. Additionally, photons have to be detected to reconstruct π^0 mesons for the $\tau \to \rho\nu$ decay mode. In the analysis, every charged track is first checked whether it is an electron. In case the check is negative, it is checked whether it is a muon and at last, a pion.

The different routines for particle identification will be described shortly in the next sections, detailed descriptions may be found in the references given at the beginning of each section.

3.3.1. Electron Identification

The methods used for electron identification, EID, are well described in Ref. [96]. Two basic approaches are used in order to distinguish electrons from muons and hadrons, namely the difference in electromagnetic showers in the `ECL` and the different velocity at the same momentum in the `CDC` and `ACC`. A total of five different discriminants is used, where for each discriminant a probability density function is prepared, out of which an electron likelihood L_e and a non-electron likelihood $L_{\bar{e}}$ can be computed. The combination of all five discriminants is done as a

combined likelihood

$$L_{\text{eid}} = \frac{\Pi_{i=1}^n L_e(i)}{\Pi_{i=1}^n L_e(i) + \Pi_{i=1}^n L_{\bar{e}}(i)}, \tag{3.1}$$

where i runs over each discriminant. Note, that since the correlation between the discriminants is not computed, this is strictly speaking not a probability, but still a useful variable for EID. The five discriminants used are described in the following.

First, for every charged track, the position of its ECL cluster is extrapolated and matched with the center of gravity of the measured ECL clusters. The expected position of the ECL cluster is derived from Monte Carlo simulations. For every cluster, the difference between its center and the extrapolated cluster position is computed in the azimuthal and polar angle, $\Delta\phi$ and $\Delta\theta$, respectively. In order to compute a likelihood for this measure, a χ^2 is defined as

$$\chi^2 = \left(\frac{\Delta\phi}{\sigma_{\Delta\phi}}\right)^2 + \left(\frac{\Delta\theta}{\sigma_{\Delta\theta}}\right)^2, \tag{3.2}$$

where $\sigma_{\Delta\phi,\theta}$ are the Gaussian widths of the $\Delta_{\phi,\theta}$ distributions for electrons, respectively. Figure 3.11a shows the χ^2 distribution for electrons and pions.

The ratio of energy measured in the ECL, E, and the momentum of the charged track by the CDC, p, is another important variable. In the relevant energy range the electron mass can be neglected and the ratio E/p should be close to one for electrons. For massive particles like pions, muons or other hadrons, E/p is usually smaller than one. This discriminant is very effective for electrons with high momentum $p \gtrsim 0.5\,\text{GeV/c}$. Lower momentum electrons lose energy in interactions with material in front of the ECL and have $E/p < 1$. The distributions of E/p for electrons and pions are shown in Figure 3.11b.

The shape of the shower in the ECL is another discriminant that can be used for electron identification. The electromagnetic and hadronic shower shapes are different in both longitudinal and transverse direction. In order to exploit the shape difference in transverse direction, the energy ratio $E9/E25$ is computed. In this ratio, $E9$ is the sum of energy in a 3×3 crystal array round the crystal at the shower center. Similarly, $E25$ is the energy in the 5×5 array around the center. The energy ratio for electrons peaks at $E9/E25 = 0.95$ with a small contribution towards lower values. Pion showers contribute more to lower ratio values, but due to minimum ionizing energy deposit, they have many events at $E9/E25 \approx 1$. The distributions are shown in Figure 3.11c.

The above mentioned variables are best for high momentum tracks. The rate of energy loss, dE/dx, is an effective discriminant for both high and low momentum. dE/dx is defined as the energy loss per distance due to inelastic scattering of the particle in question. It depends on β^{-2} and can be computed by the Bethe-Bloch equation. The dependence of dE/dx on p_{lab}, determined via MC studies, is shown in Figure 3.11d. When measuring the dE/dx of a track, large fluctuations in the Landau tail may arise. Therefore, Belle uses the so-called truncated mean method to calculate the mean dE/dx of each track. In this methods, the largest 20% of the measured dE/dx values for each track are discarded, and the remaining values are averaged.

The last discriminant is the emitted light in the ACC. This information is very useful for particles with $p < 1.0\,\text{GeV/c}$. The threshold for Cherenkov light emission for electrons is in the lower MeV region. For pions the threshold is between $0.5 - 1.0\,\text{GeV/c}$, depending on the refractive index. For this discriminant, a likelihood for the light yield of the ACC is calculated from Monte Carlo distributions for 20 velocity ranges.

(a) χ^2 distribution for `ECL` cluster matching.

(b) E/p dependent on p_{lab}

(c) $E9/E25$

(d) dE/dx dependent on p_{lab}
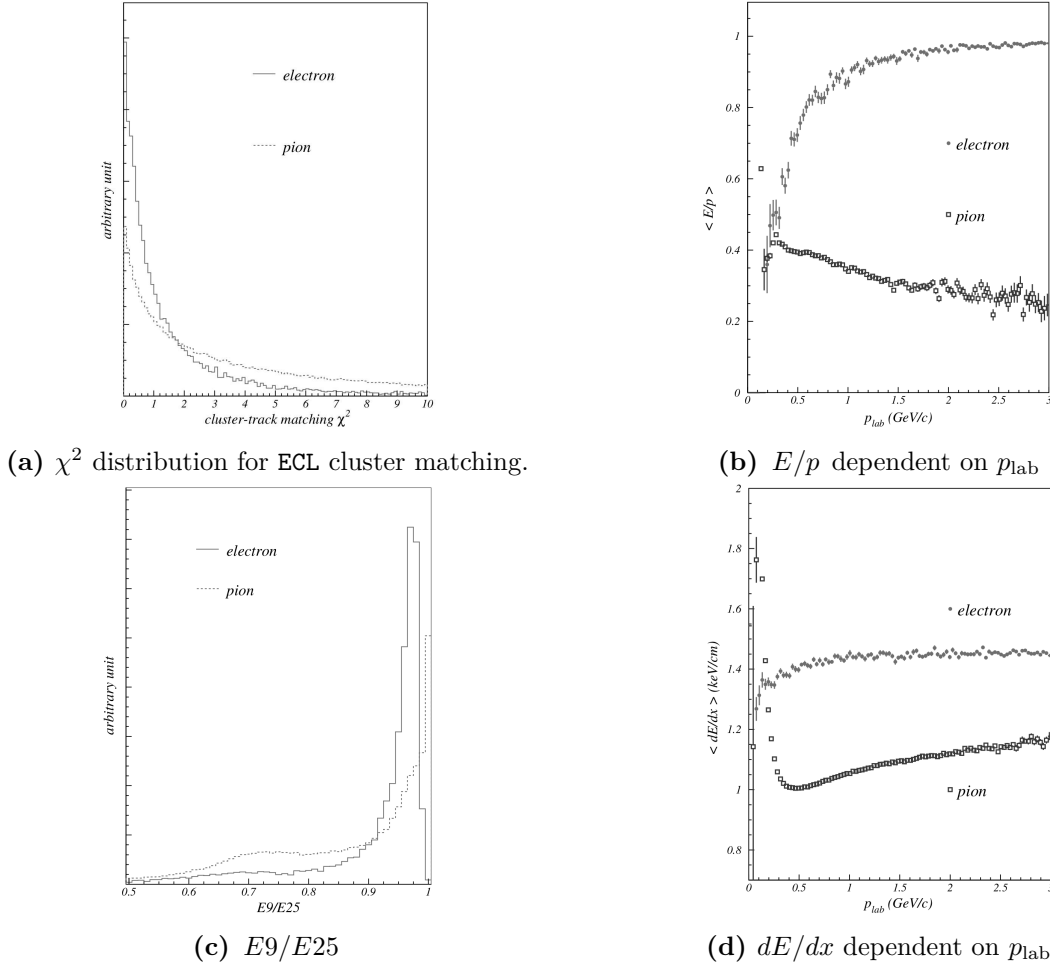
Figure 3.11.: Plots for four of the five physical measures used in EID, plotted for electrons and pions. Plots are taken from [96].

3.3.2. Muon Identification

Muon identification is described in Ref. [97] and will be described here shortly. In order to identify muons, information from the `SVD`, `CDC` and `KLM` is used. The whole process is split up into two consecutive steps.

First, the track is extrapolated from the `CDC` into the `KLM` until it is stopped there or escapes the detector. Due to performance reasons the extrapolation is done assuming the particle to be a pion without strong interactions. For every layer of the `KLM`, a hit is associated with the track if it lies within 25 cm or $5\,\sigma$ of the crossing point of the extrapolated track with the layer. For every track, the expected and measured range in terms of `KLM` layers are determined. The expected range is the given by the outermost layer the track crosses in the extrapolation. The measured range is given by the last outermost layer with of associated hit. In order for a track to be selected for the second step it must have at least two hits in the `KLM` and the difference between expected and measured range must be no more than five layers, else the track is rejected. The preselection removes about 90% of all pions but only 4.5% of all muons.

In the second step, the track is extrapolated again in the `KLM`, this time assuming a muon

hypothesis. The criteria for hit association change slightly, now being 20 cm or $5\,\sigma$. Two measures are now taken into account for computing a muon likelihood, the difference between expected and measured range ΔR, and the transverse deviation of all hits associated with the track, normalized to the number of hits, χ_r^2. Their distributions are shown in Figure 3.12. For all three possible particle types, μ, π, K, the probability density function is determined and defined. The range and transverse deviation are assumed to be uncorrelated which leads to the joint PDF to be just the product of both single PDFs $p_{\mu,\pi,K}(\Delta R, \chi_r^2) = p_{\mu,\pi,K}(\Delta R) \cdot p_{\mu,\pi,K}(\chi_r^2)$.

The muon identification package then provides this information for every track to the user, namely if the track has been rejected in the preselection step, and the normalized likelihood

$$\mathcal{L}_\mu = \frac{p_\mu}{p_\mu + p_\pi + p_K}. \tag{3.3}$$
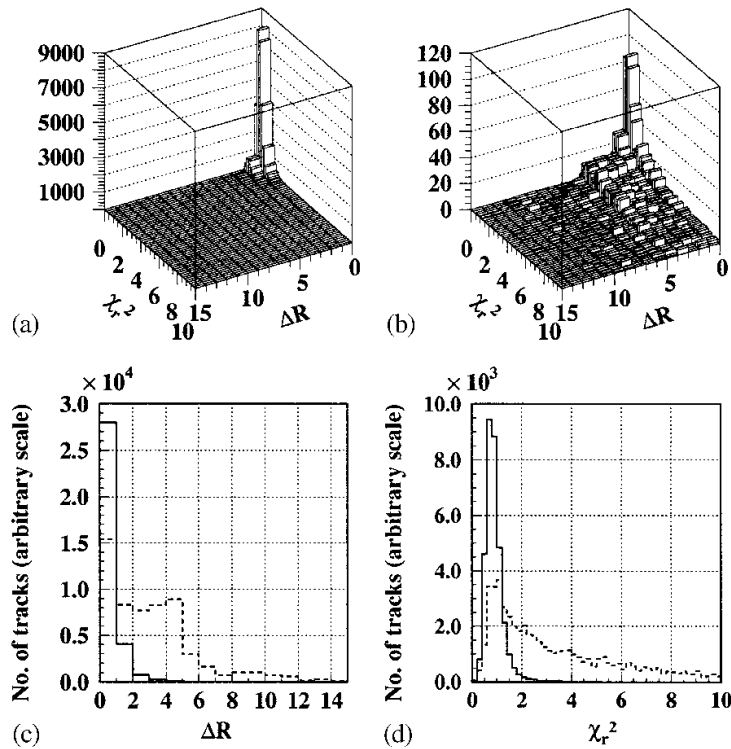
Figure 3.12.: Distributions of $\Delta R - \chi_r^2$ for (a) muons and (b) pions, and of (c) ΔR and (d) χ_r^2 for muons (*solid line*) and pions (*dashed line*) used in muon identification. Taken from [97].

3.3.3. Pion Identification

The pion identification method is described in detail in Ref. [98] and will be be described shortly here. Most of the hadronic, charged final state particles in B meson decays are pions, followed by kaons. Protons occur as well, but less frequently.

In order to separate the hadronic particles, information of several sub detectors are combined, as already seen in the leptonic case. These sub detectors are the `CDC`, `TOF` and `ACC`.

The `CDC` provides information about the energy loss dE/dx. As in the electron case, the truncated mean method is applied. The dE/dx information provides a greater than 3σ separation for tracks up to $0.8\,\text{GeV}/c$ and about 2σ for tracks above $2.0\,\text{GeV}/c$. The dE/dx distribution for the charged final particles in Belle are plotted in Figure 3.13 over the relevant momentum range.
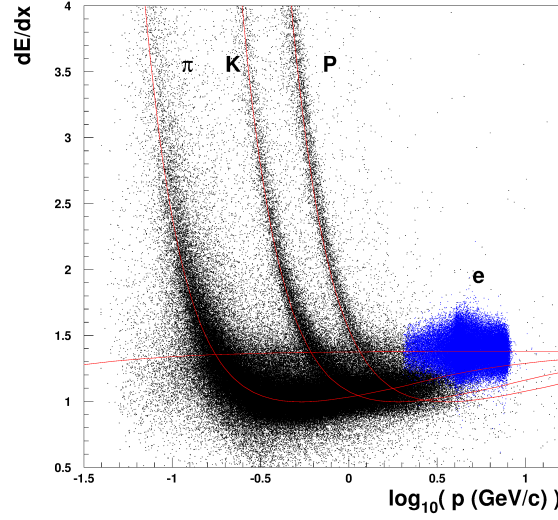
Figure 3.13.: dE/dx distribution for pions, kaons, protons and electrons, depending on the particle momentum. Taken from [98].

The TOF subdetector allows to calculate a hypothetical mass of a track from the measured momentum and its time to reach the `TOF`. Since the three hadronic final state particles in question differ significantly in mass, this information separates well between them in the low to intermediate momentum range. Figure 3.14 shows that for tracks with $p < 1.25\,\text{GeV}/c$ the
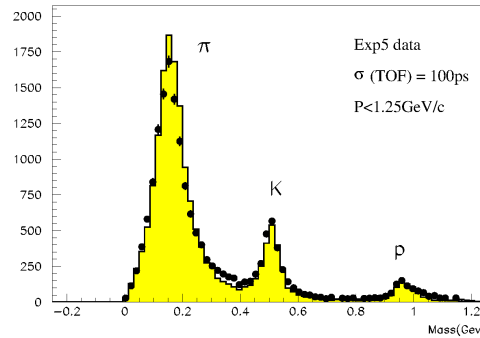
Figure 3.14.: Mass calculated for tracks from measured momentum $p < 1.25\,\text{GeV}/c$ and time to reach the `TOF` detector. Taken from [98].

time-of-flight information clearly separates the three particle types.

The Cherenkov emission in the `ACC` is the last information used to separate hadronic final states. As described above, the refractive index of the cells is not constant, but ranges from 1.010 to 1.028, covering a momentum range of 1.2 to $3.5\,\text{GeV}/c$. The choice has been optimized to separate two-body decays as $B \rightarrow \pi\pi$ and $B \rightarrow \pi K$. As can be seen in Figure 3.15, pions and kaons can be separated by the number of emitted photo electrons.
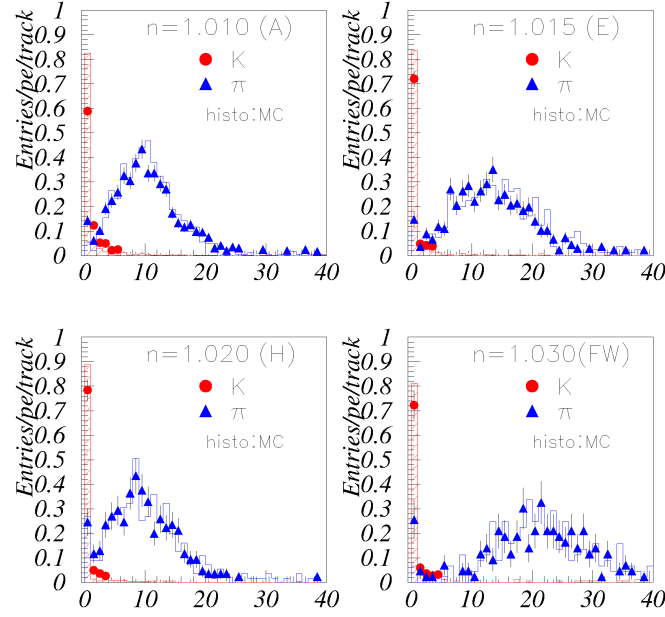
Figure 3.15.: Distributions of number of measured photo electrons in the ACC for different refractive indices n. The filled triangles and circles show numbers measured for pi and K tracks, respectively, whereas the histograms are MC predictions. Taken from [98].

The final output of the hadronic particle ID is a likelihood ratio. In contrast to the lepton particle ID, the hadron particle ID always discriminates between two particle types. The output is given as

$$Prob(i,j) = \frac{P_i}{P_i + P_j}, \tag{3.4}$$

where i and j can be any of the five particle types e, μ, π, K and p, though it usually is only used for the hadrons. $Prob(i,j)$ is lies between 0 and 1, and is therefore interpreted as a probability, though strictly speaking this is not correct. If the algorithm can not distinguish between both types, it returns the value 0.5, same as the lepton particle ID algorithms.

The likelihood P_i is composed of the three likelihoods given by the sub detectors

$$P_i = P_i^{dE/dx} \times P_i^{\texttt{TOF}} \times P_i^{\texttt{ACC}}, \tag{3.5}$$

which in turn are obtained from the characteristic behavior of each particle type in each sub detector as described above.

3.3.4. Photon Identification

Photons leave clusters in the `ECL`, but no tracks in the `SVD` or `CDC`. All clusters that are not associated with a charged track can therefore been associated with a photon. In order to improve fake rejections by noise or secondary clusters, the energy ratio $E9/E25$ and a required minimum energy are used as quality check variables. Furthermore, there are different ways a track may be associated with a cluster in the `ECL`. If the interpolated track intersects the cluster in its center, the association can be done without much doubt. In cases where the interpolated

track intersects the cluster at its border or passes closely to the cluster, the cluster itself can be checked whether it is likely to originate from a real photon or from a charged particle.

The cluster c is defined to be a good neutral cluster if $\theta_c \in [17; 150]°$ and if the deposited energy $E_c > 0.02\,\text{GeV}$. In case of $E < 0.05\,\text{GeV}$, the root-mean square of the cluster width, the $E9/E25$ and the number of crystals associated with the cluster are usually checked, too. However, these cuts are usually analysis dependent.

3.4. Dataset

The analysis makes use of the Belle dataset, recorded at the $\Upsilon(4S)$ resonance, as well as Monte Carlo simulations of resonant $e^+e^- \to \Upsilon(4S)$ and $e^+e^- \to q\bar{q}$ events. All datasets used in this analysis are introduced in this section.

3.4.1. Recorded data

For this study, the full dataset recorded with Belle at the $\Upsilon(4S)$ resonance has been used. The integrated luminosity is $711\,\text{fb}^{-1}$, corresponding to $(771.581 \pm 10.566) \times 10^6$ $B\overline{B}$ pairs. 152×10^6 $B\overline{B}$ pairs have been collected with `SVD1` and 620×10^6 $B\overline{B}$ pairs with `SVD2`.

The process $e^+e^- \to \Upsilon(4S)$ has a cross section of $1.1\,\text{nb}$. Non resonant events happen more often, and are suppressed by a filter, called skim at Belle. The HadronB skim is built to have a high efficiency for resonant events, but to reduce the number of non resonant background events. The cross sections of the various processes in e^+e^- collisions at $\sqrt{s} = 10.58\,\text{GeV}$, and their effective cross section after the skim, are shown in Table 3.2.

Process	σ [nb]	Effective σ [nb]	Skim efficiency
$b\bar{b}$	1.1	1.09	0.991
$q\bar{q}$ $(q = u, d, s, c)$	3.3	2.62	0.795
$\tau^+\tau^-$	0.93	0.05	0.049
QED $(25.551° < \theta < 159.94°)$	37.8	0.001	0.00002
$\gamma\gamma \to q\bar{q}$ $(w > 500\,\text{MeV})$	11.1	0.04	0.004

Table 3.2.: Cross section and effective cross section after applying the HadronB skim.

3.4.2. Monte Carlo samples

The analysis is first performed on Monte Carlo (MC) simulated data. The HadronB skim is applied to all samples, as well. Large MC samples have been generated for known processes at Belle, which are used to describe the expected background events, while additional signal MC has been generated to describe solely the process $B^0 \to \pi\tau\nu$. The decay process is modeled with `EvtGen` [99]. After event generation, the detector response is simulated using a package based on GEANT 3.21 [100], which describes the interaction of the decay particles with the Belle detector.

The MC samples used for this study describe the following processes: $B^0 \to X_c$, $B^\pm \to X_c$, $e^+e^- \to q\bar{q}$, $B \to X_u \ell \nu$ and rare $B \to X_{us}$ decays. Table 3.3 lists all used MC samples and their size in terms of recorded data luminosity. It is found that no further MC samples for the other types of background is needed, due to the low efficiency of the skim, and the hadronic full reconstruction used in the analysis. While the samples listed above have been prepared by the collaboration, two private signal MC samples have been generated specially for this study, containing 60×10^6 and 24×10^6 events. In each decay, one B^0-meson decays into the signal final state, while the other B-meson decays generically. The reason for two samples is the need for two statistically independent samples for the training of the Boosted Decision Tree and cut selection described in Section 4.3.

The signal decay process is described mathematically as in Section 2.3. The `EvtGen` package does not provide a calculator for the BCL model, instead the two form factors f^+ and f^0 can be described in a pole form,

$$f(q^2) = \frac{f(0)}{\left(1 + a\frac{q^2}{m_B^2} + b\frac{q^4}{m_B^4}\right)^c}, \tag{3.6}$$

where q is the momentum transfer and m_B is the cut-off value.

process	size [$\times$ data] data	process	size [$\times$ data]
$B^0 \to X_c$	10	$B \to X_u \ell \nu$	20
$B^\pm \to X_c$	10	$B \to X_{us}$	50
$e^+e^- \to q\bar{q}$	10	$B^0 \to \pi\tau\nu$	≈ 800

Table 3.3.: MC samples used for the analysis. The second column displays the size of the sample, normalized to the full recorded dataset of $711\,\mathrm{fb}^{-1}$.

.

The pole model given in Equation (3.6) is fitted to the BCL parameters listed in [8]. $f(0)$ is not fitted but taken directly from the calculation. The model prediction and the pole fit for the form factors are shown in Figure 3.16, the fit results listed in Table 3.4.

Parameter	$f_{B\pi}^+$	$f_{B\pi}^0$
$f(0)$	0.281	0.281
a	-1.444	-0.8059
b	0.4484	0.09872
c	1.0	1.0

Table 3.4.: Pole Parameters used for signal MC generation. Fitted to BCL parameters published in [8].

Efficiency Corrections

Although data are described quite well by the Belle Monte Carlo samples, the reconstruction efficiency on MC and data differ for various settings. The correction factors depend on different factors like reconstruction methods or the theory model used in the generation of decay processes.

(a)

(b)
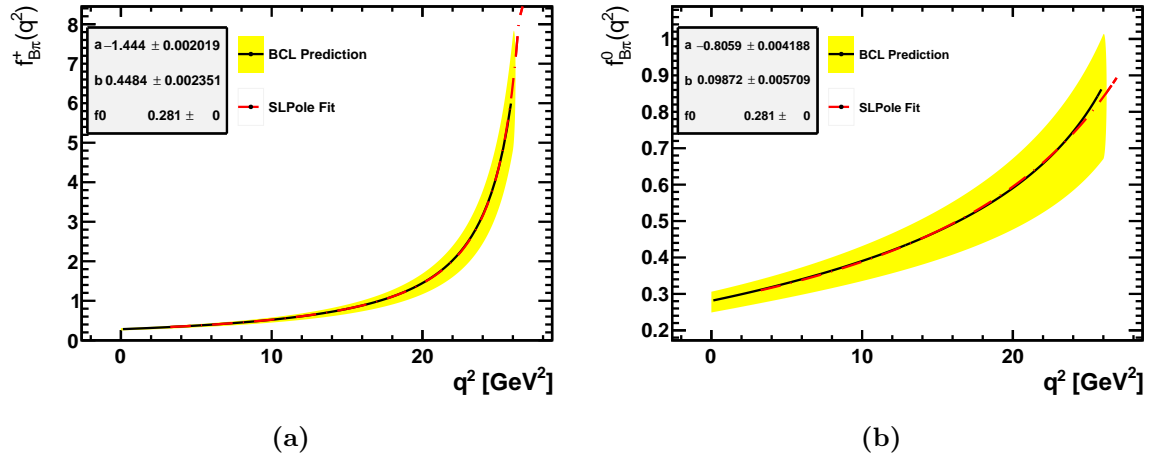
Figure 3.16.: Fit (dashed, red) of the pole model parameters to the theory prediction (black) for (a) $f^+(q^2)$ and (b) $f^0(q^2)$. The theory prediction is taken from Ref. [8].

In order to maintain a clear structure of the thesis, all analysis steps are described first in Chapters 4 and 5. Only then are the necessary correction factors introduced in Section 5.2.

4. Methods

The different analysis techniques used in this thesis are introduced here. The next two Chapters 5 and 6 describe, how these methods are applied in the analysis. A brief overview of the analysis chain is given here to motivate the use of the presented methods.

To facilitate the handling of the neutrinos present in the signal decay, events are taken into account only if one of the two B-mesons has been successfully fully reconstructed in a hadronic decay. The full reconstruction algorithm is described in Section 4.1. In Section 4.2, techniques used for continuum suppression are described, used to suppress $e^+e^- \to q\bar{q}$ events. This search suffers from large resonant $e^+e^- \to \Upsilon(4S)$ background contributions that are difficult to distinguish from the signal process. In order to enhance the separation of signal and background events, several event variables are combined in a multivariate analysis, which is described in Section 4.3. Finally, the methods used to calculate the significance level and the upper limit on the signal branching fraction are detailed in Section 4.4.

4.1. Full Reconstruction

In a lepton collider, the initial state of the process is fully defined by the four-momenta of the incoming particles. Due to conservation of momentum, the full 4-momentum of the final state is known. The process of interest at the Belle experiment is $e^+e^- \to \Upsilon(4S)$ which in turn decays into a $B\overline{B}$-pair with a branching fraction greater than $96\,\%$ [10]. Usually, only one of the B-mesons is reconstructed in the signal decay. The reconstruction in those cases is often checked via the so-called beam-constrained mass

$$M_{bc} = \sqrt{E_{\mathrm{beam}}^2 - \vec{p}_B^2}, \tag{4.1}$$

where E_{beam} is half of the beam energy and $\vec{p}_B$ is the three-momentum of the reconstructed B meson. The beam four-momentum p_{beam} is defined as

$$E_{\mathrm{beam}} = \frac{1}{2}(E_{e^+} + E_{e^-}) \tag{4.2a}$$

$$\vec{p}_{\mathrm{beam}} = \frac{1}{2}(\vec{p}_{e^+} + \vec{p}_{e^-}) \tag{4.2b}$$

$$p_{\mathrm{beam}} = (E_{\mathrm{beam}}, \vec{p}_{\mathrm{beam}})^T. \tag{4.2c}$$

The missing momentum and missing mass is defined as

$$p_{\mathrm{miss}} = 2p_{\mathrm{beam}} - p_{\mathrm{vis}} \tag{4.3a}$$

$$\mathrm{M}_{\mathrm{miss}}^2 = p_{\mathrm{miss}}^2, \tag{4.3b}$$

respectively, where p_{vis} is the visible, that is measured, sum of four-momenta in the detector. It includes all particles that leave tracks or clusters in the detector, but not neutrinos or particles that escape undetected, either through the beam pipe or through detector inefficiencies. For a maximum of one neutrino, the missing mass vanishes, while multiple neutrinos make both Equations (4.1) and (4.3) less usable due to non vanishing missing mass. A solution is to fully reconstruct one of the B-mesons, which then is called B_{tag}. All tracks and clusters left in the event are then from the second B which is tested for the signal decay and labeled B_{sig}.

This analysis uses the hadronic full reconstruction algorithm [101] based on NeuroBayes [102]. B_{tag} candidates are reconstructed in several different decay modes following in hierarchical order from track and cluster informations up to the final B candidate. The reconstructed B-meson decay modes are listed in Table 4.1, with the intermediate D and J/ψ modes listed in Table 4.2. The reconstruction principle is shown graphically in Figure 4.1.

The quality of each particle combination is estimated using neural networks which include various properties of the combined particle as well as the quality of the daughter particles. In this way, an overall quality of the full reconstruction can be estimated, which is used by the analysis as a quality requirement. Events with only poorly reconstructed B^0-candidates are disregarded for further evaluation.

The efficiency of the full reconstruction is $0.18\,\%$ for B^0 and $0.28\,\%$ for B^+ [101].
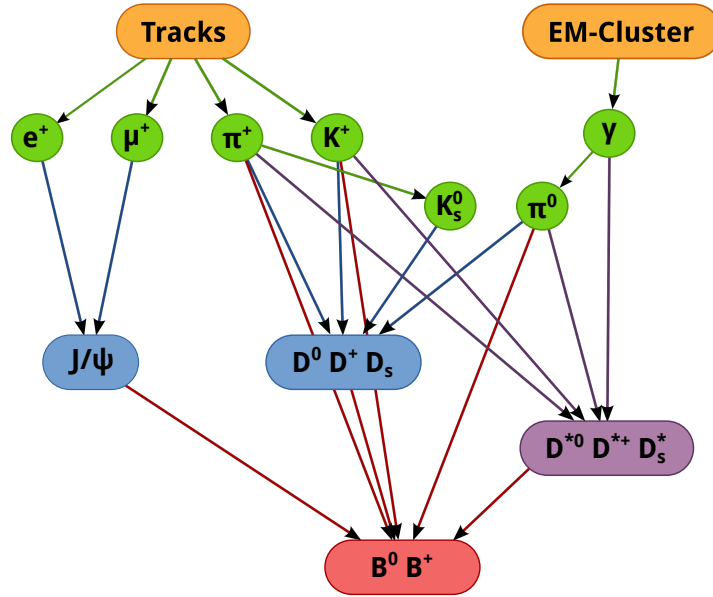
Figure 4.1.: Sketch of the hierarchical reconstruction principle used in the hadronic full reconstruction. Starting from the lightest particles, combinations are built until a B-meson candidate can be constructed, inspired by [101].

B^0 modes	B^- modes
$D^{*+}\pi^-$	$D^{*0}\pi^-$
$D^{*+}\pi^-\pi^0$	$D^{*0}\pi^-\pi^0$
$D^{*+}\pi^-\pi^+\pi^-$	$D^{*0}\pi^-\pi^-\pi^+$
$D^+\pi^-$	$D^0\pi^-$
$D^+\pi^-\pi^0$(with D^{*+} veto)	$D^0\pi^-\pi^0$(with D^{*0} veto)
$D^+\pi^-\pi^+\pi^-$	$D^0\pi^-\pi^-\pi^+$
$D^{*+}D_s^{*-}$	$D^{*0}D_s^{*-}$
$D^{*+}D_s^-$	$D^{*0}D_s^-$
$D^+D_s^{*-}$	$D^0D_s^{*-}$
$D^+D_s^-$	$D^0D_s^-$
$J/\psi K_S^0$	$J/\psi K^-$
$J/\psi K^-\pi^+$	$J/\psi K^-\pi^+\pi^-$
$J/\psi K_S^0\pi^+\pi^-$	D^0K^-
$D^0\pi^0$	$D^+\pi^-\pi^-$
$D^{*+}\pi^-\pi^-\pi^+\pi^0$	$D^{*0}\pi^-\pi^-\pi^+\pi^0$
	$J/\psi K^-\pi^0$
	$J/\psi K_S^0\pi^-$

Table 4.1.: B^0 and B^- modes reconstructed in the hadronic full reconstruction. Reconstruction modes of the combined daughter particles are listed in Table 4.2.

D^0 modes	D^+ modes	D_s^+ modes
$K^-\pi^+$	$K^-\pi^+\pi^+$	$K^+K^-\pi^+$
$K^-\pi^+\pi^0$	$K^-\pi^+\pi^+\pi^0$	$K_S^0K^+$
$K^-\pi^+\pi^+\pi^-$	$K_S^0\pi^+$	$\pi^+\pi^+\pi^-$
$K_S^0\pi^0$	$K_S^0\pi^+\pi^0$	$K^+K^-\pi^+\pi^0$
$K_S^0\pi^+\pi^-$	$K_S^0\pi^+\pi^+\pi^-$	$K_S^0K^-\pi^+\pi^+$
$K_S^0K^+K^-$	$K^+K^-\pi^+$	$K_S^0K^+\pi^+\pi^-$
K^+K^-	$K^-K^+\pi^+\pi^0$	$K^+\pi^+\pi^-$
$\pi^+\pi^-$		$K^+K^-\pi^+\pi^+\pi^-$
$\pi^+\pi^-\pi^0$		
$K_S^0\pi^+\pi^-\pi^0$		

D^{*0} modes	D^{*+} modes	D_s^{*+} modes	J/ψ modes
$D^0\pi^0$	$D^0\pi^+$	$D_s^+\gamma$	e^-e^+
$D^0\gamma$	$D^+\pi^0$		$\mu^-\mu^+$

Table 4.2.: Intermediate modes used in the hadronic full reconstruction.

4.2. Continuum Suppression

Events of type $e^+e^- \to q\bar{q}$ are called continuum events. Their cross-section is roughly three times higher than $e^+e^- \to \Upsilon(4S)$ at a center-of-mass energy of $\sqrt{s} = 10.58\,\text{GeV}$. A powerful separation between continuum and $\Upsilon(4S)$ events can be achieved by analysing the geometrical event topology. While continuum events produce jet-like shapes due to the bigger available phase-space, the B-mesons decay isotropically, forming spherical shapes. The event topology can be described mathematically by Fox-Wolfram-Moments (FWM) [103]. Belle uses an advanced implementation called super Fox-Wolfram moments (SFWM) [104], which are also included in the neural network of the full hadronic reconstruction. SFWM compute the FWM for each the B_{tag} candidate and the rest of the event individually. The full reconstruction algorithm adds a second quality variable for each event which is then used to suppress events which are most likely continuum.

