

**Measurement of the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$
decays at the Belle experiment**

Bastian Kronenbitter

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Phys. Bastian Kronenbitter
aus Karlsruhe

Tag der mündlichen Prüfung: 19. Dezember 2014
Referent: Prof. Dr. Michael Feindt
Korreferent: Prof. Dr. Günter Quast

Contents

1	Introduction	1
2	The decay $B^+ \rightarrow \tau^+ \nu_\tau$	3
2.1	Branching fraction	3
2.2	History of the measurement	7
3	The Belle experiment	9
3.1	The $\Upsilon(4S)$ resonance	9
3.2	The Belle detector	10
3.3	Data sample	17
3.4	Tool set	18
4	Methods	21
4.1	Semileptonic B reconstruction	21
4.2	Continuum suppression	28
5	Reconstruction	33
5.1	Full event reconstruction	33
5.2	Selection	34
5.3	Efficiency	39
6	Validation	41
6.1	Tag-side efficiency correction	41
6.2	E_{ECL} composition	45
6.3	Off-resonance data	48
6.4	Comparison of data and simulation in control samples	52
6.5	Charged track veto	56
7	Branching fraction extraction	59
7.1	Description of the fit procedure	59
7.2	Component description	60
7.3	Fit on simulated data	65
7.4	Stability test	68
7.5	Results on data	69
7.6	Systematic uncertainties	73

7.7	Significance of the result	77
8	A track-finding algorithm for the Belle II central drift chamber	79
8.1	The Belle II experiment	79
8.2	Track finding	81
8.3	Performance	90
8.4	Result	92
9	Conclusion	93
	Appendix	101
A	Data / MC comparisons	102
B	PDF shapes	114
C	Log of the data fit	120

1. Introduction

The Standard Model of particle physics is very successful in describing the known matter and its interactions. But we also know that it is not sufficient to explain all observed effects like dark matter, neutrino oscillation, or our sheer existence. There is currently no experimentally supported explanation why the visible universe exists in form of matter and there is no balance between matter and anti-matter. So, a tension exists in the current state of particle physics: there is a very successful model which delivered a large number of predictions and all of them which have been tested at accelerator experiments could not be falsified. But we know that it is not sufficient to describe everything we observe. Therefore, great effort has been put into finding physics beyond the Standard Model and still will be.

The Belle experiment is part of this effort and able to measure a large number of observables. Instead of reaching for high energies like the LHC experiments Atlas and CMS, it concentrates on measuring physical quantities with a very high precision, especially observables connected to the decays of B mesons. One of the primary goals of the Belle experiment was the validation of the CKM mechanism and the resulting time-dependent CP violation; another prediction of the Standard Model which was confirmed. The experimental setup of the KEKB accelerator and the Belle detector, located at the KEK research facility in Tsukuba, Japan, is not only particularly suitable for such time-dependent measurements. It also allows to test the Standard Model in various ways, including a lot of measurements which are only possible at experiments like Belle. The challenge is to test the Standard Model in the right places. A search for new physics is only promising if its potential influence is sizable enough to be measurable and the Standard Model contribution is not so large that a simple fluctuation overshadows the new physics contribution. The decay $B^+ \rightarrow \tau^+ \nu_\tau$ is a very good opportunity for such a search. Its branching fraction is often referred to as one of the current “golden modes” of particle physics. It can be calculated with very low uncertainties and the potential influence of commonly considered models of physics beyond the Standard Model can become large. It is an input to global fits of the CKM triangle as it has a direct relation to the CKM matrix element V_{ub} . And

it is experimentally accessible, although not easily. It requires the special experimental environment provided by an experiment like Belle and sophisticated methods.

The predicament lies within the fact that the presence of at least two neutrinos in the final state of the decay does not allow for a complete reconstruction of the B meson. Moreover, the decay products of the τ lepton do not deliver enough information to select the decay with a good signal to background ratio. Therefore the production mechanism of the B mesons at the Belle experiment, which are only produced in pairs of B mesons, is exploited. To be able to select the desired decay, the second B meson is reconstructed in a semileptonic decay channel. This method is often referred to as semileptonic tagging.

In this thesis, I describe the latest measurement of the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ decays. I performed it with the full Belle data set of 770 million $B\bar{B}$ pairs and it will be the last measurement of that kind until the start of the successor experiment Belle II. The previous Belle result using semileptonic tagging was performed with a data sample of 660 million $B\bar{B}$ pairs. As I not only wanted to profit from the slightly larger data set, I developed a completely distinct analysis. I employed new methods, carefully optimized the selection, and performed a thorough validation to reach a maximal sensitivity.

The thesis contains a discussion of the decay $B^+ \rightarrow \tau^+ \nu_\tau$ in Chapter 2, a brief description of the Belle experiment in Chapter 3, the introduction of the measurement-specific methods in Chapter 4, and the description of the measurement in the following chapters. The analysis consists of the reconstruction and selection of $B^+ \rightarrow \tau^+ \nu_\tau$ decays (Chapter 5), the validation of certain aspects of the measurement (Chapter 6), and the extraction of the number of reconstructed $B^+ \rightarrow \tau^+ \nu_\tau$ events (Chapter 7).

Unrelated to this measurement, a track-finding algorithm for the central drift chamber of the Belle II experiment is described. This algorithm deduces tracks of charged particles from the information provided by the drift chamber which consists of drift circles around the activated wires. A track is a circular arc tangential to all contributing drift circles and the task is to group the drift circles together to form a track candidate. As strong computational time constraints were imposed by the Belle II collaboration, the main challenge was the efficient implementation of such a method without sacrificing efficiency in finding tracks. I developed a track finder based on known formalisms, but was able to do this in an efficient manner and give a first, promising estimation of its performance. The development, consisting of the formalism, the employed methods, and the resulting performance are described in Chapter 8.

2. The decay $B^+ \rightarrow \tau^+ \nu_\tau$

The decays $B^+ \rightarrow \ell^+ \nu_\ell$, where ℓ^+ can be e^+ , μ^+ , or τ^+ , give a very good opportunity to test the Standard Model (SM), since the theoretical predictions only suffer from low uncertainties and the influence from commonly considered models of physics beyond the Standard Model can become relatively large.

2.1 Branching fraction

The tree-level Feynman graph of this decay is given in Figure 2.1.

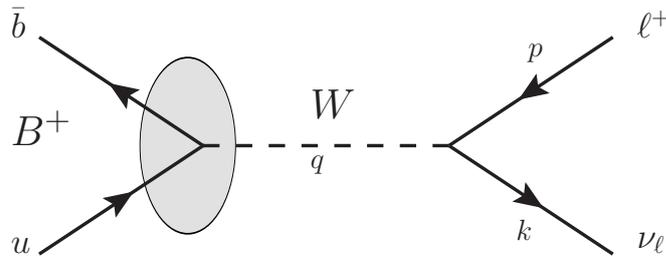


Figure 2.1: Feynman graph of the decay $B^+ \rightarrow \ell^+ \nu_\ell$ in the Standard Model.

The Feynman graph can be used to calculate the branching fraction of these decays. The matrix element is given by

$$\mathcal{M} = \sqrt{2}G_F F_{(bu)}^\mu \bar{u}(p)\gamma_\mu(1 - \gamma^5)v(k), \quad (2.1)$$

where γ^μ are the Dirac matrices and $\bar{u}(p)$ and $v(k)$ are the spinors of the particles indicated by the according momenta in Figure 2.1.

The term $F_{(bu)}^\mu$ represents the weak quark current, indicated by the ellipse in Figure 2.1. It cannot be represented by terms describing the scattering of two free quarks, since they

2. The decay $B^+ \rightarrow \tau^+ \nu_\tau$

are in a bound state. The calculation would result in a factor $|\Psi(0)|^2$, where Ψ is the wave function of the quarks in the B meson. Thus, the finite annihilation probability of the two quarks in the B meson needs to be taken into account. As the mediating interaction is the strong interaction at low energies, the wave function of the quarks in the B meson is not known. Therefore the factor $|\Psi(0)|^2$ cannot be determined analytically and is absorbed in a constant, denoted as decay constant f_B . Also the Cabibbo–Kobayashi–Maskawa (CKM) matrix element V_{ub} must be part of $F_{(bu)}^\mu$ and therefore it is written as

$$F_{(bu)}^\mu = \frac{1}{2} q^\mu f_B V_{ub},$$

with q being the overall momentum transfer in the decay.

The expression $\gamma_\mu(1 - \gamma^5)$ is related to the parity violation of the weak interaction. The term γ^μ alone would yield a vector coupling, whereas $\gamma^\mu\gamma^5$ gives an axial-vector coupling. The combination of both acts as a projection operator on the left-handed component and thus violates the conservation of parity maximally.

The Fermi coupling constant G_F is the effective coupling strength of the charged weak interaction. It is given by

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2},$$

where g is the actual coupling constant of the weak interaction, which is larger than the electromagnetic one. This description is only sufficient for values of q being significantly smaller than the mass of the W boson. Then the coupling can be reduced to a point-like four-fermion interaction. So Equation 2.1 can be illustrated by the effective Feynman graph shown in Figure 2.2.

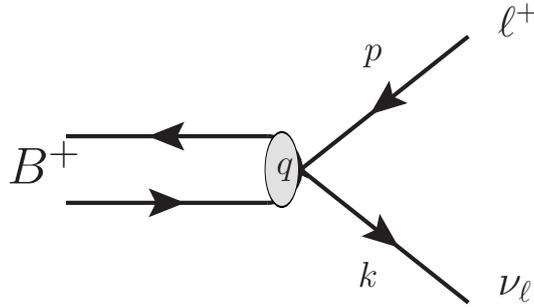


Figure 2.2: Effective Feynman graph of the decay $B^+ \rightarrow \ell^+ \nu_\ell$ in the Standard Model.

Using Equation 2.1 to calculate the branching fraction results in

$$\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell)_{\text{SM}} = \frac{G_F^2 m_B m_\ell^2}{8\pi} \left(1 - \frac{m_\ell^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B, \quad (2.2)$$

where m_B and m_ℓ are the masses of the B meson and the lepton respectively, and τ_B the lifetime of the B meson.

The branching fraction depends on the mass of the lepton weakly by the phase space factor $(1 - m_\tau^2/m_B^2)^2$ and strongly by the factor m_τ^2 because of the helicity suppression. The helicity suppression denotes the influence of the parity violating coupling of the W boson. The configuration of momentum and spin in the decay is illustrated in Figure 2.3. The neutrino is, in very good approximation, mass-less and therefore has left-handed helicity. Only the finite mass of the anti-lepton allows for the shown spin-momentum configuration as it has not a purely left-handed helicity and higher masses lead to a weaker suppression of the decay. However, it is not necessary to assume parity violation to explain the suppression of small lepton masses. Every combination of vector and axial-vector coupling leads to this, as shown in Reference [1, p. 208]. As a consequence of this suppression, the decay $B^+ \rightarrow \tau^+ \nu_\tau$ is expected to be the highest purely leptonic branching fraction of the B^+ meson.

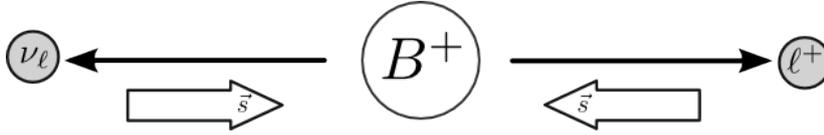


Figure 2.3: Illustration of the momentum and spin configuration in the decay $B^+ \rightarrow \ell^+ \nu_\ell$ in the center-of-mass system of the B^+ meson.

All inputs to Equation 2.2 can be measured directly, extracted from measurements, or, in case of f_B , calculated. Measuring the masses of the B meson and the leptons is an ongoing task and the relative error of all of them is already on the order of 10^{-3} . Also, the lifetime of the B meson has been measured with great precision. The Fermi decay constant is determined by measuring the lifetime of muons and is very well known. The B -meson decay constant can, up to now, only be calculated using the methods of lattice quantum-chromodynamics. The largest relative uncertainty comes from the CKM matrix element V_{ub} . It can be determined both by examining inclusive or exclusive semileptonic decay processes. The inclusive measurements reconstruct the decays $B \rightarrow X_u \ell \bar{\nu}$, where X_u denotes final states of $b \rightarrow u$ transitions. The exclusive measurements reconstruct this X_u system in specific decays like $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ or $B^0 \rightarrow \rho^- \ell^+ \nu_\ell$. The results of the two methods show only marginal agreement.

The input parameters and their current best known values are given in Table 2.1. Using these values, Equation 2.2 gives

$$\begin{aligned} \mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)_{\text{inc}} &= (1.20 \pm 0.13) \times 10^{-4} \\ \mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)_{\text{exc}} &= (0.66 \pm 0.12) \times 10^{-4}, \end{aligned}$$

where the error is dominated by the uncertainty on V_{ub} . The strong discrepancy illustrates the dependency of the calculated branching fraction on V_{ub} as input.

The CKMfitter group [2] performs a global unitarity-triangle fit and uses an average of the different V_{ub} measurements which is $V_{ub} = (3.70 \pm 0.12 \pm 0.26) \times 10^{-3}$ as input. The fit gives

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (0.753_{-0.052}^{+0.102}) \times 10^{-4}.$$

2. The decay $B^+ \rightarrow \tau^+ \nu_\tau$

To illustrate the impact of the helicity suppression also the expected branching fractions of other leptonic B^+ decays are given. It leads to significantly smaller branching fractions for the decays with lighter leptons:

$$\begin{aligned}\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu) &= (0.367_{-0.034}^{+0.026}) \times 10^{-6} \\ \mathcal{B}(B^+ \rightarrow e^+ \nu_e) &= (0.859_{-0.079}^{+0.062}) \times 10^{-11}.\end{aligned}$$

Parameter	Value	Relative uncertainty	
G_F	$1.17 \times 10^{-5} \text{ GeV}^{-2}$	5×10^{-7}	[3]
m_B	5.28 GeV	3×10^{-5}	[3]
m_τ	1.78 GeV	9×10^{-5}	[3]
f_B	186 MeV	2×10^{-2}	[4]
τ_B	$1.64 \times 10^{-12} \text{ s}$	2×10^{-3}	[3]
$ V_{ub} _{\text{inc}}$	4.41×10^{-3}	5×10^{-2}	[3]
$ V_{ub} _{\text{exc}}$	3.28×10^{-3}	9×10^{-2}	[3]

Table 2.1: Input values to the calculation of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$.

2.1.1 The influence of new physics

The presence of additional charged particles, which can replace the W Boson as propagator in the decay process, could interfere with the Standard Model weak decay process. A possible type of additional particles, which are not part of the Standard Model, are charged Higgs particles. The most general form of the two-Higgs-doublet model, known as type-III two-Higgs-doublet model, adds scalar terms to the branching fraction:

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = \mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)_{\text{SM}} \times \left| 1 + \frac{m_B^2}{m_\tau m_b} (S_R - S_L) \right|^2,$$

where m_b is the mass of the b quark and S_R and S_L are complex Wilson coefficients, that describe the contributions from new physics models [5]. The Standard Model corresponds to S_R and S_L being zero.

A more restrictive type of model is the type-II two-Higgs-doublet model. This has been studied extensively in the past, since it is the Higgs sector of the minimalistic supersymmetric Standard Model. The branching fraction becomes

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = \mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)_{\text{SM}} \times \left| 1 - \frac{\tan^2 \beta}{m_{H^\pm}^2} m_B^2 \right|^2,$$

where m_{H^\pm} is the mass of the charged Higgs particle and $\tan \beta$ a parameter related to the vacuum expectation values of the Higgs doublets. This relation can be used to constrain the parameters of this new-physics model. Figure 2.4 shows the excluded regions for the latest measurements of the Belle collaboration. The allowed band within the excluded

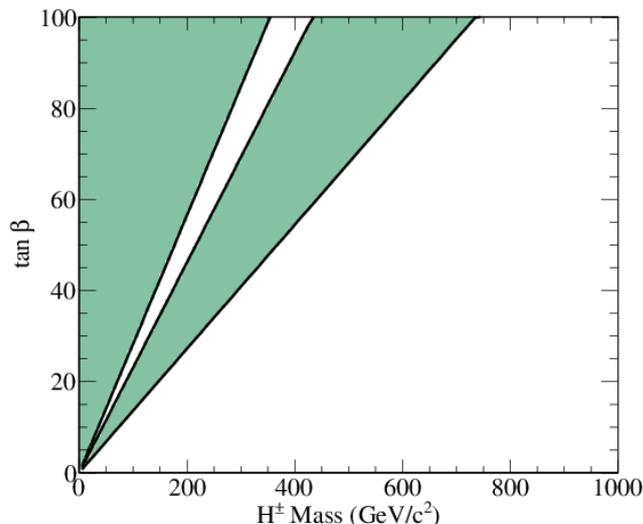


Figure 2.4: Constraint on $\tan \beta$ and $m_{H^\pm}^2$ in the type-II of two-Higgs-doublet model. The green regions indicate the excluded regions at a confidence level of 95%. Taken from Reference [6].

region is where the new-physics contribution is twice as large as the SM and therefore the term $\left|1 - \frac{\tan^2 \beta}{m_{H^\pm}^2} m_B^2\right|$ becomes one.

However, a recent measurement in the decay channel $B \rightarrow D^* \tau^+ \nu_\tau$ [7] is not compatible with a charged Higgs boson in this type of model. Assuming that this measurement is not the result of a statistical fluctuation or an underestimated theoretical or systematic uncertainty, the type-II two-Higgs-doublet model is falsified. Therefore only the more generic type-III two-Higgs-doublet model can be studied. Excluding regions in the parameter space in such a model is more complicated and their study is an active field of research [5].

2.2 History of the measurement

Several experiments have tried to measure the branching fraction of the decay $B^+ \rightarrow \tau^+ \nu_\tau$, beginning with ALEPH[8] and CLEO[9] in 1995. But only the B factories succeeded in finding evidence for this decay.

There are two methods used to search for the decay: semileptonic and hadronic tagging. Both the Belle and the BaBar collaboration used both methods in their measurements. In their latest measurements the BaBar collaboration used data samples containing $\approx 460 \times 10^6$ [10] and $\approx 470 \times 10^6$ $B\bar{B}$ pairs [11] and the Belle collaboration data samples containing $\approx 660 \times 10^6$ [12] and $\approx 770 \times 10^6$ $B\bar{B}$ pairs [13]. The results of the measurements are shown in Figure 2.5 and compared to expectation from the CKM fit.

Before the latest measurement at the Belle experiment, there was a sizable discrepancy between the theoretically determined value and the world average of the measurements, as all of them were above the expectation. But the result of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) =$

2. The decay $B^+ \rightarrow \tau^+ \nu_\tau$

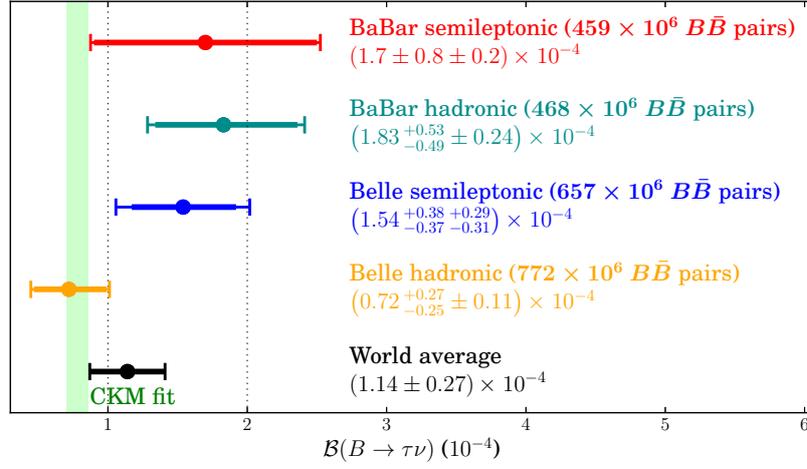


Figure 2.5: Comparison of the latest measurements of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$ of the B factories and the predicted values of the CKM fit of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (0.739^{+0.091}_{-0.071}) \times 10^{-4}$.

$(0.72^{+0.27}_{-0.25} \pm 0.11) \times 10^{-4}$ moved the world average to a value consistent with the expectation within two standard deviations. This discrepancy and the according shift when including the latest Belle measurement is illustrated in Figure 2.6.

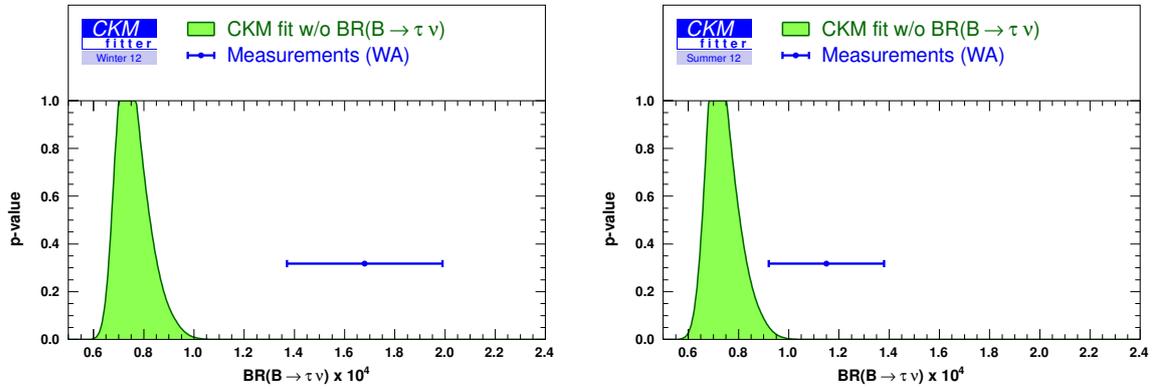


Figure 2.6: Comparison of the indirect fit prediction [2] of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$ with the average of the measurements. The left and right plots show the status before and after the latest Belle measurement [13], respectively.