4.3. Boosted Decision Trees

Multivariate techniques often yield better discrimination power than rectangular cuts. Many different methods exist, and this analysis makes use of Boosted Decision Trees (BDT). The basic principle of a BDT is to learn if a given event is signal or background, based on different variables provided to the BDT. This is called classification.

In this section, first decision trees are described, followed by a short description of boosting. This analysis uses TMVA [105] for the Boosted Decision Tree classification.

4.3.1. Decision Trees

Decision trees have, as the name suggests, a tree structure as sketched in Figure 4.2. Every event goes through the tree, starting at the root node, until it reaches an end node. Each node provides a constant prescription into which daughter node the event proceeds until the final node is reached and the event is classified. The test chain is always top to bottom, meaning a node cannot be passed twice.

The decision trees used in this analysis are binary classification trees. This means that the tree is built to assign a given event to one of two classes, namely background or signal. The tree has to be built before, with events clearly identified as signal or background. The process of building the tree with a known sample is called training. In the root node the training sample is split into two subsamples by a splitting criterion, further described below. Each of the two subsamples is now checked at the corresponding daughter node using a different splitting criterion. This procedure is continued until the end criterion is reached. At the end, each node is classified as signal or background, according to the class the majority of events belong to. By this procedure the space of variables used in the tree is divided into disjoint regions R. Figure 4.3 shows an example of final decision tree regions for a tree trained on only two variables $v1$ and $v2$.

The splitting criterion at each node is basically a threshold test. As indicated in Figure 4.2, every node checks the value of one variable. The next node is then given by the value of the variable, relative to the threshold value. In the training procedure, all available variables are
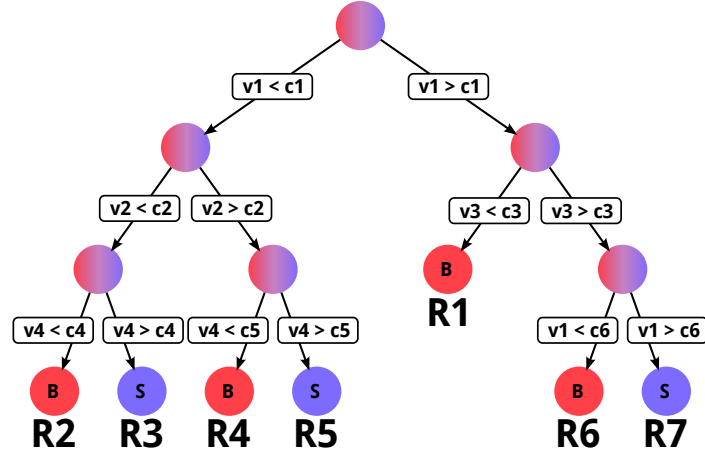
Figure 4.2.: Sketch of a binary decision tree. Decision variables are labeled v, the cut values at each node are c, and the disjoint regions the variable space is divided into R. B and S represent background and signal, respectively.

tested, and the one with the best separation power is finally selected. It is important to note that a variable may be the separation variable multiple times in a single tree.

The separation criteria used in this analysis is the Gini Index [105], defined as

$$G = p \cdot (1 - p), \tag{4.4}$$

where p is the purity of the node, defined as the ratio of signal events to all events in the node. Nodes with mostly signal events have higher purity than events with mostly background events. While TMVA allows for different separation criteria, no big difference between them was observed for this analysis. The Gini Index has a maximum for fully mixed samples with $p = 0.5$ and falls off to zero for samples of only one class.

The variable and cut value at each node is determined by the value that provides the highest decrease of the Gini index between the node and the weighted sum of the Gini indices of the two daughter nodes. The weights in the sum of the daughter indices are given by the relative fractions of events in the nodes.

The tree can be grown very deep until a certain purity is reached, for example. A single tree trained in this way is usually prone to overtraining. Overtraining describes the fact that the classifier learns random noise from the training sample, which often leads to unpredictable outcome on real data. Overtraining is reduced by using multiple decision trees for the same problem by boosting.

4.3.2. Boosting

The term boosting in regard to multivariate techniques can be described as combining multiple, so-called weak learners by weighting them to obtain a single, better separation rule for the given problem. In the case of Boosted Decision Trees, weak learner means that each single decision tree by itself does not separate well between signal and background, which is usually achieved
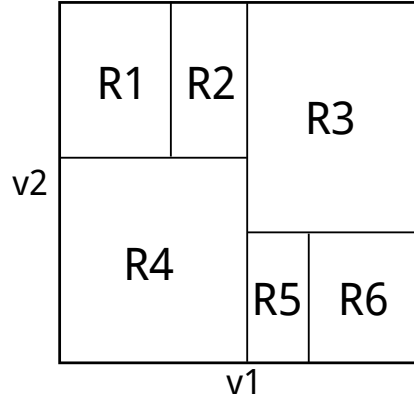
Figure 4.3.: Example of the final, disjoint regions of a decision tree trained with only two variables $v1$ and $v2$. As described in the text, multiple cuts may occur on the same variable, which is shown here, too.

by limiting the depth of each single tree. The first boosting algorithm used for decision trees is called AdaBoost [106].

In this analysis, a more general approach called Gradient Boost [107, 108] is applied. While both boosting methods are described well in literature [105–109], a short overview is given here, which follows the notation used in Ref. [109].

For this chapter, the training sample used to construct the decision trees consists of training variables **x** and the known outcome **y**. As shown in Figure 4.3, a single decision tree splits the space of all used variables into disjoint regions R_j. In the case of a binary tree which only separates signal and background, each region is assigned a constant outcome γ_1 or γ_2, either signal or background. A single tree T can then be expressed as

$$T(x; \Theta) = \sum_{j=1}^{J} \gamma_j I(x \in R_j), \tag{4.5}$$

where x is the data point, J is the number of final disjoint regions and I is the indicator function, evaluating to 1 if $x \in R_j$, and 0 elsewise. The regions and classifications are summarized in the parameter $\Theta = \{R_j, \gamma_j\}_1^J$.

Boosting M decision trees results in the classifier

$$f_M(x) = \sum_{m=1}^{M} T(x; \Theta_m), \tag{4.6}$$

which is induced in a forward stagewise manner. This means that always one tree is formed at a time, where for each step forward, the parameters Θ_m are determined by

$$\hat{\Theta}_m = \arg\min_{\Theta_m} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + T(x_i; \Theta_m)). \tag{4.7}$$

The boosting procedure is then defined by the choice of the loss function L. The loss function used in AdaBoost is the exponential loss

$$L(y, f(x)) = \exp(-yf(x)), \tag{4.8}$$

which leads directly to a prescription how to build the decision trees by applying higher weights to previously misclassified events [106]. For the analysis presented in this thesis, it was found that the binomial log-likelihood loss

$$L(y, f(x)) = \ln \left(1 + \exp(-2yf(x)) \right) \tag{4.9}$$

provides better separation.

It can be shown ([109]), that the minimization of Equation (4.7) can be replaced by the minimization

$$\tilde{\Theta}_m = \arg \min_{\Theta} \sum_{i=1}^{N} \left(-g_{im} - T(x_i; \Theta) \right)^2 , \tag{4.10}$$

which will not result in the same regions R_{jm}, but regions $\tilde{R}_{jm}$ which will result in similar separation performance. $\mathbf{g}_m$ is the gradient of $L(\mathbf{f})$ evaluated at $\mathbf{f} = \mathbf{f}_{m-1}$, where the vector $\mathbf{f}$ is given by the values of the classifier $f(x)$ defined in Equation (4.6), as $\mathbf{f} = \{f(x_1), \ldots, f(x_N)\}^{\mathrm{T}}$. The gradient components g_{im} are then given as

$$g_{im} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x_i) = f_{m-1}(x_i)} . \tag{4.11}$$

The use of the gradient in case of a general loss function is therefore called Gradient Boost.

A further addition is the so-called stochastic gradient boosting. This method uses only a random subsample of all training events for the training of each tree in each iteration. Stochastic gradient boosting is applied for some of the τ reconstruction channels used in this analysis.

4.3.3. Variable Preprocessing

BDTs, like many other multivariate techniques, usually perform best when the input variables are not correlated. By adding additional input variables that are correlated, the classifier gets less new information than would be possible in uncorrelated cases.

However, many of the input variables used in this analysis are highly correlated, but different techniques exist that try to decorrelate them. Two different methods are used which are also implemented in the TMVA toolkit, both of them linear transformations.

Deco **Transformation**

The first method is a decorrelation via the square-root of the covariance matrix. For lack of a better, distinct name and to stay in the TMVA nomenclature, this method will be called *deco* transformation from here on.

The square-root of a matrix A is the matrix A' that fulfills

$$A' \cdot A' = A. \tag{4.12}$$

The decorrelation of the input variables $\mathbf{x}$ is done by the transformation

$$\mathbf{x} \longmapsto \left(C' \right)^{-1} \mathbf{x}, \tag{4.13}$$

where C' is the square-root of the covariance matrix C of the input variables. The square-root matrix C' is calculated by diagonalising the covariance matrix in TMVA,

$$D = S^T C S$$
$$\Rightarrow C' = S\sqrt{D}S^T. \tag{4.14}$$

D is a diagonal matrix with $D = \text{diag}(d_1, \cdots, d_n)$ and $\sqrt{D} = \text{diag}(\sqrt{d_1}, \cdots, \sqrt{d_n})$ and S is a symmetric matrix.

Principal Component Analysis Transformation

The Principal Component Analysis (PCA) is the second transformation used which, as the deco transformation, eliminates linear correlations for Gaussian variables. The input variables are not Gaussian distributed, but correlations are still reduced by applying this transformation.

The PCA transformation rotates the coordinate system such that the principal component has the largest variance. Each following component has the largest variance under the constraint that it is orthogonal to the preceding components.

Given the classes $U = \{S, B\}$ for signal and background, respectively, the matrix of eigenvectors $V_U = (\mathbf{v}_U^{(1)}, \cdots, \mathbf{v}_U^n)$ and the tuple of eigenvalues D_U of the covariance matrix C is given by $C_U \cdot V_U = D_U \cdot V_U$. With the input variables $\mathbf{x}_U = (x_{U,1}(i), \cdots, x_{U,n}(i))$ for event i and their sample means $\bar{\mathbf{x}}_U$, the PCA transformed variables $\mathbf{x}_U^{\text{PCA}}$ are given by

$$x_{U,k}^{\text{PCA}}(i) = \sum_{l=1}^{n} (x_{U,l}(i) - \bar{x}_{U,l}) \, v_{U,l}^{(k)}. \tag{4.15}$$

4.4. Significance Level and Limit Calculation

The basis for both significance level and limit calculation is performing a hypothesis test. Both measures differ in what hypothesis is to be tested, but the mathematical principle is the same for both.

In Section 4.4.1, different test statistics will be introduced as the basic component of a hypothesis, followed by how exactly they can be implemented and how the underlying likelihood is defined. Next, in Section 4.4.2, the calculation of the significance level is explained, followed by the calculation of the upper limit in Section 4.4.3.

Hypothesis testing is well described in literature [110, 111]. It will be described in this section in sufficient detail for this analysis.

In order to test a hypothesis, the hypothesis needs to be exactly defined. This might be the existence of a new process that is not described in the SM, a mass of a particle, or something else. When searching for a new decay process, as in the case of this analysis, the hypothesis to be tested is usually called the null hypothesis H_0. The null hypothesis then is the known theory description of the experiment. In case of new physics searches, H_0 would be the SM. The process $B^0 \to \pi\tau\nu$ is described in the SM, but not yet discovered, so in this case, H_0 contains all SM processes but $B^0 \to \pi\tau\nu$.

As described below, it is useful to define a second hypothesis, which describes the process to be tested or searched for, usually called the alternative hypothesis H_1. In searches in high energy physics, the null hypothesis is often called the background-only hypothesis, H_B, while the alternative hypothesis is called signal-plus-background hypothesis, H_{S+B}. For better readability, and small redefinitions of the null and alternative hypothesis for the calculation of an upper limit, the notation H_0 and H_1 will be used in this document.

4.4.1. Test Statistic

When testing if the measured data $\mathbf{x} = (x_1, \dots, x_N)$ are compatible with a hypothesis, a test statistic $t(\mathbf{x})$ which depends on the data needs to be defined. The usage of a test statistic is the frequentist way to perform a hypothesis test.

The probability density function (pdf) of the test statistic t given a certain hypothesis H will be called $g(t|H)$. A good test statistic provides a clear separation of $g(t|H_0)$ and $g(t|H_1)$. There is no unique recipe to define a test statistic, as it usually depends on the hypotheses in question. Figure 4.4 shows examples for the three test statistics defined below based on pseudo data.

(a) Simple Likelihood Ratio **(b)** Ratio of Profiled Likelihoods **(c)** Profile Likelihood Ratio
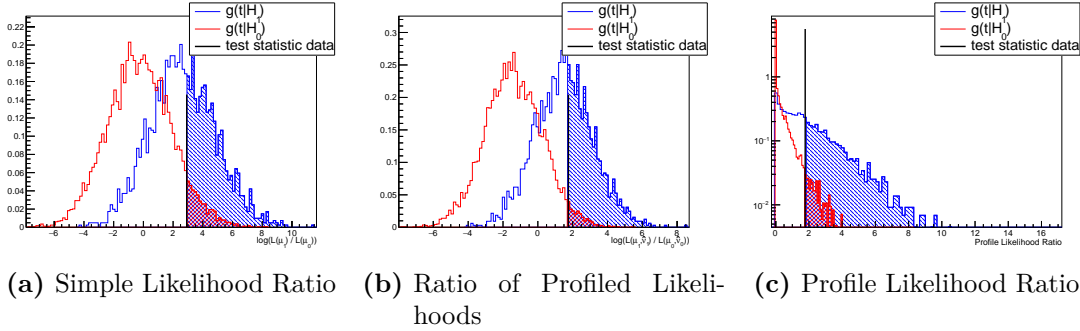
Figure 4.4.: Example plots based on pseudo data of test statistic distributions. H_0 and H_1 are the same in all plots. The black line indicates the value of the test statistic computed on the *data* sample. Note the logarithmic scale in case of the profile likelihood (c).

In the case of simple hypotheses, an optimal test statistic is given by the Neyman-Pearson lemma[112], as

$$Q(\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}|H_1)}{\mathcal{L}(\mathbf{x}|H_0)}. \tag{4.16}$$

Based on Equation (4.16), one can define three test statistics that differ only slightly in their definition. The test statistics below are all constructed using the ratio of two likelihoods. The exact form of the likelihoods will be given in the appropriate Chapter 5, but generally the likelihoods used will depend on two sets of parameters, the parameter of interest μ and the nuisance parameters θ. The parameter of interest μ is also called the signal strength parameter. It is defined as $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)/\mathcal{B}\left(B^0 \to \pi\tau\nu\right)^{\mathrm{SM}}$, such that $\mu = 0$ means no signal at all, while $\mu = 1$ represents the SM expectation of the branching fraction of the signal process. Systematic uncertainties or other floating parameters in the likelihood, that are not the parameter of interest, are called nuisance parameters and will be summarized as θ here. The maximum likelihood estimate of a given parameter or parameter set is denoted using the hat character, eg $\hat{\mu}$.

- The simple likelihood ratio is basically the ratio given in Equation (4.16), where the likelihoods are not maximized. Instead, both signal strength and nuisance parameters are fixed at their nominal value.

$$Q = \frac{\mathcal{L}\,(\mu = 1)}{\mathcal{L}\,(\mu = 0)}.$$

(4.17)

- Ratio of Profiled Likelihoods is similar to the profile likelihood ratio, but the signal strength parameter is fixed in both the nominator and denominator. The double hat $\hat{\hat{\theta}}$ means that the maximum likelihood estimate of θ in the nominator is a priori not the same as in the denominator, but estimated independently.

$$Q = \frac{\mathcal{L}\left(\mu = 1, \hat{\hat{\theta}}\right)}{\mathcal{L}\left(\mu = 0, \hat{\theta}\right)}.$$

(4.18)

- The profile likelihood ratio test statistic uses

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \hat{\hat{\theta}}\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\theta}\right)}.$$

(4.19)

  The test statistic $t = -2 \ln \lambda(\mu)$ has a known asymptotic behavior as described by Wilk [113] and further evaluated in [114]. The double hat $\hat{\hat{\theta}}$ indicates that the nuisance parameters are fitted for both hypotheses independently.

  It is important to note that the test statistic depends on μ. While μ is fixed in the nominator, it will be fitted in the denominator. For the significance calculation, the null hypothesis (background only) needs to be tested, meaning $\lambda(\mu = 0)$. In contrast to the first two definitions, the profile likelihood ratio can be used to test any value of μ.

The distribution $g(t|H)$ has to be known in order to calculate the significance level and the upper limit. Wilk's theorem [113] states that for a large data sample, the test statistic $t = -2 \ln \lambda$ with λ from Equation (4.19) is distributed asymptotically like a χ^2 distribution with 1 degree of freedom. The most straightforward yet more time-consuming strategy to get knowledge about the test statistic distribution is to use pseudo data, which is also used in this analysis.

The pseudo data is generated using the data pdf $f(\mathbf{x}|H_0)$ and $f(\mathbf{x}|H_1)$ for the null and alternative hypothesis, respectively. Special care is needed in the treatment of the systematic uncertainties, which are included as nuisance parameters in the pdf. This analysis uses the RooStats toolkit which is part of the ROOT software package [115], for the computation of significance levels and upper limits. In the toolkit, two different treatments of nuisance parameters are available, a fully frequentist way and a hybrid approach where the constrained term is interpreted as a Bayesian prior and the nuisance parameters are integrated. In the hybrid approach, the value of each nuisance parameter is randomized following its pdf in every pseudo experiment. The constrained term is usually a Gaussian distribution which expresses the variation of the nuisance parameter in terms of standard deviations around the nominal value.
The frequentist treatment introduces so-called global observables, whose value is the nominal value of a sample's contribution in case of no systematic uncertainty. In every pseudo experiment, the global observables will be randomized and held constant in the likelihood maximization. Before generating pseudo data, the maximum likelihood estimate (MLE) of each nuisance

parameter is determined by fitting to the observed data. For pseudo data generation, the nuisance parameter are set to their MLEs, but are floating parameters in the fit. This analysis uses the frequentist approach.

4.4.2. Significance Level

When trying to determine whether there is evidence or even a discovery of new physics in data, the background-only hypothesis has to be rejected with a certain confidence. The test statistics defined above tend to higher values, the less compatible the data is with the null hypothesis. A useful measure for the compatibility of data and an hypothesis is the p-value. It quantifies how often one would obtain data as incompatible or worse with the hypothesis as the observed data, if the experiment was to be repeated many times.

Given the test statistic evaluated on data t_{obs}, the p-value for a hypothesis is given as

$$p = \int_{t_{\mathrm{obs}}}^{+\infty} g(t|H)d\mathrm{t}. \tag{4.20}$$

The smaller the p-value, the less compatible is the data with the hypothesis. In this analysis, the significance of the signal hypothesis is obtained by computing the p-value of the background-only hypothesis, meaning $t = -2\ln\lambda(0)$.

In order to evaluate the p-value as a significance level, the Z-value is used in high energy physics,

$$Z = \Phi^{-1}\left(1 - p_0\right), \tag{4.21}$$

where Φ^{-1} is the cumulative distribution of the unit Gaussian and p_0 is the p-value of the null hypothesis H_0. The significance level therefore corresponds to the one-sided tail probability of a Gaussian distribution.

The likelihood ratios Q defined above have Gaussian shape if working on a sufficiently large data sample, such that a good estimate of the Z-value is

$$Z = \sqrt{-2\ln Q}. \tag{4.22}$$

4.4.3. Upper Limit

The upper limit of a process is quoted with a certain confidence level, in high energy physics usually 90%. This statement can be expressed as an exclusion of a signal hypothesis with $p < 0.1$. When speaking about limits, one usually uses the following nomenclature of CL_b and CL_{s+b},

$$\mathrm{CL}_{s+b} = \int_{t_{\mathrm{obs}}}^{+\infty} g(t|H_{\mathrm{s+b}})dt \tag{4.23a}$$

$$1 - \mathrm{CL}_b = \int_{-\infty}^{t_{\mathrm{obs}}} g(t|H_{\mathrm{b-only}})dt \tag{4.23b}$$

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b}. \tag{4.23c}$$

This means, in order to set an upper limit, one performs many hypothesis tests for different signal strength values. The upper limit is that value for which the signal hypothesis can be excluded with the wanted confidence level CL_{s+b}. Often though, CL_s is used instead of CL_{s+b}. Fluctuations in the data towards the background only hypothesis might lead to the exclusion of the signal hypothesis, although the analysis might lack sensitivity in that case. In order to correct for a downward fluctuation, the CL_s-method [116] is used, which excludes a given hypothesis with a confidence level C if $\mathrm{CL}_s < (1 - C)$ instead of $\mathrm{CL}_{s+b} < (1 - C)$.

5. Event Reconstruction

The reconstruction strategy is to fully reconstruct one of the B^0 mesons and reconstruct the signal decay in the other B^0 meson. The fully reconstructed meson will be called B_{tag}. The τ-lepton will be reconstructed in its four one-prong decay modes listed in Table 5.1, totaling a branching fraction of roughly 72 %. This means the signal signature will consist of two charged tracks, a neutral pion in case of $\tau \to \rho \nu_\tau$, and missing momentum due to two or three neutrinos. Figure 5.1 shows a sketch of a typical event with a signal decay of B_{sig} and a reconstructed B_{tag}.

Decay mode	$\mathcal{B}$ [%]
$\tau \to e \nu_e \nu_\tau$	17.83 ± 0.04
$\tau \to \mu \nu_\mu \nu_\tau$	17.41 ± 0.04
$\tau \to \pi \nu_\tau$	10.83 ± 0.06
$\tau \to \rho [\pi^\pm \pi^0] \nu_\tau$	25.52 ± 0.09

Table 5.1.: Used τ decay modes and their branching fractions [10].

While the signal is reconstructed in four separate modes, namely the decay modes of the τ-lepton with a single charged final state particle, the cross-feed through misidentification between the decay and reconstruction modes is quite high for certain modes. The reconstruction channel $\tau \to \pi \nu$ contains the most cross-feed as will be seen throughout this chapter. The aim of the analysis is to obtain the best separation of signal and background samples in the signal region and to obtain a high significance of the signal. Cross-feed is therefore accepted as long as it improves the expected significance.
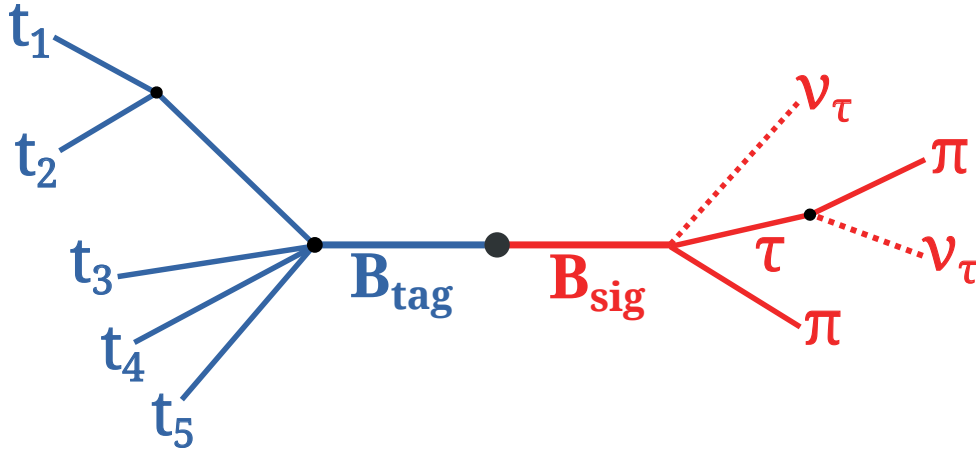
Figure 5.1.: Sketch of a $B\overline{B}$ event at Belle with a signal decay of one of the two B-mesons.

Notation For better readability, the neutrino will not be mentioned when labeling the reconstruction mode. $\tau \to e$ labels the decay reconstruction $B^0 \to \pi^+ \tau^- [e^- \bar{\nu}_e \nu_\tau] \bar{\nu}_\tau$ and the charge conjugate, as an example. It is important to note that the label refers to the reconstruction, not the physical decay for the reasons stated above.

5.1. Basic Event Selection

The event selection is divided into two parts. Events first have to pass the hadronic full reconstruction with a certain quality. Next, quality requirements are applied on the remaining charged tracks and clusters in the event. Given there are only two oppositely charged tracks remaining after the clean up, particle identification is performed. If one charged pion and one electron, muon, ρ or additional pion is left in the event on the signal side, the event is processed by boosted decision trees.

5.1.1. Tagside Reconstruction

The B_{tag} is reconstructed via the full hadronic reconstruction explained in Section 4.1. The quality of the full reconstruction with continuum suppression taken into account is described by the output variable of the neural network, called $o_{\text{tag}}^{\text{cs}}$, ranging from 0 to 1. Higher values indicate a higher probability of correctly reconstructed decay chains. Its distribution for signal and background Monte Carlo samples is shown in Figure 5.2.

5.1.2. Event Cleanup

Next, quality requirements for charged tracks and cluster without track association in the `ECL` are applied.

Tracks are required to originate from the beam interaction point, which, with a relatively short live time of the relevant particles in the signal decay, can be expressed by a requirement on the impact parameters dz and dr, which are the distances of closest approach of a track to the interaction point along the z-axis and the transverse plane, respectively. Good tracks need to have $|dr| \leq 2\,\text{cm}$ and $|dz| \leq 4\,\text{cm}$.

Tracks with low transversal momentum $p_T \leq 275\,\text{MeV}$ may curl up in the detector, which may trick the tracking algorithm to identify two separate tracks originating from two particles with opposite charge. The principle is shown in Figure 5.3. Every pair of reconstructed charged tracks with a momentum difference of $\Delta|\vec{p}| \leq 100\,\text{MeV}$ and a respective angle of $\theta \leq 15°$ for same-charged tracks and $\theta \geq 165°$ for opposite-charged tracks, is considered a candidate for this to happen. They are labeled curly double tracks. Previous studies at Belle showed that computing

$$X = |5dr|^2 + |dz|^2$$

for each of the two tracks and removing the track with the larger value from the analysis would eliminate double counting while still selecting the correct track in most of the cases.

For neutral clusters in the `ECL`, energy requirements are applied. A cluster is defined as neutral if no track can be associated with it. The clusters are required to contain a minimum energy

(a) $\tau \to e$ **(b)** $\tau \to \mu$

(c) $\tau \to \pi$ **(d)** $\tau \to \rho$
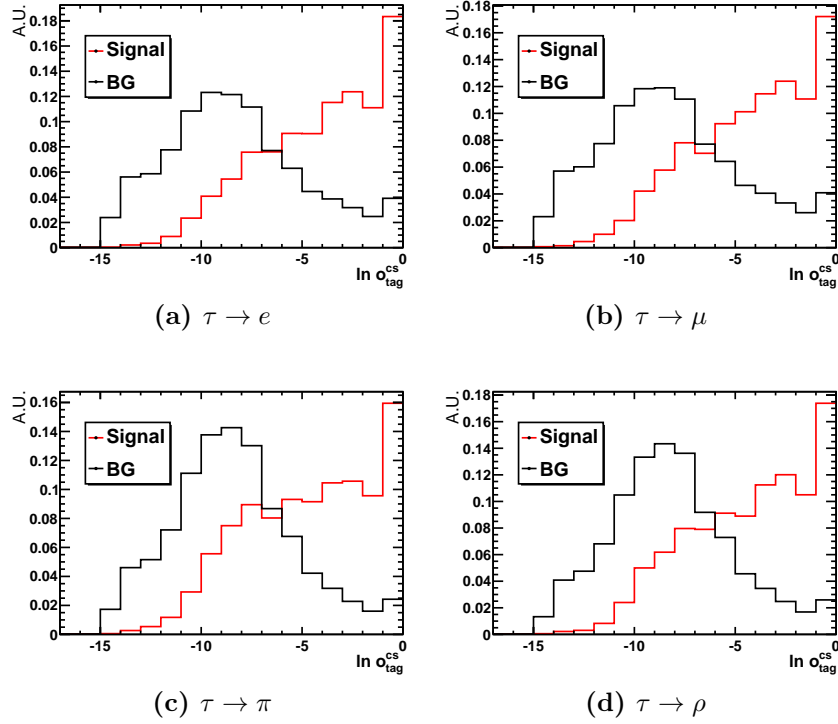
Figure 5.2.: Distributions of $\ln o_{\text{tag}}^{\text{cs}}$ for signal and background, plotted separately for the four τ decay modes analysed in this thesis. The background consists of all background MC samples.

deposition, depending on the cluster location in the detector. The requirements are listed in Table 5.2.

region	$\theta\,[°]$	$E_{\min}\,[\text{MeV}]$
barrel	$[32; 130]$	50
forward endcap	$(130; 150]$	150
backward endcap	$[17; 32)$	100

Table 5.2.: Energy requirements for neutral clusters in the `ECL`, depending on the cluster region.

After the cleanup, events with exactly two remaining good charged tracks are considered in the further event reconstruction.

5.1.3. Particle Identification

The particle identification routines are explained in detail in Section 3.3.

For each particle, the identification methods are applied in sequential order. Once a track is identified, the procedure is aborted such that one track is identified as exactly one particle type.

First, the track is checked to be an electron. Electron identification (EID) is only applied if the lab-frame momentum is higher than 400 MeV to prevent bad electron ID performance. Tracks
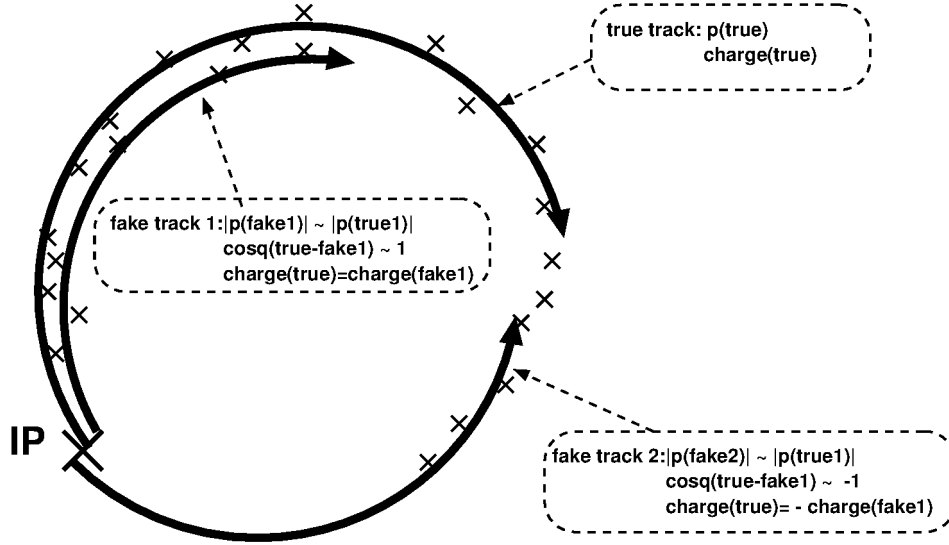
Figure 5.3.: Schematic view of duplicated tracks originating from a single real charged particle.

with an electron likelihood of $\mathcal{L} > 0.9$ are chosen to be an electron for the further analysis.
In case the EID fails, the tracks is checked to be a muon. A muon needs to have a lab-frame momentum of at least $p \geq 600\,\mathrm{MeV}$ to reach the KLM. In order to achieve a high efficiency in the muon identification, the minimum momentum required is set to $p \geq 800\,\mathrm{MeV}$. Tracks with an muon likelihood of $\mathcal{L} > 0.9$ are identified as muons.
Pion-kaon discrimination is applied to all tracks not identified as electron nor muon. If the pion-kaon likelihood is $\mathcal{L}(\pi; K) < 0.6$, the event is neglected.

All events with at least one charged pion on the signal side are further analysed. Events with two charged pions on the signal side are further tested, if one of the two pions can be combined with a neutral pion to form a $\rho^{\pm}$.

$\rho^{\pm}$ reconstruction

π^0 candidates are built by the Belle detector software by combining two photons. Combinations within the correct mass range are stored for further analysis. Monte Carlo studies show that the $\pi^{\pm}$ originating from the $\rho^{\pm}$ has usually less momentum in the center-of-mass (CM) frame than the $\pi^{\pm}$ originating directly from the $b \to u$ process. Therefore, only this $\pi^{\pm}$ is considered in the $\rho^{\pm}$ reconstruction. $\rho^{\pm}$ candidates are obtained by combining the $\pi^{\pm}$ with all valid π^0 candidates. A π^0 candidate is valid if both daughter photons fulfill the energy requirements of Table 5.2 and are not used in the reconstruction of the tagside. A mass vertex fit is performed for the $\pi^{\pm}\pi^0$ pair. Combinations are kept as $\rho^{\pm}$ candidates if the vertex fit succeeds with $\chi^2 < 20$ and the invariant mass of the candidate is in the range $m_{\pi^{\pm}\pi^0} \in [725; 925]\,\mathrm{MeV}$. In the case of multiple $\rho^{\pm}$ candidates, a best candidate selection is performed by using the candidate closest to the nominal $\rho^{\pm}$ mass $m \approx 875\,\mathrm{MeV}$ [10]. These criteria have been found to yield the highest expected significance after the final selection described in the next section.

Expected Cross-Feed

The requirements on the minimum lab-frame momenta for the electron and muon identification directly translate to the expectation of considerable cross-feed into the $\tau \to \pi$ reconstruction channel.

The simulated lab-frame momenta of the charged τ daughters are shown in Figure 5.4. The minimum momentum required for successful muon identification (μID) is indicated by the vertical black line. As can be seen, many $\tau \to \mu\nu\nu$ decays are not probed in the μID because they have not enough momentum. Instead, they proceed directly to the pion-kaon separation where many μ tracks are more pion like and the events end up in the $\tau \to \pi$ reconstruction. Almost no $\tau \to \mu\nu\nu$ are reconstructed in the $\tau \to e$ channel. While not as drastic, the same reasoning applies to electrons with $p < 400\,\text{MeV}$.

The exact cross-feed after all selection criteria applied is shown further below in Table 5.11.
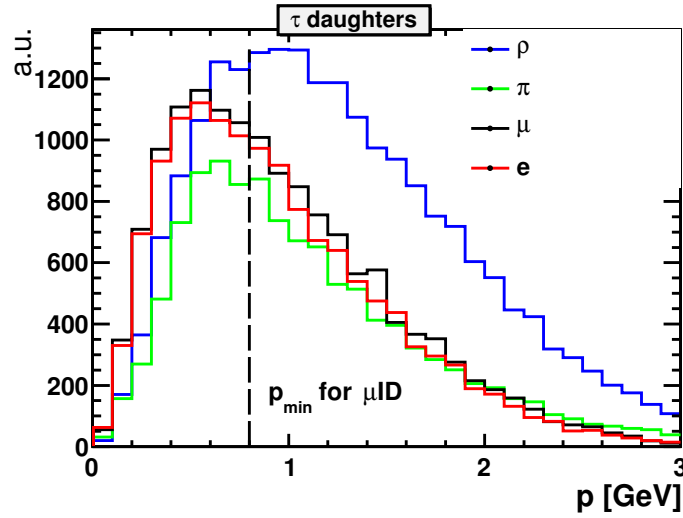
Figure 5.4.: Lab-frame momenta of the signal side τ decay daughter particles.

5.1.4. K_L veto

An important background arises from decays of type $B^0 \to D^{(*)+}\ell\nu$ with subsequent decays of $D^+ \to K_L\pi^+$. K_L mesons are not stopped in the Belle detector which results in missing momentum being observed. If decays with a K_L in the final state are not handled separately, they will display the same final state and kinematic requirements as the signal decay, namely two charged tracks on the signal side and missing momentum. The missing mass will often be higher than the K mass due to the additional neutrino present the decay, which is also the case for signal events.

A K_L meson traversing the Belle detector will sometimes deposit energy in the E_{ECL}, but not always. As explained below in Section 5.3.2, the energy deposited in the E_{ECL} is a crucial variable in this analysis as well as most τ analyses in the Belle group. A K_L meson is identified in the Belle detector as a cluster in the KLM with no associated charged track. A neutral cluster in the ECL on the extrapolated flight path from the interaction region to the K_L cluster in the

`KLM` is associated with the K_L candidate. Events that include a K_L without energy deposition in the `ECL` are vetoed in this analysis. The veto efficiency is found to be 0.860 ± 0.013 and 0.824 ± 0.005 on data and MC, respectively. A correction factor is introduced to correct for the MC versus data difference.

5.2. MC Corrections

The correction factors needed for this analysis are described in this section. First of all, the particle identification methods show a small difference in efficiency when performed on Monte Carlo simulations and data. These corrections are applied on the reconstructed particles of the signal side, only. An overall efficiency correction factor for the tagside is included separately. Furthermore, the Monte Carlo samples prepared by the Belle experiment may contain branching fractions or decay models that do not represent the current knowledge anymore. For the analysis presented in this paper, the decay model of the prominent $B \to D^{(*)}\ell\nu$ has to be corrected for.

5.2.1. Particle ID

The methods of particle identification used in this analysis are explained in detail in Section 3.3. The efficiency on simulation and data depends on the direction of the particle as well as the momentum. For tracks with a transverse momentum $p_t > 0.2\,\text{GeV}$, the efficiency is similar in both samples such that a only a small correction factor needs to be applied [96–98]. Tracks with less momentum will be in almost all cases pion because of the minimum momentum requirement of $0.4\,\text{GeV}$ and $0.8\,\text{GeV}$ for electron and muon identification, respectively, cf. Section 5.1.3. As can be seen in Figure 5.4, the momenta of the τ daughter particles on the signal side may be well below $0.2\,\text{GeV}$. In background decays such as $B^0 \to D^{*-}X$, the D^{*-} decays into $D^0\pi-$ and $D^-\pi^0$ [10], where the pions have low momenta. The difference in efficiency for these so-called slow pions has been studied in $B^0 \to D^{*-}\pi^+$ and $B^+ \to D^{*0}\pi^+$ decays [117]. Given that $B^0 \to D^{*-}\ell^+\nu$ is a non negligible background, the special treatment of slow pions has to be included in the analysis.

The product of all correction factors related to particle identification for all relevant decay types is shown in Figure 5.5. The decays are grouped by decay types.

5.2.2. Tag side

It has been found [118] that the efficiency of the hadronic full reconstruction differs between data and simulation. The difference depends on the network output, ie the reconstruction quality, and the reconstructed decay channel of the B_{tag}. A correction factor depending on these two variables is determined in decays with charmed semileptonic decays $B^0 \to D^{(*)-}\ell^+\nu_\ell$ on the signal side. It is assumed that the correction factor does not depend on the signal side, such that the factors obtained in this way can be applied to other signal side searches, too. The correction needs to be applied in for all simulations of neutral B^0 mesons and continuum $e^+e^- \to q\bar{q}$ events.

The tag side correction factor distribution is shown in Figure 5.6, splitted by decay types for all four τ reconstruction modes.
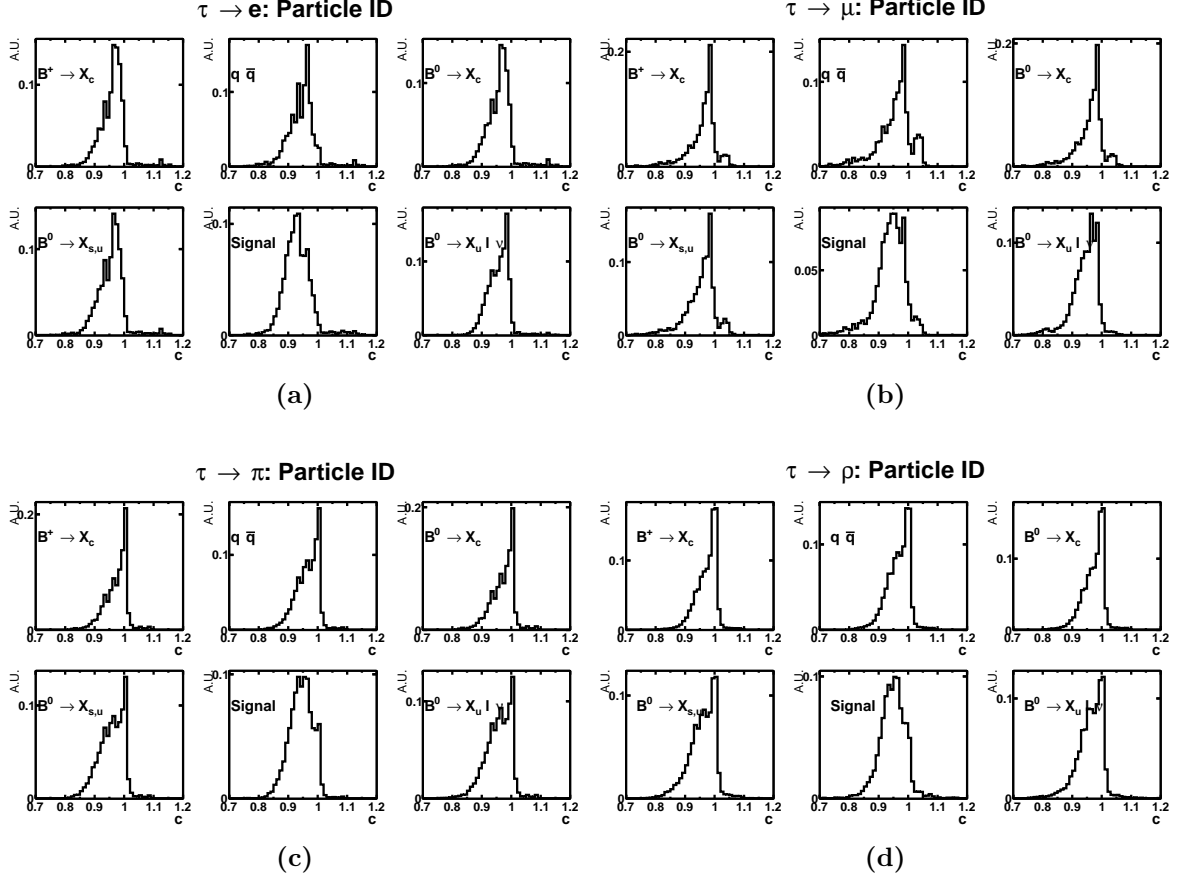
Figure 5.5.: Product of all correction factors related to particle identification. For each τ reconstruction mode, the correction factor distribution is plotted by decay process.

5.2.3. $D^{(*)}\ell\nu$ model

$B \to D^{(*)}\ell\nu$ decays are an important background, specially if the D meson decays into $K_L\pi$. If the K_L is not detected, it emulates the behavior of an additional neutrino and thus has the same final state as $B \to \pi\tau\nu$. The simulation in the Belle Monte Carlo samples of this decay is based on the heavy quark effective theory. The decay is described by the heavy quark effective theory and the `EvtGen` module takes three parameters in case of D^* and one in the case of D. The values have been updated to the current knowlegde and correction factors depending on the momentum transfer q^2 and the lepton momentum p in the center-of-mass frame of the decaying B meson have been derived [119].

Figure 5.7 shows the distribution of this correction factor for the relevant classes of B decays, namely the charmed MC samples $B^0 \to X_c$ and $B^+ \to X_c$.

5.2.4. Branching Fractions

The branching fractions of the dominant background decays have been updated to current knowledge. The correction factors are given as the ratio of the latest average branching fraction taken from the PDG report 2014 [10] over the branching fraction used in the Belle MC samples.
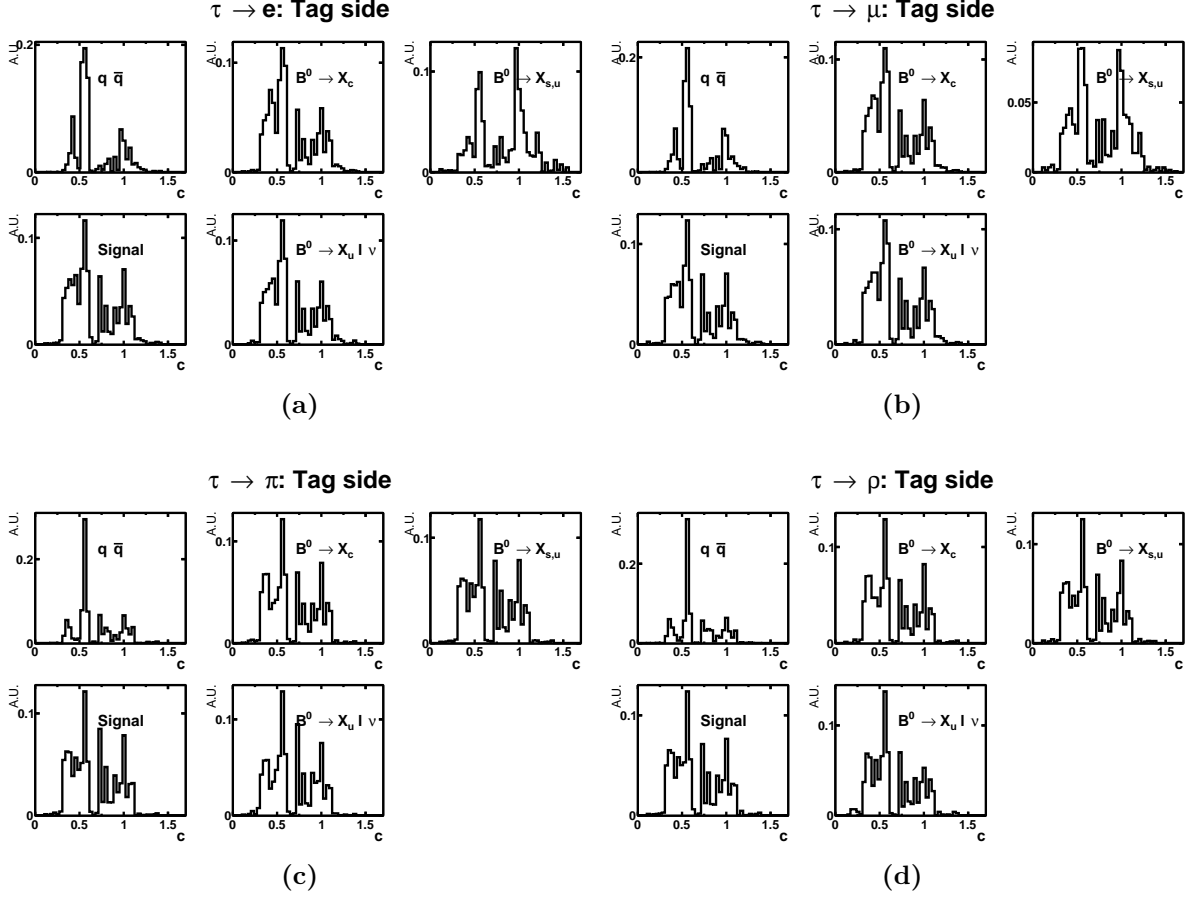
Figure 5.6.: Tag side correction factor distributions. For each τ reconstruction mode, the correction factor distribution is plotted by decay process.

Both values for the relevant decays are listed in Table 5.3. The $B \to D^{(*)}\ell\nu$ model correction introduced above does not result in a different branching fraction, but only alters the differential distribution. Therefore, both correction factors have to be applied for these decays.

Decay	$\mathcal{B}_{\mathrm{MC}}$	$\mathcal{B}_{\mathrm{PDG}}$	Decay	$\mathcal{B}_{\mathrm{MC}}$	$\mathcal{B}_{\mathrm{PDG}}$
$Dl\nu$	0.0213	0.0218	$K^0\eta$	$1.90 \cdot 10^{-6}$	$1.23 \cdot 10^{-6}$
$D^*l\nu$	0.0533	0.0493	$K^0\eta'$	$6.49 \cdot 10^{-5}$	$6.60 \cdot 10^{-5}$
$K^{*+}K^{*-}$	$1.41 \cdot 10^{-4}$	$2 \cdot 10^{-6}$			

Table 5.3.: Branching fraction corrections for the main background contributions. PDG values taken from Ref.[10].
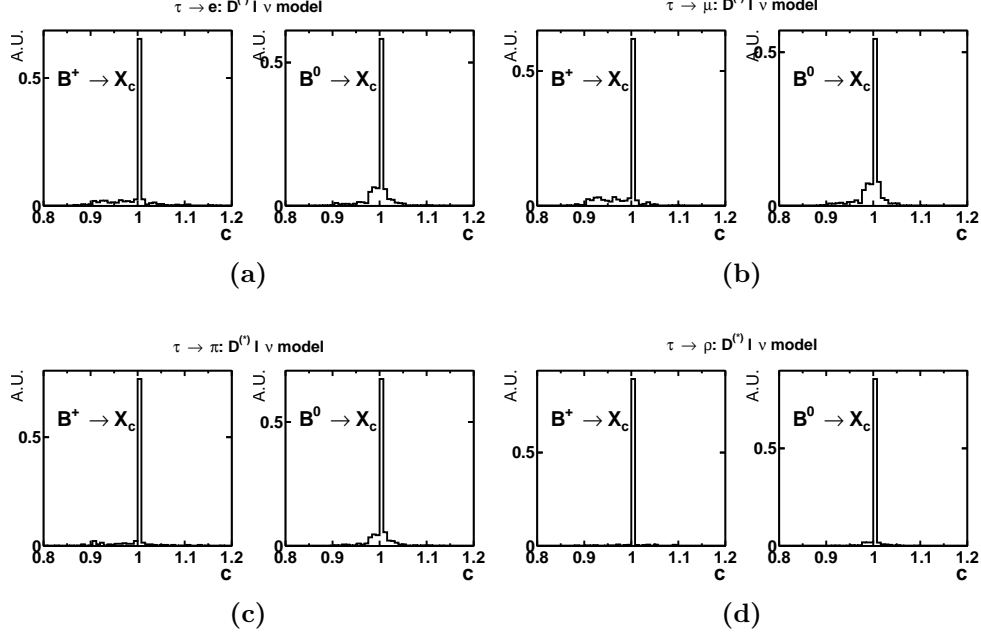
Figure 5.7.: Correction factor distributions for charmed semileptonic decays $B \to D^{(*)}\ell\nu$. For each τ reconstruction mode, the correction factor distribution is plotted by decay process.

5.3. Event Variables

This section will describe the variables and their labels used in this analysis. It is split into two sections, where first the variables used in the BDTs are described, followed by the definition of the extra energy E_{ECL} which is the discriminator variable used for the fit and limit calculation.

5.3.1. BDT Input Variables

Multiple variables are used in the BDT training and evaluation, but only a subset of them is used in the single τ decay modes. All variables are defined in this section, while their usage as well as their distributions are shown below, in the relevant description of the BDT setup in Section 5.4.

All variables are listed in Table 5.4. The subscripts of the variables have the following meaning:

It is important to note that although l stands for *lepton*, $|\vec{p}_l|$ denotes $|\vec{p}_\pi|$ in case of $\tau \to \pi$ decay. Next to obvious kinematic variables available in the process, several combinations of different 4-momentum vectors are built and called m_n with a counting index n.

5.3.2. Extra Energy

The extra energy E_{ECL} is defined as all energy deposited in the ECL that is not associated to any track or neutral cluster originating from the tag or signal side. In signal events, the distribution should peak at $E_{ECL} = 0\,\text{GeV}$. While the number of charged tracks left on the

signal side is constrained to two oppositely charged tracks, no such requirement is applied for neutral clusters. Background events usually contain additional photons or neutral pions which lead to higher values of E_{ECL}. The E_{ECL} variable is often used at Belle in rare decays with τ leptons involved.

To compute the extra energy, each cluster c in the `ECL` is run through the following tests:

1. if c is associated with B_{tag} or a charged particle on the signal side, reject it.

2. c has to contain a minimum energy depending on the cluster location

   - Forward Endcap: $E_c > 0.1\,\text{GeV}$

   - Barrel: $E_c > 0.05\,\text{GeV}$

   - Backward Endcap: $E_c > 0.15\,\text{GeV}$

If the cluster c has an associated track, the quality of the track - cluster association is checked. If the association is of good quality and the cluster has been used in the signal or tag side reconstruction, the cluster is not used in the calculation of the extra energy. The track - cluster association is defined to be of good quality if the interpolated track matches the center of the cluster. If this is not the case, the track may still be associated to the cluster if the intersection between interpolated track and cluster exists. In this case though, the cluster has to fail the definition of a good photon cluster as defined in Section 3.3.4. The distribution of E_{ECL} for the four τ reconstruction modes are shown in Figure 5.8.

(a) $\tau \to e$

(b) $\tau \to \mu$

(c) $\tau \to \pi$

(d) $\tau \to \rho$
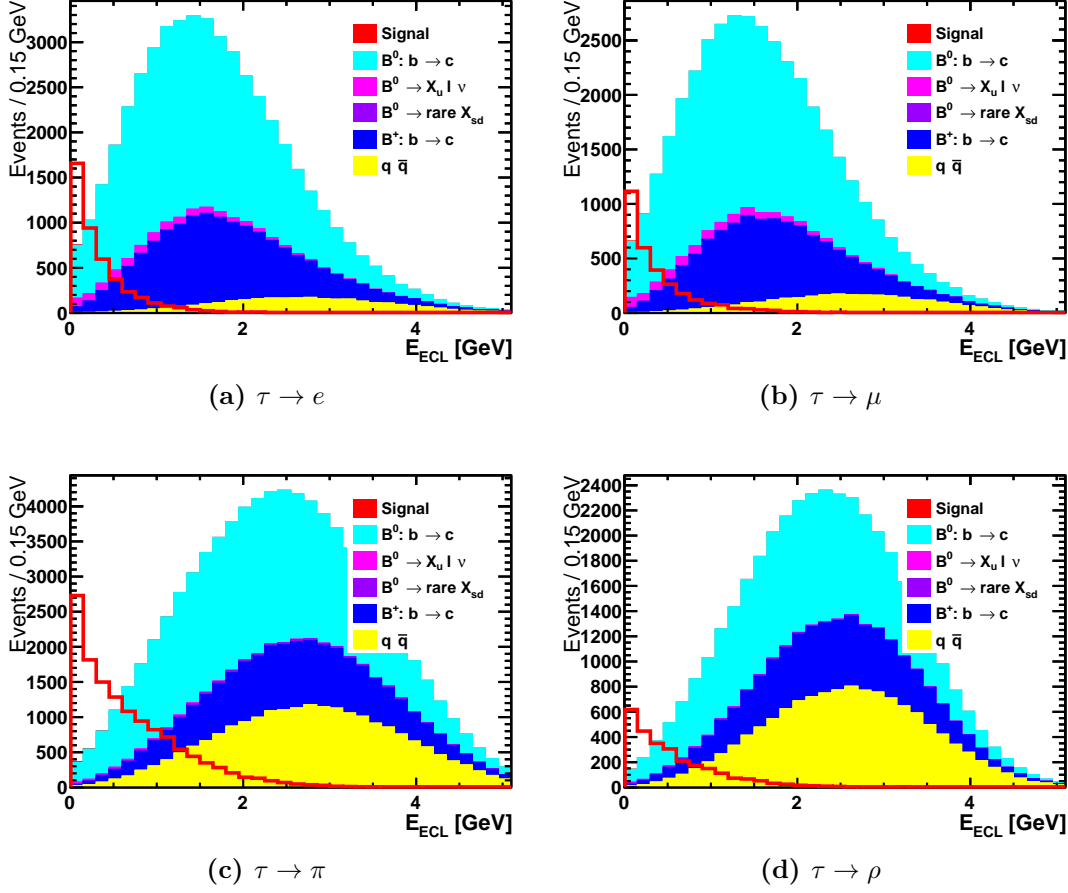
Figure 5.8.: Distributions of E_{ECL} and number of expected events. The plots are created from MC with the requirement of a successful reconstruction of a B^0 candidate on the tag side, exactly two charged oppositely tracks on the signal side which are identified to fit the corresponding τ reconstruction mode. Additionally, a quality requirement of the hadronic tag reconstruction is applied by the cut $\ln o_{\mathrm{tag}}^{\mathrm{cs}} > -7$ and $\ln o_{\mathrm{tag}}^{\mathrm{cs}} > -5$ for the leptonic and hadronic τ reconstruction modes, respectively, which is the cut used in the final selection. The signal contribution is not plotted in the stacked histogram, instead it is scaled by a factor of 200 with respect to $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.0 \times 10^{-4}$.

Variable	Definition	Description
$\|\vec{p}_l\|$		Magnitude of the 3-momentum of the τ daughter particle
$\|\vec{p}_m\|$		Magnitude of the 3-momentum of the B-meson daughter particle $\pi^{\pm}$
$p_{B_{\text{vis}}}$	$p_{B_{\text{vis}}} = p_l + p_m$	4-momentum sum of both charged tracks of the signal side
$\|\vec{p}_{B_{\text{vis}}}\|$		Magnitude of the 3-momentum of $p_{B_{\text{vis}}}$
$E_{B_{\text{vis}}}$		Visible energy on the signal side
p_{beam}	$p_{\text{beam}} = \frac{1}{2}\left(p_{e^-} + p_{e^+}\right)$	Beam momentum
p_{miss}	$p_{\text{miss}} = 2p_{\text{beam}} - p_{B_{\text{tag}}} - p_{B_{\text{vis}}}$	
M^2_{miss}	$\text{M}^2_{\text{miss}} = p^2_{\text{miss}}$	Missing mass squared
E_{miss}		Missing energy, first component of p_{miss}
$p_{(\tau\nu)}$	$p_{(\tau\nu)} = p_\tau + p_{\nu_\tau} = p_{B_{\text{sig}}} - p_m$	4-momentum of the lepton pair
$m^2_{(\tau\nu)}$		Mass squared of the lepton pair
$\|\vec{p}_{(\tau\nu)}\|$		3-momentum of the lepton pair
N_{π^0}		Number of π^0 candidates on the signal side

4-momentum combinations

$$m_4 = (E_{\text{beam}} - E_m)^2 - (\vec{p}_{\text{beam}} - \vec{p}_m)^2$$

$$m_5 = (p_{\text{beam}} - p_l)^2$$

$$m_{10} = (E_{\text{beam}} - E_m)^2 - (\vec{p}_{\text{beam}} - \vec{p}_m)^2 + (E_{\text{beam}} - E_l)^2 - (\vec{p}_{\text{beam}} - \vec{p}_l)^2$$

$$m_{11} = \frac{(p_{\text{beam}} - p_m)^2}{(p_{\text{beam}} - p_l)^2}$$

$$m_{12} = (E_{\text{miss}} + E_l + E_m)^2 - (\|\vec{p}_{\text{miss}}\| + \|\vec{p}_l\| + \|\vec{p}_m\|)^2$$

$$m_{15} = \frac{(E_{\text{miss}} - E_m)^2 - (\vec{p}_{\text{miss}} - \vec{p}_m)^2}{(E_l + E_m)^2 + (\vec{p}_l - \vec{p}_m)^2}$$

$$m_{57} = (p_{\text{beam}} - p_m)^2$$

$$m_{59} = \frac{E_{\text{miss}}}{E_m}$$

$$m_{510} = \frac{E_{\text{miss}}}{E_{B_{\text{vis}}}}$$

Table 5.4.: Variable definitions of input variables used in the Boosted Decision Trees. 4-momentum variables are not used in the BDTs but still listed to help the definition of the 4-momentum vector components which are used in the analysis.

5.4. BDT Training

The separation between signal and background is done by using Boosted Decision Trees (BDT), of which the principle is described in Section 4.3. For every τ reconstruction channel, one BDT classifier is trained. Due to the small number of expected signal events, only one classifier is used. Using different classifiers for the different background types reduces the expected significance, mainly because of the very low expected number of signal events. Due to different kinematics in the reconstruction channels because of different masses and background contributions, using one single BDT classifier for the complete τ reconstruction results in less separation power. The configuration of the BDT training is described in this section, as well as the output and result on the Monte-Carlo samples. For all BDT classifiers, many different configurations have been checked in order to establish the best set. The optimization has been done by choosing the BDT configuration which yielded the highest expected statistical significance, given by

$$\sigma = \sqrt{-2 \cdot \ln\left(\frac{\mathcal{L}_0}{\mathcal{L}_1}\right)}, \tag{5.1}$$

where the likelihoods $\mathcal{L}_0$ and $\mathcal{L}_1$ are given by

$$\mathcal{L}_k = \prod_{i=1}^{n} P\left(x_i; X_{i,k}\right), \tag{5.2}$$

with P being the Poisson function, x_i being the number of expected events, $X_{i,0}$ the number of predicted background events and $X_{i,1}$ the number of predicted background plus signal events in bin i. The number of expected events is the sum of background and signal events. The numbers are obtained from the Monte Carlo samples with the efficiency corrections described in Section 5.2 applied.

After initial trainings, a set of precuts has been determined to speed up and improve the general BDT performance. Cuts are made on the missing mass squared, the quality output of the full hadronic reconstruction and the extra energy in the ECL, specifically events are kept if

- $M_{\text{miss}}^2 > -0.5 \, \text{GeV}^2$
- $\ln o_{\text{tag}}^{\text{cs}} > -7$
- $E_{ECL} < 1.0 \, \text{GeV}$.

These cuts are not very restrictive on the signal sample, and mostly serve to reduce computing time as well restrict the training events to the signal region.

All classifiers are trained with a signal sample consisting of 30×10^6 generated signal events. Background events are taken from two streams of $B^0 \to X_c$ events in all classifiers, while additional decay types or streams are mentioned in the specific section below. The production and types of MC samples are described in Section 3.4.2. All efficiency correction factors are applied.

The BDT training and evaluation is done with the TMVA [105] toolkit. Additionally, 30×10^6 generated signal events are used as the testing sample in TMVA, which is used to check for overtraining.

The plots shown in the next sections include plots of background rejection versus signal efficiency, also called receiver operating characteristic (ROC) curve. The ROC curve can be used to

interpret the quality or performance of a classifier. The goal is to reach high background rejection while maintaining a good signal efficiency. While the ROC curve has been used for this analysis to get an idea of the BDT performance, it has not been used in order to decide which BDT configuration should be used finally. This decision has been made by calculating the expected significance defined in Equation (5.1). The expected number of events has been computing using all samples that have not been used for the BDT training. In this way, overtraining on the training sample does not propagate into the final BDT decision. The exact procedure to determine the final cut selection is described in Section 5.5.

Many different combinations of variables and BDT configurations have been tested, varying the boosting algorithms, number of trees, decorrelation methods, tree depths and node numbers, and bagging fraction. When stochastic gradient boost is used, the bagging fraction described the size of the sub sample used in each tree. The following sections describe the best BDT configuration in terms of maximized expected significance, which is used in this analysis.

5.4.1. $\tau \to e\nu\nu$

The $\tau \to e$ BDT classifier is trained using five streams of $B^0 \to X_c$ MC and half of the available $B^0 \to X_u^- \ell^+ \nu$ sample as background events. An additional signal sample is used which contains 20×10^6 generated signal events with only $\tau \to e\nu\nu$ decays.

As an additional precut in the training, the events are required to contain at maximum one charged track outside of the impact parameter cut given in Section 5.1.2. The BDT training configuration is listed in Table 5.5. The variables used for training are shown in Figure 5.9a. Since most variables are highly correlated, the Deco decorrelation algorithm is applied, which results in the variable distributions shown in Figure 5.9b. The result of the decorrelation can also be seen in Figure 5.10. Figure 5.11a shows the BDT output for both signal and background samples, Figure 5.11b plots the ROC curve.

Variables	$|\vec{p}_e|$, $|\vec{p}_\pi|$, $\mathrm{M}^2_{\mathrm{miss}}$, m_5, m_{11}, m_{57}, m_{59}, m_{510}, $m^2_{(\tau\nu)}$
Variable Transformation	Decorrelation
Number of Trees	800
Boost Type	Gradient Boost
Shrinkage	0.8
Max Depth	5

Table 5.5.: Configuration of the final BDT classifier used in $\tau \to e$.
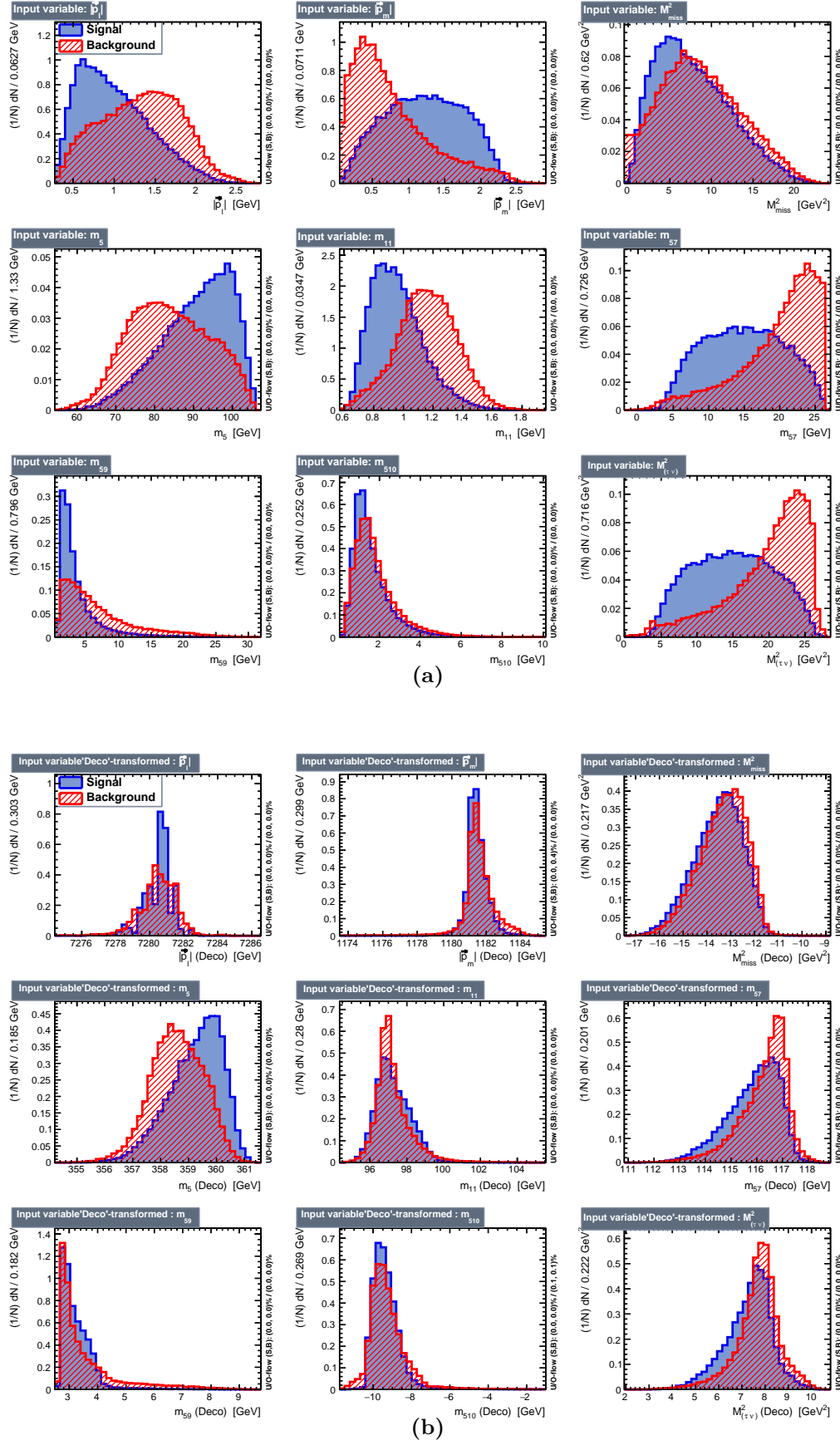
Figure 5.9.: Variables used in the training of the $\tau \to e$ BDT classifier, (a) vanilla and (b) decorrelated.