3. The Belle experiment

The Belle experiment is part of the long-lasting and world-wide effort to measure and understand the system of particle physics. The Belle detector was the only detector located at the KEKB particle accelerator, which is often referred to as a B factory. This is motivated by the fact that the accelerator runs at a center-of-mass energy which corresponds to the invariant mass of the $\Upsilon(4S)$ resonance. This resonance decays with a probability of nearly 1 to a pair of B mesons. So operating at this center-of-mass energy leads to two characteristics: The cross-section for the production of B mesons is relatively large. Only at hadron accelerators larger samples of B mesons are produced. But at a B factory nothing is produced in addition to the B mesons, which is not true at higher energies. So the Belle experiment combines two advantages: A large data sample of B mesons is produced in a very clean environment with few particles in the final state of an event. Both of them are essential for the following measurement.

3.1 The $\Upsilon(4S)$ resonance

The Υ resonances are $b\bar{b}$ mesons with an angular momentum of one. The $\Upsilon(4S)$ is the lightest resonance of that kind with a mass high enough to decay into two B mesons. Lighter bottomonium mesons are not heavy enough to form mesons out of the b quarks. It therefore has a much larger width than the lower Υ resonances as it decays strongly without the necessity to mediate the color current by three gluons (denoted as OZI suppression). The first four Υ resonances can be seen in a measurement of the CUSB experiment in Figure 3.1. The production of other hadrons have a nearly flat cross section in the shown energy range. Therefore $e^+e^- \rightarrow q\bar{q}$ processes, where q can be a $u, d, s,$ or c quark, are often referred to as “continuum” processes. The cross section for the production of an $\Upsilon(4S)$ is 1.2 nb and for hadron production from continuum processes 2.8 nb. Other processes, e.g. Bhabha scattering, can have larger cross sections but are strongly suppressed in the course of the data taking and the reconstruction. The fraction of the mass of the $\Upsilon(4S)$ and its decay products is very near to one:

3. The Belle experiment

$2m_B/m_{\Upsilon(4S)} = 0.998$. This has an important consequence for the following measurement. The remaining energy in the decay is not enough to produce additional particles in the hadronization. Thus only two B mesons and nothing else is produced in an $\Upsilon(4S)$ event.

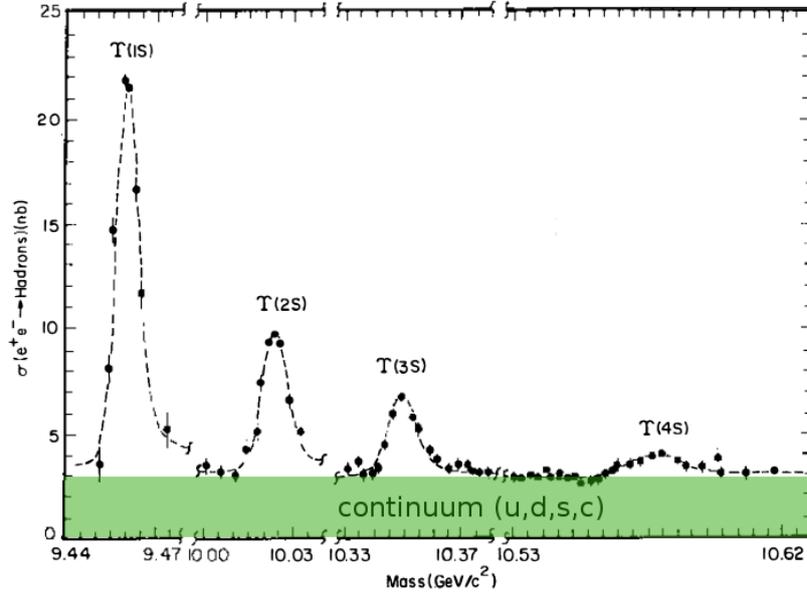


Figure 3.1: The cross section of $e^+e^- \rightarrow \text{hadrons}$ in the $\Upsilon(1S) - \Upsilon(4S)$ region. The width of the first three resonances is dominated by the measurement resolution. The green region indicates the production of $u, d, s,$ and c quarks. Taken from Reference [14].

3.2 The Belle detector

The Belle detector was built with the special requirements of a B factory in mind. It follows the typical structure of general-purpose detectors and consists of several sub-detectors, but a strong emphasis was put on the identification of particles, especially the separation of charged kaons and pions and the detection of neutral particles like photons.

In the following, a brief description of the most important components of the detector will be given. A more detailed description can be found in References [15, 16].

The purpose of certain sub-detectors can roughly be separated in two categories: the detection and measurement of particles and their identification. The measurement of neutral particles is achieved solely by the electromagnetic calorimeter (ECL), while the measurement of charged particles is performed by the silicon vertex detector (SVD) and the central drift chamber (CDC). The identification of particles is performed with information from the CDC, the ECL, and other detector components which are built especially for that purpose. These are the time-of-flight detector (TOF), the Aerogel Cherenkov counter system (ACC) and the K_L and muon system (KLM). Rarely, the KLM is also used for the detection of neutral, long-living kaons.

The sub-detectors mostly cover an angular region of 17° to 150° relative to the beam-axis, where the most forward and backward region often are referred to as end-cap regions and the innermost, cylindrical section as barrel region. An annotated side view of the Belle detector is shown in Figure 3.2.

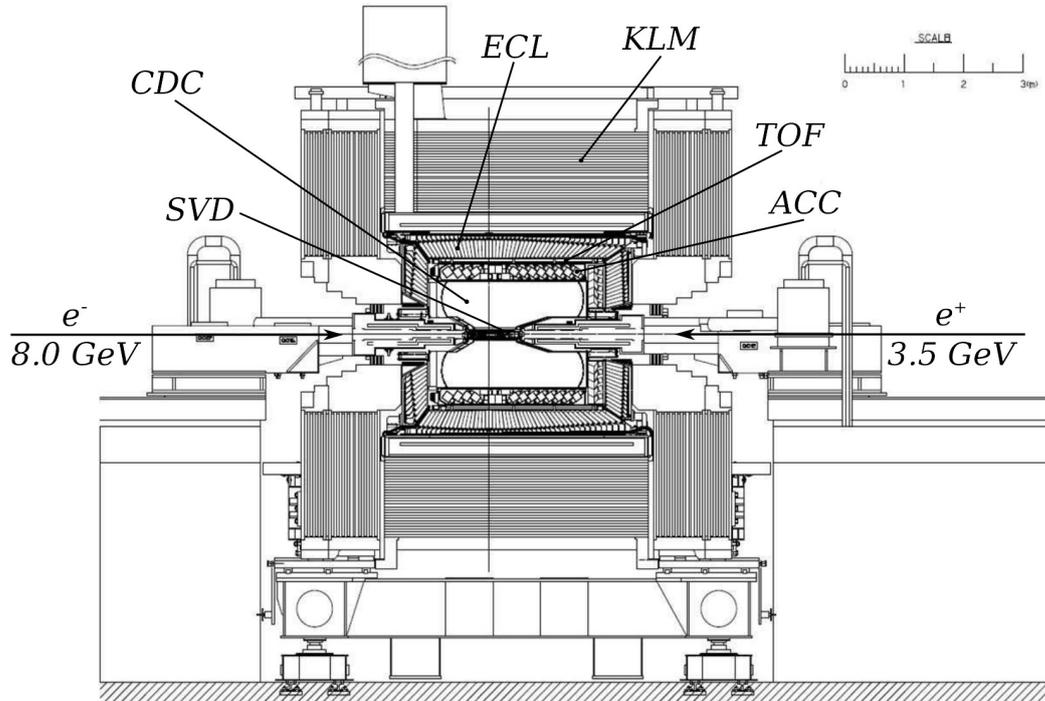


Figure 3.2: Side view of the Belle detector. Taken from Reference [16].

3.2.1 Silicon vertex detector

The SVD consisted of several layers of semi-conductor detectors. It measured the charge carriers which are freed when a charged particle passes through a depleted pn junction. The purpose of the SVD was the measurement of tracks of charged particles with a very good spatial resolution. This is important for the measurement of time-dependent CP violation, which requires the exact determination of the decay vertices of the B mesons. The SVD was the detector component nearest to the interaction region and therefore suffered from high radiation damage. This required a replacement of the whole SVD system in 2003 which was used to also upgrade it. The number of layers was increased from three to four and the inner radius was decreased from 30 to 20 mm. The upgrade significantly improved the detection capabilities of low-momentum tracks and the resolution of the spatial measurement of charged tracks.

3.2.2 The central drift chamber

In the drift chamber mainly the trajectories of the charged tracks were measured. But it also provided information for the particle identification via the measurement of the

3. The Belle experiment

energy loss of the particles. The trajectories of charged particles were determined by measuring the time and position of the ionization of the gas in CDC which occurred when a charged particle passed through the CDC.

The CDC consisted of 8400 drift cells in 50 cylindrical layers. The layers were organized in thirteen super-layers with some of them, called stereo-layers, having a small angle of 40 to 70 mrad to the beam-axis.

The charged tracks were reconstructed from the individual hits in the CDC using a pattern-recognition algorithm and the momentum was measured by determining the radius of the trajectories. This radius is directly related to the transverse momentum of the tracks by the nearly constant magnetic field which was produced by a solenoid magnet surrounding the CDC. The angle of the stereo layers allowed for a determination of the momentum component parallel to the beam-axis.

3.2.3 Electromagnetic calorimeter

The ECL was a key component in detecting and measuring electromagnetically interacting final state particles. It consisted of 8736 crystals pointing to the interaction region with a length of 30 cm, which corresponds to 16 radiation lengths. Electromagnetically interacting particles like photons or electrons caused multiple interaction within in the crystals, like pair production and bremsstrahlung. This lead to a clear signature in form of a shower spanning over multiple crystals. The ECL enabled the Belle detector to measure photons to very low energies and with good resolution. It also provided information which was crucial for the separation of electrons and hadrons.

3.2.4 Time-of-flight detector

The time-of-flight detector consisted of scintillator counters and measured the time between the collision and the arrival at the sub-detector. The time of the collision was given by a precisely synchronized reference clock. In combination with the momentum of the according particle this allowed for the determination of the mass. With a resolution of about 100 ps pion and kaon tracks with momenta up to 1.2 GeV/c could be separated.

3.2.5 Aerogel Cherenkov counter system

The Cherenkov counter was a detector component solely built for particle identification, especially for the separation of kaons and pions. This was done via detection of Cherenkov light, which is emitted when the charged particles propagate faster than the speed of light in the medium. As the detectable particles have a speed very near to the speed of light in vacuum, the used materials had refractive indices between 1.01 and 1.03. The separation between kaons and pions was achieved due to the choice of these parameters, as kaons have a higher mass, i.e. they are slower than pions of the same momentum. Therefore they emit no Cherenkov light in a range of about 1.2 to 3.5 GeV/c, while pions do.

3.2.6 Particle signatures

The above-mentioned sub-detectors all had a special purpose and only the combination of their output allowed for a meaningful interpretation. Since the particles of interest partially have very different properties, also their signature can differ. In Table 3.1 a list of the final state particles of interest for the following analysis is given together with their detectable traces in the detectors.

	Track in CDC and SVD	Shower in ECL	Cluster in KLM
π^\pm/K^\pm	✓	✓	
e^\pm	✓	✓	
μ^\pm	✓		✓
γ		✓	

Table 3.1: List of final state particles together with their detectable signature.

3.2.7 The trigger system

The trigger was a multi-level system. The lowest level, denoted as Level 1 trigger, worked with the direct output of the detector components and started the process of reconstructing an event. In the very beginning of the Belle project it was the only trigger system, but it soon turned out to be not sufficient to reach an applicable data amount. So other trigger levels (Level 3 and Level 4) were implemented to further reduce the amount of data. These higher levels were software components and used more sophisticated methods to separate the desired events from background processes.

The design frequency of beam collisions was 508 MHz, but in most of the cases, no interaction was taking place, or if e^+e^- scattering occurred, only Bhabha scattering, $\gamma\gamma$ or beam-background processes took place. These kind of events mostly were not of interest and were reduced to keep the data rate processable.

The rate of events which passed the highest level of the trigger was around 200 Hz or below in the beginning of the operation at Belle, but rose to 400 Hz and more. The trigger efficiency, hence the fraction of desired events which passed the trigger selection was near to 100 %. About half of the events passing the trigger were still beam-related processes, like interaction with the beam gas, but the obvious background was not treated further.

3.2.8 Event processing

To illustrate the process of the event reconstruction, the steps which are performed for each event independent of the following measurement are listed below:

- The processing of the event is started by a low-level trigger.
- The raw data is read out from each of the sub-detectors and processed so that it can be used by later steps of the event reconstruction.
- The detector calibration is applied. This step is necessary to compensate for the small misalignment of certain detector components and the pedestal voltages of the readout electronics. There is a large number of constants required for this (> 10.000) which are determined beforehand.
- The event time is obtained, i.e., information from the TOF and CDC is used to determine the actual bunch crossing which was recorded.
- Tracks originating from charged particles are reconstructed from hits in the CDC using a pattern recognition algorithm.
- The tracks are fitted to a helix to extract the physical information of the particle.
- The charged tracks are extrapolated to the inner part of the detector and hits in the SVD are matched to it.
- The charged tracks are extrapolated to the outer part of the detector to be matched with hits in the ECL, KLM, and the particle identification systems.
- Clusters in the ECL are formed from hits in single calorimeter crystals. Each cluster is tested if it can be matched to a charged track.
- The particle identification is applied for charged tracks.
- Pairs of charged particles which do not pass through the interaction region are formed to reconstruct long-living, neutral particles like K_S mesons or converted photons.
- The information is stored in form of a list of particles and their measured properties.

3.2.8.1 Illustration

The information delivered by the detector and the reconstruction algorithms can be visualized with an event display. An exemplary event is shown in Figures 3.3 and 3.4. In this event, four charged particles were reconstructed. Some of them can easily be identified by eye. Track 1 is an electron, depositing a significant amount of energy in the calorimeter. The trail in the KLM related to Track 2 identifies it as a muon. Track 3 and 4 do not show any strong signature in the ECL or KLM and are therefore identified as pions or kaons. Track 4 is a low momentum particle and therefore multiple curls of its helix-shaped track can be seen. Cluster 5 in the ECL which only can be seen in the side-view of the detector cannot be matched to a charged track and is therefore reconstructed as a photon.

3.3 Data sample

The Belle experiment recorded over 1 ab^{-1} of data, as shown in Figure 3.5, until the stop of data taking in 2010. Most of the data were taken at the $\Upsilon(4S)$ resonance, but there are substantial data samples recorded at other energies. The integrated luminosity and the corresponding number of produced resonances is given in Table 3.2. The off-resonance sample was recorded 60 MeV below the $\Upsilon(4S)$ resonance and allows for the investigation of continuum background without a contribution from B mesons. Over its whole lifetime the Belle experiment recorded a data sample containing $(722 \pm 11) \times 10^6 B\bar{B}$ pairs from $\Upsilon(4S)$ decays, where equal numbers of B^+B^- and $B^0\bar{B}^0$ pairs were produced.

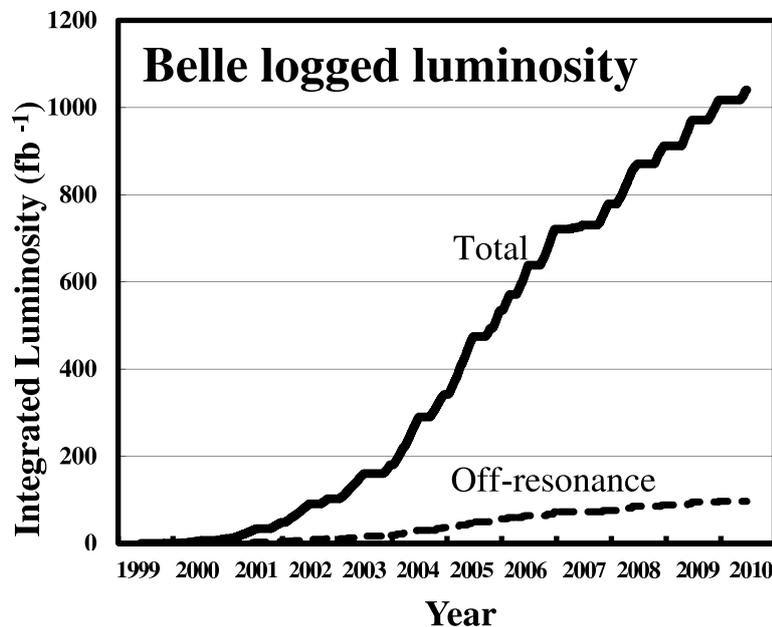


Figure 3.5: Integrated luminosity recorded by Belle. Taken from Reference [16].

Resonance	Luminosity (fb^{-1})	Produced resonances
$\Upsilon(1S)$	6	102×10^6
$\Upsilon(2S)$	25	158×10^6
$\Upsilon(3S)$	3	11×10^6
$\Upsilon(4S)$	711	772×10^6
$\Upsilon(5S)$	121	$7 \times 10^6 (B_s\bar{B}_s)$
$\Upsilon(4S)$ off-resonance	89	
Scan	28	

Table 3.2: Integrated luminosity recorded by Belle, broken down by center-of-mass energy.

3.4 Tool set

The following tools and procedures have been developed by the Belle collaboration and are frequently employed in measurements.

3.4.1 Particle identification

The measurement of charged tracks in the CDC and SVD only delivers information about the three-momentum of a particle. The mass or energy cannot be deduced from the trajectory. Therefore additional information needs to be used to identify charged particles. The number of potential particles leaving a measurable signature in the tracking detectors is limited. Only charged pions, kaons, muons, electrons, and protons are long-lived enough to reach the tracking detectors, so the particle identification algorithm concentrates on separating these particles.

The following information is combined with likelihood ratios where the identification of different particles uses different sub-sets of information:

- The relative energy loss per distance in the CDC, denoted as dE/dx .
- The time-of-flight, measured in the TOF.
- The number of photons which are produced in a Cherenkov shower and measured in the ACC.
- Information about the shower in the ECL which could be matched to the charged track.
- Information about the hits in the KLM which could be matched to the charged track.

The likelihood ratios are formed depending on the particle types to be distinguished. The likelihood ratios are formed with

$$\text{PID}_{\alpha\text{vs.}\beta} = \frac{\prod_i L_{\alpha}^i}{\prod_i L_{\alpha}^i + \prod_i L_{\beta}^i},$$

where L_{α}^i is the likelihood for the particle α from the sub-detector i . The likelihood curves are extracted from simulated data and corrected according to control-sample studies.

The following three separations are required in the course of the following measurement:

Pions vs. kaons : Because of their different masses, these two particles show different behavior in certain sub-detectors for the same value of momentum. For the separation, the information from the CDC, TOF, and ACC are combined. Each of the sub-detectors is especially powerful in separating the particles in a certain range of momentum. The lower momentum range of 0 to 1.2 GeV/c is covered by the CDC and TOF, as the relative resolution of the time-of-flight measurement is small enough and the energy loss differs significantly at this momentum. The configuration of the ACC was chosen so that it separated pions and kaons with a momentum of 1.2 to 3.5 GeV/c.

Electrons vs. rest : The electron identification includes information from the ECL, CDC, and ACC. To include the information from the calorimeter, the charged track it is matched to a shower in the ECL. To parametrize the characteristics of a shower it is described by three variables: the spatial consistency of the shower and the track, the energy of the shower and the shower shape. The shape of an electromagnetic shower induced by an electron differs significantly from a hadronic shower induced by pions or other hadrons. It is described with the ratio of the energy measured in the 3×3 and 5×5 crystals surrounding the peak of the energy deposition. The ratio of energy and the momentum is directly related to its rest mass. Although the resolution of the energy measurement is not good enough to measure the mass directly, it adds important information to the particle identification.

Muons vs. rest : The muon identification solely is based on information from the KLM. The only charged particles reaching the KLM are muons. Therefore it is only tested if the extrapolated trajectory of the charged track is near a hit in the KLM. The likelihood curve is formed from the distance to the nearest hit.

A detailed description of the particle identification can be found in Reference [17].

3.4.2 Simulation

The following analysis was developed while extensively employing simulated data. Multiple software packages are used in the course of the simulation. The decays of all particles were simulated using EvtGen [18]. In this step, the momenta and decay vertices of the primary particles are determined. This includes all intermediate particles in the decay chains and the final state particles, but not particles arising from the interaction with the detector. This interaction and the response of the detector is simulated using the software package GEANT3 [19].

The simulation only covers the interaction originating from the primary particle and its decay products. Other effects are not simulated. Therefore the so-called beam background is added to the simulated events. This denotes background originating from interaction of the electron and positron beam with itself or the beam gas. As this cannot be simulated accurately, a special data set is recorded at random points of time, where most probably no e^+e^- interaction is taking place and only beam background is occurring. One event from this randomly triggered data set is added to each simulated event.

Multiple samples of simulated data were prepared by the Belle collaboration. They often are referred to as Monte Carlo (MC) data.

Generic MC: For background studies, samples of $b \rightarrow c$ processes were prepared. They contain all known $b \rightarrow c$ transitions which are produced according to their branching fractions. Depending on the charge of the decaying B mesons, the Monte Carlo samples either are denoted as charged (B^+B^-) or mixed ($B^0\bar{B}^0$) generic MC. There are ten data samples available, each of them corresponding to the total integrated luminosity of the Belle data set.

3. The Belle experiment

Continuum MC: The background contribution from $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) is also simulated. The simulation needs to describe the hadronization of the quarks. This is performed by the software packages PYTHIA and JETSET [20].

$b \rightarrow u\ell\nu_\ell$ MC: The $b \rightarrow u\ell\nu_\ell$ MC contains $B \rightarrow X_u\ell\nu_\ell$ decays, where the X_u system is the product of a $b \rightarrow u$ transition, e.g. a pion or a ρ meson. The available data sample corresponds to 20 times the total integrated luminosity of the Belle data set.