(a)

(b)
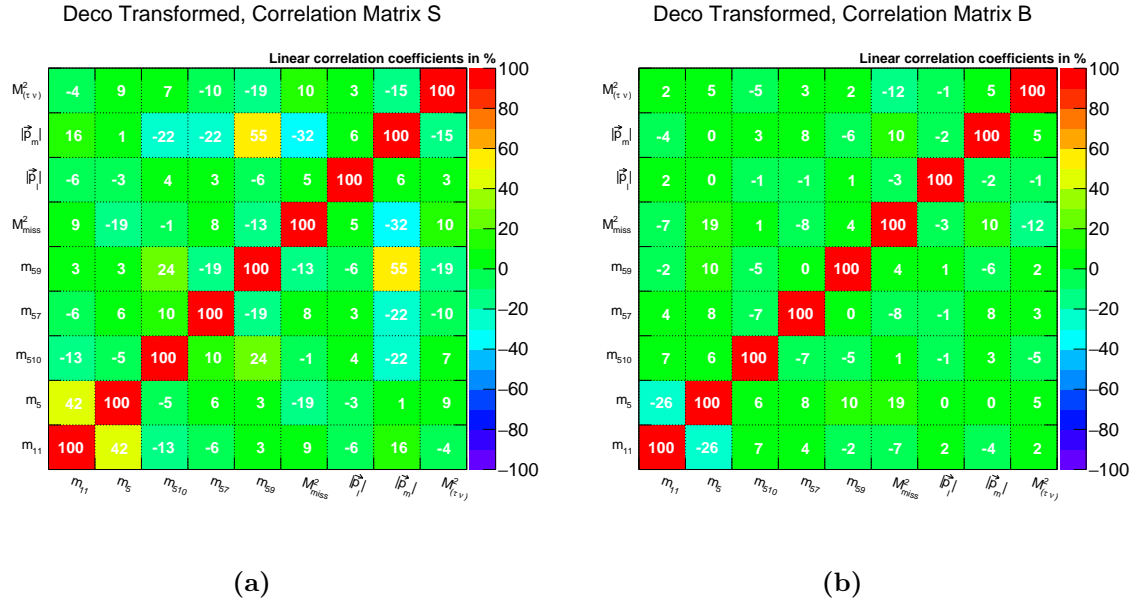
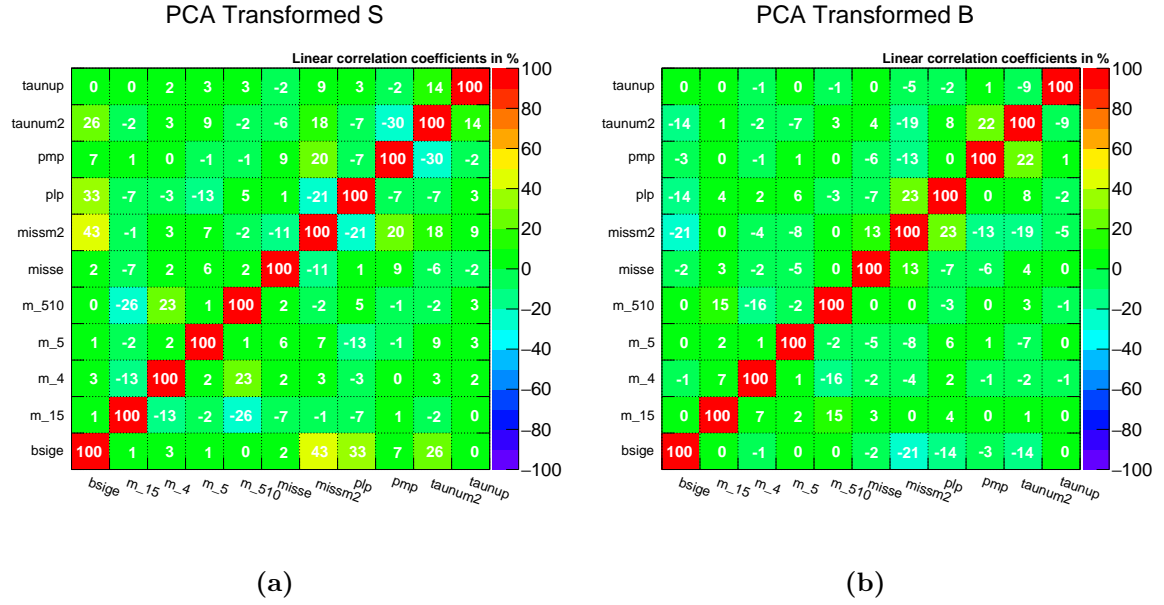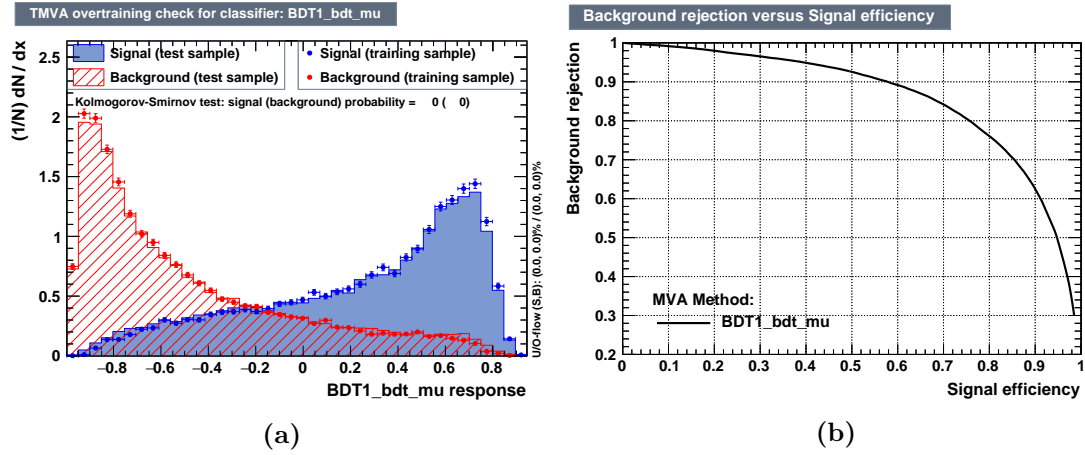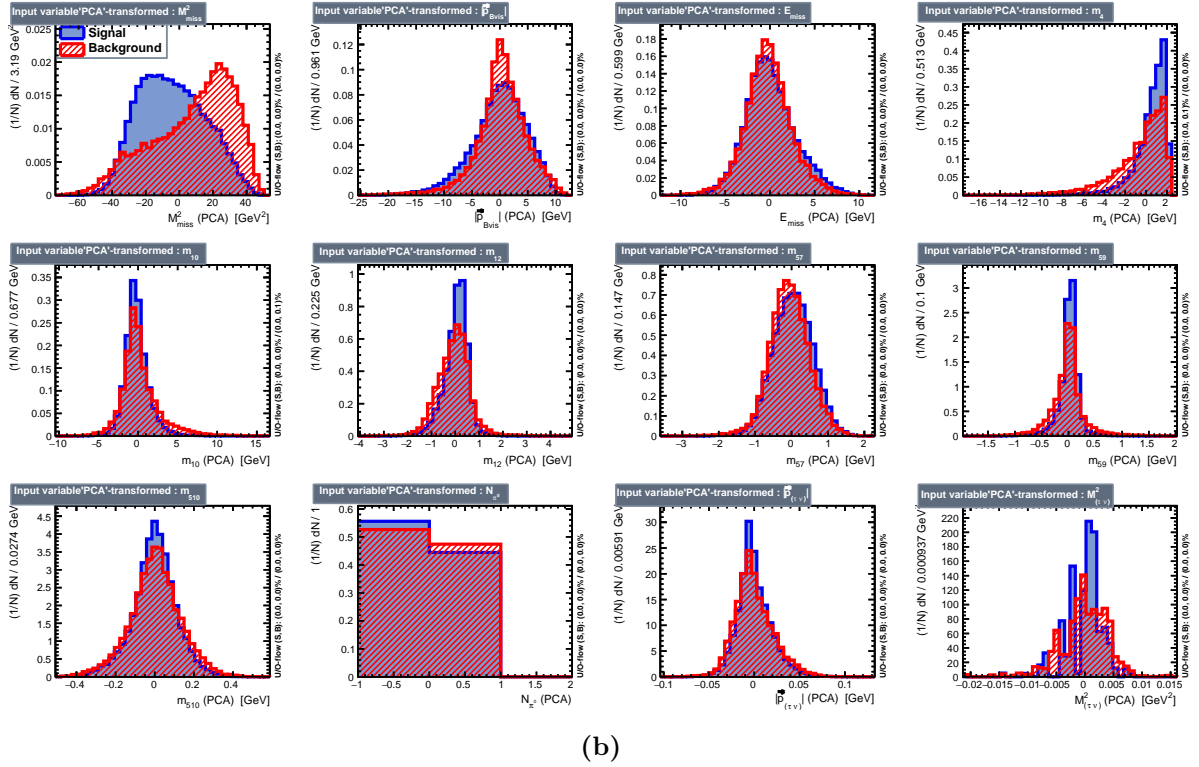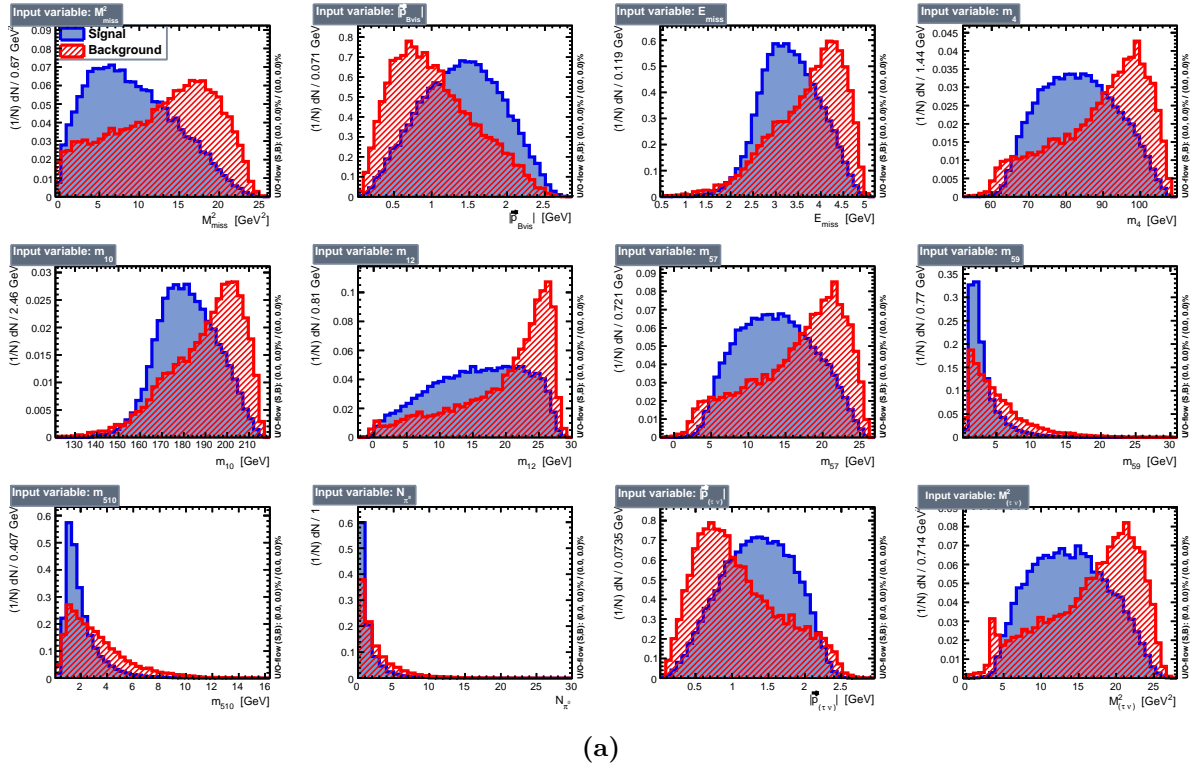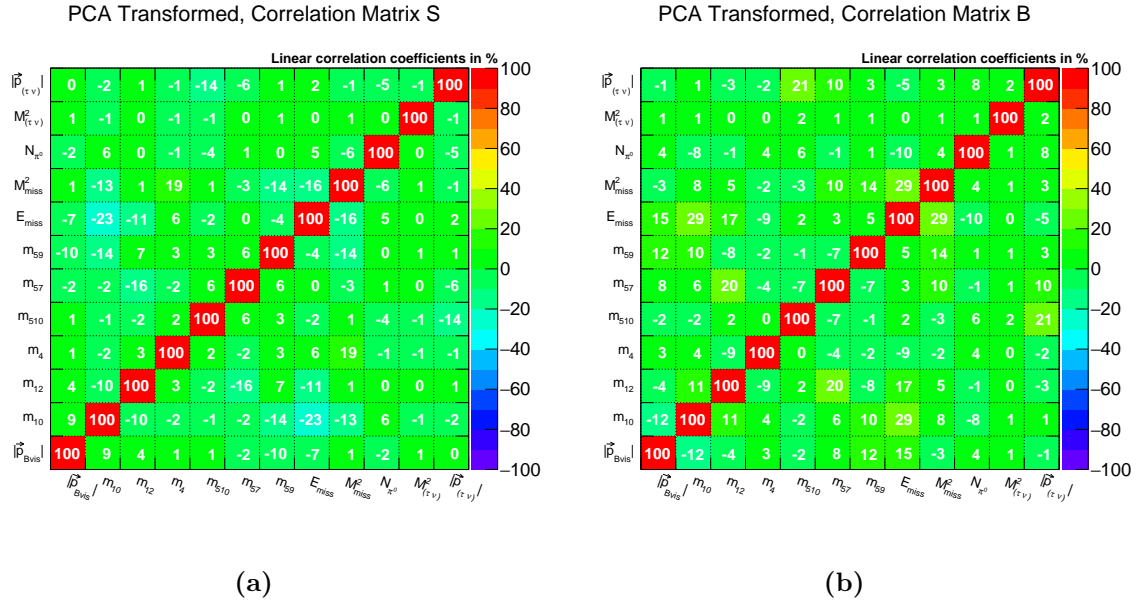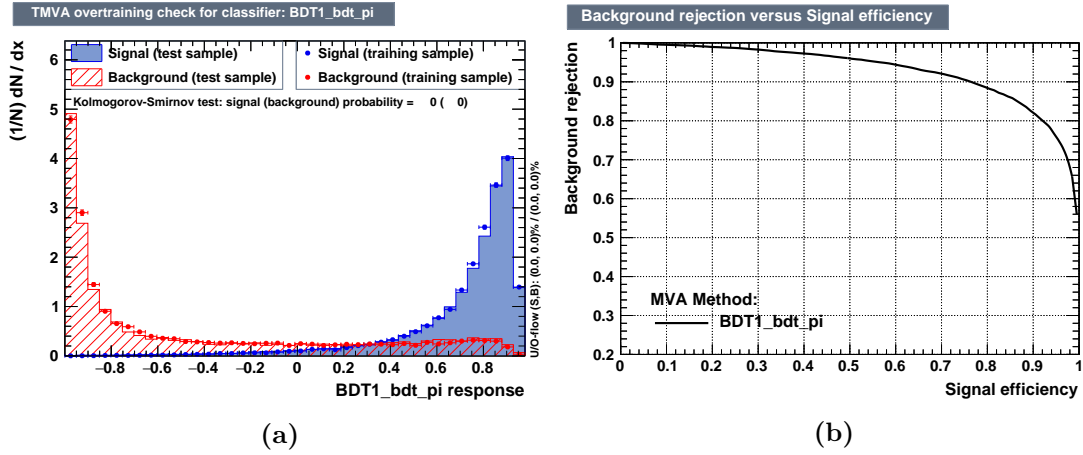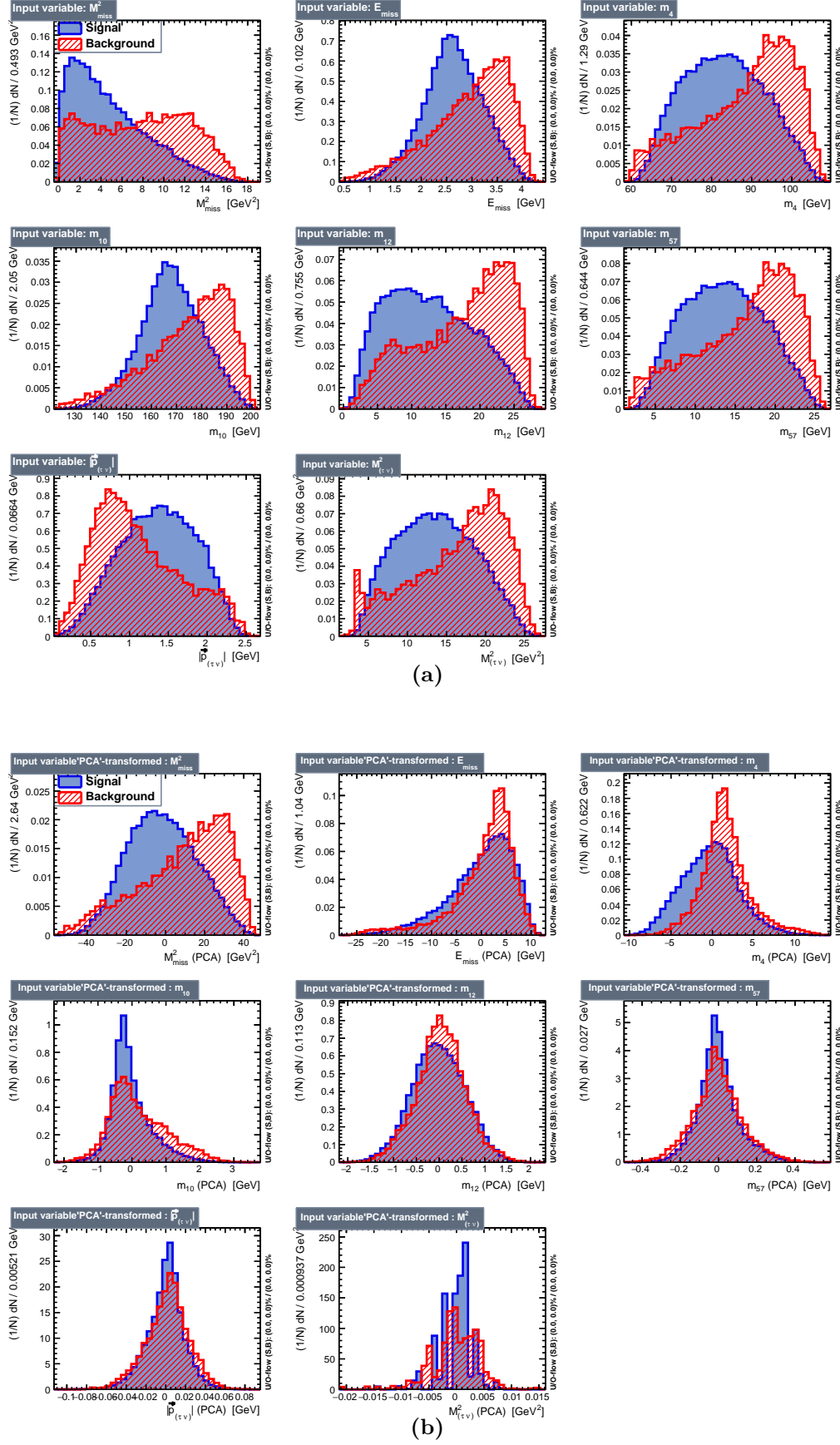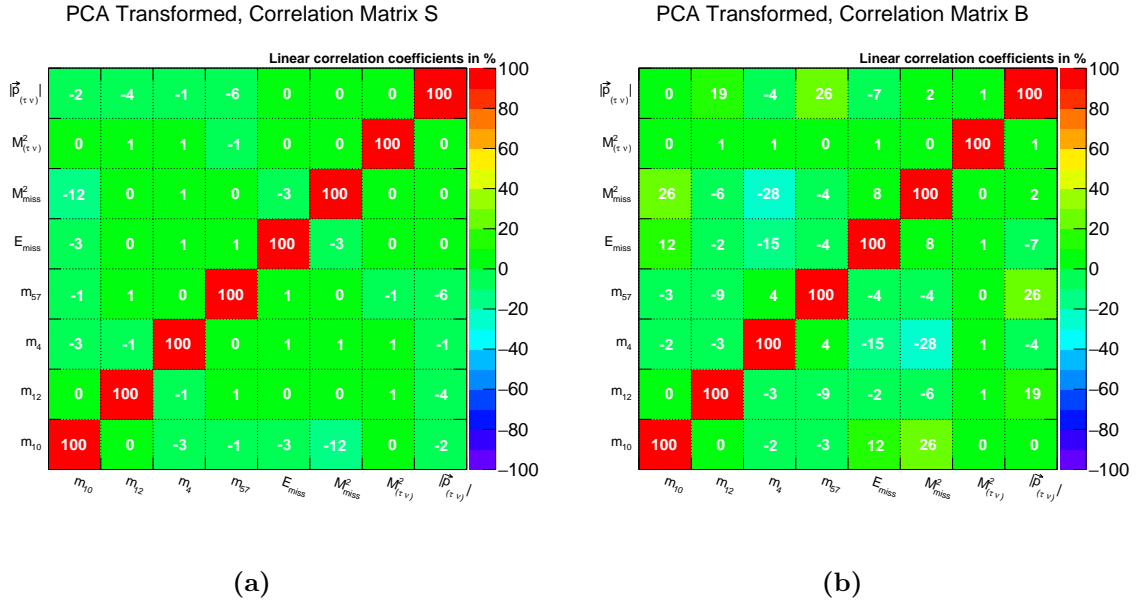**Figure 5.10.:** Correlation matrix of the $\tau \to e$ variables for (a) signal and (b) background.

(a)

(b)
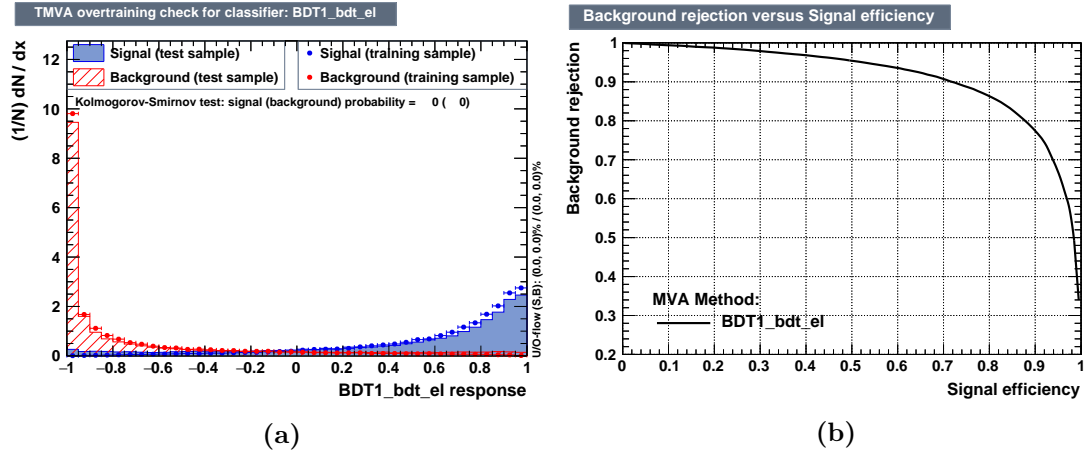
Figure 5.11.: Information and results of the $\tau \to e$ BDT classifier. (a) shows the distribution of the output variable for the TMVA training and testing samples, (b) plots the ROC curve.

5.4.2. $\tau \to \mu\nu\nu$

The $\tau \to \mu$ BDT classifier is trained using four streams of $B^0 \to X_c$ MC and half of the available $B^0 \to X_u^- \ell^+ \nu$ sample as background events. The default signal sample is used.

As an additional precut in the training the events are required to contain at maximum one charged track outside of the impact parameter cut given in Section 5.1.2. The BDT training configuration is listed in Table 5.6. The variables used for training are shown in Figure 5.12a. Since most variables are highly correlated, the PCA transformation algorithm is applied, which results in the variable distributions shown in Figure 5.12b. The result of the decorrelation can also be seen in Figure 5.13. Figure 5.14a shows the BDT output for both signal and background samples, Figure 5.14b plots the ROC curve.

Variables	$|\vec{p}_\mu|$, $|\vec{p}_\pi|$, m_4, m_5, m_{15}, m_{510}, $|\vec{p}_{(\tau\nu)}|$, $m_{(\tau\nu)}^2$, E_{B_vis}, $\mathrm{M}_\mathrm{miss}^2$, E_miss
Variable Transformation	Principle Component Analysis
Number of Trees	200
Boost Type	Gradient Boost
Shrinkage	0.1
Max Depth	5

Table 5.6.: Configuration of the final BDT classifier used in $\tau \to \mu$.

5.4.3. $\tau \to \pi\nu$

The final $\tau \to \pi$ BDT classifier is trained using four streams of $B^0 \to X_c$ MC and half of the available $B^0 \to X_u^- \ell^+ \nu$ sample as background events. The default signal sample is used.

The BDT training configuration is listed in Table 5.7. The variables used for training are shown in Figure 5.15a. Since most variables are highly correlated, the PCA transformation algorithm is applied, which results in the variable distributions shown in Figure 5.15b. The result of the decorrelation can also be seen in Figure 5.16. Figure 5.17a shows the BDT output for both signal and background samples, Figure 5.17b plots the ROC curve.

Variables	$\mathrm{M}_\mathrm{miss}^2$, E_miss, m_4, m_{10}, m_{12}, m_{57}, m_{59}, m_{510}, $|\vec{p}_{(\tau\nu)}|$, $m_{(\tau\nu)}^2$, $|\vec{p}_{B_\mathrm{vis}}|$, N_{π^0}
Variable Transformation	Principle Component Analysis
Number of Trees	1000
Boost Type	Stochastic Gradient Boost
Bagging Fraction	50%
Shrinkage	0.1
Max Nodes	10

Table 5.7.: Configuration of the final BDT classifier used in $\tau \to \pi$.

5.4.4. $\tau \to \rho\nu$

The final $\tau \to \rho$ BDT classifier is trained using four streams of $B^0 \to X_c$ MC and half of the available $B^0 \to X_u^- \ell^+ \nu$ sample as background events. The default signal sample is used.

The BDT training configuration is listed in Table 5.8. The variables used for training are shown in Figure 5.18a. Since most variables are highly correlated, the PCA transformation algorithm is applied, which results in the variable distributions shown in Figure 5.18b. The result of the decorrelation can also be seen in Figure 5.19. Figure 5.20a shows the BDT output for both signal and background samples, Figure 5.20b plots the ROC curve.

Variables	M_{miss}^2, E_{miss}, m_4, m_{10}, m_{12}, m_{57}, $|\vec{p}_{(\tau\nu)}|$, $m_{(\tau\nu)}^2$
Variable Transformation	Principle Component Analysis
Number of Trees	500
Boost Type	Stochastic Gradient Boost
Bagging Fraction	80%
Shrinkage	0.1
Max Nodes	50

Table 5.8.: Configuration of the final BDT classifier used in $\tau \to \rho$.
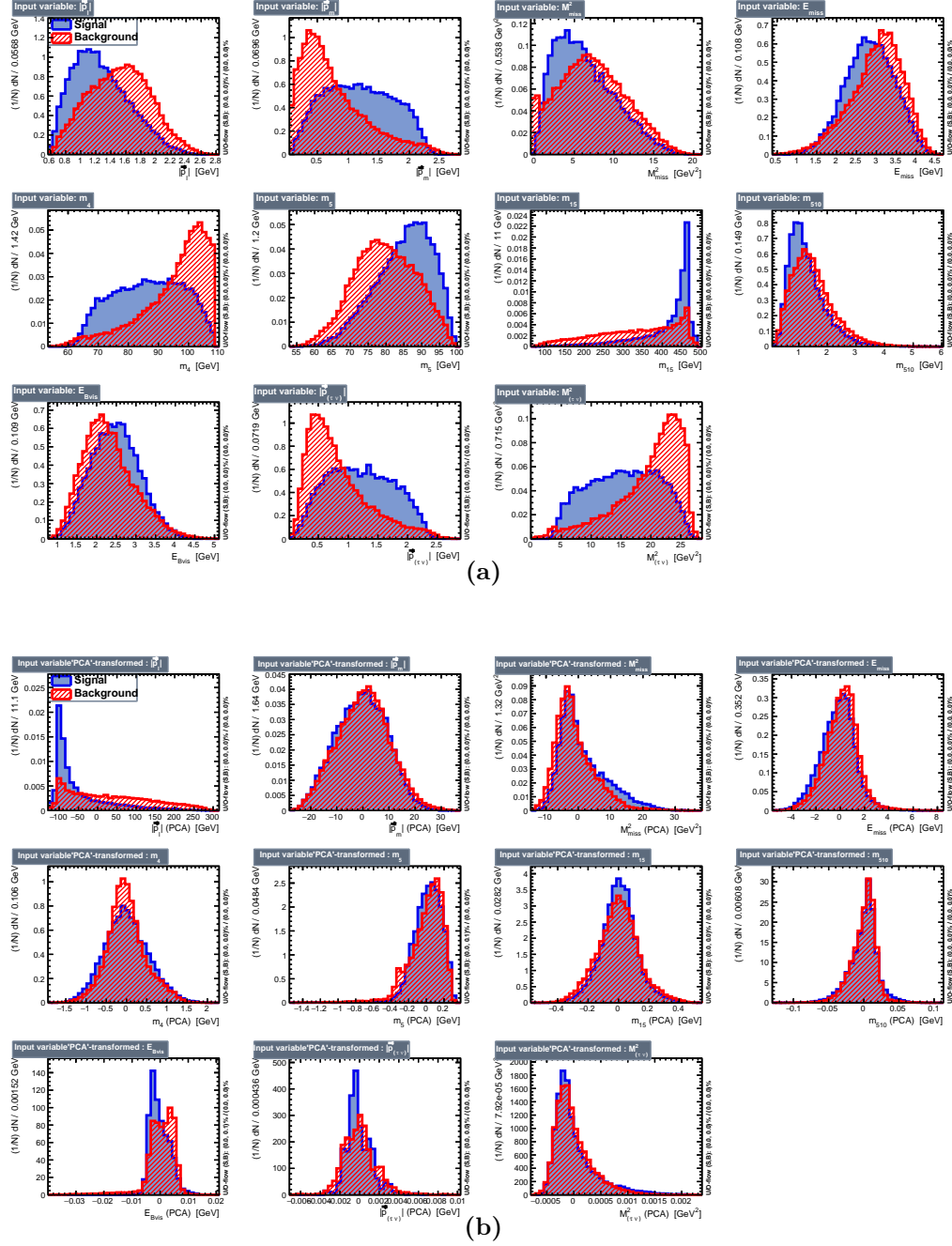
Figure 5.12.: Variables used in the training of the $\tau \to \mu$ BDT classifier, (a) vanilla and (b) PCA transformed.

Figure 5.13.: Correlation matrix of the $\tau \to \mu$ variables for (a) signal and (b) background.

Figure 5.14.: Information and results of the $\tau \to \mu$ BDT classifier. (a) shows the distribution of the output variable for the TMVA training and testing samples, (b) plots the ROC curve.

Figure 5.15.: Variables used in the training of the $\tau \to \pi$ BDT classifier, (a) vanilla and (b) PCA transformed.

(a)

(b)

Figure 5.16.: Correlation matrix of the $\tau \to \pi$ variables for (a) signal and (b) background.

(a)

(b)

Figure 5.17.: Information and results of the $\tau \to \pi$ BDT classifier. (a) shows the distribution of the output variable for the TMVA training and testing samples, (b) plots the ROC curve.

Figure 5.18.: Variables used in the training of the $\tau \to \rho$ BDT classifier, (a) vanilla and (b) PCA transformed.

Figure 5.19.: Correlation matrix of the $\tau \to \rho$ variables for (a) signal and (b) background.
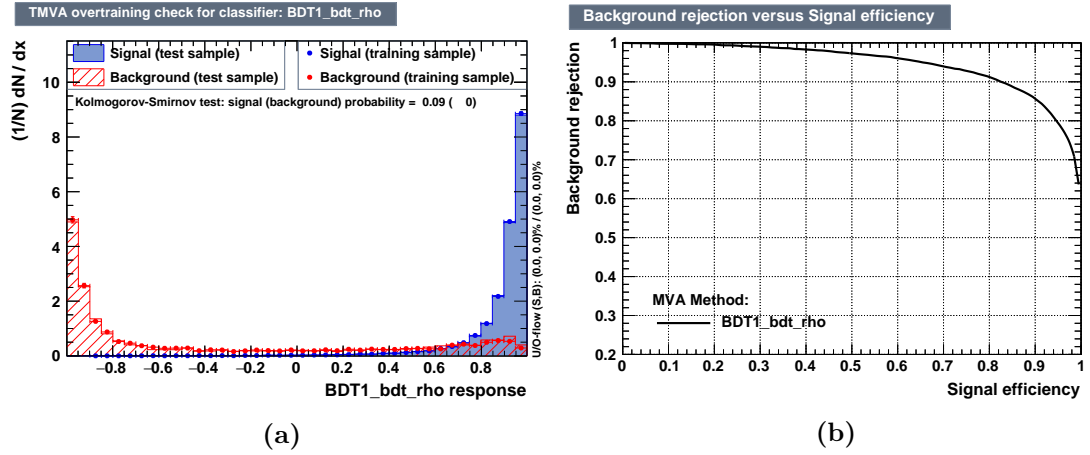
Figure 5.20.: Information and results of the $\tau \to \rho$ BDT classifier. (a) shows the distribution of the output variable for the TMVA training and testing samples, (b) plots the ROC curve.