Rare MC: The rare MC contains all B decays not covered by the two types of MC above. This includes semileptonic decays like $B^+ \rightarrow K^+\tau^+\tau^-$ but also hadronic decays like $B^+ \rightarrow K^+\pi^0$ or decays with a photon in the final state as $B^+ \rightarrow K^{*+}\gamma$. These decays typically have a very small branching fraction lower than 10^{-4} . The available data sample corresponds to 50 times the total integrated luminosity of the Belle data set. In principle, also the signal decay $B^+ \rightarrow \tau^+\nu_\tau$ is part of this sample, but it is removed as the signal contribution is treated separately.

Signal MC: The signal decay $B^+ \rightarrow \tau^+\nu_\tau$ was simulated 10 million times. This large number is necessary as different samples are required for different steps of the analysis and a signal sample must not be used twice to avoid the influence of fluctuations in a single data sample.

4. Methods

Many of the methods used in this measurement have been used before. Generic techniques like the identification of charged particles, the determination of decay vertices, or the simulation of production and decay of particles, are the product of collaboration-wide efforts and are used frequently in different kinds of measurements. The following methods, though not developed especially for this measurement, have been studied and modified to take into account the peculiarities of the studied process. The fact that this is the analysis of a rare decay with a very specific detectable signature makes it necessary to adjust the following techniques. Or the first usage and the resulting validation is an integral part of the improvement of this measurement over previous ones.

4.1 Semileptonic B reconstruction

The detectable signature of a $B^+ \rightarrow \tau^+ \nu_\tau$ decay is often only one charged track. To reconstruct a single charged track from a certain decay and distinguish it from tracks originating from other processes is experimentally challenging, especially if this track is the product of a multi-body decay with no particular momentum or other properties. So to be able to identify the desired decay additional information is necessary. This information can be gained by reconstructing the other B meson in the event, often referred to as tag B meson or B_{tag} , which is produced accompanying to the signal B meson in the decay of the $\Upsilon(4S)$ resonance.

The combination of both B mesons in an event is identical to the reconstruction of the $\Upsilon(4S)$. That's why this technique is often called *full reconstruction* or *full event interpretation*.

In this thesis, a B_{tag} reconstruction algorithm is used, which is similar to the already existing hadronic full reconstruction algorithm [21]. This was used in a variety of analysis (e.g. References [13, 22, 23]). The newly developed tagging algorithm reconstructs the

semileptonic decay channels

$$B^0 \rightarrow D^{(*)-} \ell^+ \nu_\ell \quad \text{and}$$

$$B^+ \rightarrow \bar{D}^{(*)0} \ell^+ \nu_\ell,$$

where ℓ can be either an electron or a muon and $D^{(*)}$ is either a D meson or the excited D^* meson. These decay channels have a very large branching fraction 0.16 in sum, but the fact that the neutrino cannot be reconstructed reduces the number of possible kinematic constraints. For this analysis, only the reconstruction of charged B mesons is used.

The semileptonic B reconstruction follows some basic principles: A high number of decay channels are reconstructed; the selection is performed as loose as possible taking into account the limited computing resources; the selection is performed using the multivariate selection package NeuroBayes, following the same hierarchical structure as the reconstruction, illustrated in Figure 4.1.

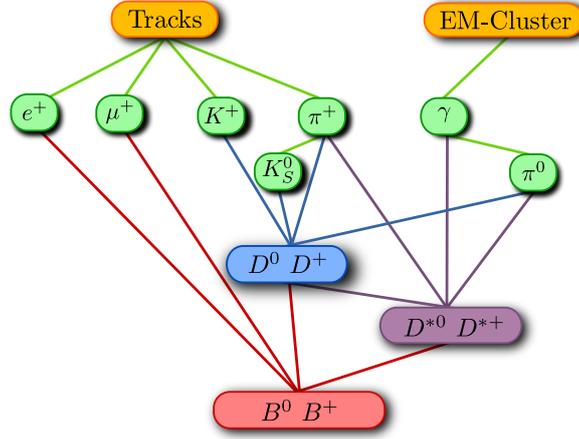


Figure 4.1: Illustration of the hierarchical structure of the semileptonic B meson reconstruction. Taken from Reference [24].

The whole algorithm, including a description of the reconstructed decay channels and the variables used for the training of the multivariate selection is described in great detail in Reference [24].

4.1.1 Selection and decay channels

The primary way of selecting intermediate particles and B -meson candidates is a cut on the product of the output of the multivariate discriminator (\mathcal{N}) of the daughter particles. It is applied to reduce the amount of reconstructed candidates to a degree, where the required computing time is reasonable. It is chosen in such a way that a maximum of 10 % of the signal is lost, but the signal to background ratio is increased by at least a factor of ≈ 20 .

Other selection is also applied as listed below. Only cuts which have a signal efficiency of nearly one are applied in addition to the multivariate selection.

4.1.1.1 Final state particles

The following particles have a relatively long lifetime and are thus considered stable in the frame of experimental particle physics. They can be directly detected by the Belle detector and are often referred to as final state particles.

Photons: Photons which are used to form π^0 and D^* candidates are required to have an energy larger than 30 MeV and 50 MeV, respectively.

Charged particles: Tracks which are measured in the CDC are produced by long-living charged particles. They are required to have a distance to the interaction point along and perpendicular to the beam direction smaller than 2 cm and 4 cm, respectively. Tracks which are assumed to be kaons must be identified as such by the particle identification. This is only required for kaons since they suffer from the large background from charged pions.

4.1.1.2 Intermediate particles

π^0 : Neutral pions are reconstructed as $\pi^0 \rightarrow \gamma\gamma$. The spatial resolution of photons is very bad compared to charged tracks due to the limited spatial measurement capabilities of the ECL. To compensate for this, the tracks of the two photons are fitted to originate from a common vertex under the constraint of the known invariant mass of the π^0 . The invariant mass of the two-photon system before this fit, $M_{\gamma\gamma}$, must satisfy $|M_{\gamma\gamma} - m_{\pi^0}| < 19 \text{ MeV}/c^2$, where m_{π^0} is the invariant mass of the π^0 [3].

K_S^0 : K_S^0 candidates are reconstructed as $K_S^0 \rightarrow \pi^+\pi^-$. The K_S candidates must lie within 30 MeV of the nominal K_S mass [3].

4.1.1.3 D mesons

The D^0 and D^{0*} mesons are reconstructed in different decay channels, which are listed in Table 4.1 and 4.2, respectively.

The selection criteria again are very soft to maximize the potential reconstruction efficiency. The momentum of the D^0 meson in the $\Upsilon(4S)$ center-of-mass system (CMS) is required to be smaller than $2.6 \text{ GeV}/c$. This suppresses background from $e^+e^- \rightarrow c\bar{c}$ events, where the D mesons tend to have larger momenta. Further, the invariant mass of the D^0 -meson candidate must lie within 40 or $60 \text{ GeV}/c^2$ of the mass of the D meson, depending on the decay channel.

To increase the resolution of the D^{*0} -mass measurement, not the mass itself but the difference of the mass of the D^{*0} and the D^0 candidate is considered. This mass difference is denoted as ΔM . It is required to satisfy $132 \text{ MeV}/c^2 < \Delta M < 152 \text{ MeV}/c^2$.

4.1.1.4 B mesons

Finally, B mesons are reconstructed in the decay channels listed in Table 4.3. No further selection is applied to the B mesons.

4. Methods

Decay channel	Branching fraction (%)
$D^0 \rightarrow K^- \pi^+ \pi^0$	13.9 ± 0.5
$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$	8.1 ± 0.2
$D^0 \rightarrow K_S^0 \pi^+ \pi^- \pi^0$	5.2 ± 0.6
$D^0 \rightarrow K^- \pi^+$	3.9 ± 0.1
$D^0 \rightarrow K_S^0 \pi^+ \pi^-$	2.8 ± 0.2
$D^0 \rightarrow \pi^+ \pi^- \pi^0$	1.4 ± 0.1
$D^0 \rightarrow K_S^0 \pi^0$	1.2 ± 0.1
$D^0 \rightarrow K_S^0 K^+ K^-$	0.5 ± 0.03
$D^0 \rightarrow K^+ K^-$	0.4 ± 0.01
$D^0 \rightarrow \pi^+ \pi^-$	0.1 ± 0.003
Sum	37.5 ± 0.85

Table 4.1: List of reconstructed D^0 -decay channels and their branching fractions.

Decay channel	Branching fraction (%)
$D^{*0} \rightarrow D^0 \pi^0$	61.9 ± 2.9
$D^{*0} \rightarrow D^0 \gamma$	38.1 ± 2.9
Sum	100 ± 0

Table 4.2: List of reconstructed D^{*0} -decay channels and their branching fractions.

Decay channel	Branching fraction (%)
$B^+ \rightarrow D^{*0} \mu^+ \nu_\mu$	5.70 ± 0.19
$B^+ \rightarrow D^{*0} e^+ \nu_e$	5.70 ± 0.19
$B^+ \rightarrow D^0 \mu^+ \nu_\mu$	2.26 ± 0.11
$B^+ \rightarrow D^0 e^+ \nu_e$	2.26 ± 0.11
Sum	15.92 ± 0.44

Table 4.3: List of reconstructed B^+ -decay channels and their branching fractions.

4.1.2 Training of multivariate discriminator

The training of the multivariate discriminators (MVD) is performed using a simulated data sample of 150×10^6 $B\bar{B}$ pairs, which decay generically, thus in all known $b \rightarrow c$ decay channels.

The variables used as input are chosen to be uncorrelated to the angle between the reconstructed B -meson candidate and the $D\ell$ system, $\cos \theta_{B,D^{(*)}\ell}$. In this way, it can be used for additional selection or for the extraction of the number of correctly reconstructed B mesons at a later stage.

The trainings of the MVDs related to final state particles, D mesons, and D^* mesons are the same as in the existing hadronic tagging algorithm. Nearly all available information is used in the training. Commonly used variables are: the reconstructed energy, mass, and momentum of the particle candidates; the masses and outputs of the MVDs of daughter particles; and spacial information like the distance to the interaction point. Additionally, particle-type specific information like particle identification information and the output of certain sub-detectors is used.

The variables used in the training of the MVDs related to the tag B mesons are as follows:

$\mathcal{N}_{\text{children}}$: The output of the MVDs of the daughter particles.

$M_{D^{(*)}}$: The reconstructed mass of the $D^{(*)}$ meson.

D -meson decay channel: A discrete identifier for the decay channel of the D meson.

$\cos \theta_B$: The angle of the reconstructed B -meson candidate with respect to the beam in the CMS.

Pseudo helicity angle of the $D^{(*)}$ meson: The cosine of the angle between the $D^{(*)}$ meson boosted in the direction of the B meson and the B meson in the CMS.

$D\ell$ distance: The spatial distance between the reconstructed D -meson vertex and the helix of the lepton.

$D\ell$ angle: The angle between the $D^{(*)}$ meson and the lepton in the CMS.

In the case where the B meson decays to an excited D meson, the following variables are used also:

D^* -meson decay channel: A discrete identifier for the decay channel of the D^* meson.

ΔM : The difference between the reconstructed masses of the D^* - and the D -meson candidate.

4.1.3 Reconstructed B mesons

To illustrate the performance of the tagging algorithm, an exemplary sample of reconstructed events is shown. Throughout the following analysis, the B_{tag} is never reconstructed alone, rather the event is reconstructed as a whole. Therefore the output of the tagging algorithm is shown for a sample with a preselection motivated by the analysis. Namely, the

presence of no additional charged tracks in the event, no additional π^0 candidates, and a maximal amount of additional energy deposited in the calorimeter. For such a sample of simulated data, which consists of a B_{tag} and charged track, consistent with the hypothesis to be a muon, the distribution of $\cos \theta_{B,D^{(*)}\ell}$ is shown in Figure 4.2. The variable $\cos \theta_{B,D^{(*)}\ell}$ is the angle between the momentum of the B meson and the $D^{(*)}\ell$ system, calculated under the assumption, that only one massless particle is not reconstructed. It is given by

$$\cos \theta_{B,D^{(*)}\ell} = \frac{2E_{\text{beam}}E_{D^{(*)}\ell} - M_B^2 - M_{D^{(*)}\ell}^2}{2p_B^*p_{D^{(*)}\ell}^*}, \quad (4.1)$$

where E_{beam} is the energy of the beam in the CMS, $E_{D^{(*)}\ell}$, $M_{D^{(*)}\ell}^2$ and $p_{D^{(*)}\ell}^*$ are the energy, mass and momentum of the $D^{(*)}\ell$ system in the CMS, respectively, M_B is the nominal B -meson mass [3], and p_B^* is the nominal B -meson momentum in the CMS, calculated from the beam energy and the nominal mass.

For correctly reconstructed B mesons, $\cos \theta_{B,D^{(*)}\ell}$ has values between -1 and 1 , as is the case for the majority of the green component in Figure 4.2. Signal B mesons, which are partially correctly reconstructed (e.g. B mesons where the γ or π^0 from the D^{*0} decay is not reconstructed) still peak around zero but are more wide distributed. The background is distributed over a much wider region, especially towards high values of $\cos \theta_{B,D^{(*)}\ell}$. The peaking structure in the background comes from the selection on $p_{D_{\text{cand.}}}^*$, which is correlated to $\cos \theta_{B,D^{(*)}\ell}$.

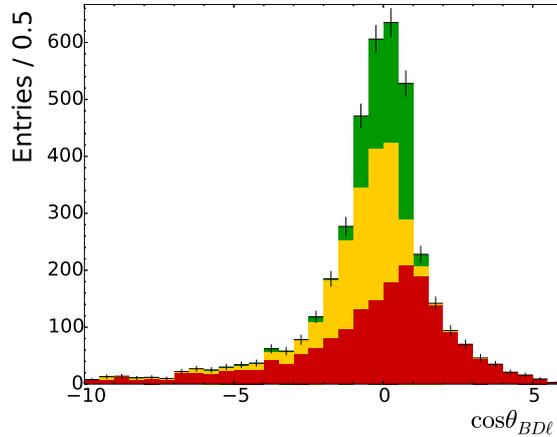


Figure 4.2: Distribution of $\cos \theta_{B,D^{(*)}\ell}$ for a sample with a preselection motivated by the following analysis. The red component shows the background, the green component the signal, and the orange component the partially correct reconstructed B mesons.

The output of the MVD algorithm is shown in Figure 4.3. To illustrate the composition of the distribution, it is also plotted separately for $B^+ \rightarrow \bar{D}^0 \ell \nu_\ell$ and $B^+ \rightarrow \bar{D}^{*0} \ell \nu_\ell$ decays. The loss of signal efficiency (including partially correctly reconstructed signal) and the background suppression is shown in Figure 4.4. It is clearly viable, that a selection on \mathcal{N}_{tag} reduces the background stronger than the signal. Choosing the optimal requirement in the context of this measurement will be described later.

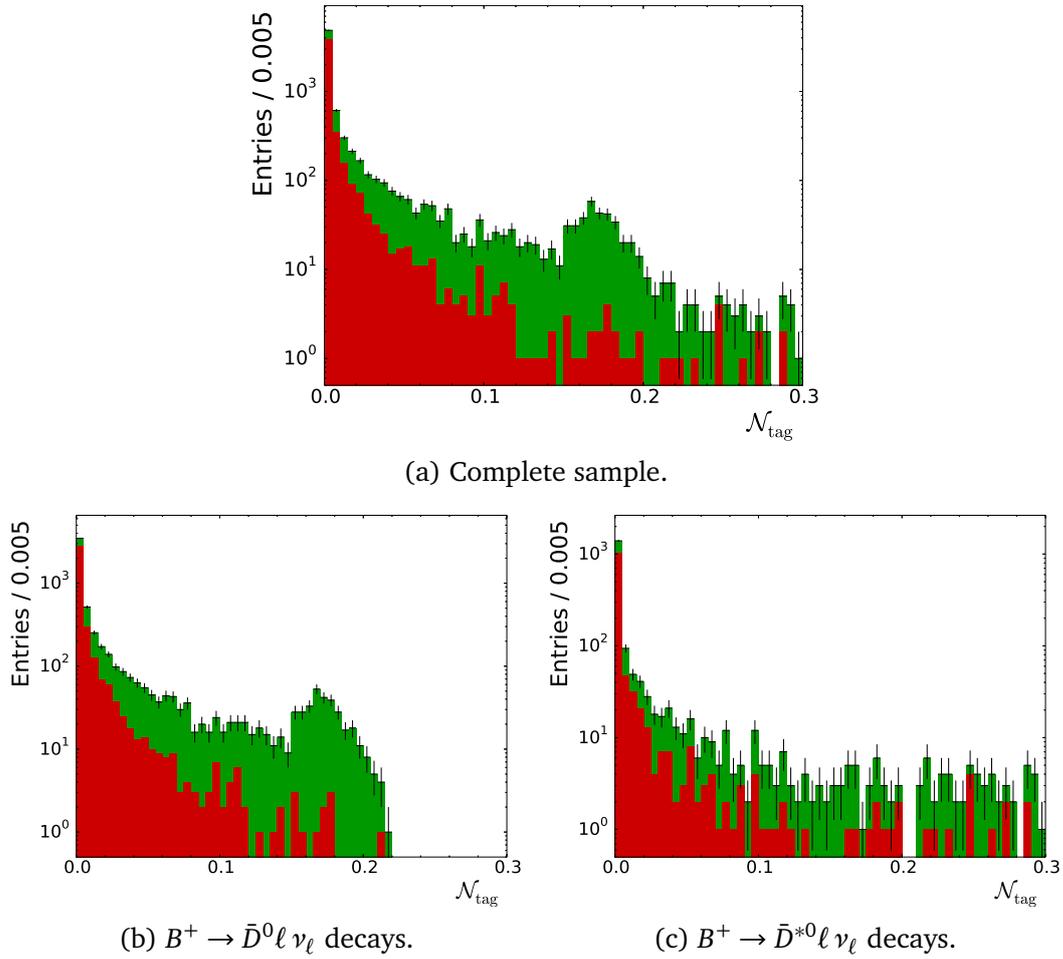


Figure 4.3: Distribution of the output of the MVD for different samples. The red component shows the background and the green component the signal including partially correctly reconstructed signal.

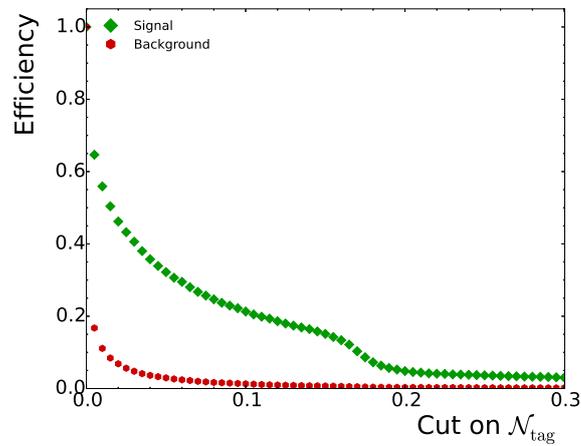


Figure 4.4: Signal and background efficiency for a wide range of cuts on \mathcal{N}_{tag} .

4.2 Continuum suppression

The background arising from $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) events is often referred to as continuum. Its characteristics, like the spatial distribution of the final state particles, differ significantly from events containing B mesons. On the other hand, its theoretical description and therefore also the simulation of such events suffer from high theoretical uncertainties related to the fragmentation processes. They are difficult to describe in the frame of quantum chromo-dynamics. So the suppression of such background is both possible and necessary.

Therefore, an algorithm, which suppresses continuum background based on a hierarchical system of multivariate selection algorithms, is implemented. This kind of algorithm was already used in similar form in other analysis [25, 26].

In events, where two B mesons are produced, they are nearly at rest in the CMS. As the decays of B mesons have no preferred spacial distribution, they decay nearly isotropically in the CMS. In contrast to this, lighter quarks have a significant momentum in the CMS and therefore the hadronization products fly back-to-back and tend to form jet-like structures. So several variables, describing the shape of the event, are calculated and combined with multivariate selection methods, based on the NeuroBayes algorithm [27]. For some of the variables, the information originating from one reconstructed B meson is used. Since more information is available from the B_{tag} meson, this is used. In contrast to the case, where one B meson is fully reconstructed, here some kinematic information is missing, due to the presence of the neutrinos. Therefore some of the variables have less separation power than in other analysis, which use a similar approach.

The following variables are used:

$\cos \theta_B$: The angle θ_B is the angle between the z axis (given by the beam direction) and the flight direction of the B meson. The spin $1 \rightarrow 0$ decay $\Upsilon(4S) \rightarrow B^+B^-$ leads to a $\sin^2 \theta_B$ distribution, where this is flat for background, which just consists of randomly combined tracks.

Δz : The decay vertex position difference in z direction is determined, using kinematic fits to extract the individual decay vertices of the two B mesons. All tracks, originating from the B_{tag} meson are fitted to a common vertex, taking into account the fact that a B meson most likely decays in a tube around the interaction region.

All remaining tracks again are fitted to a common vertex, where most of the time only the signal tracks contribute to the fit, so again the constraint related to the tube around the interaction region is used. The value of Δz then is given by $z_{\text{sig}} - z_{\text{tag}}$.

Since the distribution of correctly reconstructed B mesons follows the exponential decay curve with the significant B -meson lifetime, it is different from the background distribution, which tends to peak at zero.

Second Fox-Wolfram moment R_2 : The Fox-Wolfram moments are variables, which take into account all charged tracks in the event and describe the shape of the event using Legendre polynomials. So in principle, there is an infinite number of Fox-Wolfram moments,

but in the past, the second moment proved to be especially good in separating B from continuum events. It is calculated with

$$H_k = \sum_{i,j}^N \frac{|\vec{p}_i||\vec{p}_j|P_k(\cos \theta_{ij})}{E_{\text{vis}}^2}$$

$$R_k = \frac{H_k}{H_0},$$

where N is the number of charged tracks in the event, \vec{p}_i is the three-momentum of the i th charged particle, P_k is the k th Legendre polynomial, θ_{ij} the angle between the momenta of the i th and j th particle, and E_{vis} the sum of all measured energy in the event.