5.5. Final Event Selection

In order to determine the final selection, the expected significance given in Equation (5.1) is calculated and the cut yielding the highest value is chosen. The calculation is performed on all MC samples that have not been used in the BDT training. The size of these is five times the data sample size for the $B^0 \to X_c$, $B^\pm \to X_c$ and $e^+e^- \to q\bar{q}$ samples, ten times the size of data for the $B \to X_u\ell\nu$ samples and 50 times the size of the data sample for the $B \to X_{us}$ sample which has not been used in the BDT training. The signal contribution is computed using a sample containing 24×10^6 generated events, scaled to $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.0 \times 10^{-4}$. All background samples are scaled to their respective luminosity given the Belle dataset.

The expected significance has been computed varying cuts on four variables simultaneously. The cuts are $\ln o_{\mathrm{tag}}^{\mathrm{cs}} > C1$, $\mathrm{M}_{\mathrm{miss}}^2 > C2$, BDT $> C3$, where BDT is the BDT output and the maximum number of charged tracks allowed outside the impact parameter range $dr < 4\,\mathrm{cm}$ and $dz < 2\,\mathrm{cm}$. The cut ranges have been $C1 \in [-7; -1]$, $C2 \in [0.0; 2.0]\,\mathrm{GeV}^2$ and $C3 \in [-1; 1]$. Since a scan over four variables is not easily plotted and the cut on the BDT output had the most impact, a plot of the expected significance depending on the BDT output cut is shown in Figure 5.21 for different signal branching fractions, where both other variables are fixed at their optimal value.

A complete list of the three values for all four reconstruction modes is given in Table 5.9.

Variable	$\tau \to e$	$\tau \to \mu$	$\tau \to \pi$	$\tau \to \rho$
$M_{\mathrm{bc,tag}}$	$\in (5.27; 5.29)\,\mathrm{GeV}$	$\in (5.27; 5.29)\,\mathrm{GeV}$	$\in (5.27; 5.29)\,\mathrm{GeV}$	$\in (5.27; 5.29)\,\mathrm{GeV}$
K_L veto	✓	✓	✓	✓
$\ln o_{\mathrm{tag}}^{\mathrm{cs}}$	> -7	> -7	> -5	> -5
N_{extra}	< 2	< 2	< 2	< 2
$\mathrm{M}_{\mathrm{miss}}^2$	$> 2.2\,\mathrm{GeV}$	$> 0.8\,\mathrm{GeV}$	$> 0.0\,\mathrm{GeV}$	$> 0.6\,\mathrm{GeV}$
BDT	> 0.48	> 0.21	> 0.8	> 0.9

Table 5.9.: Selection requirements for all four reconstruction modes.

5.6. Extra Energy

The corrections described in Section 5.2 are applied to the Monte Carlo samples. They are then scaled to the luminosity recorded with Belle at the $\Upsilon(4S)$ resonance, $\mathcal{L} = 711\,\mathrm{fb}^{-1}$, and the E_{ECL} distributions are obtained which reflect the expected number of events in the Belle data sample.

The E_{ECL} distributions are shown in Figure 5.22, separately for the four τ reconstruction modes, grouped by signal and different background types.

It is clearly visible that the signal peaks at very low values, as it is expected. The background distributions peak at higher values for all background types, such that a fit can discriminate well between signal and background.

Secondly, the number of expected background events differs clearly between the τ reconstruction modes. The dominant background over the whole E_{ECL} region is in all cases the $b \to c$

(a) $\tau \to e$

(b) $\tau \to \mu$

(c) $\tau \to \pi$

(d) $\tau \to \rho$
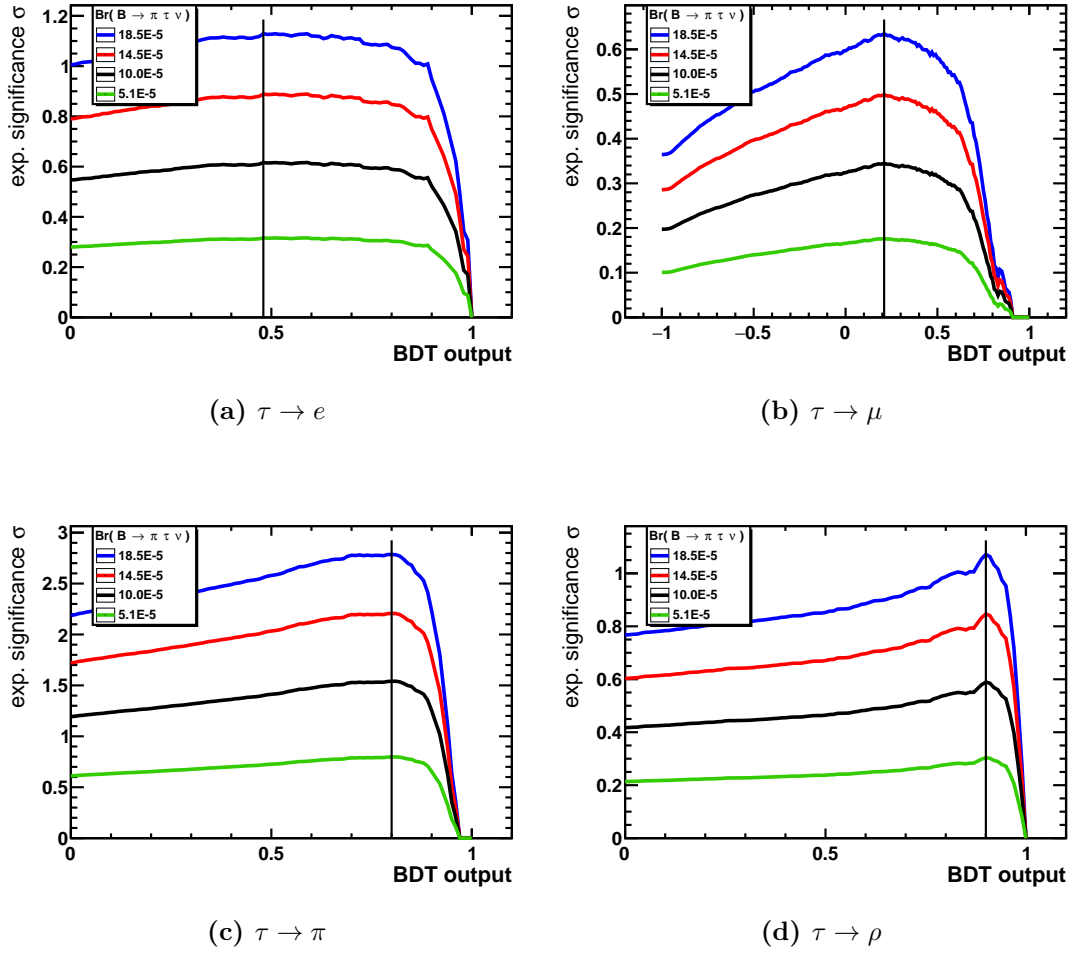
Figure 5.21.: Expected significance as a function of the BDT output. In the plot, the other three cut variables are fixed at their respective optimal value. The vertical black line indicates the optimal, final cut on the BDT output.

transitions of neutral B mesons, which is expected due to the big value of $|V_{cb}|$ compared to $|V_{ub}|$. Semileptonic $b \to u$ decays of neutral B mesons are an important background in the low E_{ECL} region, as they result in similar final states as the signal decay. The exact background composition is detailed below in Section 5.8.

While the expected signal yield in the $\tau \to \pi$ channel is the highest, followed by the $\tau \to e$ channel, the $\tau \to \mu$ and $\tau \to \rho$ channel yield only small number of events after the selection. A big difference between the two modes is the number of expected background events, which is much higher in the $\tau \to \mu$ channel than in the hadronic τ channels. As will be seen later in the combination of the modes, the $\tau \to \mu$ channel does not increase the expected significance and will not be included in the final measurement.

(a) $\tau \to e$

(b) $\tau \to \mu$

(c) $\tau \to \pi$

(d) $\tau \to \rho$
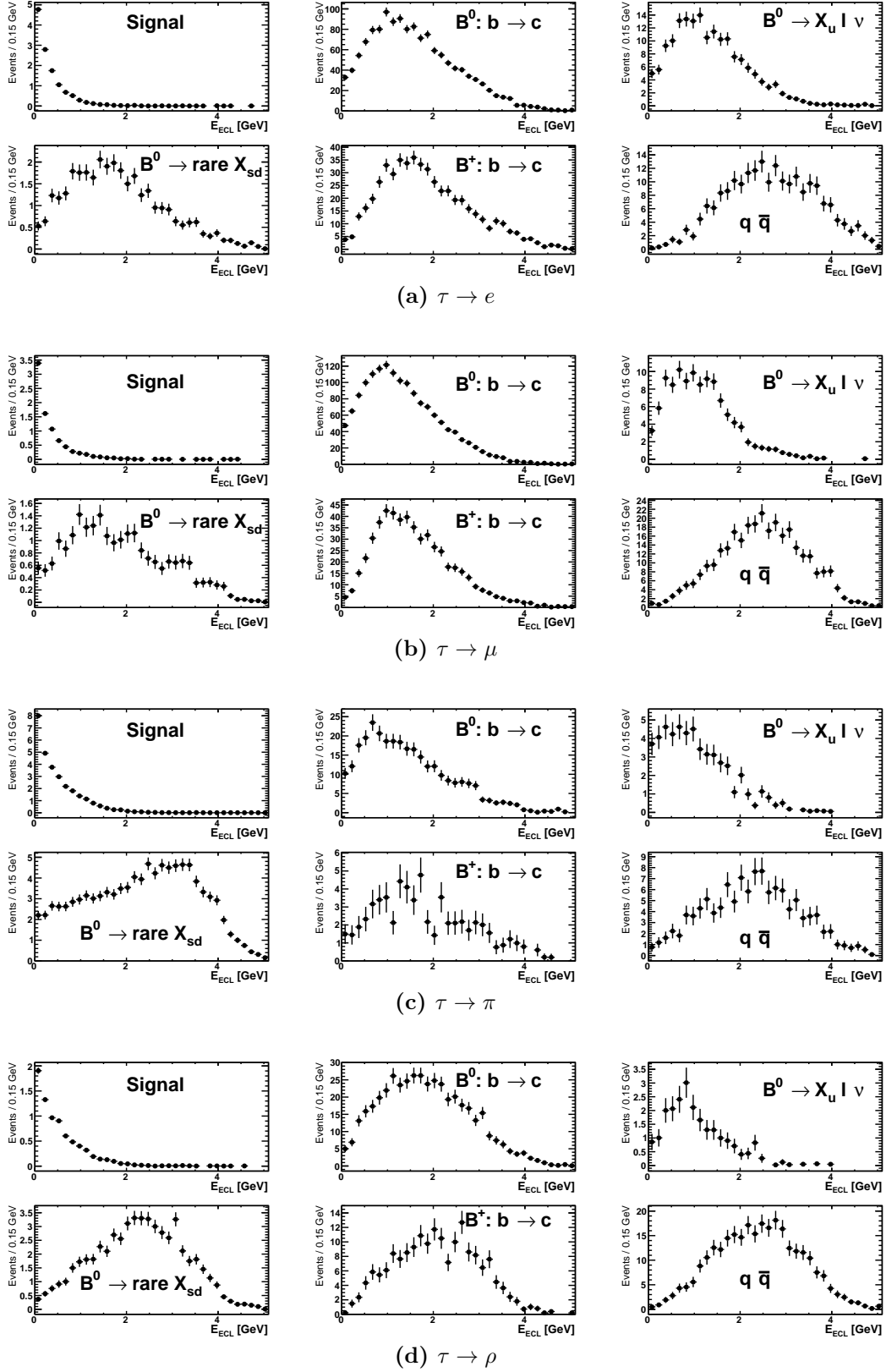
Figure 5.22.: E_{ECL} distributions scaled to the recorded luminosity. (a)-(d) show the distribution for signal and groups of background types for the four reconstruction modes.

5.7. Signal Efficiency and Crossfeed

The effect of each cut in all four reconstruction modes is shown in Table 5.10. For each reconstruction channel, the number of events in the channel and the percentage is shown. The number of events is based on a Monte Carlo sample with 24×10^6 generated $B^0 \to \pi\tau\nu$ events with generic τ decays. The first line shows the number of events after successful full reconstruction with $M_{\text{bc,tag}} \in (5.27; 5.29)\,\text{GeV}$ and all efficiency corrections applied. While the K_L veto is a binary cut, the actual cut values for the variables in the four lower rows differ between the reconstruction channels as detailed in Section 5.5. The requirements on the full reconstruction quality and the BDT output contribute most to the reduction of final signal events. Furthermore, the $\tau \to \pi$ reconstruction channel contains the most events at all selection requirements although the $\tau \to \pi\nu_\tau$ branching fraction is the lowest compared to the other three channels. The reason is that the purity of the $\tau \to \pi$ channel is quite low as can be seen in Table 5.11.

It is important to note that the table shows numbers grouped by τ reconstruction channel, not by actual τ decay. The actual τ decay composition of each reconstruction channel is shown in Table 5.11.

| | $\tau \to e$ | | $\tau \to \mu$ | | $\tau \to \pi$ | | $\tau \to \rho$ | |
Selection	Events	[%]	Events	[%]	Events	[%]	Events	[%]
no cut	7,347.3	100.0	5,270.5	100.0	33,390.0	100.0	6,150.1	100.0
K_L veto	6,403.6	87.2	4,496.0	85.3	28,515.3	85.4	5,280.0	85.9
$\log(cont\_nbout)$	4,871.2	66.2	3,395.2	64.4	14,932.4	44.6	2,999.2	48.8
$N_{\text{extra}} < 2$	4,626.6	62.9	3,221.8	61.1	13,950.5	41.7	2,830.9	46.0
M_{miss}^2	4,297.4	58.4	3,175.2	60.2	13,929.5	41.6	2,704.6	44.0
BDT	2,789.5	37.9	1,947.3	37.0	6,458.7	19.3	1,711.1	27.8

Table 5.10.: Signal selection cutflow for all τ reconstruction modes. The cuts are additive, meaning all cuts mentioned in upper rows are included. The second column of each τ decay mode shows the percentage relative to *no cut*. The number are based on the second signal MC sample, containing $24M$ generated signal events. The numbers in the first row are the remaining numbers after successful hadronic tag reconstruction and basic event selection.

(a) $\tau \to e$		(b) $\tau \to \mu$		(c) $\tau \to \pi$		(d) $\tau \to \rho$	
Decay	[%]	Decay	[%]	Decay	[%]	Decay	[%]
$e^-\bar{\nu}_e\nu_\tau$	98.43	$\mu^-\bar{\nu}_\mu\nu_\tau$	88.04	$\rho^-\nu_\tau$	33.35	$\rho^-\nu_\tau$	62.37
$\rho^-\nu_\tau$	0.55	$\rho^-\nu_\tau$	5.32	$\pi^-\nu_\tau$	27.02	$a_1^-\nu_\tau$	25.29
$a_1^-\nu_\tau$	0.40	$\pi^-\nu_\tau$	3.52	$\mu^-\bar{\nu}_\mu\nu_\tau$	21.58	$\pi^-\nu_\tau$	4.73
$\mu^-\bar{\nu}_\mu\nu_\tau$	0.27	$a_1^-\nu_\tau$	2.29	$a_1^-\nu_\tau$	7.90	$\mu^-\nu_\tau\nu_\mu$	2.76
$\pi^-\nu_\tau$	0.18	$e^-\bar{\nu}_e\nu_\tau$	0.21	$e^-\bar{\nu}_e\nu_\tau$	7.27	$\pi^-\pi^0\pi^0\pi^0\nu_\tau$	2.25
$K^{*-}\nu_\tau$	0.04	$K^{*-}\nu_\tau$	0.13	$K^{*-}\nu_\tau$	0.77	$e^-\bar{\nu}_e\nu_\tau$	0.66
$K^-\pi^0\nu_\tau$	0.04	$K^-\nu_\tau$	0.11	$\pi^-\pi^0\pi^0\pi^0\nu_\tau$	0.57	$\omega\pi^-\nu_\tau$	0.40
$\pi^-\pi^0\pi^0\pi^0\nu_\tau$	0.03	$\pi^-\pi^0\pi^0\pi^0\nu_\tau$	0.10	$\pi^-\pi^0\nu_\tau$	0.48	$\pi^-\pi^0\nu_\tau$	0.37
$\pi^-\pi^-\pi^+\pi^0\nu_\tau$	0.02	$\pi^-\pi^0\nu_\tau$	0.08	$\omega\pi^-\nu_\tau$	0.23	$K^0\pi^-\pi^0\nu_\tau$	0.28
$\omega\pi^-\nu_\tau$	0.02	$K^-\pi^0\nu_\tau$	0.06	$K^0\pi^-\pi^0\nu_\tau$	0.22	$K^{*-}\nu_\tau$	0.26
$K^-K^0\pi^0\nu_\tau$	0.02	$K^-K^+\pi^-\nu_\tau$	0.05	$\pi^-\pi^+\pi^-\pi^0\nu_\tau$	0.18	$\eta\pi^-\pi^0\nu_\tau$	0.19
$K^-K^0\nu_\tau$	0.00	$K^-\pi^0\pi^0\nu_\tau$	0.04	$K^-\nu_\tau$	0.13	$K^-\pi^0\nu_\tau$	0.12
		$\pi^-K^0\pi^0\nu_\tau$	0.03	$\eta\pi^-\pi^0\nu_\tau$	0.11	$\omega\pi^-\pi^0\nu_\tau$	0.10
		$\pi^+\pi^-\pi^+\pi^0\nu_\tau$	0.03	$\omega\pi^-\pi^0\nu_\tau$	0.09	$\pi^-\pi^+\pi^-\pi^0\nu_\tau$	0.09
				$K^-\pi^0\nu_\tau$	0.03	$K^-\pi^0\pi^0\pi^0\nu_\tau$	0.06
				$K_S^0K_L^0\pi^-\nu_\tau$	0.03	$\pi^-\pi^0\pi^0\eta\nu_\tau$	0.05
				$K^0\pi^-\pi^0\pi^0\nu_\tau$	0.02	$K_S^0K_L^0\pi^-\nu_\tau$	0.03
				$K^-K^0\pi^0\nu_\tau$	0.01	$K^0\pi^-\pi^0\pi^0\nu_\tau$	0.00
				$K^-\pi^-\pi^-\pi^0\nu_\tau$	0.00	$K^-\nu_\tau$	0.00
				$K^-\pi^0\pi^0\nu_\tau$	0.00		
				$K_L^0K_L^0\pi^-\nu_\tau$	0.00		

Table 5.11.: Composition of τ decays.

5.8. Background Composition

The background in the signal region $E_{ECL} < 1.5\,\mathrm{GeV}$ and the complete E_{ECL} region is listed in Table 5.12 for all four τ reconstruction modes. The dominant background clearly is $D^{(*)}\ell\nu$ and $D^{(*)}\rho$. This is easily explained when looking into the decay of the $D^{(*)}$ meson. In many cases, the D meson decays into $K_L\pi$. While a K_L veto is applied, it is only applied to K_L without contribution to E_{ECL} because the E_{ECL} distribution is used to fit the signal yield. Due to inefficiencies in the K_L detection of the Belle detector, some K_L escape the detector undetected. $B \to D(K_L\pi)\ell\nu$ events then have exactly the same visible final state as the signal decay.

Furthermore, wrongly reconstructed B_{tag} mesons and inefficiencies in the particle detection due to inactive detector regions add to the background.

$\tau \to e\nu\nu$		$\tau \to \mu\nu\nu$		$\tau \to \pi\nu\nu$		$\tau \to \rho\nu\nu$	
decay	Occ [%]	decay	Occ [%]	decay	Occ [%]	decay	Occ [%]
$D^+ e^- \nu_e$	12.86	$D^+ \mu^- \nu_\mu$	17.33	$D^+ \mu^- \nu_\mu$	8.45	$D^+ \rho^+$	6.15
$D^{*+} e^- \nu_e$	10.93	$D^{*+} \mu^- \nu_\mu$	12.73	$D^{*+} \mu^- \nu_\mu$	6.36	$D^{*+} \mu^- \nu_\mu$	4.63
$D^+ \rho^+$	5.41	$D^{*+} \rho^+$	5.84	$D^+ \rho^+$	5.75	$D^{*+} \rho^+$	4.60
$D^{*+} \rho^+$	3.68	$D^+ \rho^+$	5.69	$D^{*+} \rho^+$	4.64	$D^+ \mu^- \nu_\mu$	3.33
$D^{*0} e^- \nu_e$	3.56	$\rho^+ \mu^- \nu_\mu$	3.07	$D^{*+} e^- \nu_e$	2.92	$a_1^+ D^+$	2.66
$\rho^+ e^- \nu_e$	2.93	$D^{*0} \mu^- \nu_\mu$	2.97	$D^+ e^- \nu_e$	2.62	$D^{*+} a_1^+$	2.41
$D^{*+} \mu^- \nu_\mu$	2.39	$D^{*+} a_1^+$	2.62	$\rho^+ \mu^- \nu_\mu$	2.51	$D^{*+} e^- \nu_e$	2.26
$D^+ \tau^- \nu_\tau$	2.05	$D^{*+} \rho^0 \pi^+$	2.07	$D^+ \tau^- \nu_\tau$	2.26	$D^+ \tau^- \nu_\tau$	1.85
$D^+ \mu^- \nu_\mu$	2.02	$D^{*+} \mu^- \nu_\mu$	2.03	$D^{*+} a_1^+$	2.17	$D^{*+} \rho^0 \pi^+$	1.73

(a) $E_{ECL} < 1.5\,\mathrm{GeV}$

$\tau \to e\nu\nu$		$\tau \to \mu\nu\nu$		$\tau \to \pi\nu\nu$		$\tau \to \rho\nu\nu$	
decay	Occ [%]	decay	Occ [%]	decay	Occ [%]	decay	Occ [%]
$D^+ e^- \nu_e$	8.54	$D^+ \mu^- \nu_\mu$	12.94	$D^+ \mu^- \nu_\mu$	4.58	$D^+ \rho^+$	4.73
$D^{*+} e^- \nu_e$	7.31	$D^{*+} \mu^- \nu_\mu$	9.06	$D^+ \rho^+$	4.32	$D^{*+} \rho^+$	3.60
$D^+ \rho^+$	4.63	$D^+ \rho^+$	5.20	$D^{*+} \mu^- \nu_\mu$	4.00	$D^{*+} a_1^+$	2.24
$D^{*+} \rho^+$	3.64	$D^{*+} \rho^+$	4.94	$D^{*+} \rho^+$	3.65	$D^{*+} \mu^- \nu_\mu$	2.03
$D^{*0} e^- \nu_e$	2.95	$D^{*0} \mu^- \nu_\mu$	2.56	$D^{*+} e^- \nu_e$	2.16	$a_1^+ D^+$	1.89
$D^{*+} a_1^+$	1.98	$D^{*+} a_1^+$	2.45	Xsd	2.04	$D^{*0} \rho^+$	1.63
$\rho^+ e^- \nu_e$	1.78	$D^{*+} \mu^- \nu_\mu$	2.00	$D^{*+} a_1^+$	1.77	$D^{*+} \rho^0 \pi^+$	1.40
$a_1^+ D^+$	1.74	$\rho^+ \mu^- \nu_\mu$	1.97	$D^+ e^- \nu_e$	1.67	$D^{*+} \pi^0 \pi^+$	1.32
$D^{*+} e^- \nu_e$	1.70	$a_1^+ D^+$	1.96	$D^{*+} \rho^0 \pi^+$	1.43	$D^+ \mu^- \nu_\mu$	1.24

(b) Complete E_{ECL} region

Table 5.12.: Background composition in (a) the region $E_{ECL} < 1.5\,\mathrm{GeV}$ and (b) the complete final selection. The contribution relative to all background in the respective region of the dominant decays are listed for all four τ reconstruction modes.

5.9. Sideband Verification

As can be seen in Figure 5.22, the E_{ECL} distribution in the signal decay vanishes for higher values. We define the sideband by $E_{ECL} > 1.35\,\mathrm{GeV}$, to make sure basically no signal contribution is left here.

To show that the simulation describes the data well, both distributions are compared in the sideband region, shown in Figure 5.23. For completeness, the MC distributions are not blinded.

No second sideband region is available for this selection as the signal does not show a peaking structure in any other kinematic variable or other distribution. As a second check, MC vs data comparison plots are produced for a lot of other variables applying the E_{ECL} sideband cut. After measuring, MC vs data plots without the sideband cut are produced, too. Disagreements between data and simulation in the lower E_{ECL} region should translate differently into different variables and should be visible in these plots, which is not the case. The additional

sideband comparison plots are shown in Appendix A.2.1 and the unblinded comparison plots in Appendix A.2.2.

(a) $\tau \to e$

(b) $\tau \to \mu$

(c) $\tau \to \pi$

(d) $\tau \to \rho$
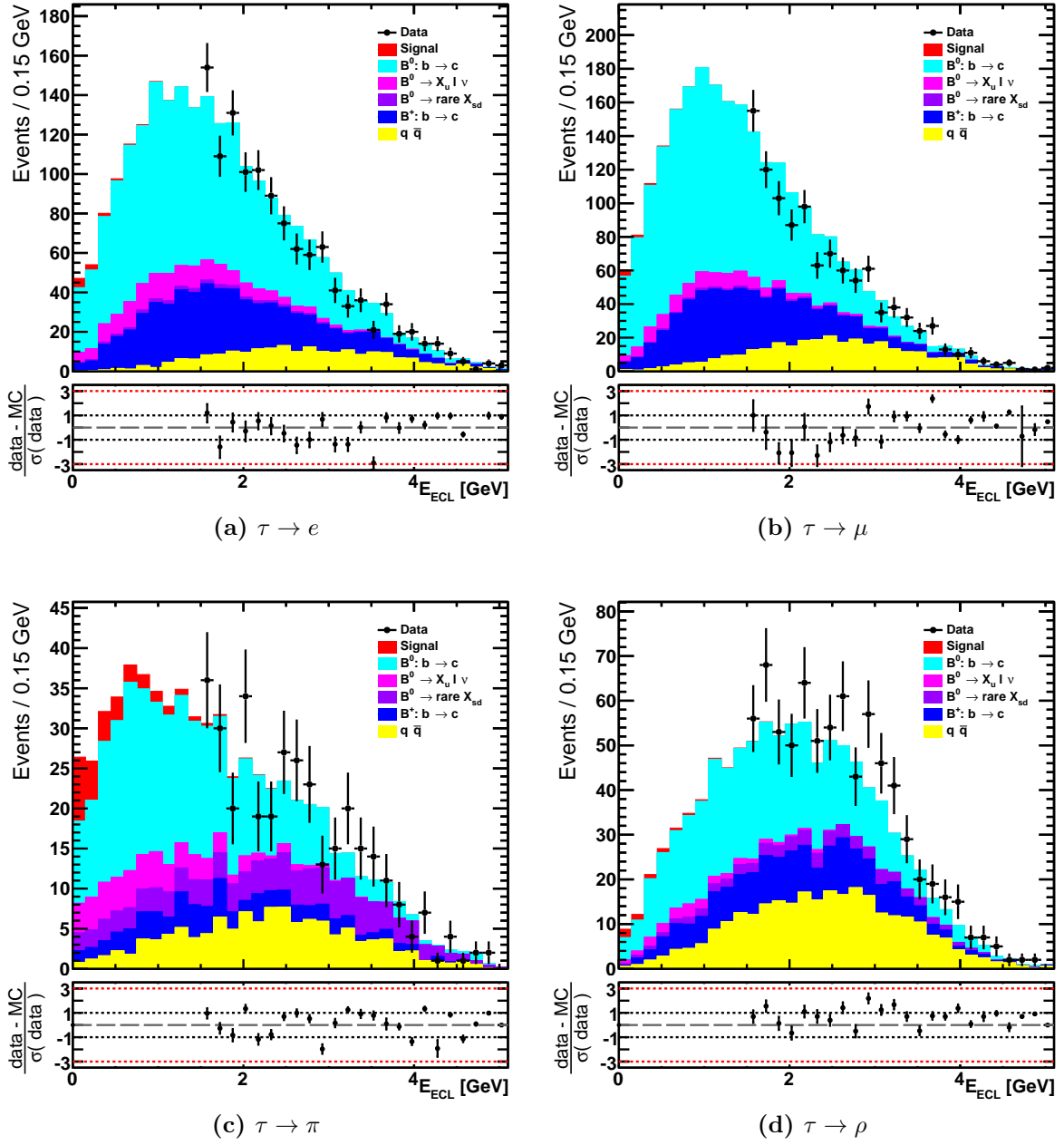
Figure 5.23.: Comparison of data and Monte Carlo simulation in the sideband $E_{ECL} > 1.35\,\mathrm{GeV}$. The Monte Carlo expectations are drawn in the complete region assuming a $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.0 \times 10^{-4}$.

6. Measurement

The measurement and the final result of the analysis are presented in this chapter. In order to reduce the systematic uncertainty on the $B^0 \to X_c$ MC prediction, a fit is first performed. The fit also produces the number of signal events in the Belle data sample, as well as the branching fraction, presented in Section 6.1. As no evidence is obtained, the significance level and upper limit are computed in Sections 6.4 and 6.5, respectively, using Monte Carlo techniques explained in Section 4.4. The construction of the likelihood is described in Section 6.2, followed by the description of the systematic uncertainties included in the final result, in Section 6.3.

6.1. Fit

A binned maximum likelihood fit is performed in E_{ECL}. The number of signal events is extracted from the fit. Additionally, the fit is used to determine a scaling factor for the $b \to c$ contribution in order to reduce the systematic uncertainty of this contribution due to finite MC sample size. Please note, that in this chapter, $b \to c$ labels only the dominant $B^0 \to X_c$ decays, not the less relevant $B^+ \to X_c$ decays. In order to improve readability, the $b \to c$ label has been chosen. Due to low statistics and similar shapes, all background contributions are fixed except for the dominant $b \to c$ transitions. The pdfs are built from the Monte Carlo predictions in E_{ECL} that are shown above in Figure 5.22. The fit is performed using the RooFit framework.

The likelihood function of a single reconstruction channel c used for the fit can then be described as

$$
\begin{aligned}
\mathcal{L} = N_{tot} \cdot \mathrm{PDF}_c = {} & (\mu \cdot N_{c,\mathrm{sig}}^{\mathrm{MC}}) \cdot \mathrm{PDF}_{c,sig} \\
& + (f_{c,b\to c} \cdot N_{c,b\to c}^{\mathrm{MC}}) \cdot \mathrm{PDF}_{c,b\to c} \\
& + (f_{c,const} \cdot N_{c,const}^{\mathrm{MC}}) \cdot \mathrm{PDF}_{c,const}
\end{aligned}
\tag{6.1}
$$

with the coefficients μ, $f_{b\to c}$ and f_{const} and the number of events per contribution determined from MC simulation N^{MC}. As there are only rough calculation available for the expected branching fraction $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$ in the SM, we define the $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)^{\mathrm{MC}} = 1.0 \times 10^{-4}$ in this analysis. The coefficients μ, $f_{b\to c}$ and f_{const} are the fitting parameters with $f_{const} = 1.0$ fixed. The single PDFs are normalized to 1.

The coefficient μ will be called the signal strength and is defined as

$$
\mu = \frac{\mathcal{B}\left(B^0 \to \pi\tau\nu\right)}{\mathcal{B}\left(B^0 \to \pi\tau\nu\right)^{\mathrm{SM}}}.
\tag{6.2}
$$

Same as above, we define

$$
\mathcal{B}\left(B^0 \to \pi\tau\nu\right)^{\mathrm{SM}} = 1.0 \times 10^{-4},
\tag{6.3}
$$

which also allows easy conversion the measured signal strength to the measured branching fraction.

Before the fit is performed on data, the stability of the fit is tested using pull distributions and a linearity test, which are described in Sections 6.1.1 and 6.1.2, respectively. The fit result on data is then shown in Section 6.1.3.

The final measurement is performed as a simultaneous fit in the three τ reconstruction channels $\tau \to e, \tau \to \pi$ and $\tau \to \rho$, as the $\tau \to \mu$ reconstruction does not improve the expected significance. This will be shown below in Section 6.4. It is already noted here because only stability tests and fit results for the combination of these three modes are shown in this section for better readability. The results of the test as well as the results of the fit are shown in Appendix A.1.

6.1.1. Pull Distribution

Pull distributions allow to investigate the fit results on whether the fit produces a bias and if the errors are estimated correctly. The pull p of a fit is defined as the difference between input value f_{in} and fit value f_{fit}, normalized to the fit error σ_{fit},

$$p = \frac{f_{\mathrm{in}} - f_{\mathrm{fit}}}{\sigma_{\mathrm{fit}}}. \tag{6.4}$$

The pull distribution is obtained by repeating the fit multiple times on pseudo data which is generated based on the fit pdf with the parameter values set to f_{in}. A robust fit usually results in a pull distribution in form of a normal Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. A different mean value indicates a bias in the fit results while smaller or higher standard deviations indicates that the errors are over- or underestimated by the fit, respectively.

To avoid confusion due to the multiple meanings of the letter μ, the signal strength parameter will be labeled with an additional superscript μ^{sig} for this section only. The difference between the muon and the mean of the Gaussian distribution should be clear from context.

The pull distributions for μ^{sig} and $f_{b \to c}$ for all reconstruction modes and the combination is shown in Figure 6.1. The pull distribution for this fit is obtained for $\mu = f_{b \to c} = 1$ for all background parameters. Pull distribution parameters have been determined for different input values of the signal strength parameter and background $b \to c$ contributions. The results are summarized in Table 6.1. The table contains the parameters without a quoted error to improve readability. The size of the error is the same for all values of input parameters because the same number of pulls has been computed. As quoted in Figure 6.1, we obtain $\sigma_\mu = 0.01$ and $\sigma_\sigma = 0.007$. Pull distributions fits in the single modes and in a simultaneous fit in all four reconstruction modes have been computed, too. The results are summarized in Appendix A.1.1.