Thrust Axis Angle: The thrust axis \vec{T} is the axis, which maximizes the expression

$$T = \frac{\sum_i |\vec{p}_i \vec{T}|}{\sum_i |\vec{p}_i|}.$$

The index i runs over one of the following sets of particles: those used in the reconstruction of the B_{tag} meson; the particles which are not used; or all particles in the event.

For $q\bar{q}$ continuum events, this axis more likely points into the direction of the beam, due to the jet-like structure of such events. In contrast, B mesons decay nearly isotropically, so the thrust axis is distributed more uniformly. To separate $B\bar{B}$ events from $q\bar{q}$ continuum events, both the angle between the thrust axes of the reconstructed B meson and the rest of the event and angle between the B -meson thrust axis and the z axis are calculated.

Super Fox-Wolfram moments K_i : The Super Fox-Wolfram moments are a modification of the normal Fox-Wolfram moments. Instead of summing over all detected particles, a combination of final-state particles of the reconstructed B meson (s) and the remaining particles in event (o) is formed. Three possible combinations exist: R_k^{oo} , R_k^{so} , and R_k^{ss} . Several of the theoretically infinite number of moments are picked to deliver the best signal and background separation and not to influence other variables, which are important in later stages of the analysis.

CLEO Cones: The CLEO Collaboration introduced the CLEO cones variables to suppress background from $e^+e^- \rightarrow q\bar{q}$ events in charmless hadronic B decays [28], very similar to this analysis. The thrust axis of the B_{tag} meson is calculated and the space around is divided in nine polar angle intervals of 10° each. For each of them, the scalar momentum flow through the cone of all particles is calculated.

4.2.1 Combination

The CLEO cones, the thrust axis angle, and the modified Fox-Wolfram moments are separately combined with a multivariate selection algorithm. Afterwards they are combined together with R_2 , $\cos \theta_B$, and Δz to a single separation variable, further denoted with \mathcal{N}_{CS} . The training of the selection algorithm is performed using a sample of simulated $B^+ \rightarrow \tau^+ \nu_\tau$ decays and a randomly decaying second B meson as signal and continuum events as background. Both these data samples were processed according to the following analysis. The comparison of the resulting performance on a sample with arbitrary signal to background ratio can be seen in Figure 4.5. The combination of the CLEO cones and the modified Fox-Wolfram moments gives a better separation than the individual methods, and adding $\cos \theta_B$ and R_2 also improves the result. However, adding Δz only results in a negligible improvement. This is not that unexpected, since both on the tag and the signal side substantial information is missing due to the presence of neutrinos in the decay.

The combination without the Δz information is chosen as the selection variable to avoid possible additional systematic uncertainties.

In Figure 4.6, the distribution both of the signal and the background contribution is shown, which illustrates the achieved separation.

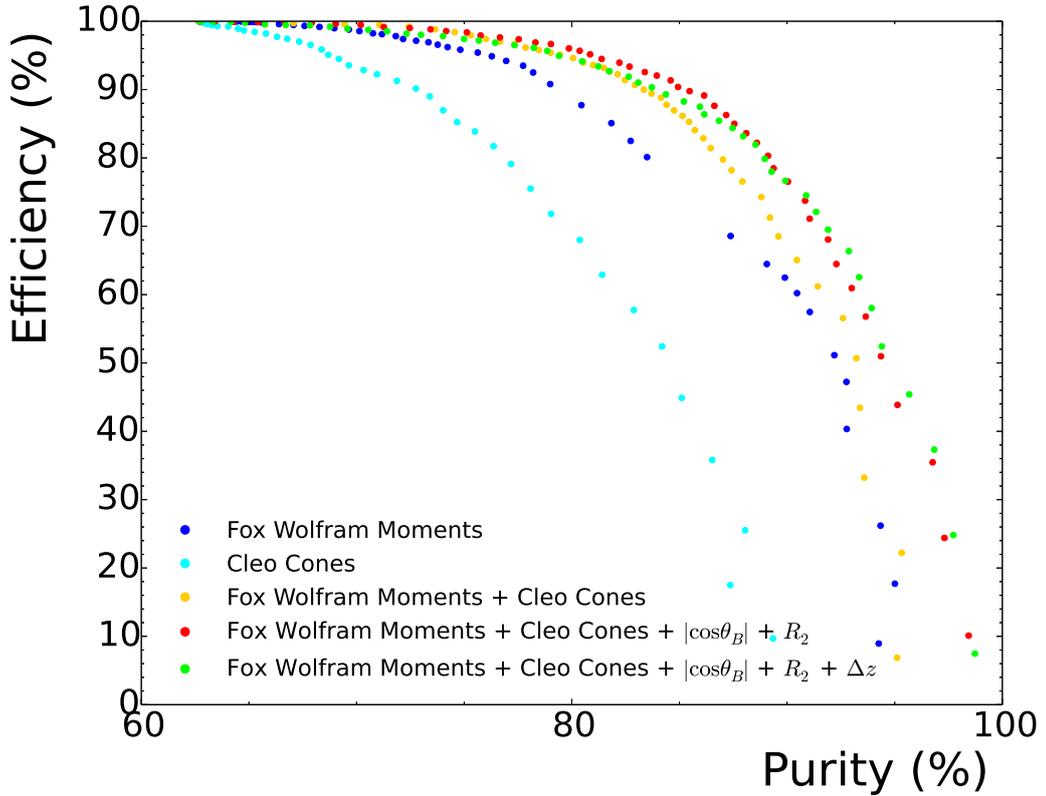


Figure 4.5: Comparison of different combination of continuum suppression variables on a sample with arbitrary signal to background ratio.

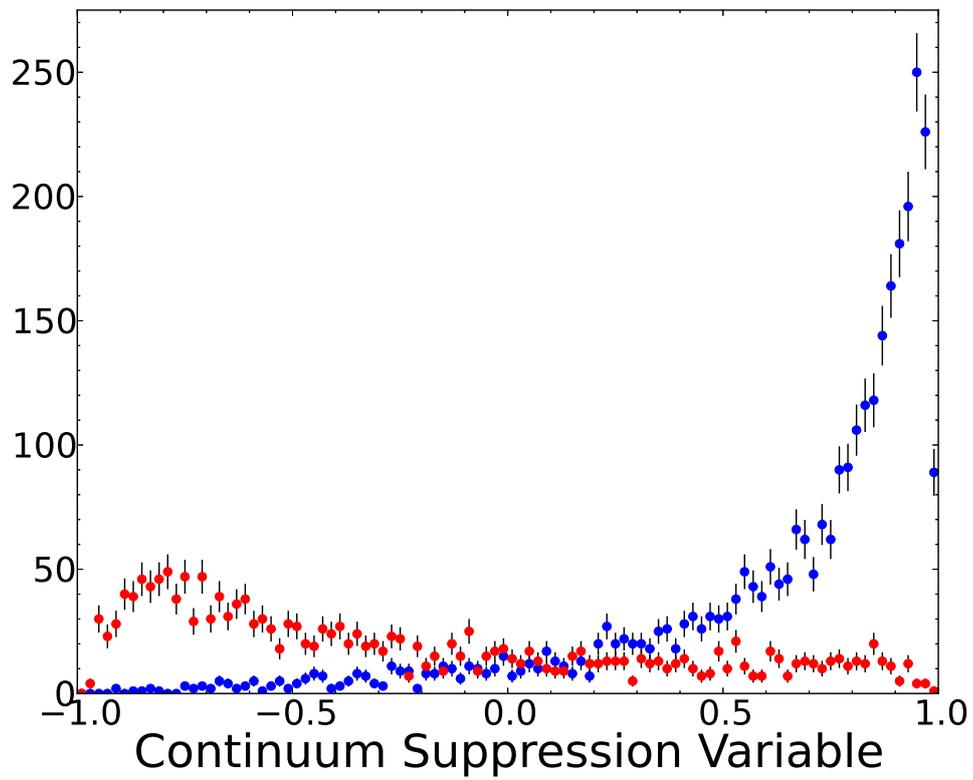


Figure 4.6: Distribution of the continuum suppression variable on a sample with arbitrary signal to background ratio. The red marker show the background and the blue ones the signal distribution.

5. Reconstruction

5.1 Full event reconstruction

The reconstructed decay channel $B^+ \rightarrow \tau^+ \nu_\tau$ leads to very few detectable particles in the final state. Depending on the decay channel of the τ lepton, this is often only a single charged track or a charged track and an additional π^0 . Since exclusively reconstructing these particles would lead to an immense amount of background, the full reconstruction technique is used. Here, the semileptonic tagging algorithm described in Chapter 4.1 is used to reconstruct the event as a whole. Since the initial state of the event is very well known, additional constraints can be applied. In essence, the $\Upsilon(4S)$ is reconstructed and selected. In the following, the B meson reconstructed in the the decay $B^+ \rightarrow \tau^+ \nu_\tau$ is called B_{sig} and the one reconstructed by the tagging algorithm B_{tag} .

The τ from the signal-side decay is reconstructed in the decay channels listed in Table 5.1. The charged particles in the final states of these decays are reconstructed from tracks in the CDC and the SVD and the π^0 is reconstructed as $\pi^0 \rightarrow \gamma\gamma$. The photons are reconstructed from clusters in the ECL which could not be matched to charged tracks. The $B^+ \rightarrow \tau^+ \nu_\tau$ candidates are combined with the B_{tag} candidates to ensure that no track or cluster is used twice in the reconstruction of the full event.

Decay channel	Branching fraction (%)
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	17.4 ± 0.4
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	17.8 ± 0.4
$\tau^+ \rightarrow \pi^+ \nu_\tau$	10.8 ± 0.6
$\tau^+ \rightarrow \rho^+ (\rightarrow \pi^+ \pi^0) \nu_\tau$	25.5 ± 0.9
Sum	72.5 ± 1.2

Table 5.1: List of reconstructed τ -decay channels and their branching fractions.

5.1.1 Background composition

After this reconstruction is performed, the data sample consists of both the desired $B^+ \rightarrow \tau^+ \nu_\tau$ decays and incorrectly reconstructed background events originating from different kinds of processes:

Generic $b \rightarrow c$ events: Semileptonic decays like $B^+ \rightarrow D^0 \ell^+ \nu_\ell$ lead to a significant amount of background. Their branching fraction is high, as already discussed in the explanation of the semileptonic tagging, and a lepton with a similar signature as in the signal decay is produced. There are decay modes of the D meson with final state particles which are hard to detect; additionally long living, neutral kaons are hardly detectable with the Belle detector and also neutral pions are missed regularly in the reconstruction, especially low energetic ones.

Rare B decays: Rare decays denote all kinds of B -meson decays without a $b \rightarrow c$ transition. Their branching fractions typically are very small, but they can produce final states, which are easily reconstructed as a signal event. For instance the neutral pion from the decay $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ can be missed in the reconstruction of the event. Other similar decays also contribute to this kind of background.

Continuum events: Continuum events contain background from $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) and two-photon events. The fact that no B meson is produced requires the coincidental combination of tracks to a $B_{\text{sig}}, B_{\text{tag}}$ combination. Pairs of leptons are often produced by the conversion of photons, which then can be taken as candidates for the signal-side particle and the lepton on the tag side. Additionally, other tracks in the event can be combined to a D -meson candidate by chance.

5.2 Selection

Since the presence of more than one neutrino in the decay of the τ lepton does not allow for strong kinematic constraints, the selection is performed mainly by vetoes and other kinds of selection variables. Only weak kinematic constraints are applied. The actual values for the selection requirements are optimized using Monte Carlo simulated data, which will be described later on.

The following variables are used for selection:

\mathcal{N}_{tag} : The output of the multivariate selection algorithm of the B_{tag} . Since the network output is related to the probability of the B -meson candidate to be reconstructed correctly, all candidates are treated equally, independent of their decay channel.

$\cos \theta_{B,D^{(*)}\ell}$: The angle between B meson and the $D\ell$ system on the tag side under the assumption, that only one massless particle is missing in the reconstruction of the decay, defined in Equation 4.1.

E_{ECL} : All depositions in the electromagnetic calorimeter are tested if they fulfill the following conditions:

- They do not originate from a photon used in the reconstruction of the $\Upsilon(4S)$.
- They are either not matched to a charged track at all or if they lie near to another cluster matched to a charged track, the ratio of the energy in the nine calorimeter cells to the 25 cells around the center of mass is greater than 0.94.
- Their energy is greater than 50, 100, or 150 MeV if they lie in the barrel, forward end cap, or backward end cap region, respectively.

The energy of all photons fulfilling these criteria is summed up.

Impact parameters dr and dz : The transverse (longitudinal) distance of the interaction point and of the point of closest approach of the according track in the plane orthogonal to the beam direction.

$p_{D_{\text{tag}}^{(*)}}^*$: The absolute value of the momentum of the D , or D^* meson if present, in the CMS.

p_{sig}^* : The absolute value of the momentum of the signal-side particle in the CMS.

$p_{\ell_{\text{tag}}}^*$: The absolute value of the momentum of the tag-side lepton in the CMS.

$\text{PID}_{e,\mu,\pi \text{ vs. } K^+_{\text{sig}}}$: The particle identification variables described in Section 3.4.1.

$M_{\pi^+\pi^0}$: For the decay channel $\tau^+ \rightarrow \rho^+ \nu_\tau$, the invariant mass of the $\pi^+\pi^0$ combination.

\mathcal{N}_{CS} : The output of the continuum suppression MVD, described in Chapter 4.2.

$M_{\ell_{\text{sig/tag}}^X}$: In the decay channel $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$, a significant background arises from events containing converted photons. To suppress this, the electron used in the reconstruction of either B_{sig} or B_{tag} is combined with every other oppositely charged track in the event, where the unspecific tracks are assumed to be electrons. The minimal invariant mass of all two-track combinations in the event is taken as selection variable.

5.2.1 Signal-side separation

The three reconstructed signal-side decay channels with only one charged track in the final state may contain a different amount of background candidates. Therefore they are separated using the particle identification variables. This is done in a way, that a candidate is never reconstructed in two signal-side decay channels simultaneously. The selection is applied for the charged track in the final state of the τ decay. The separation is done with

$$\begin{aligned} \tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu: & \quad \text{PID}_\mu > 0.9 \\ \tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e: & \quad \text{PID}_\mu < 0.9 \text{ and } \text{PID}_e > 0.9 \\ \tau^+ \rightarrow \pi^+ \nu_\tau: & \quad \text{PID}_\mu < 0.9 \text{ and } \text{PID}_e < 0.9. \end{aligned}$$

5.2.1.1 Vetoes

The strongest constraints, which only are possible due to the semileptonic tagging, are the vetoes. One of them is the π^0 veto and the other one the charged track veto.

The criterion for an event to be rejected by the π^0 veto is the presence of a good π^0 candidate, which is neither used in the signal or the tag side reconstruction. A good π^0 candidate is the combination of two photons with an invariant mass of $|M_{\gamma\gamma} - m_{\pi^0}| < 16 \text{ MeV}/c^2$.

Since there are a significant amount of incorrectly reconstructed charged tracks, excluding all candidates with an additional charged track might not be the optimal choice. A way to distinguish between correctly and falsely reconstructed tracks is their impact parameter, since fake tracks tend to have, more so than real tracks, a tendency to high impact parameters.

So the impact parameter cut is varied in dr and dz . For each set of cuts, the number of events passing the veto is evaluated and the figure-of-merit $N_S/\sqrt{N_S + N_B}$ is calculated, where N_S is the number of signal events and N_B the number of background events after the according selection is applied. The veto rejects all candidates with at least one charged track passing the impact parameter cuts. So a wider impact parameter cut leads to a more restrictive veto, since more tracks are taken into account.

The efficiency reduction for signal and background events is shown in Figure 5.1 and the resulting figure-of-merit in Figure 5.2. This gives an optimal set of cuts of $dr < 20 \text{ cm}$ and $dz < 100 \text{ cm}$.

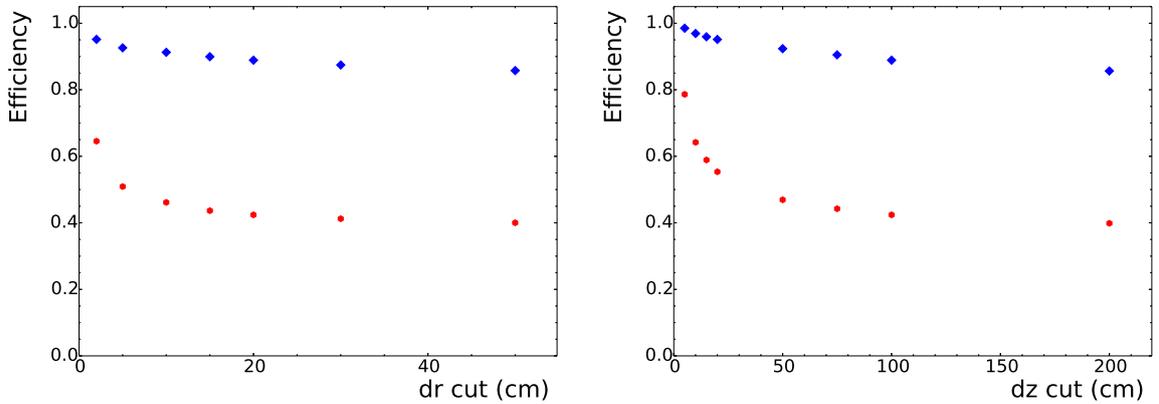


Figure 5.1: Efficiency reduction for signal and background, dependent on the impact parameter cut. The blue (red) dots show signal (background). The left (right) plots shows dr (dz).

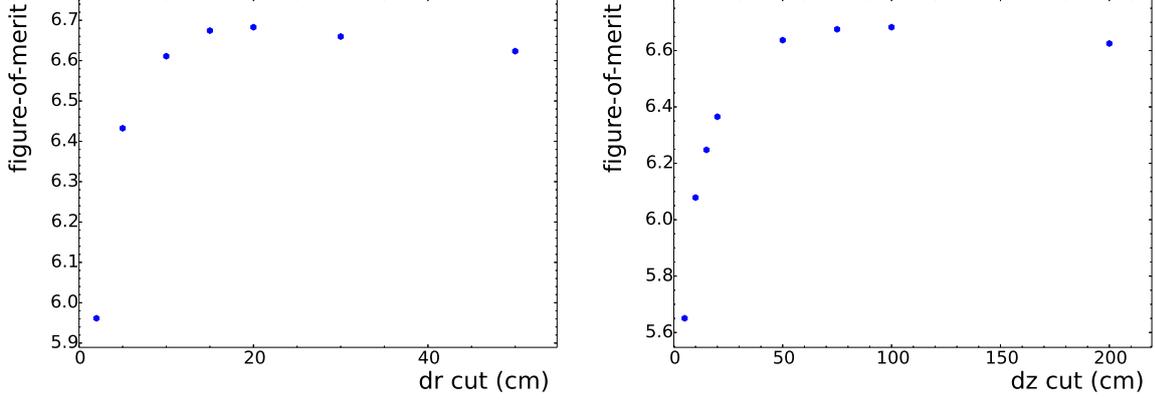


Figure 5.2: The figure-of-merit, dependent on the impact parameter cut. The left (right) plots shows dr (dz), where the selection on dz (dr) is fixed to the optimal value.

5.2.2 Selection optimization

The selection on \mathcal{N}_{tag} , $\cos\theta_{B,D^{(*)}\ell}$, $\text{PID}_{\pi \text{ vs. } K \text{ sig}}$, $M_{\pi^+\pi^0}$, and \mathcal{N}_{cs} is optimized. The other selection criteria are motivated by standard procedures in the Belle collaboration or constraints from later stages of the measurement. The optimization is performed by simultaneously varying the selection requirements and thereby maximizing the figure-of-merit. Both numbers are extracted from Monte Carlo simulated data samples in the region of $E_{\text{ECL}} < 0.2 \text{ GeV}$. As background sample, generic $b \rightarrow c$ decays and $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) continuum events are used. As signal, a sample of 2.5×10^6 simulated signal events is used and weighted according to the world average of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (1.05 \pm 0.25) \times 10^{-4}$ [3]. The optimization is illustrated in Figure 5.3 for two of the variables. The optimization leads to the selection criteria listed in Table 5.2.

	$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	$\tau^+ \rightarrow \pi^+ \nu_\tau$	$\tau^+ \rightarrow \rho^+ \nu_\tau$
E_{ECL}			$E_{\text{ECL}} > 1.2 \text{ GeV}$	
P_{sig}^*			$0.5 \text{ GeV}/c < P_{\text{sig}}^* < 2.4 \text{ GeV}/c$	
$P_{\ell \text{ tag}}^*$			$P_{\ell \text{ tag}}^* > 0.3 \text{ GeV}/c$	
$P_{D^{(*)} \text{ tag}}^*$			$P_{D^{(*)} \text{ tag}}^* < 2.5 \text{ GeV}/c$	
dr_{sig}			$dr < 2 \text{ cm}$	
dz_{sig}			$dz < 4 \text{ cm}$	
\mathcal{N}_{tag}	$\mathcal{N}_{\text{tag}} > 0.0066$	$\mathcal{N}_{\text{tag}} > 0.0075$	$\mathcal{N}_{\text{tag}} > 0.02$	$\mathcal{N}_{\text{tag}} > 0.009$
$\cos\theta_{B,D^{(*)}\ell}$	$-1.7 < \cos\theta_{B,D^{(*)}\ell} < 1$	$-1.9 < \cos\theta_{B,D^{(*)}\ell} < 1$	$-1.3 < \cos\theta_{B,D^{(*)}\ell} < 1$	$-2.6 < \cos\theta_{B,D^{(*)}\ell} < 1$
$\text{PID}_{\pi \text{ vs. } K \text{ sig}}$			$\text{PID}_{\pi \text{ vs. } K \text{ sig}} > 0.2$	$\text{PID}_{\pi \text{ vs. } K \text{ sig}} > 0.6$
$M_{\pi^+\pi^0}$				$ M_{\pi^+\pi^0} - m_{\rho^+} < 0.195 \text{ GeV}/c^2$
\mathcal{N}_{cs}	$\mathcal{N}_{\text{cs}} > -0.5$	$\mathcal{N}_{\text{cs}} > -0.5$	$\mathcal{N}_{\text{cs}} > 0.75$	$\mathcal{N}_{\text{cs}} > 0$
$M_{\ell \text{ sig}^X}$		$M_{\ell \text{ sig}^X} > 0.2 \text{ GeV}/c^2$		
$M_{\ell \text{ tag}^X}$		$M_{\ell \text{ tag}^X} > 0.2 \text{ GeV}/c^2$		

Table 5.2: Applied selection criteria.