As can be seen, no substantial bias is observed and all pull distributions fit well to a unit Gaussian.

6.1.2. Linearity

In order to investigate the stability of the fit over a broad range of possible parameter values, the linearity of the fit is tested. For a range of input parameter f_{input}, the fitted value f_{fit} is plotted against the input value. For a stable fit without bias, this should produce a straight line through the origin with a slope equal to one. The distribution of f_{fit} versus f_{input} is shown in Figure 6.2 for all four parameters in the combined fit of $\tau \to e, \pi, \rho$. The parameters of a straight line fit
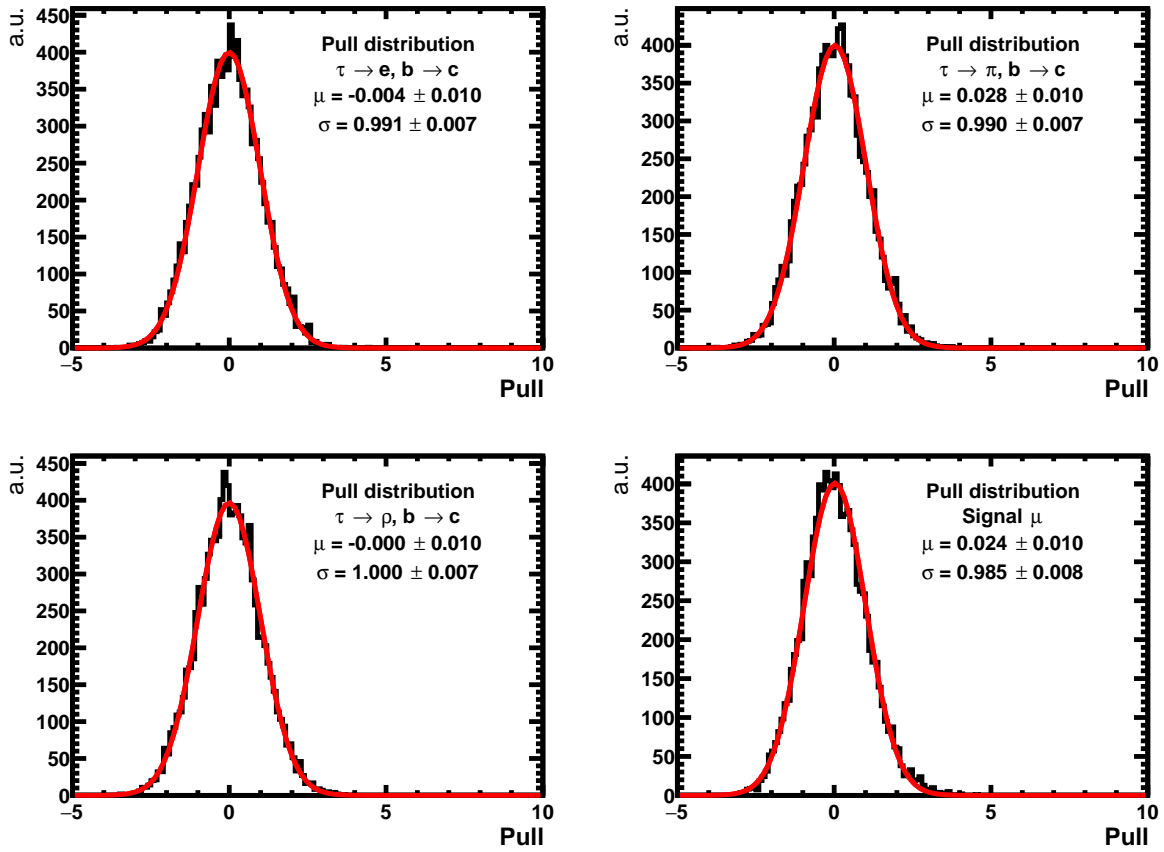
Figure 6.1.: Pull distributions in the combined fit of $\tau \to e, \pi, \rho$ for the signal (lower right) and the three $b \to c$ contributions.

are shown in the plots and all agree well with $f_{\text{fit}}(f_{\text{input}}) = 0 + 1 \cdot f_{\text{input}}$. The linearity tests for the fits in the single reconstruction modes and in all four reconstruction channels combined are shown in Appendices A.1.2 and A.1.3.

6.1.3. Fit on Data

The fit on data in the combined modes $\tau \to e, \pi, \rho$ results in a signal strength parameter of $\mu = 1.52 \pm 0.72$, which is equal to 51.9 ± 24.3 signal events. The results are shown graphically in Figure 6.3 and listed in Table 6.2. From the fit, the significance σ of the signal in terms of standard deviations can be computed ignoring systematic uncertainties by

$$\sigma = \sqrt{-2 \ln \frac{\mathcal{L}_0}{\mathcal{L}_{\text{fit}}}} = 2.70, \tag{6.5}$$

where $\mathcal{L}_{\text{fit}}$ and $\mathcal{L}_0$ are the likelihood values evaluated after the fit and at $\mu = 0$, respectively.

The fit has been performed in the single decay channels, too, the results are listed in Appendix A.1.4.

$\mu_{\text{in}}^{\text{sig}}$	μ^{sig}		$f_{b \to c}(\tau \to e)$		$f_{b \to c}(\tau \to \pi)$		$f_{b \to c}(\tau \to \rho)$	
$[\times 10^{-4}]$	μ	σ	μ	σ	μ	σ	μ	σ
0.5	0.05	0.983	-0.01	0.991	0.01	0.985	0.02	0.981
1.0	0.02	0.985	0.00	0.991	0.03	0.990	0.00	1.000
1.5	0.04	0.985	0.00	0.997	0.02	1.000	0.02	0.989
2.0	0.05	0.994	0.02	1.010	-0.01	0.988	0.01	0.991
$f_{b \to c, \text{in}}$								
0.8	0.04	0.969	0.01	0.996	0.01	0.995	0.03	0.974
0.9	0.02	0.996	0.02	0.995	0.03	0.994	0.00	0.994
1.0	0.02	0.994	0.00	1.010	0.02	0.987	0.03	0.990
1.1	0.04	0.993	0.01	0.989	0.01	0.996	0.01	0.989
1.2	0.04	1.010	0.01	0.990	0.02	0.983	0.01	0.987

Table 6.1.: Pull distributions for all fitted parameters in the combined fit of $\tau \to e, \pi, \rho$ for different input values of signal strength paramer $\mu_{\text{in}}^{\text{sig}}$ in the upper half and the $b \to c$ contribution $f_{b \to c, \text{in}}$ in the second half. No substantial bias is observed and the errors are estimated correctly in the fit.
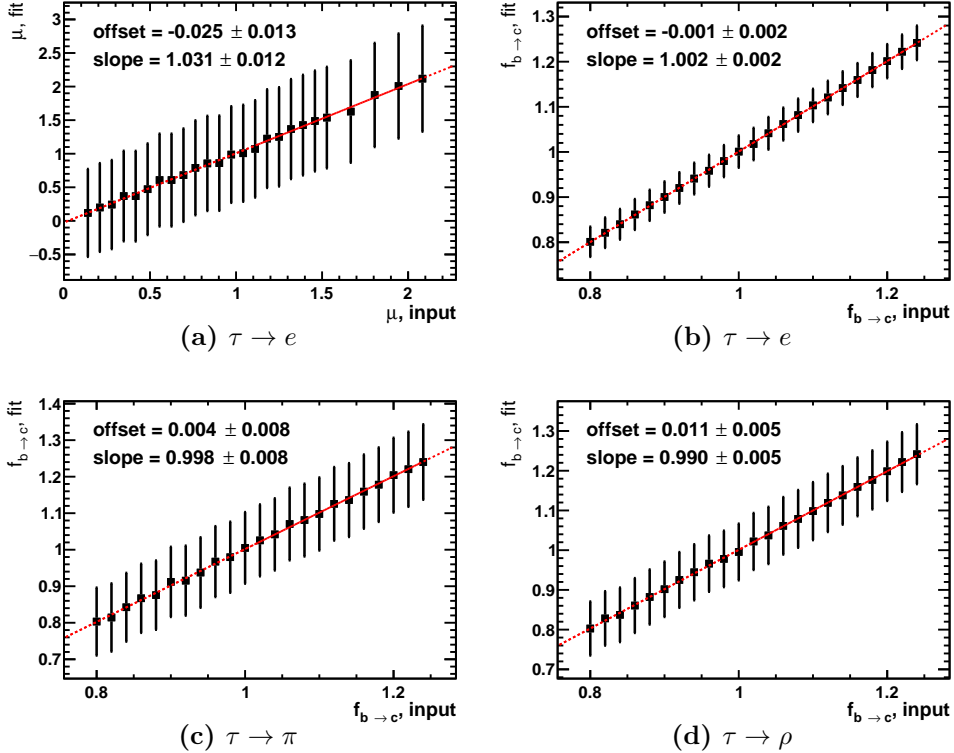
Figure 6.2.: Linearity test for signal strength μ and $f_{b \to c}$ in the combined fit $\tau \to e, \pi, \rho$.
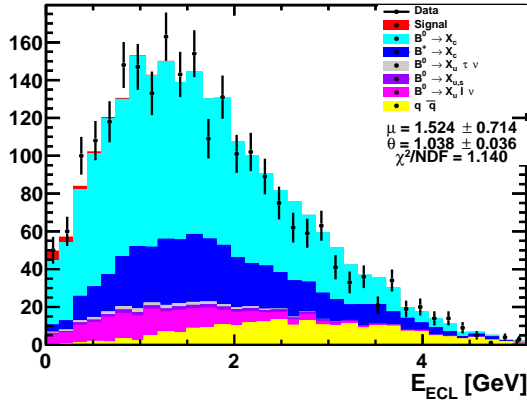
(a) Simultaneous fit to all three modes, results for $\tau \to e$.
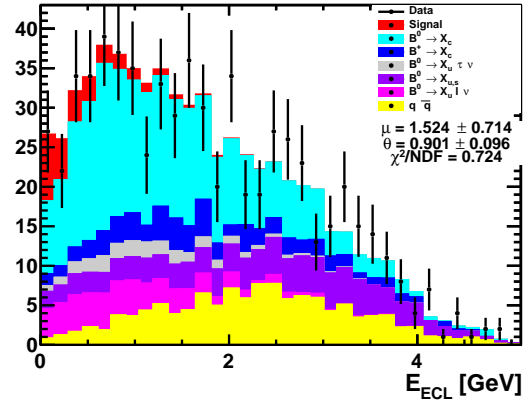
(b) Simultaneous fit to all three modes, results for $\tau \to \pi$.
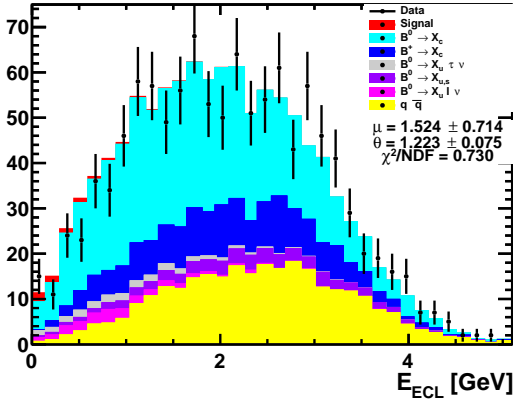
(c) Simultaneous fit to all three modes, results for $\tau \to \rho$.

Figure 6.3.: E_{ECL} distribution in the combined fits $\tau \to e, \pi, \rho$.

| mode | Combined Fit | | |
	θ	μ	N_{sig}
e	1.038 ± 0.036		13.2 ± 6.2
π	0.901 ± 0.096	1.524 ± 0.714	30.6 ± 14.3
ρ	1.223 ± 0.075		8.1 ± 3.8

Table 6.2.: Fit results for single mode and combined fit. $\mu = 1$ denotes $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.00 \times 10^{-4}$.

.

6.2. Likelihood Construction

As in the fit, a binned likelihood approach is used in this analysis. The likelihood is then used to compute the signal significance and the upper limit of the signal branching fraction, by methods explained in detail in Section 4.4. Systematic uncertainties p are included in the likelihood in the

form of nuisance parameters θ_p and their respective nominal values $\hat{\theta}_p$. With the subscripts c, b, and p to label the τ reconstruction channel, the bin in E_{ECL} and the systematic uncertainty, respectively, the likelihood $\mathcal{L}$ is given as

$$\mathcal{L}\left(\mu, \vec{\theta}\right) = \prod_c \prod_b \text{Pois}\left(n_{cb}|\nu_{cb}(\mu, \vec{\theta})\right) \cdot \prod_p f_p\left(\hat{\theta}_p|\theta_p\right),\qquad(6.6)$$

where n is the number of observed events in data and $\nu(\mu, \vec{\theta})$ is the number of expected events. For every nuisance parameter θ_p, a constraint term f_p is included in the likelihood. Most constraint terms are included in the form of a Gaussian constraint $G(\hat{\theta}|\theta, 1)$ where $\hat{\theta}$ is the nominal value for the corresponding nuisance parameter. The parameters $\hat{\theta}_p$ are called global observables from here on. A special case is the uncertainty of the luminosity, which is included as a term $G(L_0|\lambda, \Delta L)$ where $L_0 = 1$ and ΔL the relative uncertainty.

With these parameters, the expected number of events in each channel in each bin is the sum of the number of expected events in each sample. In this nomenclature, each background contribution that is treated as a single sample s. The number of expected events in each bin is then given by

$$\nu_{cb} = \sum_s \lambda \cdot \gamma_{cb}\left(\vec{\theta}\right) \cdot \phi_{cs} \cdot \kappa_{csb}\left(\vec{\theta}\right).\qquad(6.7)$$

The factor $\gamma_{cb}\left(\vec{\theta}\right)$ is a relative factor which includes the statistical error on the Monte Carlo prediction due to finite Monte Carlo statistics. The next factor, ϕ_{cs}, is a scaling factor for each sample in each mode. For the signal sample, this is the already introduced signal strength parameter μ, for the $b \to c$ sample it is $f_{b \to c}$ which has been determined in Section 6.1 and for all other samples it is fixed to 1.0. The number of events predicted from Monte Carlo simulation, combined with the effect of the systematic uncertainties, is included in the factor $\kappa_{csb}(\vec{\theta})$.

6.3. Systematic Uncertainties

This section lists the systematic uncertainties on the signal and background expectation values.

The effect of each systematic on the significance level on data is listed Section 6.3.1 further below.

- Particle ID. The uncertainty of the correction factor introduced due to the difference of particle identification on data and MC. For the lepton ID, a 3% error is included, while for the π, K discrimination, an error of 1% is included.

- Tracking efficiency. This has also been studied by the Belle collaboration and a constant uncertainty of 0.35% per track on signal side will be applied. This correction is valid for tracks with $p > 0.2\,\text{GeV}$. For slow tracks, as well as slow π^0, below this threshold, a flat uncertainty is not valid. These tracks and π^0 are evaluated following the results of [117]. This systematic uncertainty results in a different shape of the histogram and is implemented in a bin-by-bin deviation.

- Luminosity. The uncertainty on the online luminosity is also published on the Belle homepage with an uncertainty of 1.4%.

- K_L veto weighting. The K_L veto is described in Section 5.1.4. The uncertainty introduced by this method is due to the difference in K_L reconstruction efficiency on data and Monte

Carlo. The reconstruction efficiency is varied at the 1σ level and the procedure to obtain the K_L veto weight is repeated for both the upper and lower variation.

- Tag calibration. The handling of this error will follow the procedure in [118] and a flat uncertainty of 4.5% is included.

- Background branching fraction uncertainty. The branching fraction of the most prominent background decay modes will be varied by 1σ. This will be implemented in a bin-by-bin fashion, as the most important decays are $B \to D^{(*)}\ell\nu$ and $B \to D^{(*)}\rho$ in the signal region.

- Finite MC sample size. The pdf shape is known up to a certain degree due to the fact that the MC sample is limited to five times the data luminosity for the $B^0 \to X_c$, $B^\pm \to X_c$ and $e^+e^- \to q\bar{q}$ samples. The uncertainty is taken into account by introducing a Poisson constraint on the number of predicted events in each bin. The exact procedure is similar to the publication of Barlow and Beeston [120] and is described below.

- Fit result for the $B^0 \to X_c$ sample. The sample is scaled by the fit result in the likelihood calculation, the error of the fit is included as the 1σ deviation with Gaussian constraint

- Signal model. The signal model is determined by both form factors $f^+(q^2)$ and $f^0(q^2)$, both of which have theoretical errors as shown in Figure 3.16. New signal MC has been generated with new pole parameters fitted to the upper and lower error band of both form factors. The deviation in the signal expectation due to the new pole parameters is used to compute the uncertainty.

- $|V_{ub}|$. The $B^0 \to X_u^- \ell^+ \nu$ MC is scaled to match the latest $\mathcal{B}\left(B^0 \to X_u^+ \ell\nu\right) = (2.16 \pm 0.31) \times 10^{-3}$, published by the Heavy Flavor Averaging Group (HFAG) [121]. Due to the gap between inclusive and exclusive $|V_{ub}|$ measurements, an asymmetric uncertainty is introduced. A small positive uncertainty and a larger negative uncertainty is used which at the 1σ level is a relative uncertainty of $+5\%$, -15%.

- The $B^0 \to X_u^- \tau^+ \nu$ branching fractions have not been measured yet. To account for this, a reasonable branching fraction is assumed and a large systematic uncertainty is applied. Calculations in the quark model [122] are used in the event generator for these events. While this decay type contributes mostly in the higher E_{ECL} region, the $B^0 \to \rho^- \tau^+ \nu$ decays with $\tau \to \pi\nu$ will show up in the low E_{ECL} region. Not many events are expected to pollute the $\tau \to \pi$ or $\tau \to \rho$ reconstruction because of different kinematics in the $\rho^- \to \pi^-\pi^0$ decay and the fact that the $\pi^\pm$ with the lowest momentum is used in this analysis, cf. Section 5.1.3. The scaling is done such that $B \to \rho\tau\nu \approx 1.5 \times 10^{-4}$, and a flat systematic uncertainty for the whole sample is applied where 1σ equals $\pm 50\%$.

- The $B^0 \to X_{us}$ sample of rare B decays contains too many decays to be updated manually. The contribution of this sample is not negligible, but neither dramatically important. We assume 10% uncertainty on this sample.

All systematic effects are included in the computation of the upper limit, while particle ID and tracking efficiencies are not included in the computation of the significance level.

Beeston-Barlow lite

Barlow and Beeston[120] described a method to include the limitations of MC samples in the fit, which can be written in a binned example likelihood as

$$\mathcal{L}(\vec{N}|\vec{s},\vec{b}) = \prod_{\text{bins}} P(N_i|s_i + b_i) \prod_{\text{bins}} P(\widetilde{s}_i|s_i) \prod_{\text{bins}} P(\widetilde{b}_i|b_i). \tag{6.8}$$

This approach leads to many nuisance parameters in case of many background samples, as in our case, namely one constrained term per sample per bin, which is computationally time consuming when generating toy MC and fitting with the systematic uncertainties included in the likelihood. Instead, only a single constrained term for the total number of events in each bin is used. This transforms the upper equation to

$$\mathcal{L}(\vec{N}|\vec{\gamma}) = \prod_{\text{bins}} P(N_i|n_i) \prod_{\text{bins}} P(\widetilde{s}_i + \widetilde{b}_i|n_i) \tag{6.9}$$

where $\widetilde{s}_i$ and $\widetilde{b}_i$ are implemented as global observables as described in Section 6.2.

6.3.1. Effect of Systematic Uncertainties

The relative change in the measured branching fraction due to the different systematic uncertainties is listed in Table 6.3. The numbers are obtained by fitting the data with the pdfs modified by the systematic effect at the 1σ level. For each systematic effect, two fits are performed, one for up and down variation, respectively. The maximum deviation is quoted in the table in order to provide a conservative estimate.

systematic	relative error [%]
e ID	1.4
π ID	1.6
π^0 ID	1.0
Track efficiency	0.7
N($B\overline{B}$)	1.4
K_L veto	3.2
BG $\mathcal{B}$	2.8
$D^{(*)}\ell\nu$ model	0.5
Tag calibration	4.6
$|V_{ub}|$	2.8
Rare MC	2.0
$B \to X_u\tau\nu$	2.2
Background fit	0.2
Signal model	1.8
total	8.3

Table 6.3.: Effects of the single systematic effects on the branching fraction.

6.4. Significance Level

The significance level of the signal process is determined as described in Section 4.4.2 using the likelihood given in Equation (6.6). All systematic effects which do have an effect on the pdf shape are included while efficiency related systematic uncertainties are not included in the calculation. First, the expected significance is calculated for all single modes and their combination in order to determine the combination with the highest expected significance. This combination is then evaluated on the data sample.

The expected significance is calculated on the Asimov dataset assuming $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.0 \times 10^{-4}$. The Asimov dataset is a pseudo dataset built by setting the number of observed events to the expected number from Monte Carlo simulation assuming the data luminosity and a given signal branching fraction.

The results are summarized in Table 6.4. As can be seen, adding the explicit $\tau \to \mu$ reconstruction does not improve the expected significance. Therefore, it has been decided to discard this reconstruction channel in the analysis. It is important to note though, that this does not exclude all physical $\tau \to \mu\nu\nu$ decays as has been shown in Table 5.11.

mode / combination	Expected Significance
e	0.50 ± 0.02
μ	0.29 ± 0.02
π	1.34 ± 0.03
ρ	0.37 ± 0.02
$e + \pi$	1.38 ± 0.03
$e + \mu + \pi$	1.40 ± 0.03
$e + \pi + \rho$	1.45 ± 0.03
$e + \mu + \pi + \rho$	1.45 ± 0.03

Table 6.4.: Exected significance (Z-value) on the Asimov dataset, assuming $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.0 \times 10^{-4}$, for all modes and combinations. The values are obtained from $10\,000$ pseudo experiments for the background only hypothesis.

On data, we observe a significance level of $2.38 \pm 0.03\,\sigma$ which is not enough to claim evidence for the signal process. Instead we proceed to compute an upper limit of the signal branching fraction.

6.5. Upper Limit

The upper limit of the signal branching fraction is computed by Monte Carlo technique as described in Section 4.4.3. The scan has been performed in the signal strength parameter μ with a step width of 0.1. For both the null and alternative hypothesis, $10\,000$ pseudo experiments have been evaluated at each scan step. We obtain an upper limit of $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) < 2.5 \times 10^{-4}$ at the 90% confidence level and $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) < 2.8 \times 10^{-4}$ at the 95% confidence level.

7. Summary and Outlook

The search for the semileptonic decay $B^0 \to \pi\tau\nu$ at the Belle experiment has been presented in this thesis. The analysis and the results will be summarized shortly in the first section, followed by an outlook regarding further measurements of the decay.

7.1. Summary

This thesis presents the first published search for the decay $B^0 \to \pi\tau\nu$ [123]. It has been performed on the huge Belle data sample, containing 772×10^6 $B\overline{B}$ pairs. The lifetime of the τ lepton is short enough for it to decay inside of the detector, such that it has to be reconstructed from its decay products. In this analysis, the τ lepton is reconstructed in its decays with one charged particle in the final state, namely $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$, $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$, $\tau^+ \to \pi^+\bar{\nu}_\tau$, and $\tau^+ \to \rho^+\bar{\nu}_\tau$. The final state contains 2-3 neutrinos, which makes it impossible to fully reconstruct the τ lepton or the B_sig meson. This analysis exploited the fact that the initial state of a collision at Belle is exactly known and defined by the sum four-momenta of the colliding leptons, as is the four-momentum of the final state. A complex, multivariate full reconstruction algorithm, based on NeuroBayes, is used to reconstruct one of the two B mesons in an event, the B_tag. With the knowledge of all tracks that belong to the B_tag, signal candidates can be selected. A very important event variable is the deposited energy in the ECL, which is not assigned to neither the reconstructed B_tag, nor the two charged particles from the B_sig. For the separation between signal and background, boosted decision trees are used, where one BDT classifier is trained for each τ reconstruction channel. While the full hadronic reconstruction allows to perform this analysis, it is also the dominant, limiting factor on the reconstruction efficiency.

While the decay $B^0 \to \pi^-\tau^+\nu_\tau$ has been proposed by theorists as a test of the SM for a long time, as has been shown in Section 2.3, no analysis has been able to obtain a measurement, yet. This first search for the decay has reached a significance of $2.4\,\sigma$ and has finally shown, that the decay is experimentally accessible. The analysis requires advanced reconstruction techniques, at the price of a very small reconstruction efficiency. While the systematic uncertainty is 8.3%, the statistical uncertainty on the branching fraction is roughly 47%. The numerical results of the analysis are summarized in Table 7.1. More data or a combination of multiple analyses will be needed to obtain evidence for the decay, which will be the topic of Section 7.2.

Number of signal events:	51.9 ± 24.3 (stat. only)
Significance level	2.4σ
$\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$	$< 2.5 \times 10^{-4}$ @ 90% CL
	$< 2.8 \times 10^{-4}$ @ 95% CL

Table 7.1.: Summary of the results of this analysis.

The results are well compatible with the SM prediction of $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = (0.935 \pm 0.038) \times 10^{-4}$ [69]. First exclusions on $\tan\beta$ and m_H^+ in the MSSM, based on this result, have been published in Ref. [124] and are shown in Figure 7.1. The exclusion limits are to be taken with a word of warning, though. They are obtained using the branching fraction obtained in this analysis, $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = (1.52 \pm 0.75) \times 10^{-4}$, to compute $R(\pi)$ defined in Equation (2.31). As no evidence is obtained, the branching fraction result of this analysis is stated with the statistical uncertainty, only. Furthermore, the exclusion limits are obtained under the assumption, that the reconstruction efficiency and overall acceptance are the same for MSSM kinematics.

The obtained limit is lower than some of the values presented in Section 2.3.2, which might result in the exclusion of certain parameter space regions of certain NP scenarios. Several input factors have been used in those calculations, though, and no explicit dependence of $\mathcal{B}\left(B^0 \to \pi\tau\nu\right)$ on the input parameters are published. No constraints on other NP models are therefore published in this thesis.

(a)   **(b)**
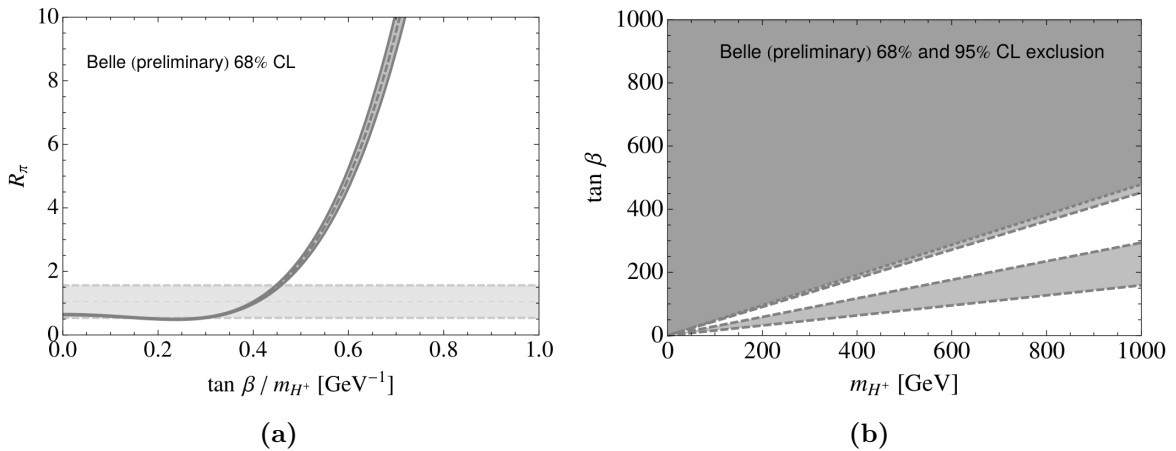
Figure 7.1.: Exclusion of MSSM parameter points under the assumption that the efficiencies and acceptances of the reconstruction are the same for MSSM scenarios. (a) The evolution of $R(\pi)$ as a function of $\tan\beta/m_H^+$ is shown in dark gray, the experimental 68% CL of $R(\pi)$ is shown in light gray. (b) Excluded MSSM parameter points in the $\tan\beta - m_H^+$ plane at the 68% CL (light gray) and 95% (dark gray). Taken from [124]

7.2. Outlook

While this analysis was not able to obtain evidence for the decay on the full Belle data sample, it still shows that evidence is to be expected in the future. First of all, in the recent past, an algorithm for the semileptonic tagging has been developed at Belle. The B_{tag} is reconstructed into the semileptonic decays $B \to D^{(*)}\ell\nu_\ell$, with light leptons ℓ [125]. The efficiency of this method is roughly two to three times higher than for the hadronic full reconstruction. A disadvantage is, however, the additional neutrino on the tag side, which means that the four-momentum of neither the tag, nor the signal side, can be determined. The search for $B^0 \to \pi\tau\nu$ using the semileptonic tag has been started in Belle by two Master theses, one in Göttingen by Harrison Schreeck [126], and one in Bonn, by Stephan Duell. Due to time constraints, the analyses have been performed on simulated data, only, and no evidence but a similar significance level is ex-

pected. As different decay channels are used to reconstruct the tag side in both analyses, the events of both analyses can be expected to be statistically independent. Therefore, evidence and a measurement of the branching fraction could be obtained in a combination of both analyses. Furthermore, Belle and BaBar recently published their first joint analysis, measuring the time-dependent CP violation in $B^0 \rightarrow D_{CP}^{(*)}h^0$ [127]. Combining the Belle and BaBar dataset for evidence of $B^0 \rightarrow \pi\tau\nu$ might be another option.

While LHCb is also specialized in B decays, it is not expected to be able to measure the decay $B^0 \rightarrow \pi\tau\nu$, though, as the B_{sig} can not be constrained by fully reconstructing the B_{tag}.

The upgrade of the Belle experiment, Belle II, will be able to obtain evidence for the decay, and shortly after, observe the decay. The Belle II experiment is scheduled to start taking data in 2018, and to collect a total integrated luminosity on the $\Upsilon(4S)$ resonance of $50\,\text{ab}^{-1}$, roughly 70 times as much data as this analysis was performed on. Further improvements of the detector and tracking algorithms are expected to increase the tracking efficency. This would improve all full reconstruction methods, and of course, also the reconstruction of the signal decay.

A. Appendix

A.1. Fit Stability

A.1.1. Pull distributions

This section shows the pull distribution parameter for all single mode fits and the combined fit to all four reconstruction channels. Both the signal strength parameter $\mu_{\text{in}}^{\text{sig}}$ and the $b \to c$ contribution parameters are varied. For both variations, the paramters of the fitted Gaussian to the pull distribution, mean μ and standard deviation σ are shown. The upper half of each table displays the results for different signal strength parameters μ_{in} while the lower half shows the results for different values of background paramters $f_{b \to c, \text{in}}$. Due to the multiple meanings of μ in this section, the fitted signal strength parameter is labeld μ^{sig}, while the meaning of the muon and the mean of the Gaussian distribution should be clear from context.

The tables contains the parameters without a quoted error to improve readability. The size of the error is the same for all values of input parameters because the same number of pulls has been computed. Compared to Figure 6.1, the uncertainty is slightly higher for both parameters because less pulls have been computed for this section. We obtain $\sigma_\mu = 0.015$ and $\sigma_\sigma = 0.14$.

$\mu_{\text{in}}^{\text{sig}}$ $[\times 10^{-4}]$	μ^{sig}		$f_{b \to c}(\tau \to e)$		$f_{b \to c}(\tau \to \mu)$		$f_{b \to c}(\tau \to \pi)$		$f_{b \to c}(\tau \to \rho)$	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
0.5	0.035	0.999	0.004	1.000	0.003	0.986	0.017	0.995	0.007	0.993
1.0	0.037	0.988	0.018	0.989	0.004	0.980	0.015	0.996	0.018	0.993
1.5	0.037	0.996	0.013	0.994	0.004	0.993	-0.001	1.000	0.022	0.999
2.0	0.037	1.000	0.005	1.010	0.018	0.988	0.013	0.991	0.009	1.000
$f_{b \to c, \text{in}}$										
0.8	0.020	0.983	0.017	1.000	-0.013	0.990	0.019	0.993	0.022	0.985
0.9	0.033	0.983	0.025	0.999	0.009	0.997	0.008	0.989	0.013	0.992
1.0	0.050	0.991	0.007	0.999	-0.013	0.982	0.003	0.995	0.025	0.999
1.1	0.024	1.000	0.019	0.993	0.023	0.986	0.020	1.000	0.014	0.986
1.2	0.047	0.994	0.003	0.996	0.005	0.985	0.003	1.000	0.021	0.992

Table A.1.: Combined fit $\tau \to e, \mu, \pi, \rho$.

$\mu_{\text{in}}^{\text{sig}}$ $[\times 10^{-4}]$	μ^{sig}		$f_{b \to c}(\tau \to e)$	
	μ	σ	μ	σ
0.5	0.036	0.991	0.030	0.971
1.0	0.054	0.999	0.035	0.980
1.5	0.034	0.975	0.024	0.986
2.0	0.037	0.991	0.036	0.980
$f_{b \to c,\text{in}}$				
0.8	0.028	0.983	0.011	0.980
0.9	0.005	0.990	0.035	0.970
1.0	0.045	1.000	0.018	0.973
1.1	0.049	0.997	0.034	0.986
1.2	0.033	0.995	−0.003	0.975

Table A.2.: Fit $\tau \to e$.

$\mu_{\text{in}}^{\text{sig}}$ $[\times 10^{-4}]$	μ^{sig}		$f_{b \to c}(\tau \to \mu)$	
	μ	σ	μ	σ
0.5	0.033	0.994	0.027	0.997
1.0	0.036	0.997	0.001	1.010
1.5	0.040	0.996	0.011	1.000
2.0	0.043	1.000	0.021	1.010
$f_{b \to c,\text{in}}$				
0.8	0.074	0.974	0.011	0.986
0.9	0.067	0.979	0.021	0.973
1.0	0.037	0.998	0.005	0.999
1.1	0.059	0.988	−0.012	0.983
1.2	0.043	1.000	−0.024	0.991

Table A.3.: Fit $\tau \to \mu$.

$\mu_{\text{in}}^{\text{sig}}$ $[\times 10^{-4}]$	μ^{sig}		$f_{b \to c}(\tau \to \pi)$	
	μ	σ	μ	σ
0.5	0.065	0.990	−0.019	1.000
1.0	0.075	1.000	−0.011	1.000
1.5	0.049	0.980	0.041	0.994
2.0	0.062	1.010	0.024	0.987
$f_{b \to c,\text{in}}$				
0.8	0.072	1.000	0.011	1.000
0.9	0.059	0.970	0.011	0.973
1.0	0.052	0.985	0.011	0.989
1.1	0.050	0.994	0.030	1.000
1.2	0.047	0.971	0.031	0.994

Table A.4.: Fit $\tau \to \pi$.

$\mu_{\text{in}}^{\text{sig}}$ $[\times 10^{-4}]$	μ^{sig}		$f_{b \to c}(\tau \to \rho)$	
	μ	σ	μ	σ
0.5	0.048	0.978	−0.012	0.992
1.0	0.086	0.970	0.007	1.010
1.5	0.101	0.989	0.013	0.999
2.0	0.080	0.995	0.021	0.995
$f_{b \to c,\text{in}}$				
0.8	0.082	0.971	−0.001	1.010
0.9	0.067	0.990	0.036	0.999
1.0	0.093	1.010	−0.003	1.000
1.1	0.099	0.973	−0.005	0.991
1.2	0.095	1.000	0.028	0.977

Table A.5.: Fit $\tau \to \rho$.

A.1.2. Linearity Signal Component

(a) $\tau \to e$

(b) $\tau \to \mu$

(c) $\tau \to \pi$

(d) $\tau \to \rho$

(e) Combination $\tau \to e, \mu, \pi, \rho$

(f) Combination $\tau \to e, \pi, \rho$
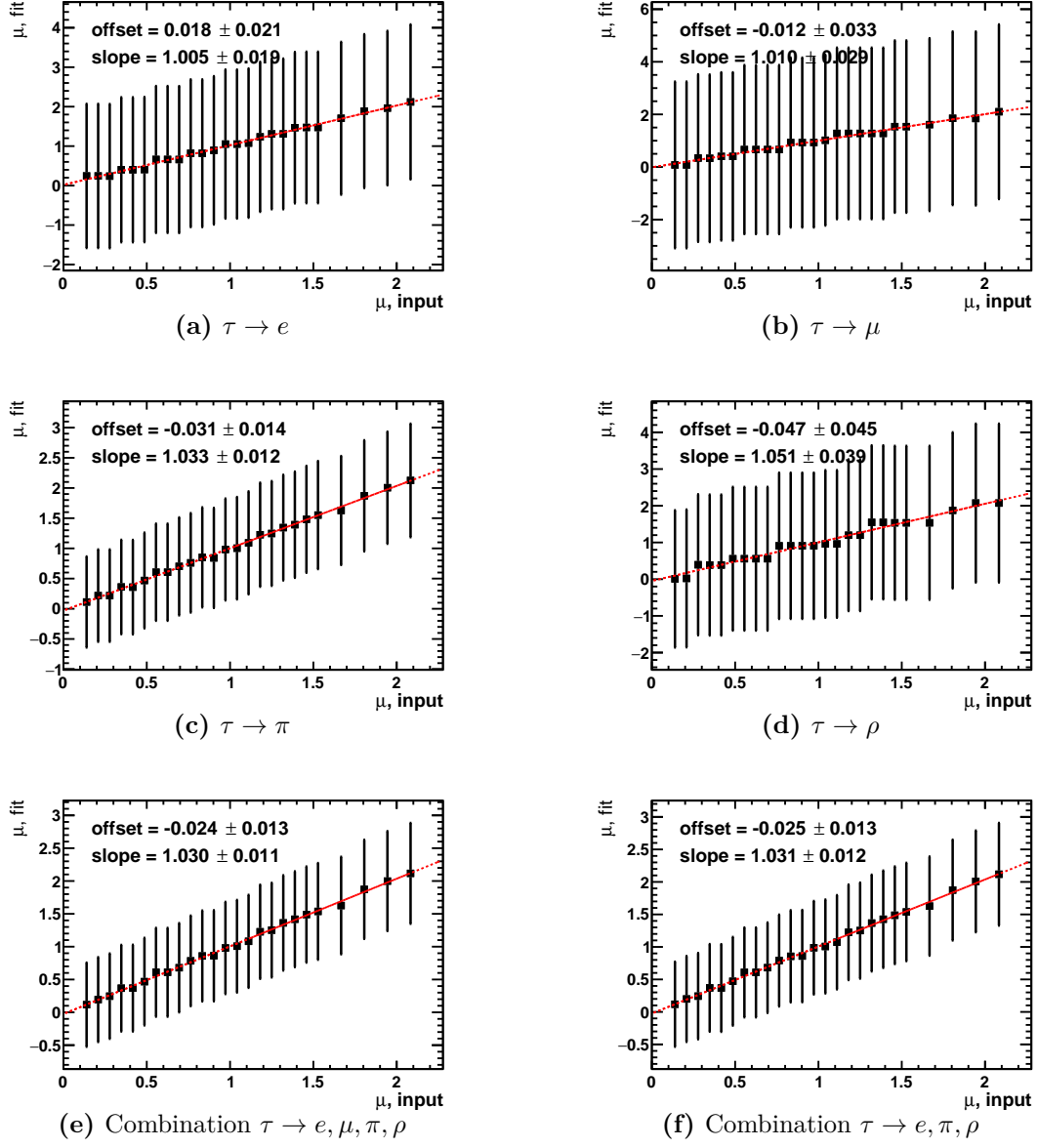
Figure A.1.: Linearity test for signal.

A.1.3. Linearity, $b \to c$ component

(a) $\tau \to e$

(b) $\tau \to \mu$

(c) $\tau \to \pi$

(d) $\tau \to \rho$
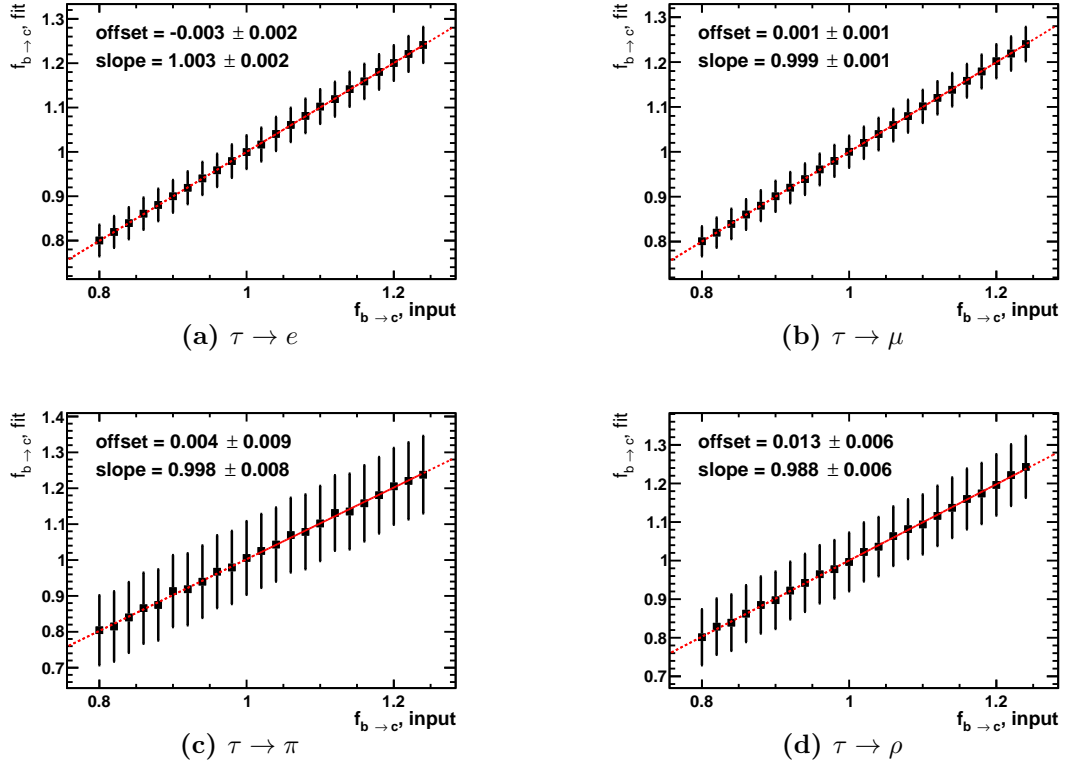
Figure A.2.: Linearity test for $b \to c$ single fits.

A.1.4. Fit on Data

	Single Mode Fit		
mode	θ	μ	N_{sig}
e	1.024 ± 0.038	3.432 ± 2.059	29.7 ± 17.9
π	0.922 ± 0.100	1.223 ± 0.803	24.5 ± 16.2
ρ	1.225 ± 0.079	1.335 ± 1.954	7.1 ± 10.5

Table A.6.: Fit results for single mode fits. $\mu = 1$ denotes $\mathcal{B}\left(B^0 \to \pi\tau\nu\right) = 1.00 \times 10^{-4}$.

.
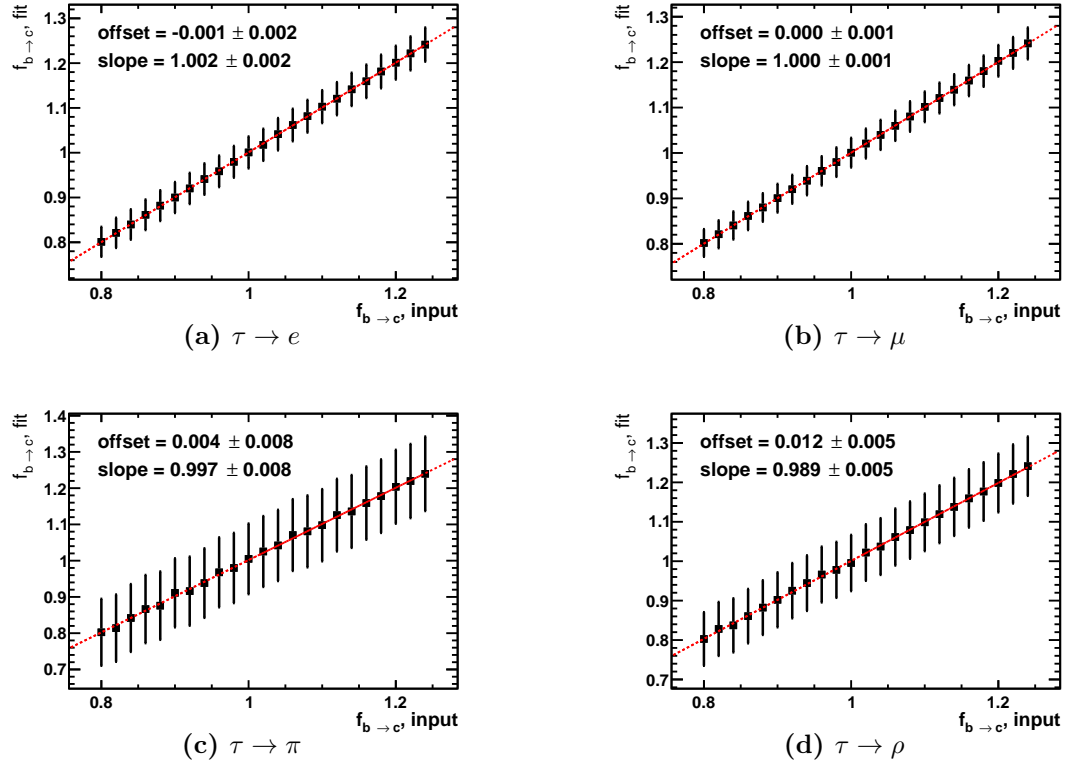
Figure A.3.: Linearity test for $b \to c$ in the combined fit $\tau \to e, \mu, \pi, \rho$.
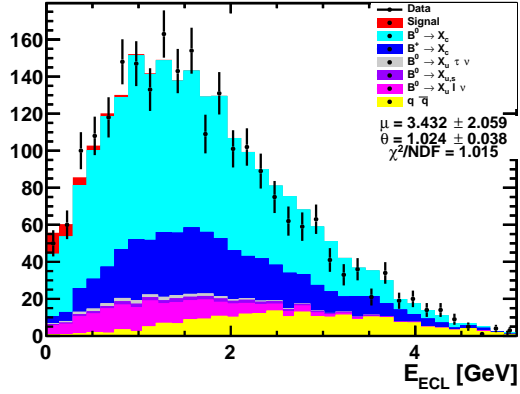
(a) Fit to only $\tau \to e$.
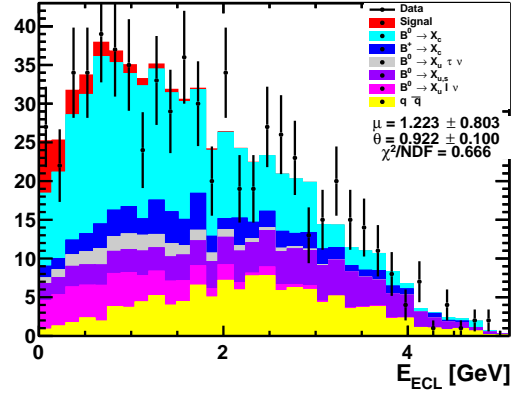
(b) Fit to only $\tau \to \pi$.
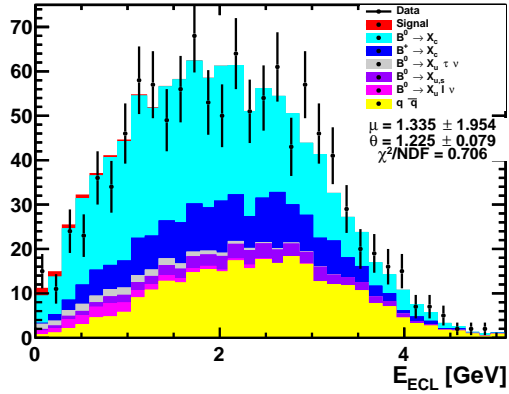
(c) Fit to only $\tau \to \rho$.

Figure A.4.: Fit to data for all three modes. For every mode, one single fit is performed.

A.2. MC / Data comparison

The comparison between MC prediction and data is presented in this section. First, both samples are compared in different variables, applying the sideband region cut of $E_{ECL} > 1.35\,\text{GeV}$. Next, the agreement of the two samples is shown without a sideband restriction, and with all MC samples scaled to their fitted value, obtained in Section 6.1.3.

For better readability, the three relevant E_{ECL} distribution already shown in Section 5.9 are repeated here. The legend for all plots is the same as in the E_{ECL} plots on this page.

(a) $\tau \to e$

(b) $\tau \to \pi$

(c) $\tau \to \rho$
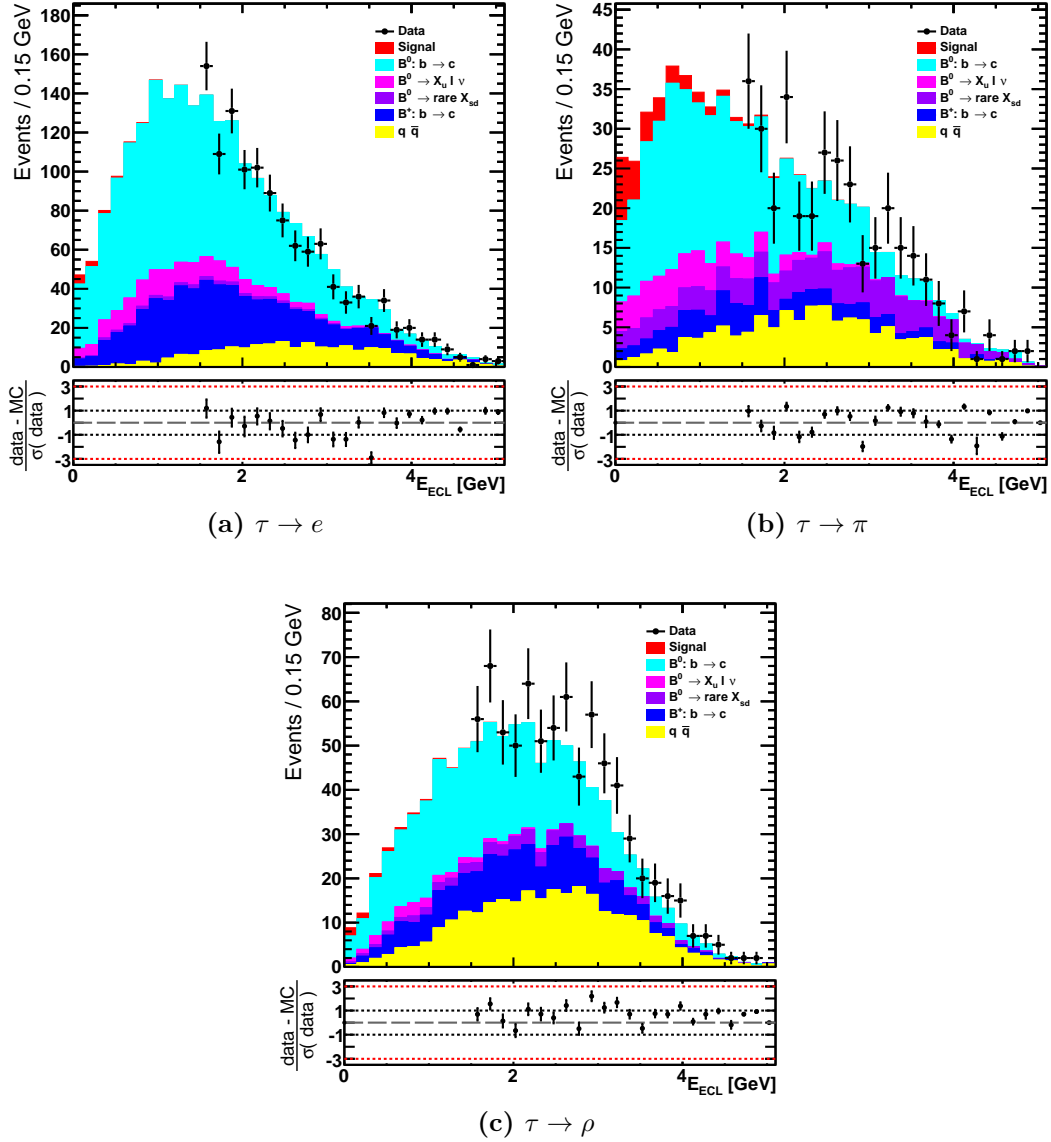
Figure A.5.: Comparison of data and Monte Carlo simulation in the sideband $E_{ECL} > 1.35\,\text{GeV}$.

A.2.1. Sideband Region

$\tau \to \mathbf{e}$
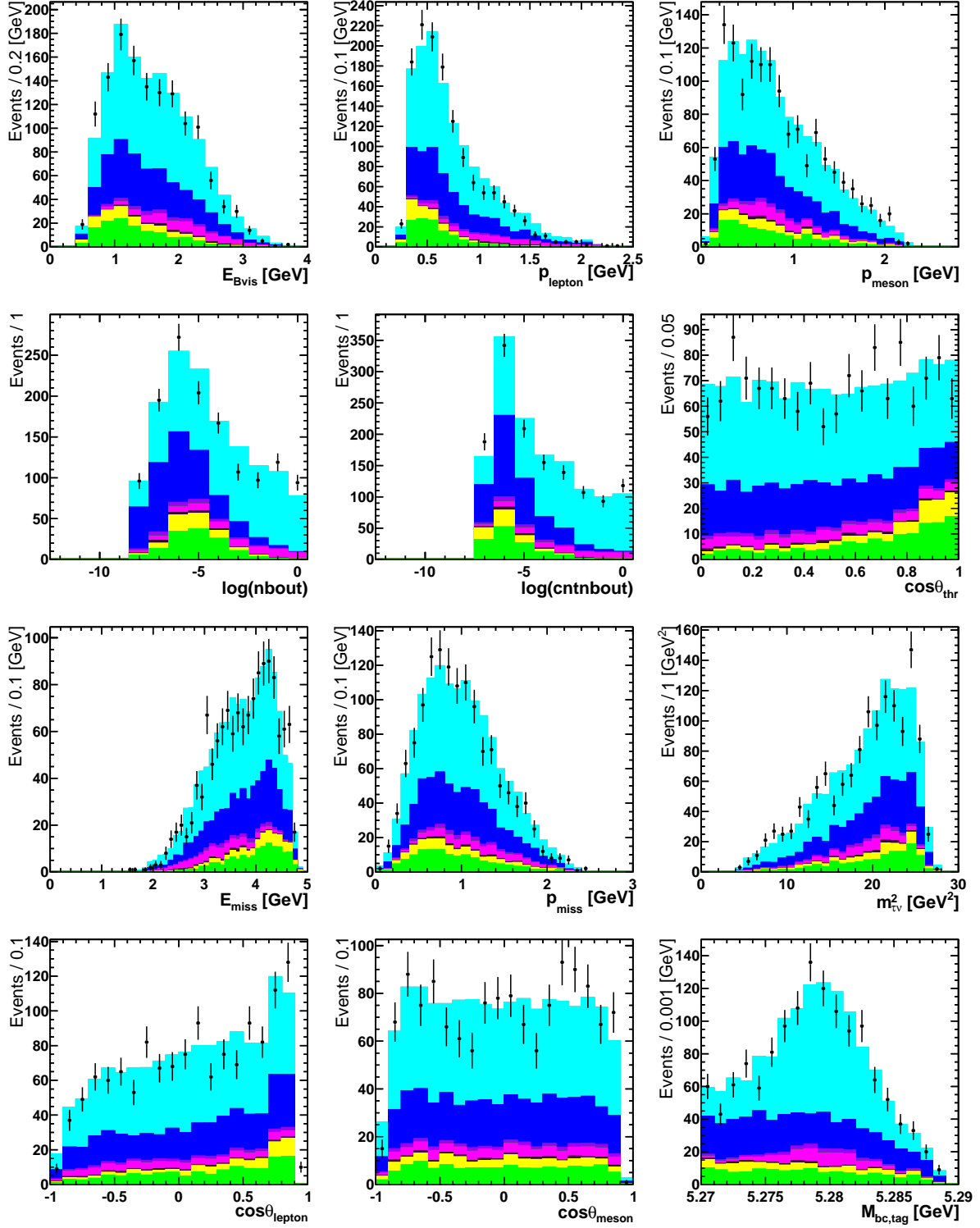
Figure A.6.: Distributions with the sideband cut for the $\tau \to e$ samples.

$\tau \to \pi$
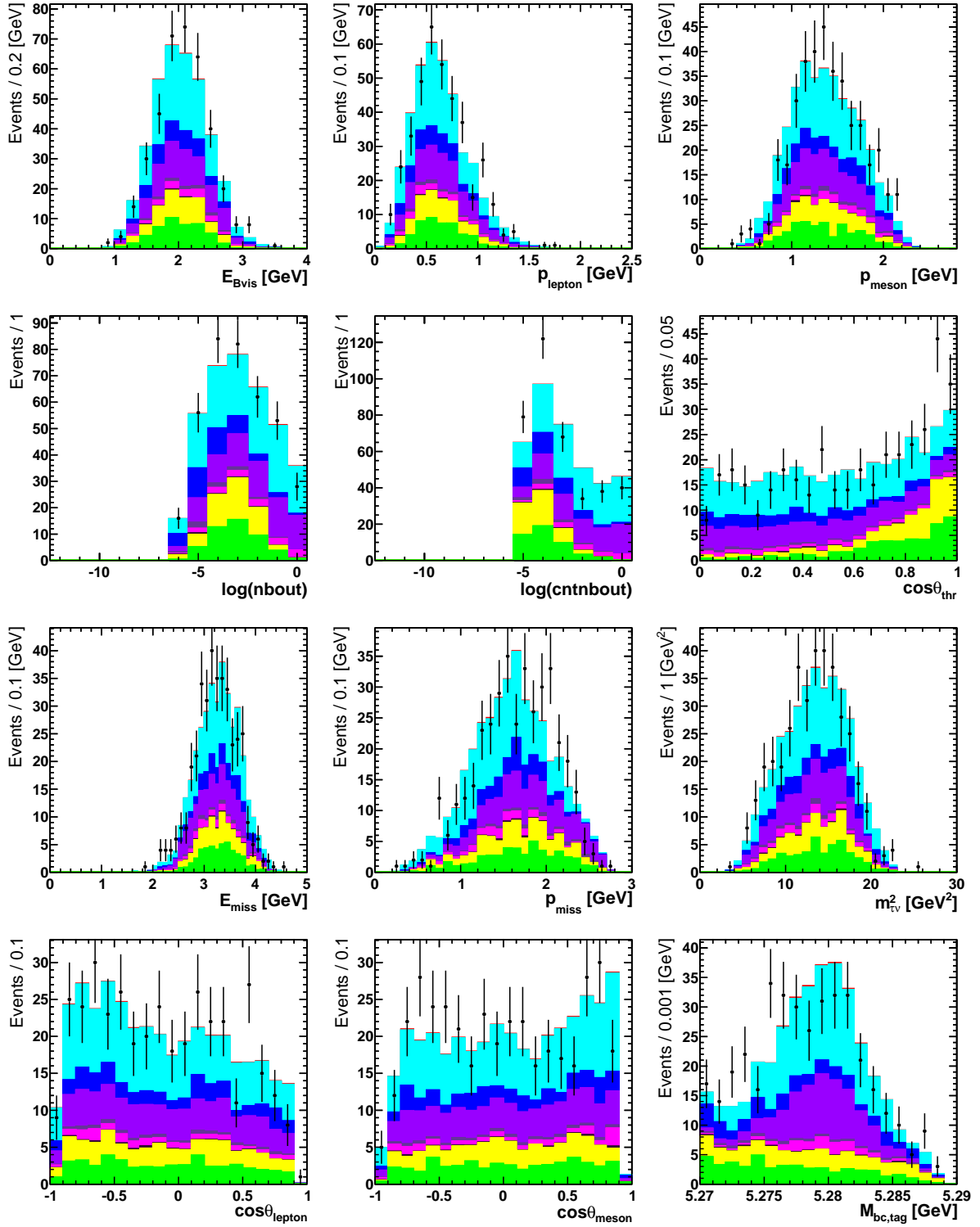
Figure A.7.: Distributions with the sideband cut for the $\tau \to \pi$ samples.

$\tau \to \rho$
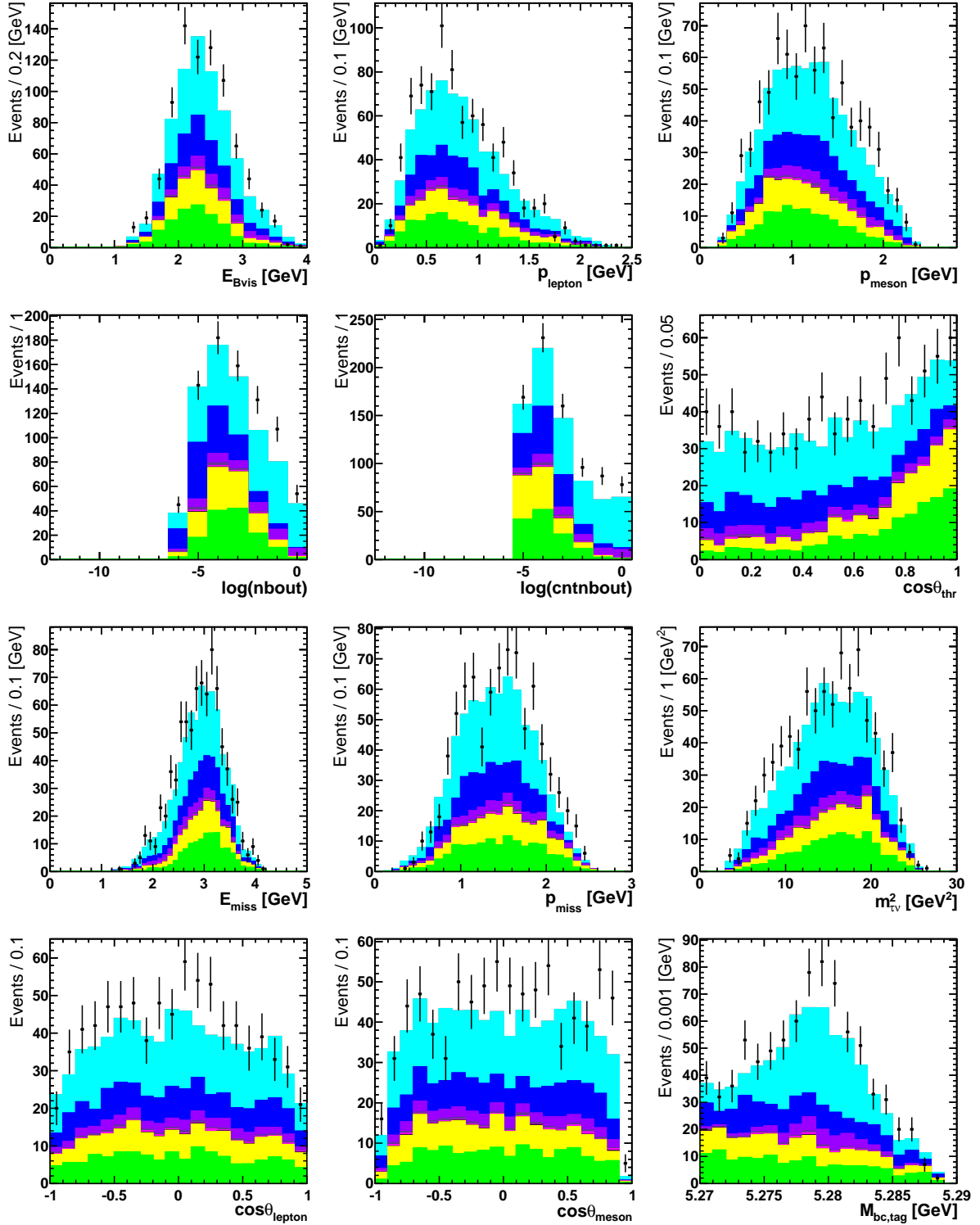
Figure A.8.: Distributions with the sideband cut for the $\tau \to \rho$ samples.

A.2.2. Complete Samples, Scaled to Fit Result

$\tau \to e$
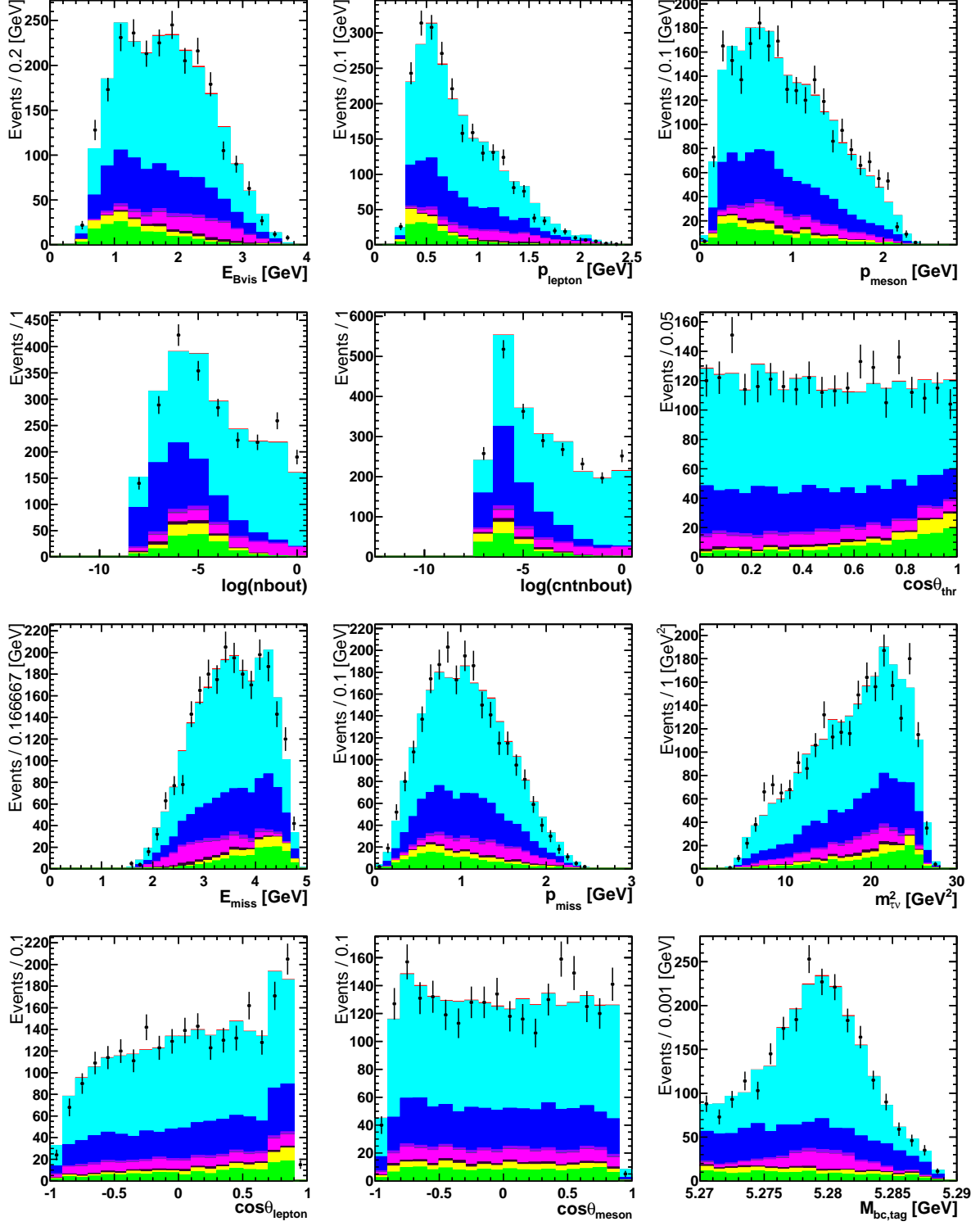
Figure A.9.: Distributions scaled to fit result for the complete $\tau \to e$ samples.