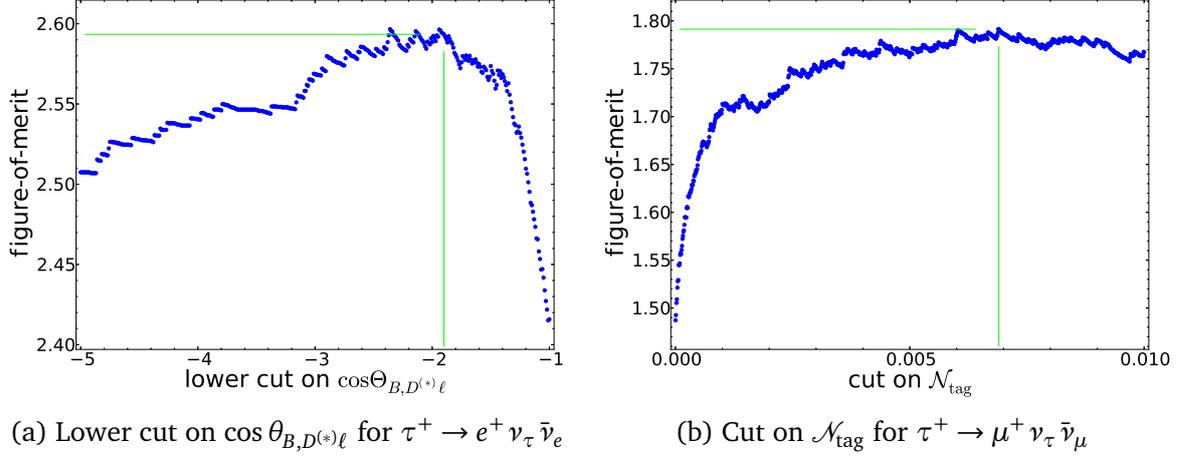


Figure 5.3: Illustration of the optimization of the selection criteria. The selection criteria not shown are fixed to their optimal value. The lines indicate the chosen value for the selection.

5.2.3 Best candidate selection

After all selection, there still are multiple candidates in 7% of all signal events. Most of them only differ in the particle hypothesis of the lepton on the tag side. Different ways of selecting the best candidate are tested and compared. They are listed in Table 5.3 together with the fraction of correctly picked candidates. The compared methods are:

- Picking the candidate by random.
- Picking the candidate with the highest output of the tag-side MVD.
- Picking the candidate with the value of $\cos \theta_{B,D^{(*)}\ell}$ closest to zero.
- Picking the candidate with D candidate on the tag-side closest to the nominal D -meson mass.

Since \mathcal{N}_{tag} contains the full information of the tag-side lepton, it delivers the best result and is taken as best candidate criterion.

Variable	Correct Ratio
Random	0.50
\mathcal{N}_{tag}	0.70
$\cos \theta_{B,D^{(*)}\ell}$	0.57
$M_{D_{\text{tag}}}$	0.51

Table 5.3: Different ways of best candidate selection and the corresponding rate of correctly picked candidates.

5.3 Efficiency

The signal reconstruction efficiency is determined using a sample of 2.5×10^6 simulated $B^+ \rightarrow \tau^+ \nu_\tau$ decays. In the simulation, the B_{tag} meson always decays in one of the reconstructed B -decay channels. The resulting reconstruction efficiencies (including the corresponding branching fractions) are given in Table 5.4. The errors are determined with

$$\sigma_\epsilon = \sqrt{\frac{N_{\text{rec}}(N_{\text{gen}} - N_{\text{rec}})}{N_{\text{gen}}^3}}, \quad (5.1)$$

where N_{gen} is the number of generated and N_{rec} the number reconstructed signal events. These errors are later on treated as systematic errors.

The contribution from incorrectly reconstructed signal is given by $\Upsilon(4S)$ candidates, which are reconstructed in the correct signal decay channel, but at least one of the contributing particles is reconstructed falsely. In the predominant part of these cases, the γ or π^0 from the D^{*0} decay on the tag side is either incorrectly reconstructed, or not reconstructed at all. This is reflected in the low fraction of this component in the tag-side decays without a D^* meson. This component is treated as signal, since the difference in E_{ECL} and other distributions is very small and it cannot be separated.

The cross-feed component denotes candidates which are reconstructed in the wrong signal decay channel. For the signal decays $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$ and $\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$, this fraction is low. But for the other two decay channels, there is a considerable amount of cross feed. The dominant contribution in $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$ comes from $\tau^+ \rightarrow a_1^+ (\rightarrow \pi^+ \pi^0 \pi^0) \nu_\tau$ decays, where in $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ more channels contribute. The efficiencies, by simulated and reconstructed decay channel are listed in Table 5.5.

5. Reconstruction

Signal Decay	Tag Decay	correct (inc BR)	incorrect (inc BR)	cross feed (inc BR)	overall (inc BR)
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	$B^+ \rightarrow \bar{D}^0 e^+ \nu_e$	1.12 ± 0.03	0.17 ± 0.01	0.05 ± 0.01	1.34 ± 0.03
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	$B^+ \rightarrow \bar{D}^0 \mu^+ \nu_\mu$	0.95 ± 0.02	0.08 ± 0.01	0.01 ± 0.00	1.04 ± 0.03
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	$B^+ \rightarrow \bar{D}^{*0} e^+ \nu_e$	0.64 ± 0.02	1.93 ± 0.03	0.09 ± 0.01	2.67 ± 0.04
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	$B^+ \rightarrow \bar{D}^{*0} \mu^+ \nu_\mu$	0.63 ± 0.02	1.71 ± 0.03	0.03 ± 0.00	2.37 ± 0.04
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	- sum -	3.34 ± 0.05	3.90 ± 0.05	0.18 ± 0.01	7.42 ± 0.07
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	$B^+ \rightarrow \bar{D}^0 e^+ \nu_e$	0.87 ± 0.02	0.10 ± 0.01	0.06 ± 0.01	1.03 ± 0.03
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	$B^+ \rightarrow \bar{D}^0 \mu^+ \nu_\mu$	0.59 ± 0.02	0.06 ± 0.01	0.08 ± 0.01	0.73 ± 0.02
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	$B^+ \rightarrow \bar{D}^{*0} e^+ \nu_e$	0.55 ± 0.02	1.47 ± 0.03	0.10 ± 0.01	2.12 ± 0.04
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	$B^+ \rightarrow \bar{D}^{*0} \mu^+ \nu_\mu$	0.39 ± 0.02	1.07 ± 0.03	0.17 ± 0.01	1.62 ± 0.03
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	- sum -	2.41 ± 0.04	2.70 ± 0.04	0.40 ± 0.02	5.51 ± 0.06
$\tau^+ \rightarrow \pi^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^0 e^+ \nu_e$	0.31 ± 0.01	0.04 ± 0.01	0.53 ± 0.02	0.89 ± 0.02
$\tau^+ \rightarrow \pi^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^0 \mu^+ \nu_\mu$	0.27 ± 0.01	0.02 ± 0.00	0.43 ± 0.02	0.71 ± 0.02
$\tau^+ \rightarrow \pi^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^{*0} e^+ \nu_e$	0.19 ± 0.01	0.43 ± 0.02	1.02 ± 0.03	1.65 ± 0.03
$\tau^+ \rightarrow \pi^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^{*0} \mu^+ \nu_\mu$	0.21 ± 0.01	0.41 ± 0.02	0.85 ± 0.02	1.48 ± 0.03
$\tau^+ \rightarrow \pi^+ \nu_\tau$	- sum -	0.98 ± 0.02	0.91 ± 0.02	2.84 ± 0.04	4.73 ± 0.05
$\tau^+ \rightarrow \rho^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^0 e^+ \nu_e$	0.58 ± 0.02	0.16 ± 0.01	0.29 ± 0.01	1.03 ± 0.03
$\tau^+ \rightarrow \rho^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^0 \mu^+ \nu_\mu$	0.55 ± 0.02	0.13 ± 0.01	0.28 ± 0.01	0.96 ± 0.02
$\tau^+ \rightarrow \rho^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^{*0} e^+ \nu_e$	0.35 ± 0.01	1.48 ± 0.03	0.92 ± 0.02	2.75 ± 0.04
$\tau^+ \rightarrow \rho^+ \nu_\tau$	$B^+ \rightarrow \bar{D}^{*0} \mu^+ \nu_\mu$	0.37 ± 0.02	1.44 ± 0.03	0.79 ± 0.02	2.60 ± 0.04
$\tau^+ \rightarrow \rho^+ \nu_\tau$	- sum -	1.85 ± 0.03	3.21 ± 0.05	2.27 ± 0.04	7.33 ± 0.07
- sum -	- sum -	8.58 ± 0.07	10.72 ± 0.08	5.69 ± 0.06	24.98 ± 0.12

Table 5.4: Reconstruction efficiencies (10^{-4}), including the relevant branching fractions.

Final State	$e^+ \nu_e \bar{\nu}_\tau$	$\mu^+ \nu_\mu \bar{\nu}_\tau$	$\pi^+ \bar{\nu}_\tau$	$\pi^+ \pi^0 \bar{\nu}_\tau$
$e^+ \nu_e \bar{\nu}_\tau$	7.2 ± 0.1	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
$\mu^+ \nu_\mu \bar{\nu}_\tau$	0.1 ± 0.0	5.1 ± 0.1	0.7 ± 0.0	0.2 ± 0.0
$\pi^+ \bar{\nu}_\tau$	0	0.1 ± 0.0	1.9 ± 0.0	0.5 ± 0.0
$\pi^+ \pi^0 \bar{\nu}_\tau$	0	0.1 ± 0.0	1.6 ± 0.0	5.1 ± 0.1
$\pi^+ \pi^0 \pi^0 \bar{\nu}_\tau$	0	0	0.2 ± 0.0	1.3 ± 0.0
Other	0	0	0.1 ± 0.0	0.2 ± 0.0
All	7.4 ± 0.1	5.5 ± 0.1	4.7 ± 0.1	7.3 ± 0.1
Total		25.0 ± 0.1		

Table 5.5: Reconstruction efficiency (10^{-4}) for each τ decay mode, determined from MC. The row denotes the generated decay mode, and the columns represent the reconstructed final state. The off-diagonal entries reflect the cross-feed between channels.

6. Validation

In this chapter, the different validations and tests, which are performed in the course of the measurement, are described. This analysis requires an thorough study of certain aspects, since it relies, more than other measurements, on the correct reproduction of physical processes in their simulation. The individual sections in the chapter are independent of each other as each of them describes the investigation of a different aspect of the analysis.

6.1 Tag-side efficiency correction

The semileptonic tagging method heavily relies on the correctness of the simulation of the physics processes and the detector response. Mismodeling of variables, which are used in the trainings of the according MVDs may lead to differences in the output distributions between data and MC.

A good understanding of potential differences between data and MC is crucial for the following analysis for multiple reasons: The efficiency in reconstructing $B^+ \rightarrow \tau^+ \nu_\tau$ decays is obtained using simulated data and directly enters the determination of the final result of this analysis. Also discrepancies in the distribution of certain observables need to be understood to be able to use simulated data for the determination of the number of reconstructed signal decays.

To be able to study the signal component of any decay which is reconstructed using the semileptonic tagging, a sizeable data sample of correctly reconstructed decays is required. So other decay channels, in addition to $B^+ \rightarrow \tau^+ \nu_\tau$, with higher branching fractions are reconstructed. Since the chosen decay channels are similar to the ones used in the semileptonic tagging algorithm, this is often referred to as double-tagging.

These double-tagged samples are used to estimate a potential correction of the reconstruction efficiencies of the tagging algorithm. To do this, the branching fraction of the signal-side decay is measured and compared with the world average of the corresponding independent measurements. Under the assumption, that the efficiency of the signal-side

reconstruction is correctly modeled by the MC simulation, this comparison can be used to obtain the efficiency correction.

6.1.1 Double-tagged reconstruction

For the signal-side reconstruction, the following decays, denoted as control channels, were chosen:

- $B^+ \rightarrow D^{*0}(\rightarrow D^0 \pi^0) \ell^+ \nu_\ell$
- $B^+ \rightarrow D^{*0}(\rightarrow D^0 \gamma) \ell^+ \nu_\ell$
- $B^+ \rightarrow D^0 \pi^+$.

The D^0 meson is always reconstructed in the decay channel $D^0 \rightarrow K^- \pi^+$.

These decay channels were chosen because of their high branching fraction ($B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$) or because they allow for a very good background suppression ($B^+ \rightarrow D^0 \pi^+$).

6.1.1.1 Selection

The following selection criteria are applied to the particles on the signal side:

- $B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$:
 - $\cos \theta_{B, D^{(*)} \ell \text{ sig}}$: $-5 < \cos \theta_{B, D^{(*)} \ell \text{ sig}} < 1.5$
 - $M_{K\pi, \text{sig}}$: $|M_{K\pi, \text{sig}} - m_{D^0}| < 35 \text{ MeV}/c^2$
 - ΔM : $|\Delta M - \Delta M_{\text{PDG}}| < 8 \text{ MeV}/c^2$
 - $PID_\ell(\mu)$: $PID_\mu > 0.9$
 - $PID_\ell(e)$: $PID_\mu < 0.9$ and $PID_e > 0.9$

- $B^+ \rightarrow D^0 \pi^+$:
 - $M_{K\pi, \text{sig}}$: $|M_{K\pi, \text{sig}} - m_{D^0}| < 35 \text{ MeV}/c^2$
 - ΔE : $|\Delta E| < 40 \text{ MeV}$
 - M_{bc} : $M_{\text{bc}} > 5.2 \text{ GeV}/c^2$,

where ΔE is the difference between the energy of the B_{sig} candidate and the beam energy. The beam-constrained mass M_{bc} is the reconstructed mass of the B_{sig} candidate with the energy substituted by the beam energy (E_{beam}^*) and is defined as

$$M_{\text{bc}} = \sqrt{(E_{\text{beam}}^*)^2 - (\vec{p}_B^*)^2} \text{ with } \vec{p}_B^* \text{ being the momentum of the } B_{\text{sig}} \text{ candidate.}$$

Tighter selection criteria have been tested, but there are significant discrepancies in the signal distributions between simulated and real data, as can be seen in Figure 6.1. These differences potentially introduce additional inconsistencies which counteract the aim to be as independent as possible from the reconstructed signal side. So the tighter selection criteria are discarded.

The further selection is related to the nominal selection, listed in Table 5.2. All selection related to the B_{tag} and the event-wide vetoes are applied in addition to the signal-side selection. The selection related to $\cos \theta_{B, D^{(*)} \ell}$ and the two MVDs depends on the τ -decay

channel in the nominal reconstruction. A set of selection criteria is applied for each of the τ -decay channels to each of the double-tagged samples, resulting in four samples for each B -decay channel which only differ in the tag-side related selection.

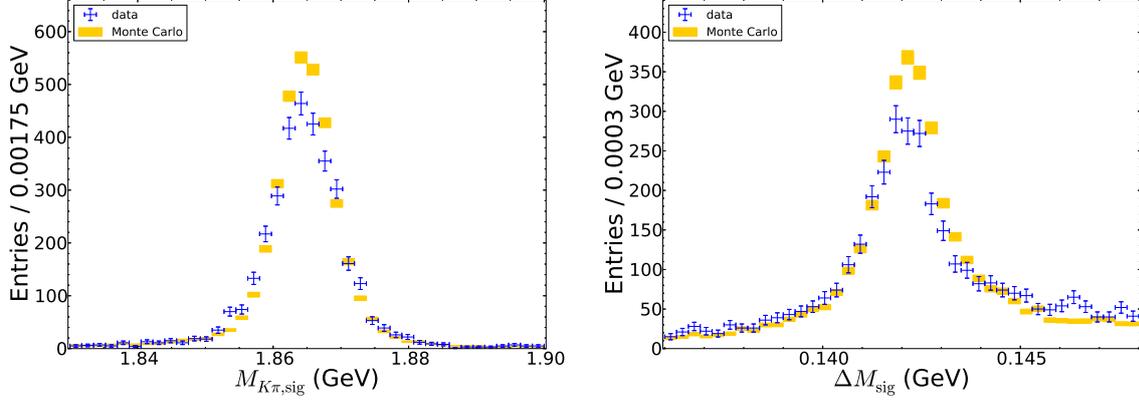


Figure 6.1: Comparison of real and simulated data in the mass of D -meson candidate (left) and the mass difference of the D^* - and the D -meson candidates (right) on the signal side in $B^+ \rightarrow D^{*0}(\rightarrow D^0 \pi^0) \ell^+ \nu_\ell$ decays.

6.1.2 Determination of number of reconstructed signal events

$B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$ decays: The sample of reconstructed $B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$ decays still contains a significant amount of background events. To extract the amount of correctly reconstructed $D^* \ell$ combinations, the $\cos \theta_{B, D^{(*)} \ell \text{ sig}}$ distribution is fitted. The shape of both the signal and the background component is determined using simulated data.

A binned maximum likelihood fit is performed for each of the $B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$ double-tagged samples. An exemplary fitted distribution is given in Figure 6.2.

$B^+ \rightarrow D^0 \pi^+$ decays: This decay channel is background free. There is a clear peak in M_{bc} at the mass of the B meson without any background contribution and studies of the simulated data samples show no peaking background component. The M_{bc} distributions of the samples related to the $\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$ and $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ decay channels are shown in Figure 6.3. Due to the much tighter $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ selection, there are considerably less signal events.

Since there is no background, the number of reconstructed signal events can be determined simply by counting.

6.1.3 Branching fractions

The extracted number of signal events N_{sig} is used to calculate the corresponding branching fraction with

$$\mathcal{B} = \frac{N_{\text{sig}}}{N_{B\bar{B}} \times \epsilon},$$

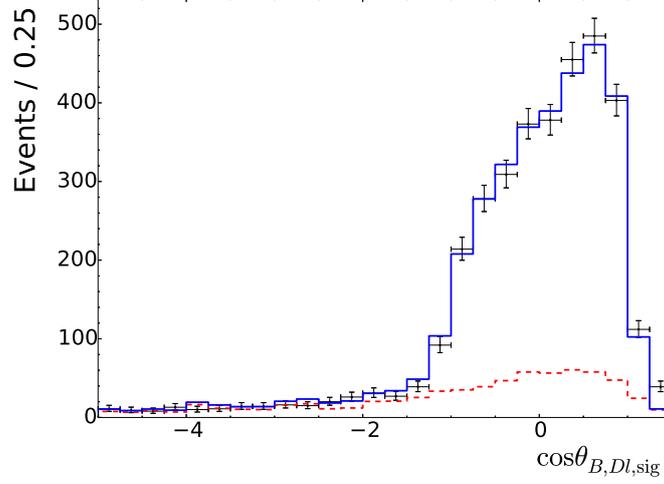


Figure 6.2: Fitted $\cos \theta_{B,D^{(*)}l_{sig}}$ distribution of $B^+ \rightarrow D^{*0}(\rightarrow D^0 \pi^0) \ell^+ \nu_\ell$ decays with the selection related to the $\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$ decay channel. The data distribution is given by the black markers, the total fitted distribution by the blue line, and the fitted background distribution by the dashed, red line.

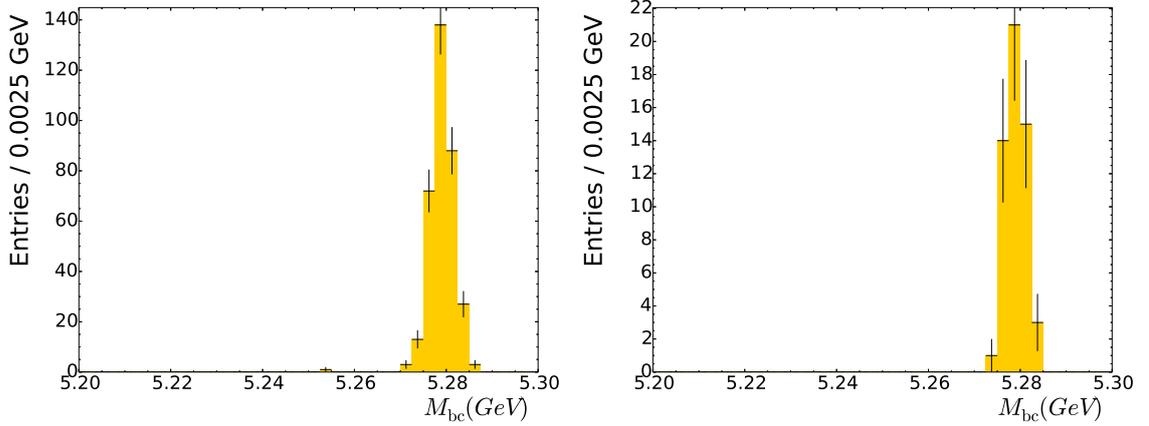


Figure 6.3: M_{bc} distribution for the $B^+ \rightarrow D^0 \pi^+$ double-tagged sample. The plots show the sample with the selection related to the $\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$ (left) and the $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ (right) decay channels.

where $N_{B\bar{B}}$ is the number of B -meson pairs in the data sample and ϵ is the reconstruction efficiency, taking into account the corresponding branching fractions of the intermediate particles. The reconstruction efficiency is determined using a sample of 5×10^6 simulated signal events.

The ratios of the measured branching fractions and their world averages are shown in Table 6.1. The factor for the same selection but different decay channels on the signal side are in relatively good agreement. The small differences can be explained by statistical fluctuations and the discrepancies in the $\Delta M_{D^{*},sig}$ distribution which is used for the

selection. Although the applied selection in this variable is very loose, it still might lead to deviations. Therefore the selection will be varied and the differences taken as a systematic error.

Comparing the different sets of selection, a clear trend is visible: The tighter the selection, the larger the deviation from one. The deviation is significantly larger for decays containing an excited D^* meson on the tag side.

This can be explained by the distributions of variables which are used in the trainings of the MVDs for the tagging algorithm, especially variables related to photons and the slow π^0 originating from the D^* decay. If these distributions differ in simulated and in real data, as was already observed on the signal side in Figure 6.1, this can lead to a wrong efficiency estimation on real data.

Weighted averages are formed from these factors and shown in Table 6.2. The errors are statistical only. These averages will be used later on to correct the reconstruction efficiencies which were obtained from MC.

Selection	Tag side Decay	$B^+ \rightarrow D^{*0}(\rightarrow D^0\pi^0)\ell^+\nu_\ell$	$B^+ \rightarrow D^{*0}(\rightarrow D^0\gamma)\ell^+\nu_\ell$	$B^+ \rightarrow D^0\pi^+$
$\tau^+ \rightarrow \mu^+\bar{\nu}_\tau\nu_\mu$	-all-	1.02 ± 0.02	0.90 ± 0.04	0.92 ± 0.05
	$B^+ \rightarrow D^0\ell^+\nu_\ell$	1.08 ± 0.03	0.90 ± 0.05	0.91 ± 0.05
	$B^+ \rightarrow D^{*0}\ell^+\nu_\ell$	0.79 ± 0.05	0.90 ± 0.12	0.94 ± 0.11
$\tau^+ \rightarrow e^+\bar{\nu}_\tau\nu_e$	-all-	1.01 ± 0.02	0.89 ± 0.04	0.91 ± 0.05
	$B^+ \rightarrow D^0\ell^+\nu_\ell$	1.07 ± 0.03	0.90 ± 0.05	0.90 ± 0.05
	$B^+ \rightarrow D^{*0}\ell^+\nu_\ell$	0.78 ± 0.05	0.87 ± 0.12	0.95 ± 0.11
$\tau^+ \rightarrow \pi^+\bar{\nu}_\tau$	-all-	0.85 ± 0.06	0.94 ± 0.11	0.88 ± 0.27
	$B^+ \rightarrow D^0\ell^+\nu_\ell$	0.90 ± 0.07	0.93 ± 0.12	0.69 ± 0.27
	$B^+ \rightarrow D^{*0}\ell^+\nu_\ell$	0.69 ± 0.10	0.99 ± 0.26	1.38 ± 0.68
$\tau^+ \rightarrow \rho^+\bar{\nu}_\tau$	-all-	0.98 ± 0.03	0.89 ± 0.05	0.94 ± 0.10
	$B^+ \rightarrow D^0\ell^+\nu_\ell$	1.04 ± 0.03	0.88 ± 0.05	0.92 ± 0.11
	$B^+ \rightarrow D^{*0}\ell^+\nu_\ell$	0.75 ± 0.05	0.94 ± 0.12	1.03 ± 0.24

Table 6.1: Ratio of measured branching fraction and the according world average for the three control channels, determined in different tag-side decay channels and with different selections, related to the reconstructed τ -decay channels.

6.2 E_{ECL} composition

The variable E_{ECL} is crucial for the separation of signal and background. In contrast to the naive expectation, the signal component is distributed over a wide range of E_{ECL} and can take values of up to a GeV. To understand the distribution, it is plotted for a sample of simulated signal decays shown in Figure 6.4. The distributions are separated by the source of photons which enter the calculation of E_{ECL} . The signal-side and tag-side split off contains photons which are the products of physics processes related to the B_{sig} or B_{tag} , respectively. The beam background consists of photons which are emitted by beam-related processes as beam-beam scattering or synchrotron radiation.

The dominant contribution in all signal decay channels is the tag-side split off. The B_{tag}

6. Validation

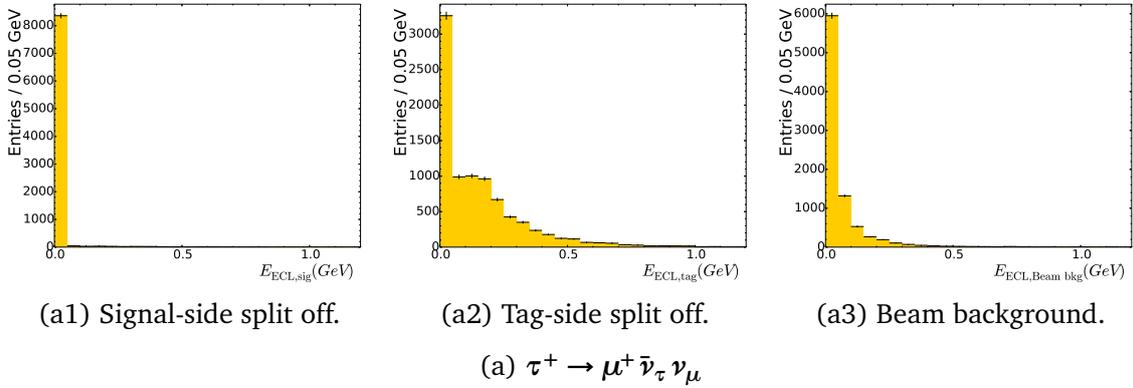
Selection	Tag side Decay	Correction factor
$\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$	$B^+ \rightarrow D^0 \ell^+ \nu_\ell$	1.01 ± 0.02
$\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$	$B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$	0.83 ± 0.04
$\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$	$B^+ \rightarrow D^0 \ell^+ \nu_\ell$	1.00 ± 0.02
$\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$	$B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$	0.82 ± 0.04
$\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$	$B^+ \rightarrow D^0 \ell^+ \nu_\ell$	0.90 ± 0.06
$\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$	$B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$	0.74 ± 0.09
$\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$	$B^+ \rightarrow D^0 \ell^+ \nu_\ell$	0.99 ± 0.02
$\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$	$B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$	0.79 ± 0.05

Table 6.2: Weighted average of the different control channels of the ratio of measured branching fraction and the according world average, shown in Table 6.1. The errors are statistical only.

candidate often decays into a D^{*0} meson, which further decays to a D^0 meson and a π^0 or γ . The following can happen: The π^0 is not reconstructed; or the D^* meson candidate is discarded due to the high background level in the D^* decays and the decay of the B_{tag} directly to a D^0 meson is taken as candidate.

This component and the beam-background related contribution are constant in the signal decay channel, where the contribution from the signal side differs between the τ decay channel. The electron has a higher probability of synchrotron radiation and therefore additional photons are accounted for in E_{ECL} . Also the amount of cross-feed differs between the signal decay channels. For $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ and $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$, there is a significant amount of signal events, which contain an additional, undetected π^0 . For instance, in the $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$ channel, the real decay could be $\tau^+ \rightarrow \pi^+ \pi^0 \pi^0 \bar{\nu}_\tau$ and one of the neutral pions was not reconstructed. One or both of the photons from the π^0 decays then are accounted for in the calculation of E_{ECL} .

None of these observations stands in contradiction to the expectations based on previous studies, for instance the investigation of the reconstruction efficiency in Section 5.3.



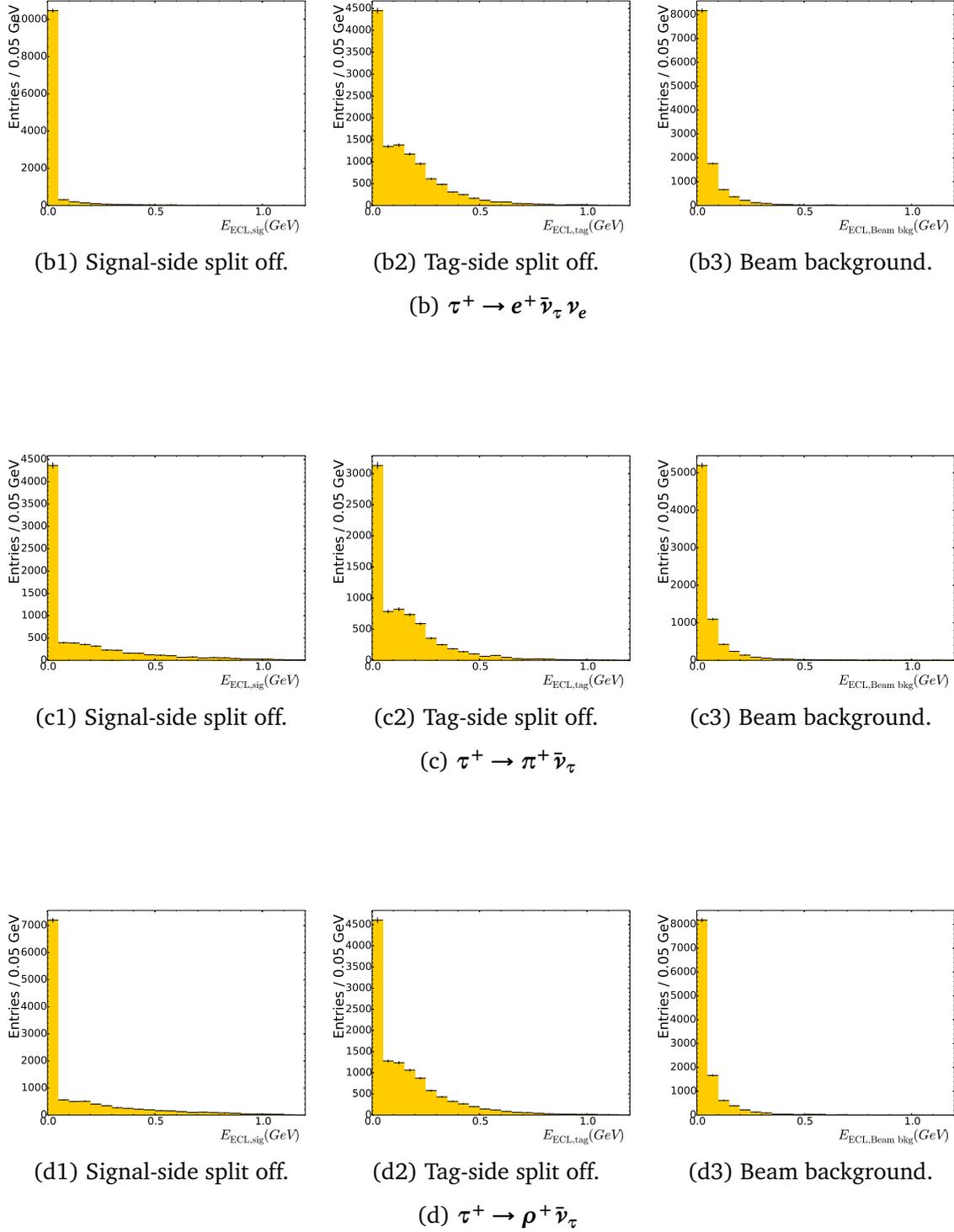


Figure 6.4: E_{ECL} distributions divided by the source of the additional photons and signal decay channel.

6.3 Off-resonance data

A frequently used method to describe background originating from continuum processes, $\tau^+\tau^-$, and two-photon events is to model its distributions using samples of simulated events. It is tested, if this is possible for this analysis. Therefore the simulation of these processes is investigated and validated. For this validation, a dedicated data sample is available. This was recorded at an energy below the $\Upsilon(4S)$ resonance at 10.52 GeV. It is reconstructed in the same way as the nominal data sample and contains all background processes without a B meson, where the corresponding cross sections differ only slightly due to the small relative difference in the center-of-mass energy. The total integrated luminosity of this so called off-resonance data sample is 79.37fb^{-1} , which corresponds to 11.3% of the total on-resonance data.

To get a more statistically significant comparison, the selection criteria for \mathcal{N}_{tag} and $\cos\theta_{B,D^{(*)}\ell}$ are released a little bit. Also the selection related to $M_{\ell_{\text{sig/tag}}X}$ and the continuum suppression is not applied for this study.

The reconstructed data sample is compared with a Monte Carlo sample, which contains $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) events. The comparison in E_{ECL} is shown in Figure 6.5. The normalization disagrees in all signal decay channels, but only in the $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$ channel also the shape is significantly different. There however, the shape of the Monte Carlo data does not nearly describe the actual data, it even peaks at low values of E_{ECL} like the signal component.

6.3.1 Min $M_{\ell_{\text{sig/tag}}X}$

Several variables are investigated to find the origin of this disagreement. In one quantity, the disagreement is especially prominent: The minimal invariant mass of the combination of the lepton on either the signal or the tag side and any other particle. To extract this information, the lepton is combined with all other oppositely charged tracks in the event and the minimal invariant mass of all possible combinations is recorded. Often the combination of the signal- and the tag-side lepton yields the minimal mass, but not always. The data and MC comparison is shown in Figure 6.6.

6.3.2 Event investigation

The presence of a large number of events with low invariant masses hints at the presence of converted photons. To further investigate this hypothesis, some of the events with a low invariant mass are examined, using the Belle event display. Some of the corresponding figures can be seen in Figures 6.7 and 6.8. They both show two charged particles, which have an opening angle of nearly zero. This also is a hint for converted photons.

The possible corresponding processes like two-photon events or $e^+e^- \rightarrow \gamma\gamma$ events are not part of the investigated MC sample, so it is not surprising that there is a discrepancy.

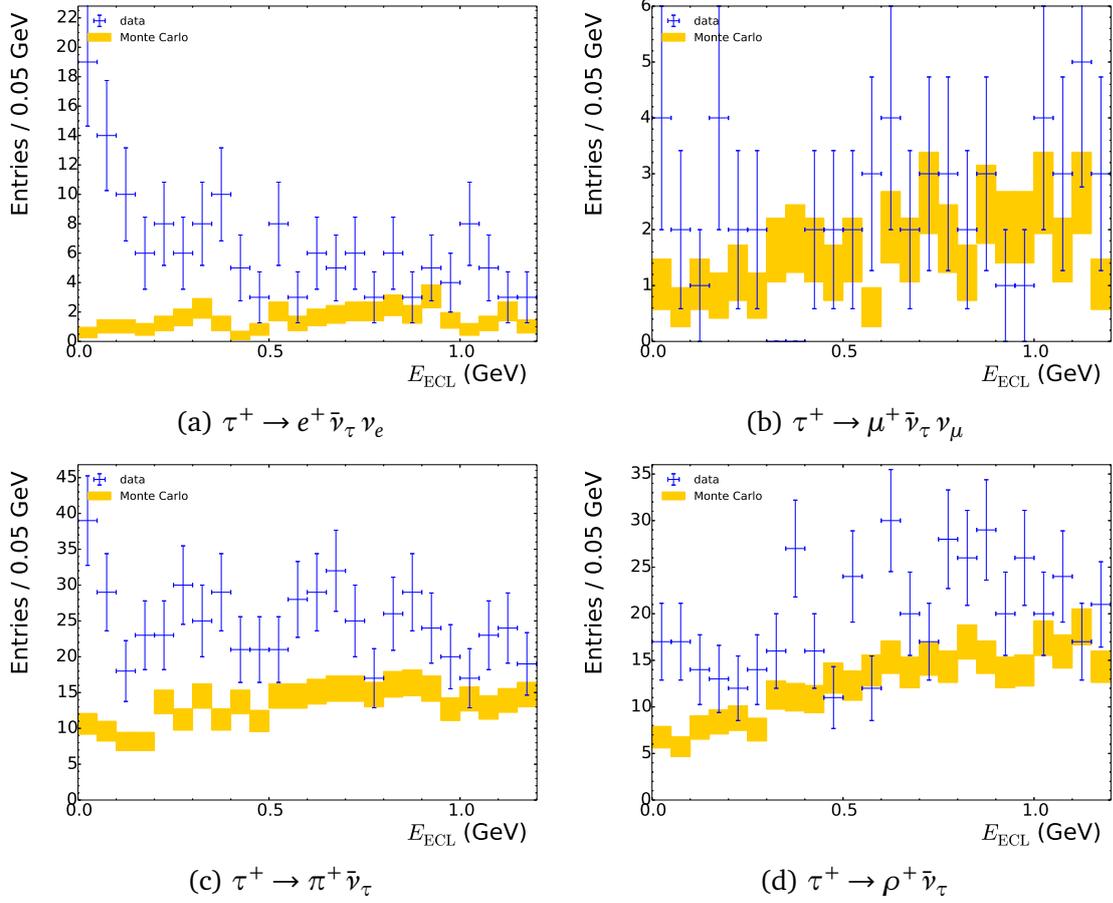


Figure 6.5: Comparison of data and Monte Carlo simulation in E_{ECL} , using the off-resonance region without the selection related to $M_{\ell_{\text{sig/tag}}^X}$ and the continuum suppression. The data distribution is represented by the blue markers and the Monte Carlo expectation by the orange boxes.

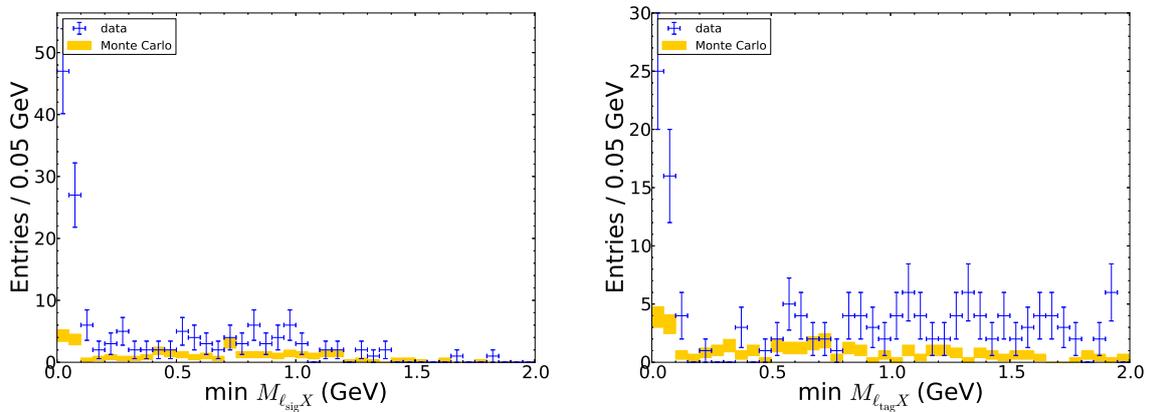


Figure 6.6: Comparison of data and Monte Carlo simulation in $\text{Min } M_{\ell_{\text{sig}}^X}$ (left) and $M_{\ell_{\text{tag}}^X}$ (right).

6. Validation

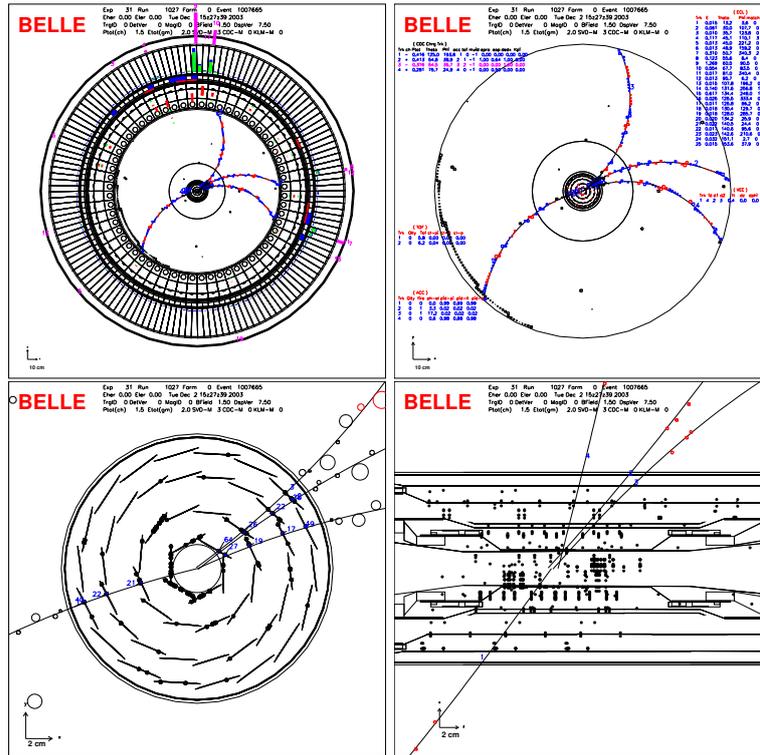


Figure 6.7: Event display of Experiment 31, Run 1027, Event 1007665.

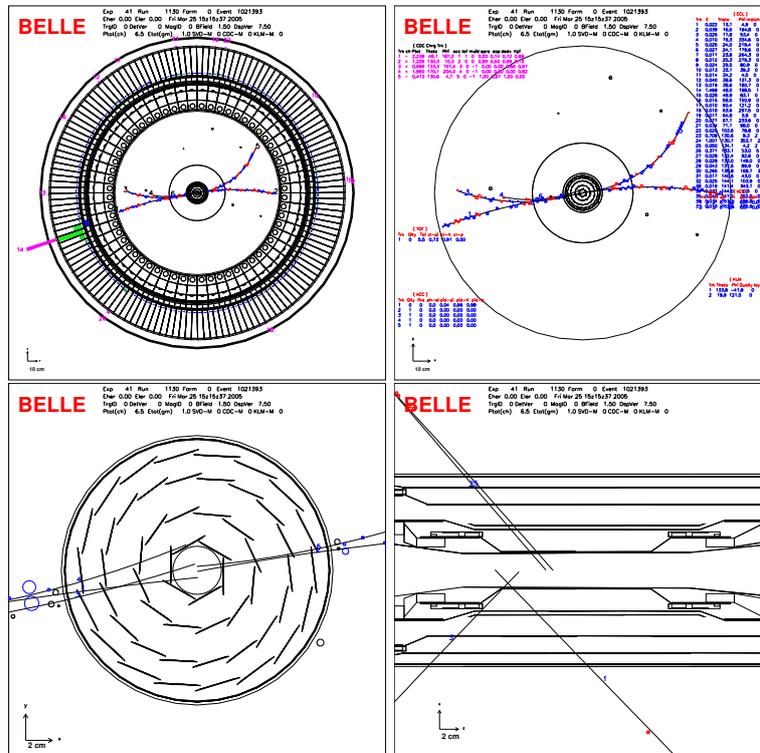


Figure 6.8: Event display of Experiment 41, Run 1130, Event 1021393.