$\tau \to \pi$
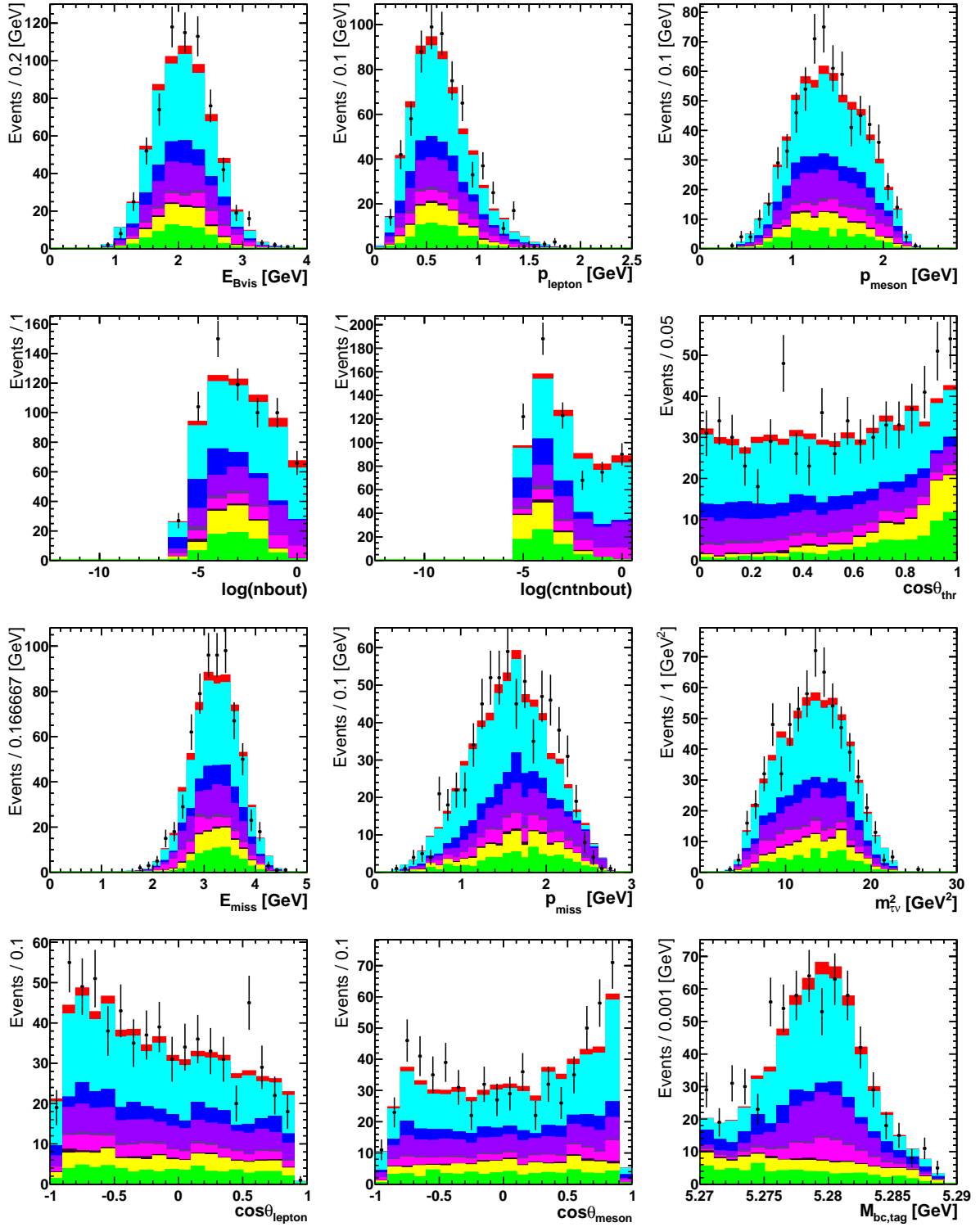
Figure A.10.: Distributions scaled to fit result for the complete $\tau \to \pi$ samples.

$\tau \to \rho$
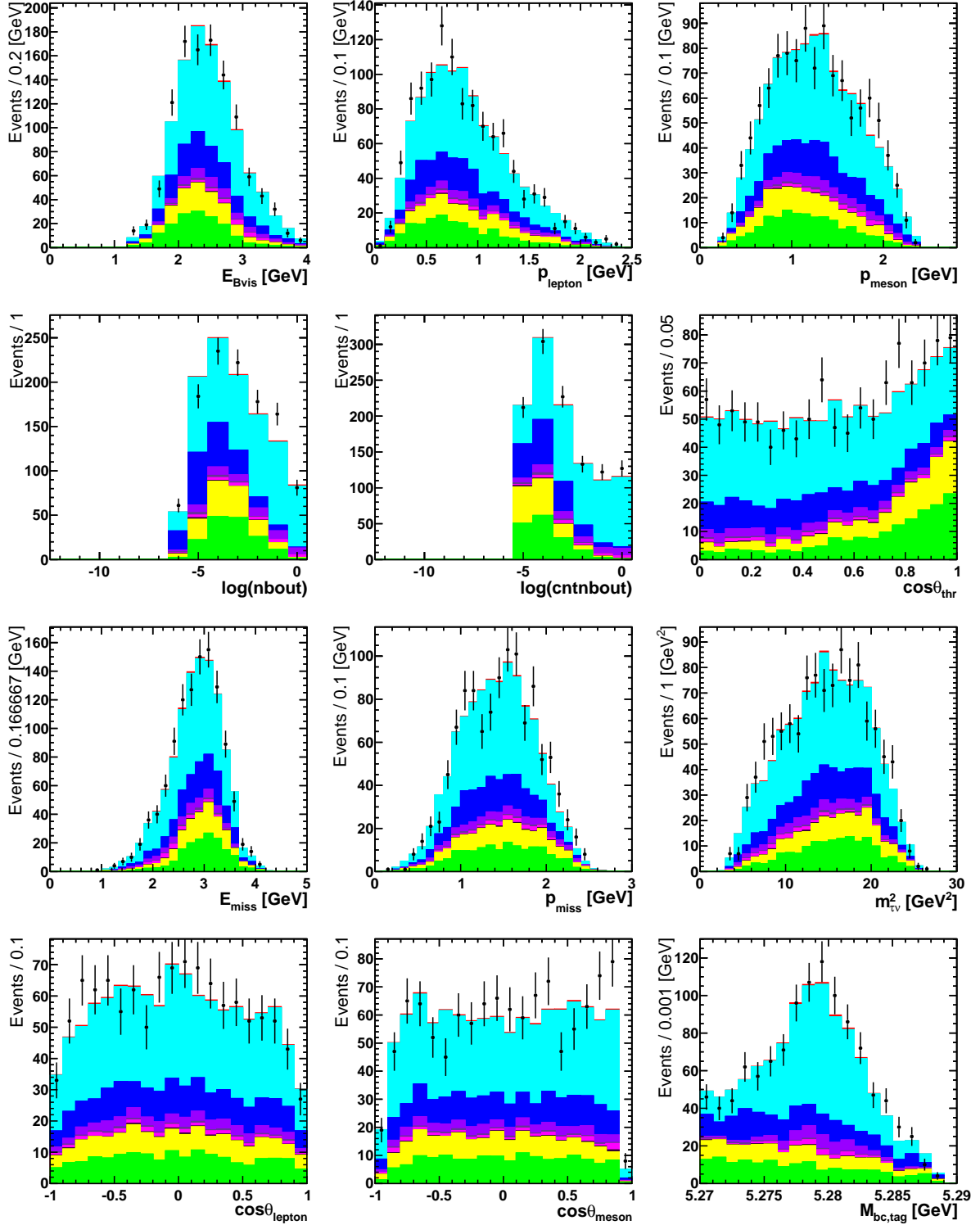
Figure A.11.: Distributions scaled to fit result for the complete $\tau \to \rho$ samples.

Bibliography

[1] F. HALZEN and A. MARTIN, *Quarks and leptons: an introductory course in modern particle physics*, Wiley, 1984, ↑ pp. 1, 5.

[2] D. GRIFFITHS, *Introduction to Elementary Particles*, WILEY-VCH, 2008, ↑ pp. 1, 5, 17.

[3] G. AAD et al. (The ATLAS collaboration), *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012), 1, ↑ pp. 1, 13.

[4] S. CHATRCHYAN et al. (The CMS collaboration), *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012), 30, ↑ pp. 1, 13.

[5] P. ADE et al. (The Planck collaboration), *Planck 2013 results. I. Overview of products and scientific results*, Astron. Astrophys. 571 (2014), A1, ↑ pp. 1, 14.

[6] NOBELPRIZE.ORG NOBEL MEDIA AB 2014, *The 2008 Nobel Prize in Physics - Press Release*, URL: http://www.nobelprize.org/nobel_prizes/physics/laureates/2008/press.html (visited on 08/21/2015), ↑ pp. 2, 25.

[7] HEAVY FLAVOR AVERAGING GROUP (HFAG), *Average of $R(D)$ and $R(D^*)$ for EPS-HEP 2015*, Aug. 2015, URL: https://www.slac.stanford.edu/xorg/hfag/semi/eps15/eps15_dtaunu.html, ↑ pp. 2, 15, 17.

[8] A. KHODJAMIRIAN et al., *$B \to \pi l \nu_l$ Width and $|V_{ub}|$ from QCD light-cone sum rules*, Phys. Rev. D 83 (2011), 094031, ↑ pp. 2, 18–22, 40, 41.

[9] M. PESKIN and D. SCHROEDER, *An Introduction to Quantum Field Theory*, Advanced book classics, Addison-Wesley Publishing Company, 1995, ISBN: 9780201503975, ↑ p. 5.

[10] K. OLIVE et al. (The Particle Data Group collaboration), *Review of Particle Physics*, Chin. Phys. C 38 (2014), 090001, ↑ pp. 7–14, 19, 20, 25, 43, 55, 58, 60–62.

[11] S. GLASHOW, *Partial Symmetries of Weak Interactions*, Nucl. Phys. 22 (1961), 579, ↑ p. 8.

[12] S. WEINBERG, *A Model of Leptons*, Phys. Rev. Lett. 19 (1967), 1264, ↑ pp. 8, 11.

[13] A. SALAM and J. C. WARD, *Electromagnetic and weak interactions*, Phys. Lett. 13 (1964), 168, ↑ p. 8.

[14] E. FERMI, *An attempt of a theory of beta radiation. 1.* Z. Phys. 88 (1934), 161, ↑ p. 10.

[15] N. CABIBBO, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. 10 (1963), 531, ↑ p. 10.

[16] M. KOBAYASHI and T. MASKAWA, *CP-Violation in the Renormalizable Theory of Weak Interaction*, PTEP 49.2 (1973), 652, ↑ p. 10.

[17] B. PONTECORVO, *Mesonium and anti-mesonium*, Sov. Phys. JETP 6 (1957), 429, ↑ p. 11.

[18] B. PONTECORVO, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge*, Sov. Phys. JETP 26 (1968), 984, ↑ p. 11.

[19] Z. MAKI, M. NAKAGAWA, and S. SAKATA, *Remarks on the Unified Model of Elementary Particles*, PTEP 28.5 (1962), 870, ↑ p. 11.

[20] F. ENGLERT and R. BROUT, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. 13 (1964), 321, ↑ p. 11.

[21] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. 13 (1964), 508, ↑ p. 11.

[22] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. 13 (1964), 585, ↑ p. 11.

[23] G. Aad et al. (The ATLAS and CMS collaborations), *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. 114 (2015), 191803, ↑ p. 13.

[24] G. Aad et al. (The ATLAS collaboration), *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. B 726 (2013), 120, ↑ p. 13.

[25] S. Chatrchyan et al. (The CMS collaboration), *Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs*, Phys. Rev. Lett. 110.8 (2013), 081803, ↑ p. 13.

[26] S. Fukuda et al. (The Super-Kamiokande collaboration), *Solar B-8 and hep neutrino measurements from 1258 days of Super-Kamiokande data*, Phys. Rev. Lett. 86 (2001), 5651, ↑ p. 13.

[27] Q. Ahmad et al. (The SNO collaboration), *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. 87 (2001), 071301, ↑ p. 13.

[28] Q. Ahmad et al. (The SNO collaboration), *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. 89 (2002), 011301, ↑ p. 13.

[29] M. E. Peskin, *The Matter with antimatter*, Nature 419 (2002), 24, ↑ p. 13.

[30] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, Helv. Phys. Acta 6 (1933), 110, ↑ p. 13.

[31] P. Peebles and B. Ratra, *The Cosmological constant and dark energy*, Rev. Mod. Phys. 75 (2003), 559, ↑ p. 13.

[32] R. Glattauer, *Semileptonic B and B_s decays at Belle*, EPS-HEP 2015, paper in preparation, URL: https://indico.cern.ch/event/356420/session/3/contribution/306/attachments/1128040/1622054/2015_07_EPSHEP_Glattauer.pdf, ↑ p. 14.

[33] R. Aaij et al. (The LHCb collaboration), *Test of lepton universality using $B^+ \to K^+ \ell^+ \ell^-$ decays*, Phys. Rev. Lett. 113 (2014), 151601, ↑ p. 14.

[34] J. T. Wei et al. (The Belle collaboration), *Measurement of the Differential Branching Fraction and Forward-Backword Asymmetry for $B \to K^{(*)} \ell^+ \ell^-$*, Phys. Rev. Lett. 103 (2009), 171801, ↑ p. 14.

[35] J. P. Lees et al. (The BaBar collaboration), *Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)} \ell^+ \ell^-$*, Phys. Rev. D 86 (2012), 032012, ↑ p. 14.

[36] S. Descotes-Genon et al., *Implications from clean observables for the binned analysis of $B \to K^* \mu^+ \mu^-$ at large recoil*, JHEP 01 (2013), 048, ↑ p. 15.

[37] S. Descotes-Genon et al., *On the impact of power corrections in the prediction of $B \to K^* \mu^+ \mu^-$ observables*, JHEP 12 (2014), 125, ↑ p. 15.

[38] The LHCb collaboration, *Angular analysis of the $B^0 \to K^{*0} \mu^+ \mu^-$ decay*, 2015, eprint: https://cds.cern.ch/record/2002772, ↑ p. 15.

[39] W. Altmannshofer and D. M. Straub, *New physics in $b \to s$ transitions after LHC run 1*, Eur. Phys. J. C 75.8 (2015), 382, ↑ p. 15.

[40] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V \ell^+ \ell^-$ in the Standard Model from Light-Cone Sum Rules*, 2015, arXiv: 1503.05534 [hep-ph], ↑ p. 15.

[41] THE LHCB COLLABORATION, *An interesting result presented at the LHCP conference.* Sept. 2015, URL: http://lhcb-public.web.cern.ch/lhcb-public/Welcome.html#RK, ↑ p. 16.

[42] THE LHCB COLLABORATION, $B^0 \to K^* \mu^+ \mu^-$: *new analysis confirms old puzzle.* Sept. 2015, URL: http://lhcb-public.web.cern.ch/lhcb-public/Welcome.html#P5p, ↑ p. 16.

[43] B. AUBERT et al. (The BaBar collaboration), *A Search for $B^+ \to \ell^+ \nu_\ell$ Recoiling Against $B^- \to D^0 \ell^- \bar{\nu} X$,* Phys. Rev. D 81 (2010), 051101, ↑ p. 15.

[44] K. HARA et al. (The Belle collaboration), *Evidence for $B^- \to \tau^- \bar{\nu}$ with a Semileptonic Tagging Method,* Phys. Rev. D 82 (2010), 071101, ↑ p. 15.

[45] J. P. LEES et al. (The BaBar collaboration), *Evidence of $B^+ \to \tau^+ \nu$ decays with hadronic B tags,* Phys. Rev. D 88.3 (2013), 031102, ↑ p. 15.

[46] I. ADACHI et al. (The Belle collaboration), *Evidence for $B^- \to \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle,* Phys. Rev. Lett. 110.13 (2013), 131801, ↑ p. 15.

[47] B. KRONENBITTER et al. (The Belle collaboration), *Measurement of the branching fraction of $B^+ \to \tau^+ \nu_\tau$ decays with the semileptonic tagging method,* accepted by Phys. Rev. D, 2015, arXiv: 1503.05613 [hep-ex], ↑ p. 15.

[48] J. P. LEES et al. (The BaBar collaboration), *Evidence for an excess of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ decays,* Phys. Rev. Lett. 109 (2012), 101802, ↑ p. 15.

[49] M. HUSCHLE et al. (The Belle collaboration), *Measurement of the branching ratio of $\bar{B} \to D^{(*)} \tau^- \bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle,* (2015), accepted by Phys. Rev. D, ↑ p. 15.

[50] R. AAIJ et al. (The LHCb collaboration), *Measurement of the ratio of branching fractions $\mathcal{B}(\overline{B}^0 \to D^{*+} \tau^- \bar{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+} \mu^- \bar{\nu}_\mu)$,* Phys. Rev. Lett. 115.11 (2015), 111803, ↑ p. 15.

[51] J. D. RICHMAN and P. R. BURCHAT, *Leptonic and semileptonic decays of charm and bottom hadrons,* Rev. Mod. Phys. 67 (1995), 893, ↑ p. 17.

[52] M. WIRBEL, B. STECH, and M. BAUER, *Exclusive Semileptonic Decays of Heavy Mesons,* Z. Phys. C 29 (1985), 637, ↑ p. 18.

[53] M. NEUBERT, *Heavy quark symmetry,* Phys. Rept. 245 (1994), 259, ↑ p. 18.

[54] E. DALGIC et al., *B meson semileptonic form-factors from unquenched lattice QCD,* Phys. Rev. D 73 (2006), [Erratum: Phys. Rev. D 75,119906(2007)], 074502, ↑ pp. 18, 19.

[55] J. A. BAILEY et al., *The $B \to \pi \ell \nu$ semileptonic form factor from three-flavor lattice QCD: A Model-independent determination of $|V_{ub}|$,* Phys. Rev. D 79 (2009), 054507, ↑ pp. 18, 19.

[56] J. A. BAILEY et al. (The Fermilab Lattice and MILC collaborations), *$|V_{ub}|$ from $B \to \pi \ell \nu$ decays and (2+1)-flavor lattice QCD,* Phys. Rev. D 92.1 (2015), 014024, ↑ pp. 18–20.

[57] R. GUPTA, *Introduction to lattice QCD: Course,* in: *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2,* 1997, pp. 83–219, arXiv: hep-lat/9807028 [hep-lat], ↑ p. 18.

[58] G. P. LEPAGE, *Lattice QCD for novices,* in: *Strong interactions at low and intermediate energies. Proceedings, 13th Annual Hampton University Graduate Studies, HUGS'98, Newport News, USA, May 26-June 12, 1998,* 1998, pp. 49–90, arXiv: hep-lat/0506036 [hep-lat], ↑ p. 18.

[59] P. BALL and R. ZWICKY, *New results on $B \to \pi, K, \eta$ decay formfactors from light-cone sum rules,* Phys. Rev. D 71 (2005), 014015, ↑ p. 18.

[60] G. DUPLANCIC et al., *Light-cone sum rules for $B \to \pi$ form factors revisited*, JHEP 04 (2008), 014, ↑ p. 18.

[61] M. A. SHIFMAN, A. I. VAINSHTEIN, and V. I. ZAKHAROV, *QCD and Resonance Physics. Theoretical Foundations*, Nucl. Phys. B 147 (1979), 385, ↑ p. 18.

[62] A. KHODJAMIRIAN and R. RUCKL, *QCD sum rules for exclusive decays of heavy mesons*, in: *Heavy Flavours II*, chap. 5, pp. 345–401, arXiv: hep-ph/9801443 [hep-ph], ↑ pp. 18, 20.

[63] V. M. BRAUN, *Light cone sum rules*, in: *Progress in heavy quark physics. Proceedings, 4th International Workshop, Rostock, Germany, September 20-22, 1997*, 1997, arXiv: hep-ph/9801222 [hep-ph], ↑ p. 18.

[64] P. COLANGELO and A. KHODJAMIRIAN, *QCD sum rules, a modern perspective*, in: *At the frontier of particle physics*, ed. by M. SHIFMAN, vol. 3, 2000, arXiv: hep-ph/0010175 [hep-ph], ↑ p. 18.

[65] C. G. BOYD and M. J. SAVAGE, *Analyticity, shapes of semileptonic form-factors, and $\bar{B} \to \pi \ell \bar{\nu}$*, Phys. Rev. D 56 (1997), 303, ↑ p. 18.

[66] C. BOURRELY, I. CAPRINI, and L. LELLOUCH, *Model-independent description of $B \to \pi \ell \nu$ decays and a determination of $|V_{ub}|$*, Phys. Rev. D 79 (2009), [Erratum: Phys. Rev. D82, 099902(2010)], 013008, ↑ p. 19.

[67] R. DUTTA, A. BHOL, and A. K. GIRI, *Effective theory approach to new physics in $b \to u$ and $b \to c$ leptonic and semileptonic decays*, Phys. Rev. D 88.11 (2013), 114023, ↑ pp. 19, 20, 23.

[68] C. A. DOMINGUEZ, J. G. KORNER, and K. SCHILCHER, *The scalar form-factor in the exclusive semileptonic decay of $B \to \pi \tau \nu_\tau$*, Phys. Lett. B 248 (1990), 399, ↑ p. 20.

[69] R. van de WATER, D. DAPING, and R. ZHOU, private communication, ↑ pp. 20, 98.

[70] T. D. LEE, *A Theory of Spontaneous T Violation*, Phys. Rev. D 8 (1973), 1226, ↑ p. 20.

[71] G. C. BRANCO et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. 516 (2012), 1, ↑ p. 20.

[72] Yu. A. GOLFAND and E. P. LIKHTMAN, *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*, JETP Lett. 13 (1971), [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)], 323, ↑ p. 20.

[73] J. WESS and B. ZUMINO, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. B 70 (1974), 39, ↑ pp. 20, 23.

[74] H. E. HABER and G. L. KANE, *The Search for Supersymmetry: Probing Physics Beyond the Standard Model*, Phys. Rept. 117 (1985), 75, ↑ p. 20.

[75] H. MURAYAMA, *Supersymmetry phenomenology*, in: *Particle physics. Proceedings, Summer School, Trieste, Italy, June 21-July 9, 1999*, 2000, pp. 296–335, arXiv: hep-ph/0002232 [hep-ph], ↑ pp. 20, 23.

[76] THE ATLAS COLLABORATION, *ATLAS Supersymmetry (SUSY) searches*, Sept. 2015, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults, ↑ p. 20.

[77] G. AAD et al. (The ATLAS collaboration), *Summary of the ATLAS experiment's sensitivity to supersymmetry after LHC Run 1 - interpreted in the phenomenological MSSM*, 2015, arXiv: 1508.06608 [hep-ex], ↑ p. 20.

[78] THE CMS COLLABORATION, *CMS Supersymmetry Physics Results*, Sept. 2015, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS, ↑ p. 20.

[79] R. D. PECCEI and H. R. QUINN, *CP Conservation in the Presence of Instantons*, Phys. Rev. Lett. 38 (1977), 1440, ↑ p. 21.

[80] M. TRODDEN, *Electroweak baryogenesis: A Brief review*, in: *33rd Rencontres de Moriond: Electroweak Interactions and Unified Theories Les Arcs, France, March 14-21, 1998*, 1998, arXiv: `hep-ph/9805252 [hep-ph]`, ↑ p. 21.

[81] N. TUROK and J. ZADROZNY, *Electroweak baryogenesis in the two doublet model*, Nucl. Phys. B 358 (1991), 471, ↑ p. 21.

[82] L. FROMME, S. J. HUBER, and M. SENIUCH, *Baryogenesis in the two-Higgs doublet model*, JHEP 11 (2006), 038, ↑ p. 21.

[83] Y. GROSSMAN and Z. LIGETI, *The Inclusive $\bar{B} \to \tau \bar{\nu}_\tau X$ decay in two Higgs doublet models*, Phys. Lett. B 332 (1994), 373, ↑ p. 21.

[84] C. S. KIM and R.-M. WANG, *Study of the exclusive $b \to u\ell^- \bar{\nu}_\ell$ decays in the MSSM with and without R-parity violation*, Phys. Rev. D 77 (2008), 094006, ↑ pp. 22–24.

[85] B. AUBERT et al. (The BaBar collaboration), *A Search for $B^+ \to \tau^+ \nu$*, Phys. Rev. D 76 (2007), 052002, ↑ p. 23.

[86] K. IKADO et al. (The Belle collaboration), *Evidence of the Purely Leptonic Decay $B^- \to \tau^- \bar{\nu}_\tau$*, Phys. Rev. Lett. 97 (2006), 251802, ↑ p. 23.

[87] C.-H. CHEN and C.-Q. GENG, *Charged Higgs on $B^- \to \tau \bar{\nu}_\tau$ and $\bar{B} \to P(V)\ell\bar{\nu}_\ell$*, JHEP 10 (2006), 053, ↑ p. 23.

[88] A. J. BEVAN et al. (The Belle and BaBar collaborations), *The Physics of the B Factories*, Eur. Phys. J. C 74 (2014), 3026, ↑ pp. 25, 29.

[89] S. KUROKAWA and E. KIKUTANI, *Overview of the KEKB accelerators*, Nucl. Instr. and Meth. A 499.1 (2003), 1, ISSN: 0168-9002, ↑ p. 25.

[90] T. SUWADA et al., *First application of a tungsten single-crystal positron source at the KEK B factory*, Phys. Rev. ST Accel. Beams 10 (7 July 2007), 073501, ↑ p. 25.

[91] KEKB, *The Luminosity Records*, 11/2013, URL: `http://www-acc.kek.jp/kekb/Commissioning/Record/Luminosity_record.html`, ↑ p. 25.

[92] J. BRODZICKA et al., *Physics achievements from the Belle experiment*, Progr. Theor. Exp. Phys. 2012.1 (2012), ↑ p. 25.

[93] A. ABASHIAN, K. GOTOW, N. MORGAN, et al., *The Belle detector*, Nucl. Instr. and Meth. A 479.1 (2002), Detectors for Asymmetric B-factories, 117, ISSN: 0168-9002, ↑ p. 27.

[94] H. AIHARA et al., *Belle SVD2 vertex detector*, Nucl. Instr. and Meth. A 568.1 (2006), 269, ISSN: 0168-9002, ↑ p. 28.

[95] P. A. CHERENKOV, *Visible emission of clean liquids by action of γ radiation*, in: *Doklady Akademii Nauk SSSR*, vol. 2, 1934, p. 451, ↑ p. 29.

[96] K. HANAGAKI et al., *Electron identification in Belle*, Nucl. Instr. and Meth. A 485 (2002), 490, ↑ pp. 33, 35, 60.

[97] A. ABASHIAN, *Muon identification in the Belle experiment at KEKB*, Nucl. Instr. and Meth. A 491 (2002), 69, ↑ pp. 35, 36, 60.

[98] K. GROUP, *Kaon Identification in Belle*, BN321 (), ↑ pp. 36–38, 60.

[99] D. J. LANGE, *The EvtGen particle decay simulation package*, Nucl. Instr. and Meth. A 462.1-2 (2001), 152, ISSN: 0168-9002, ↑ p. 39.

[100] R. BRUN et al., *GEANT 3.21*, CERN Report DD/EE/84-1, 1984, ↑ p. 39.

[101] M. FEINDT et al., *A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories*, Nucl. Instr. and Meth. A 654.1 (2011), 432, ISSN: 0168-9002, ↑ p. 44.

[102] M. FEINDT and U. KERZEL, *The NeuroBayes neural network package*, Nucl. Instr. and Meth. A 559 (2006), 190, ↑ p. 44.

[103] G. C. FOX and S. WOLFRAM, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. 41 (23 Dec. 1978), 1581, ↑ p. 46.

[104] S. H. Lee et al. (The Belle collaboration), *Evidence for $B^0 \to \pi^0\pi^0$*, Phys. Rev. Lett. 91 (26 Dec. 2003), 261801, ↑ p. 46.

[105] A. Hoecker et al., *TMVA: Toolkit for Multivariate Data Analysis*, PoS ACAT (2007), 040, ↑ pp. 46–48, 67.

[106] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences 55.1 (1997), 119, issn: 0022-0000, ↑ pp. 48, 49.

[107] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, Annals of Statistics 29 (2000), 1189, ↑ p. 48.

[108] J. H. Friedman, *Stochastic gradient boosting*, Computational Statistics & Data Analysis 38.4 (2002), Nonlinear Methods and Data Mining, 367, issn: 0167-9473, ↑ p. 48.

[109] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc. 2001, ↑ pp. 48, 49.

[110] O. Behnke, K. Kröninger, G. Schott, T. Schörner-Sadenius, Ed. *Data Analysis in High Energy Physics*, WILEY-VCH, 2013, ↑ p. 50.

[111] R. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Manchester Physics Series, Wiley, 1989, isbn: 9780471922957, ↑ p. 50.

[112] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 231.694-706 (1933), 289, issn: 0264-3952, ↑ p. 51.

[113] S. Wilk, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. 9 (1938), ↑ p. 52.

[114] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C 71.2 (2011), ↑ p. 52.

[115] I. Antcheva et al., *ROOT - A C++ framework for petabyte data storage, statistical analysis and visualization*, Computer Physics Communications 180.12 (2009), 40 YEARS OF CPC: A celebratory issue focused on quality software for high performance, grid and novel computing architectures, 2499, issn: 0010-4655, ↑ p. 52.

[116] A. L. Read, *Presentation of search results: the CL_s technique*, Journal of Physics G: Nuclear and Particle Physics 28.10 (2002), 2693, ↑ p. 54.

[117] W. Dungel et al. (The Belle collaboration), *Measurement of the form factors of the decay $B^0 \to D^{*-}\ell^+\nu$ and determination of the CKM matrix element $|V_{cb}|$*, Phys. Rev. D 82 (2010), 112007, ↑ pp. 60, 92.

[118] A. Sibidanov et al. (The Belle collaboration), *Study of Exclusive $B \to X_u\ell\nu$ Decays and Extraction of $|V_{ub}|$ using Full Reconstruction Tagging at the Belle Experiment*, Phys. Rev. D 88.3 (2013), 032005, ↑ pp. 60, 93.

[119] C. Oswald et al. (The Belle collaboration), *Semi-inclusive studies of semileptonic B_s decays at Belle*, submitted to Phys. Rev. D, 2015, arXiv: 1504.02004 [hep-ex], ↑ p. 61.

[120] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. 77 (1993), 219, ↑ pp. 93, 94.

[121] Y. Amhis et al. (The Heavy Flavor Averaging Group (HFAG) collaboration), *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2014*, (2014), ↑ p. 93.

[122] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: An update*, Phys. Rev. D 52 (1995), 2783, ↑ p. 93.

[123] P. Hamer et al. (The Belle collaboration), *Search for $B^0 \to \pi^-\tau^+\nu_\tau$ with hadronic tagging at Belle*, submitted to Phys. Rev. D, 2015, arXiv: 1509.06521 [hep-ex], ↑ p. 97.

[124] F. U. Bernlochner, *The $B \to \pi\tau\nu$ decay in the context of the 2HDM type II*, 2015, arXiv: 1509.06938 [hep-ph], ↑ p. 98.

[125] K. KIRCHGESSNER, *Semileptonic Tag Side Reconstruction*, MA thesis, Karlsruher Institut
       für Technologie, 2012, URL: http://ekp-invenio.physik.uni-karlsruhe.de/record/
       48181/files/iekp-ka2012-12.pdf, ↑ p. 98.

[126] H. SCHREECK, *Search for the decay $B^0 \to \pi\tau\nu$ with semileptonic tag using data from the
       Belle experiment*, II.Physik-UniGö-MSc-2015/03, MA thesis, Georg-August Universität
       Göttingen, 2015, ↑ p. 98.

[127] A. ABDESSELAM et al. (The Belle and BaBar collaborations), *First Observation of CP
       Violation in $\overline{B}^0 \to D^{(*)}_{\mathrm{CP}} h^0$ Decays by a Combined Time-Dependent Analysis of BABAR
       and Belle Data*, Phys. Rev. Lett. 115.12 (2015), 121604, ↑ p. 99.

List of Figures

List of Tables

Acknowledgements

On these last pages, I would like to thank everyone who, in some way or other, helped and supported me during my time as a PhD student.

First of all, I would like to thank my supervisor Ariane Frey for the opportunity to start working in the field of particle physics. It was a really exciting and instructive time, helping to build a new working group for data analysis in the Belle collaboration. In this regard, I would also like to thank all my wonderful colleagues in the Belle group at Göttingen, Benjamin, Elisabeth, César, Uwe, Ulf, Harrison and Philipp. Many thanks also to Kevin and Matthias for the instructive and long discussions about different statistical issues. I also would express my gratitude to my physics group convenor Christoph Schwanda, who helped me a lot during the analysis process and always had time to answer my questions!
Furthermore, I would like Stan Lai who agreed to be the co-referee of this thesis, and Arnulf Quadt for agreeing to be a member of my thesis committee.
A very big thank you to Andrea and Matthias, who helped proofreading this thesis, this thesis would be a lot harder to read without you!
While particle physics is a very interesting topic, you can't do research 24/7. Therefore, many thanks goes to all my friends and colleagues for the activities outside of work, specially to Andi, Andrea, Bashi, Benne, Jens, Sarah, Steffi, Theresa, Matze, Matthias, Barti, Franzi, Fabian, Martina and Elisabeth and many more for all the wonderful evenings watching movies, having dinner, playing games or climbing.
My biggest thanks goes to my parents and my brothers Matthias and Simon. You were always there for me and supported me without question. My final thank you is to Linda, for your continuous support during this time, for having patience whenever there was another delay, and simply for being there for me. Thank you!