6.3.3 Conclusion

The background originating from converted photons is suppressed by introducing the selection $M_{\ell_{\text{sig}}X} > 0.2 \text{ GeV}/c^2$ and $M_{\ell_{\text{tag}}X} > 0.2 \text{ GeV}/c^2$ in the signal decay channel $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$. The loss in signal efficiency is about five percent in this channel. Additionally the continuum suppression is applied.

The resulting comparison of real and simulated off-resonance data, again with a wider selection in \mathcal{N}_{tag} and $\cos \theta_{B,D^{(*)}\ell}$, are shown in Figure 6.9.

The data and Monte Carlo distributions agree better than before the additional selection was introduced but there still is a significant discrepancy. It is concluded, that the corresponding background component cannot be described by the sample of simulated data and it is necessary to use the off-resonance data for that purpose.

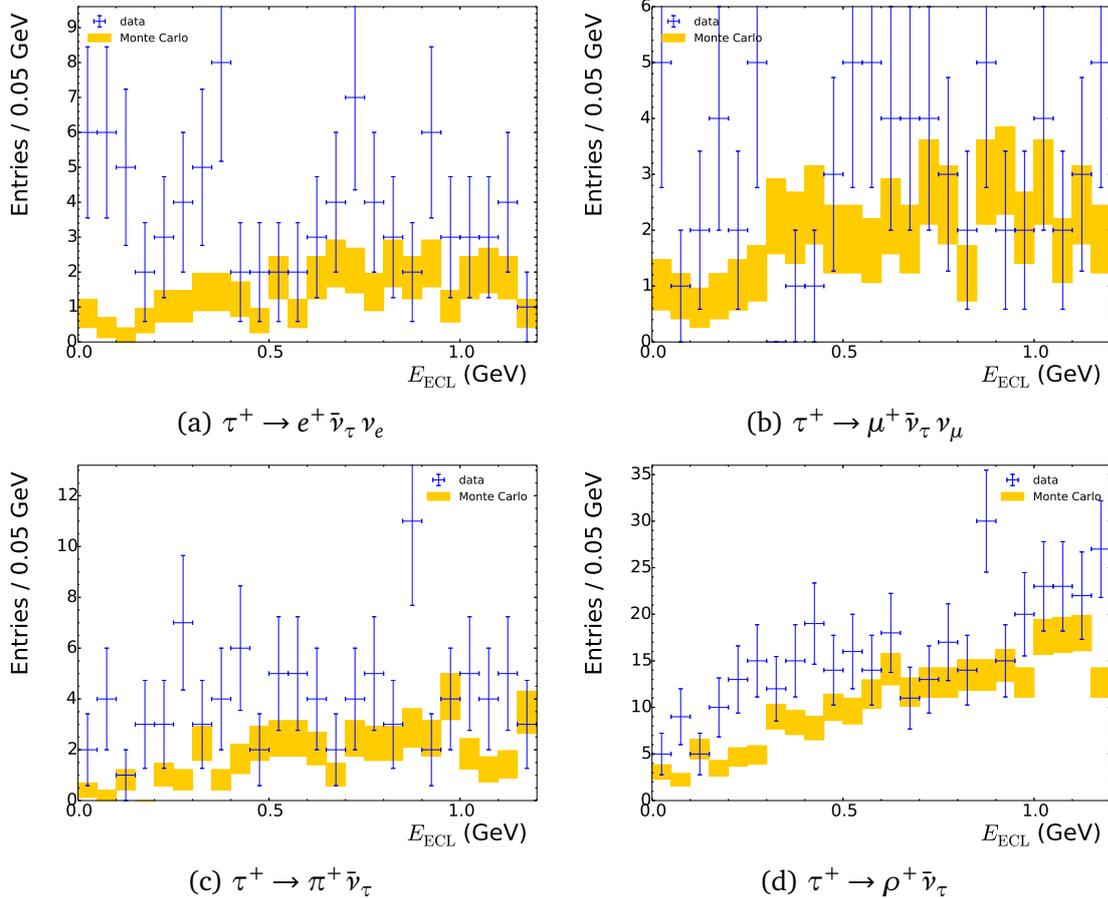


Figure 6.9: Comparison of data and Monte Carlo simulation in E_{ECL} , using the off-resonance region/after the adjustment of the selection. The data distribution is represented by the blue markers and the Monte Carlo expectation by the orange boxes.

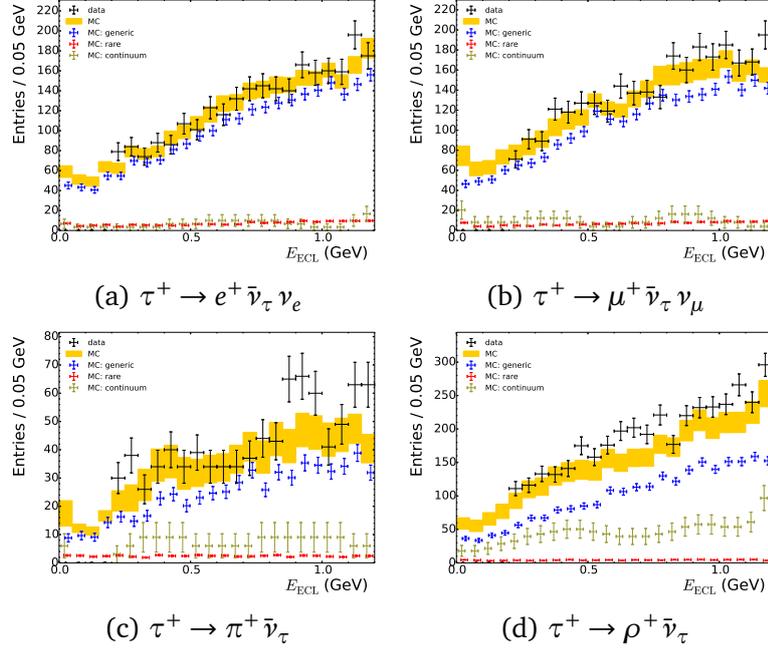
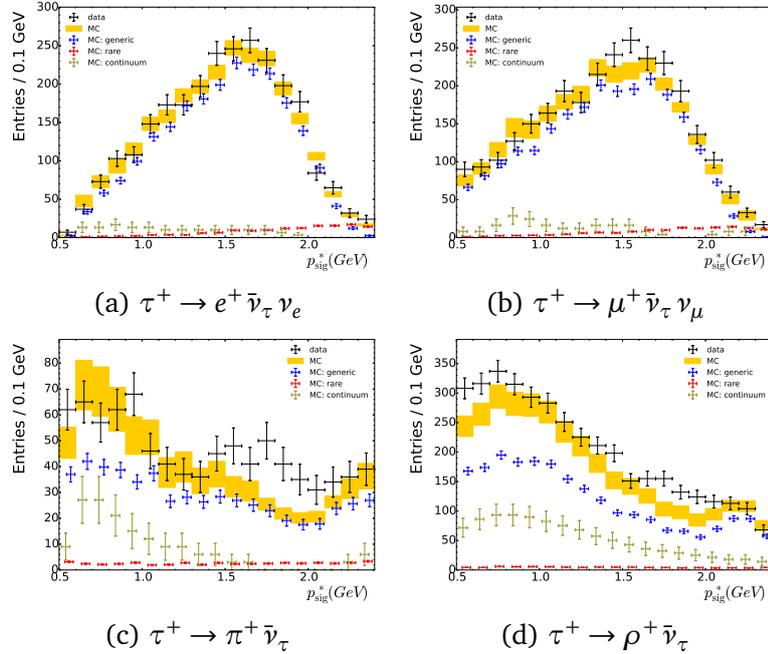
6.4 Comparison of data and simulation in control samples

Another way of validating the simulation of the relevant processes is the comparison of simulated and real data in sidebands or control samples. Three samples are chosen for that:

- The nominal reconstruction with a cut $E_{\text{ECL}} > 0.2 \text{ GeV}$.
- The combination of a B_{tag} which is reconstructed in $B^0 \rightarrow D^{(*)-} \ell^+ \nu_\ell$ with a normally reconstructed B_{sig} .
- One of the double-tagged sample, where the signal side is reconstructed in $B^+ \rightarrow D^0 \pi^+$.

In the case of the first two samples, event-by-event weights are applied to account for known discrepancies between MC and data, for instance the efficiency correction determined in Section 6.1. The double-tagged sample is not weighted since the tagging efficiency correction is extracted from it.

The comparison in the E_{ECL} sideband is shown in Figures 6.10 and 6.11, the comparison for the B^0 -tagged sample in Figures 6.12 and 6.13, and the comparison for the double-tagged sample in Figures 6.14 and 6.15. More comparisons are shown in Appendix A.

6.4.1 E_{ECL} sideband

 Figure 6.10: Comparison of the E_{ECL} distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

 Figure 6.11: Comparison of the p_{sig}^* distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

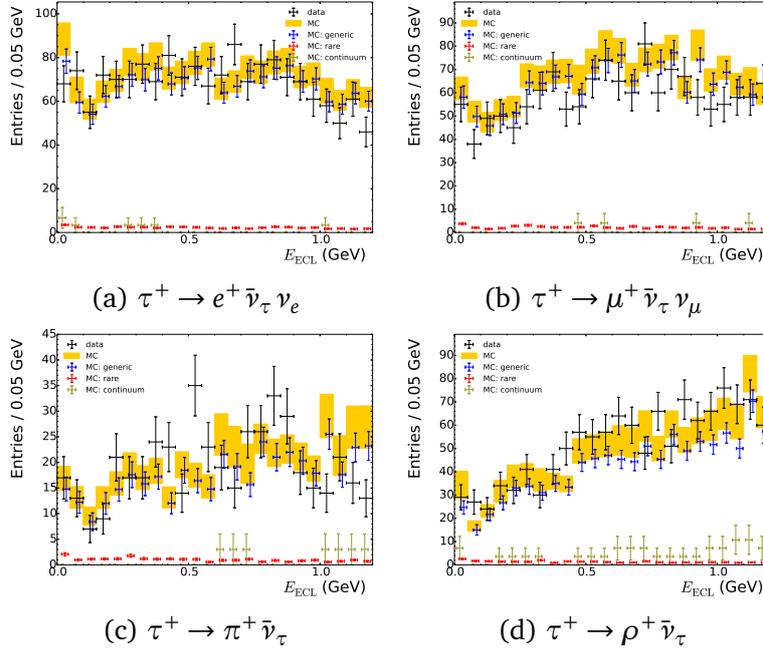
6.4.2 B^0 -tagged sample

Figure 6.12: Comparison of the E_{ECL} distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

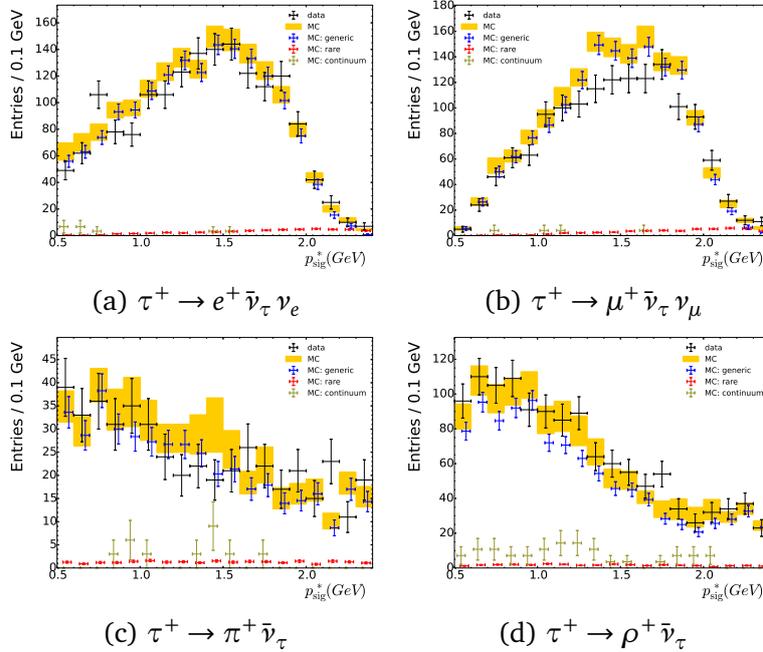


Figure 6.13: Comparison of the p_{sig}^* distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

6.4.3 Double-tagged sample

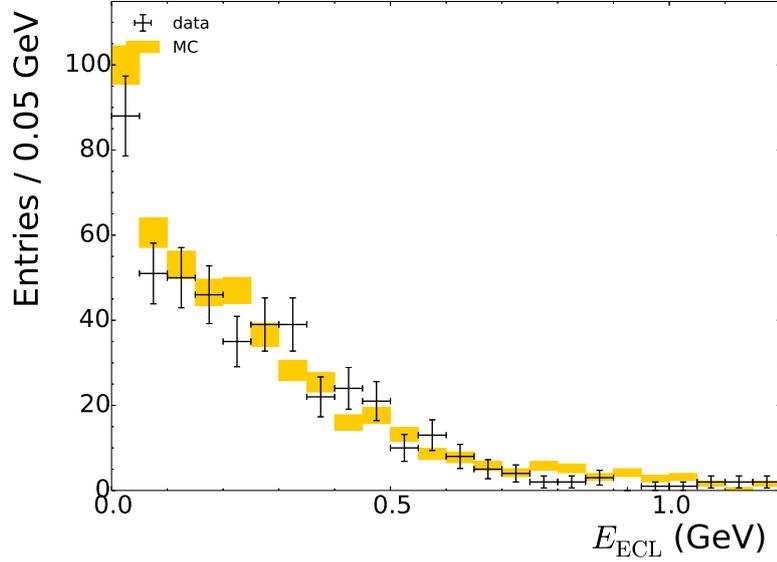
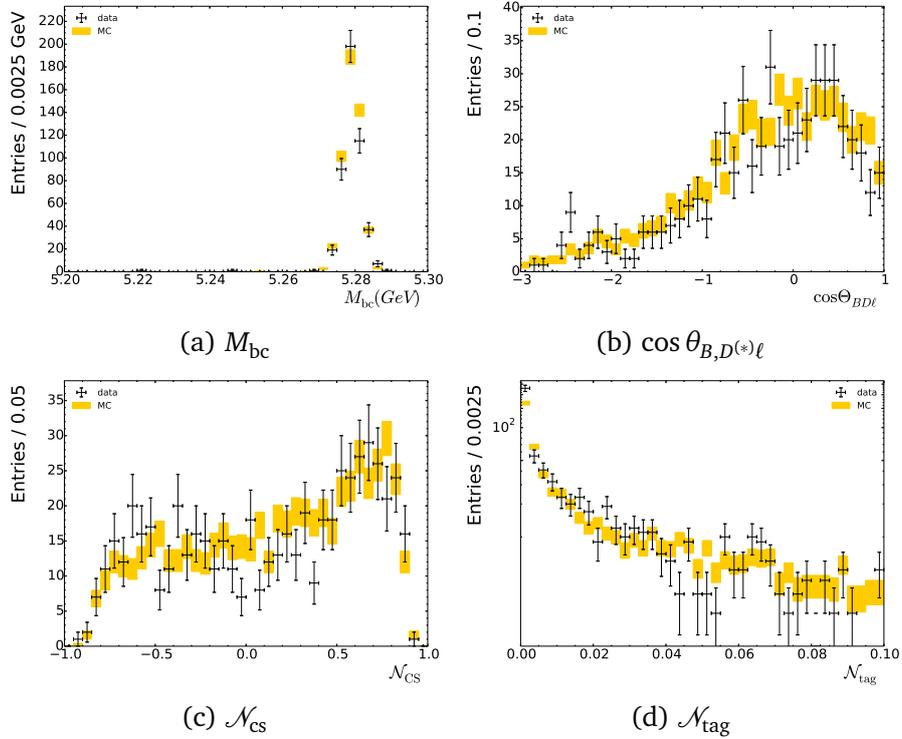

 Figure 6.14: Comparison of the E_{ECL} distribution of simulated and real data in the double-tagged sample.


Figure 6.15: Comparison of different distributions of simulated and real data in the double-tagged sample.

6.5 Charged track veto

To better understand the charged track veto, described in Section 5.2.1.1, the additional tracks are investigated. Therefore the selection is changed to veto only events with additional charged tracks with impact parameters of $dr < 2$ cm and $dz < 5$ cm. The tracks which could cause the event to be vetoed for the nominal selection are collected and divided by their type and their origin. This is done for samples of simulated signal and background events.

The tracks stem in almost all of the cases from charged pions or kaons, while the origin of these particles widely differ. The particle types of the mother particles, scaled to the overall number of events, are shown in Figure 6.16. Since there is often more than one charged track in an event, the sum of the shown numbers needs not to be smaller than one. Most of the tracks are decay products of short living particles like D or ρ mesons. They should have smaller impact parameters, so these tracks are either mismeasured or they are found twice by the track finding algorithm. This happens frequently for particles with a momentum small enough that they do not leave the CDC but their trajectory forms a helix with multiple curls. The other source of additional tracks are long living hadrons, like K_S , which are expected to produce tracks with large impact parameters. Charged pions or kaons can produce additional tracks by hadronic interaction with the detector material. A breakdown to the source of the track is shown in Figure 6.17.

Overall, there are two dominant reasons for the larger number of charged tracks with large impact parameters in background events: Background events most often contain more particles which are not reconstructed or are incorrectly reconstructed. The overall higher number of particles and mismeasured particles leads to a higher number of tracks with high impact parameters. And background events often contain neutral kaons. They decay in about two thirds of all cases to charged tracks which can have very high impact parameters.

To take into account possible differences in the signal efficiency between real and simulated data, a systematic error will be applied later on.

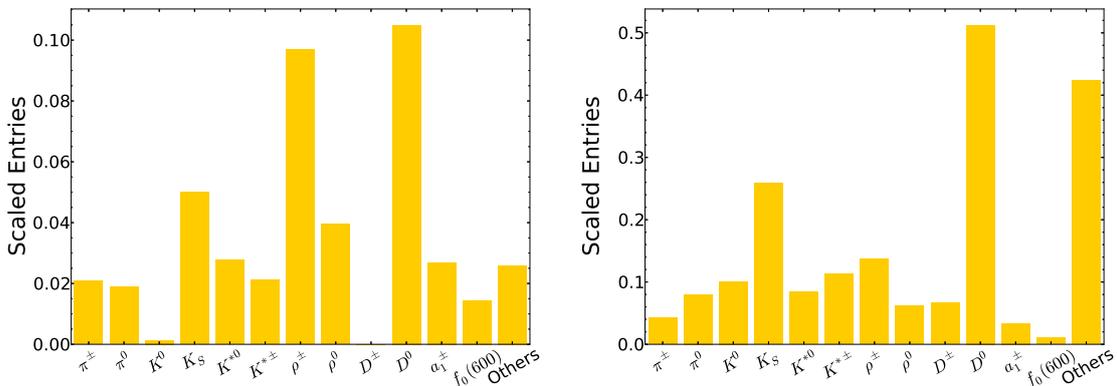


Figure 6.16: Particle type of the mother particle of additional tracks. The left plot shows the distribution for signal and the right one for background events.

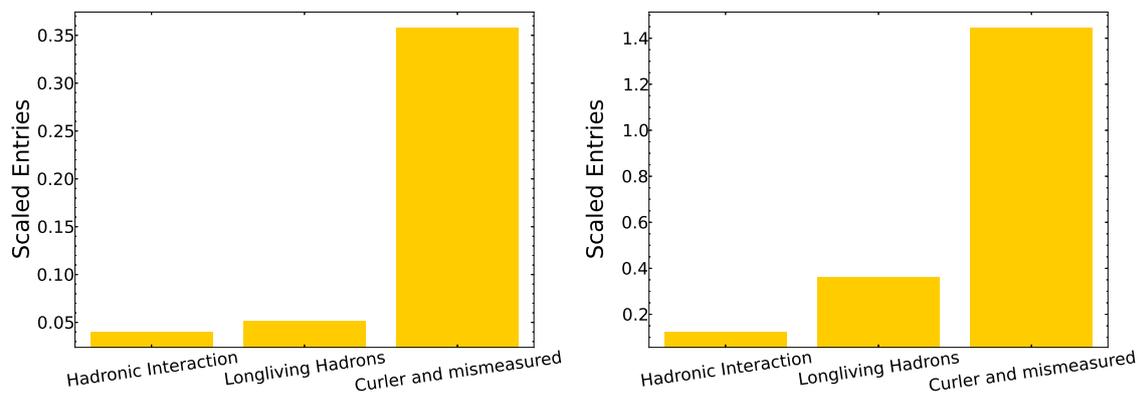


Figure 6.17: Origin of the additional tracks. The left plot shows the distribution for signal and the right one for background events.

7. Branching fraction extraction

7.1 Description of the fit procedure

To determine the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ an extended, unbinned, simultaneous maximum-likelihood fit in two dimensions is performed. Additional to E_{ECL} the momentum of the signal-side particle, p_{sig}^* , is used as observable in the fit. The shapes of all contributions are determined using simulated or real data. To do this, known discrepancies between simulated and real data are accounted for by weighting the Monte Carlo sample event-by-event.

Since the individual background components cannot be separated based on the available information, their relative normalization is fixed before the fit. For generic $b \rightarrow c$ or rare processes containing b quarks the available sample is scaled to the luminosity of the real data sample. The background from non- b processes, like continuum or two-photon events is described using the off-resonance data sample and scaled as described in Section 6.3.3. The branching fraction is calculated in each of the signal decay channels d with

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = \frac{N_{\text{sig},d}}{N_{B\bar{B}} \times \epsilon_d},$$

where $N_{B\bar{B}}$ is the number of B -meson pairs and ϵ_d the overall reconstruction efficiency in the signal decay channel d . The number of signal events, $N_{\text{sig},d}$, is constrained by the known relative reconstruction efficiencies of the different signal decay channels to fit the branching fraction simultaneously. So there are five free parameters in the fit: The signal branching fraction and the normalizations of the overall background components in the four signal decay channels.

Summarizing, the fit procedure works as following:

- All simulated events are weighted to account for data / MC differences.
- The shapes of the individual components are determined, the off-resonance component in E_{ECL} by a fit, the others using histogram probability density functions (PDFs).
- The relative fractions of the background components are fixed.
- The signal decay channels are fitted individually to get a first estimate of the background normalizations.
- The final, simultaneous fit is performed and the branching fraction is determined.

7.2 Component description

The overall PDF is defined as

$$P(E_{\text{ECL}}, p_{\text{sig}}^*) = \sum_{d=e,\mu,\pi,\rho} \left[N_{\text{sig},d} P_{\text{sig},d}(E_{\text{ECL}}, p_{\text{sig}}^*) + N_{\text{bkg},d} \sum_i c_{d,i} P_{\text{bkg},d,i}(E_{\text{ECL}}, p_{\text{sig}}^*) \right],$$

where d denotes the reconstructed signal decay channel, N_{sig} and N_{bkg} the number of signal and background events, respectively, c the relative fraction of the background components which are constrained by $\sum_i c_i = 1$, and P_{sig} and P_{bkg} the individual PDFs of the signal and background components, respectively. For all components, except the signal in $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ and $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$ the PDFs are the product of two one-dimensional PDFs. This assumes that the distributions in the two fitting dimensions are independent. This assumption has been tested using large samples of simulated data. For small ($< 0.5 \text{ GeV}/c$) and large ($> 2.4 \text{ GeV}/c$) values of p_{sig}^* , there is a sizeable correlation between E_{ECL} and p_{sig}^* in some of the components. Therefore the requirement $0.5 \text{ GeV}/c < p_{\text{sig}}^* < 2.4 \text{ GeV}/c$ has been introduced. The signal component in $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ and $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$ contains a significant amount of cross-feed from decay channels with additional, undetected neutral pions. Their contribution to E_{ECL} and the missing momentum in p_{sig}^* leads to a correlation. This is taken into account by using two-dimensional histogram PDFs to describe these two signal components.

The relative values of N_{sig} are constrained by the relative reconstruction efficiencies, so that effectively the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ is fitted instead of four independent numbers of signal events.

7.2.1 Shapes described by simulated data

All shapes, which are extracted from simulated data, including the signal component and various background components, are described by histogram PDFs. The simulated data is reweighted to account for known discrepancies between MC and data. The following weights are applied on an event-by-event basis:

- The tagging efficiency correction, shown in Table 6.1.
- A significant background arises from events containing semileptonic B decays, where the D meson further decays into hard-to-detect particles and only the lepton is reconstructed (see Sec. 5.1.1). The values of the important D -meson branching fraction in the simulation are reweighted to the best known values [3], given in Table 7.1.

Decay Process	Simulation (10^{-2})	World average (10^{-2})
$D^0 \rightarrow K_L K_L$	0.09	0.09 ± 0.01
$D^0 \rightarrow K_L K_L K_L$	0.04	0.02 ± 0.00
$D^0 \rightarrow K_L \pi^0$	1.13	1.00 ± 0.07
$D^+ \rightarrow K_L e^+ \nu_e$	3.4	4.42 ± 0.11
$D^+ \rightarrow K_L \mu^+ \nu_\mu$	3.4	4.70 ± 0.34
$D^0 \rightarrow K^- e^+ \nu_e$	3.41	3.55 ± 0.05
$D^0 \rightarrow K^- \mu^+ \nu_\mu$	3.31	3.41 ± 0.13

Table 7.1: Dominant background processes and their corresponding branching fractions in the simulation and the world average. The decays with only K_L in the final state have not been measured. There, the branching fractions of the decays to multiple K_S are shown.

The simulated distribution is smoothed with an algorithm described in Reference [29]. The shapes of all components before and after the smoothing for the signal decay $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$ are shown in Figure 7.1 for E_{ECL} and in Figure 7.2 for p_{sig}^* . The other decay channels are shown in Appendix B.

7.2.2 Continuum component shape

The continuum component shape in E_{ECL} is extracted from the off-resonance data. Since the available amount of off-resonance data is limited, it is described with a linear function. The choice of the shape is motivated both by the shape of the distribution itself and the large relative uncertainty of the data points which does not allow a more complicated description. The linear fit is shown in Figure 7.3. The uncertainty on the slope and the normalization of the component will be treated as systematic error. Also an alternative description will then be tested.

7. Branching fraction extraction

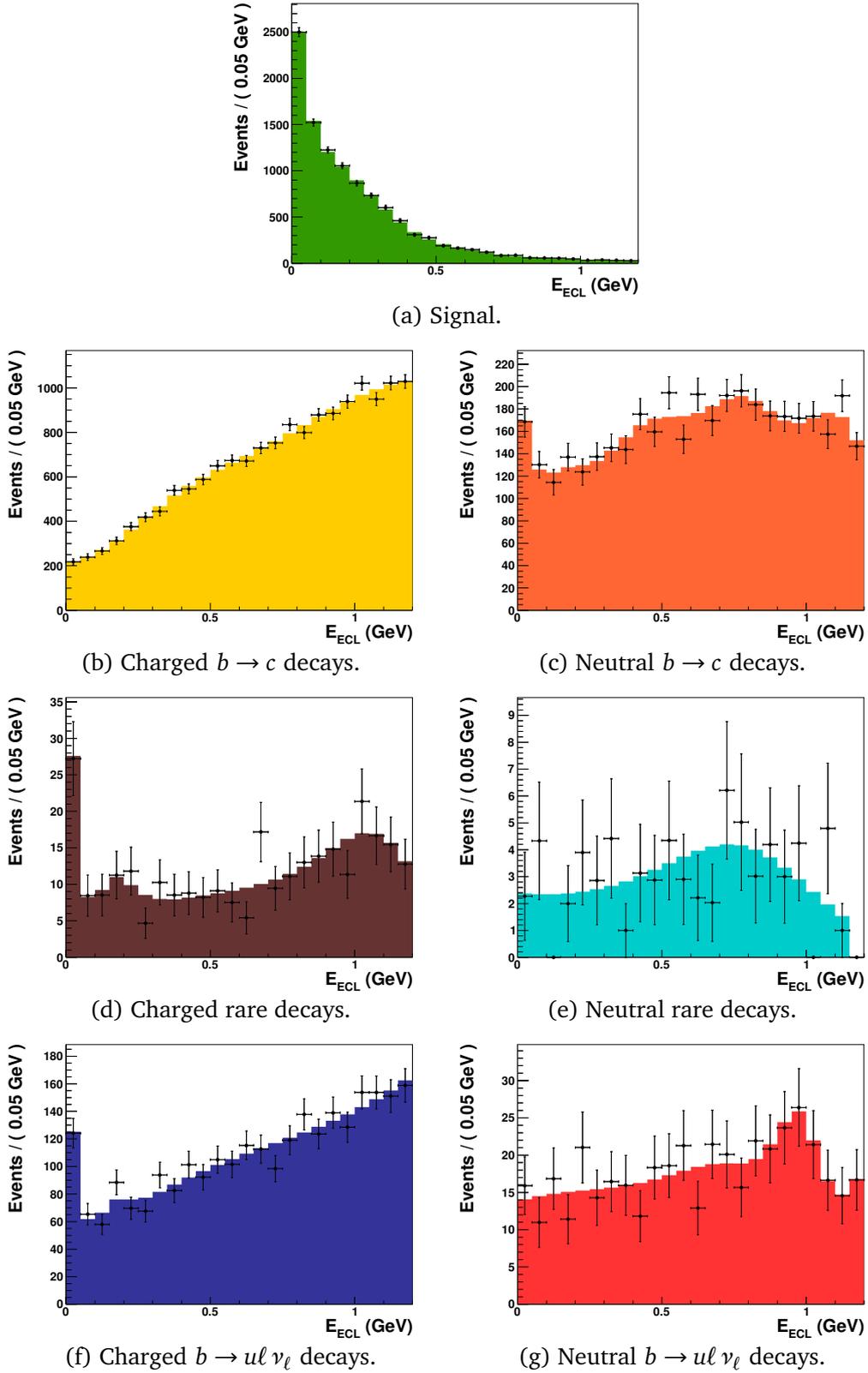


Figure 7.1: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$ in E_{ECL} . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled to the corresponding luminosity.

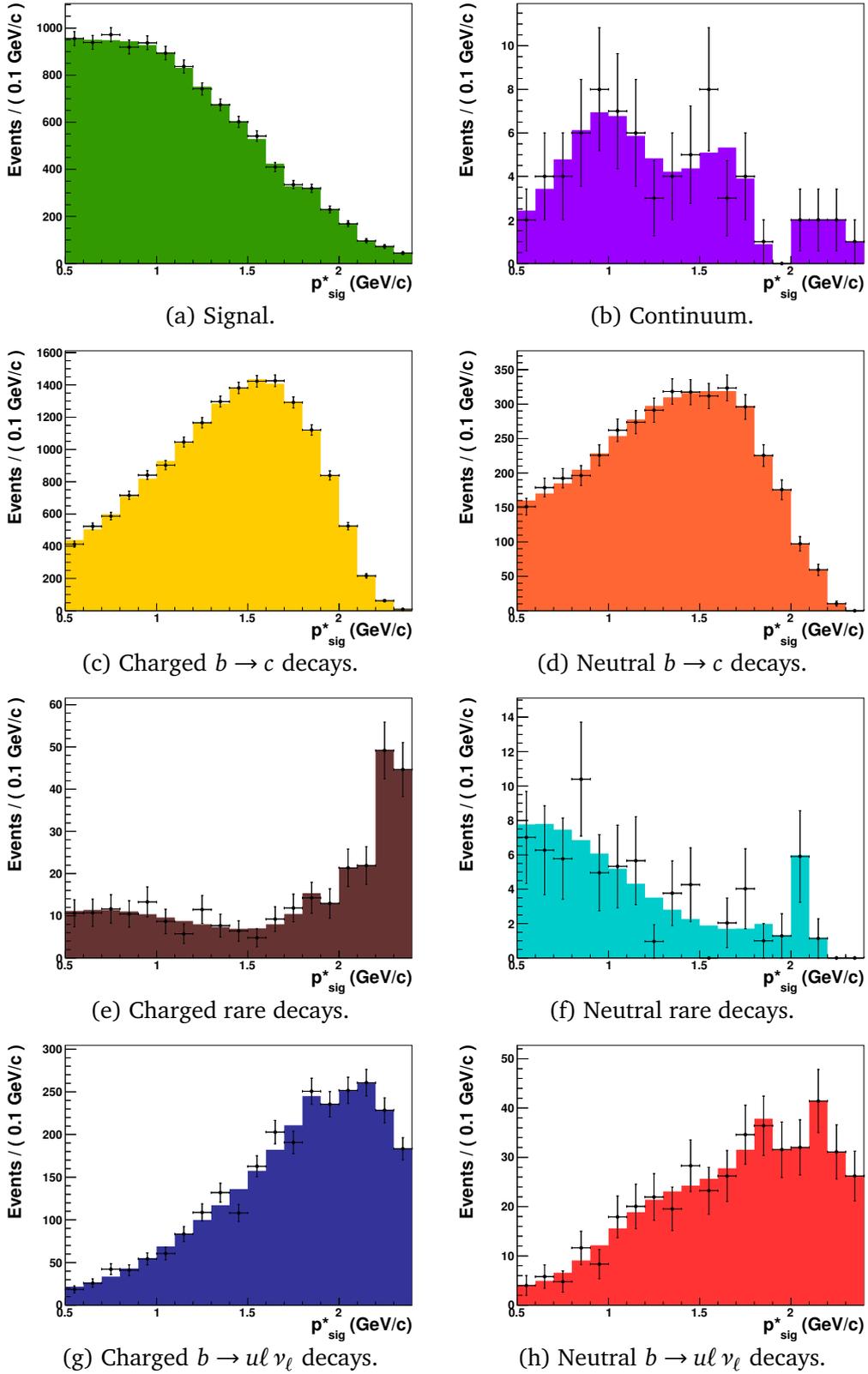


Figure 7.2: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$ in p_{sig}^* . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled to the corresponding luminosity.

7. Branching fraction extraction

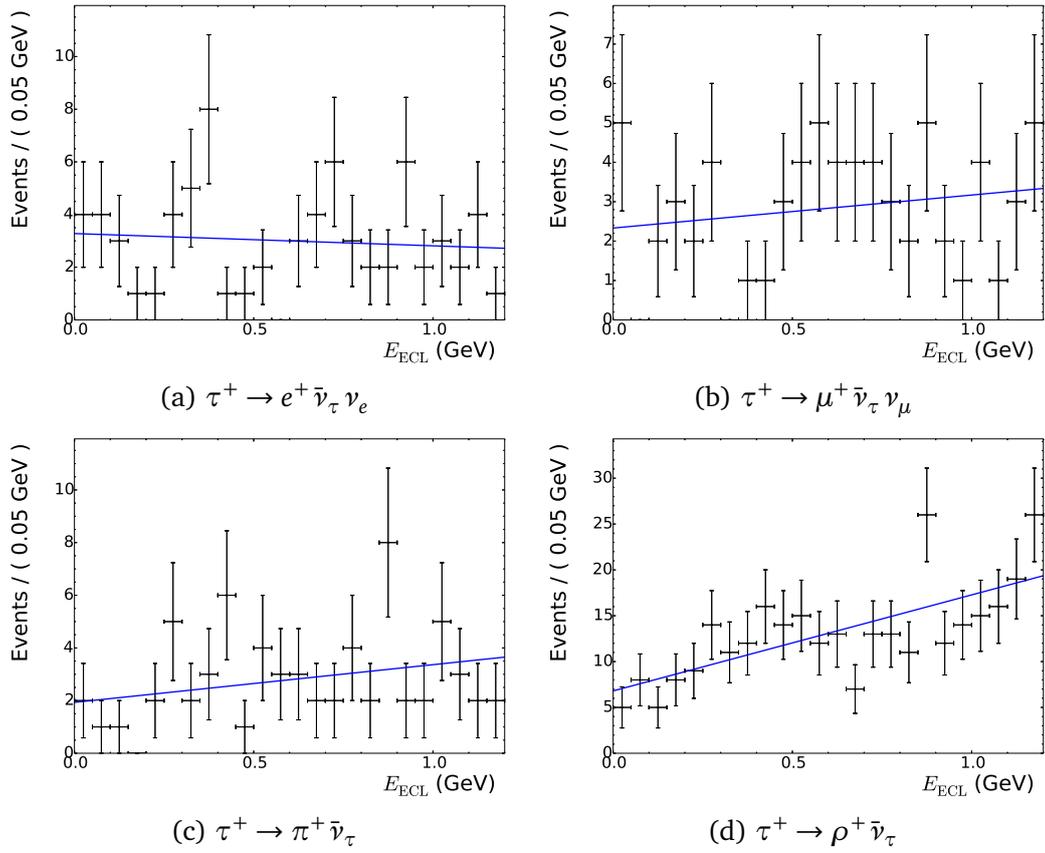


Figure 7.3: Linear fit, which is performed to determine the shape of off-resonance background components in E_{ECL} . The black markers show the data distribution and the blue line the fitted, linear PDF.

7.3 Fit on simulated data

The fit procedure is first tested and validated on MC data. To generate data samples, which represent all relevant processes, the existing samples of generic $b \rightarrow c$ MC, rare MC, and the signal component are combined with a sample generated following the fitted distribution of the continuum component. These contributions are added up according to their cross section. The limiting factor in the available amount of simulated data is the generic $b \rightarrow c$ MC. There are ten data samples available with each of them having an integrated luminosity corresponding to the real data sample. One of them has been used for the optimization of the selection so there are nine samples left to test the fitting procedure. Each of them is used to form one data sample, which is then used to fit, taking the place of the real data sample. The remaining eight generic MC data samples are used to determine the shape of the corresponding background component. The results of the fits are shown in Figure 7.4. The fit reproduces the input value within the uncertainties. An exemplary fit on Sample 4 is shown in Figures 7.5 and 7.6.

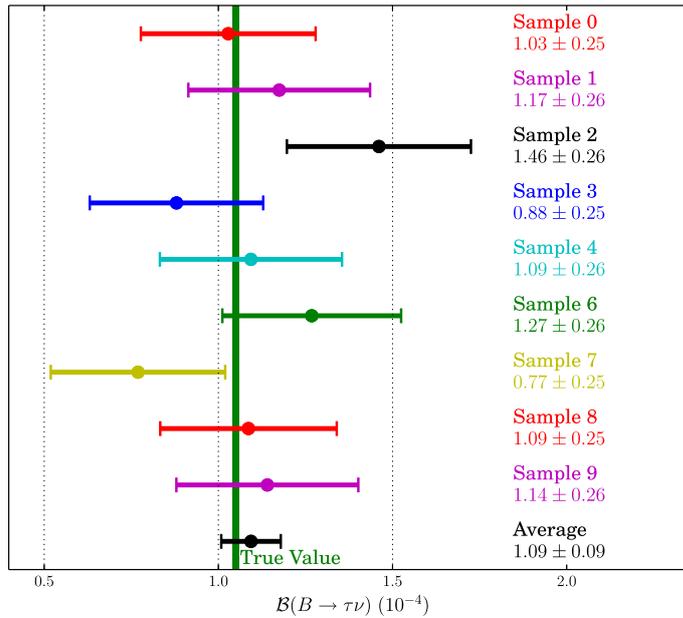


Figure 7.4: Fitted branching fractions in comparison to the input value represented by the green bar. The error bars only show the statistical uncertainty.

7. Branching fraction extraction

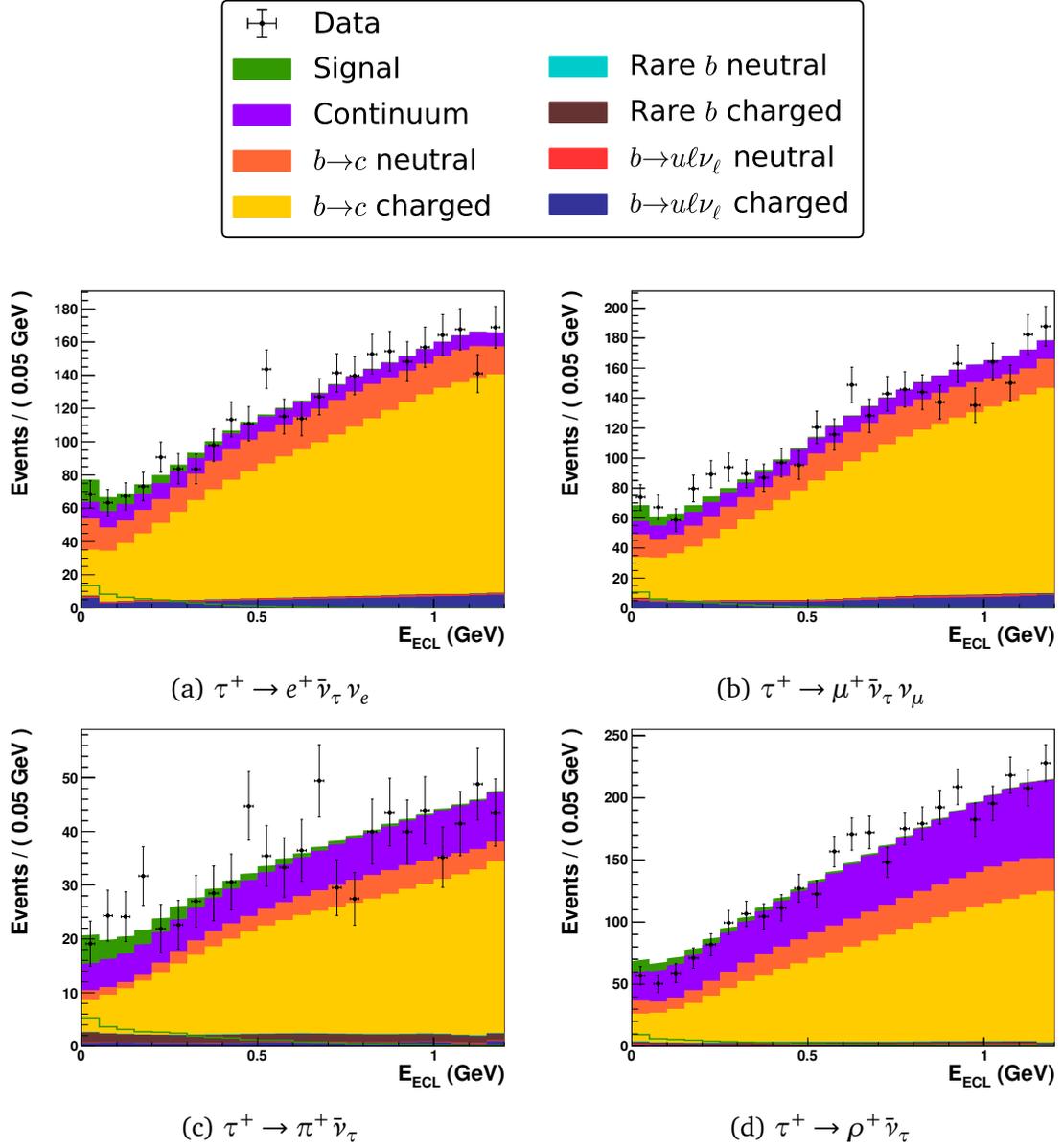


Figure 7.5: Result of the fit on sample 4 in E_{ECL} .

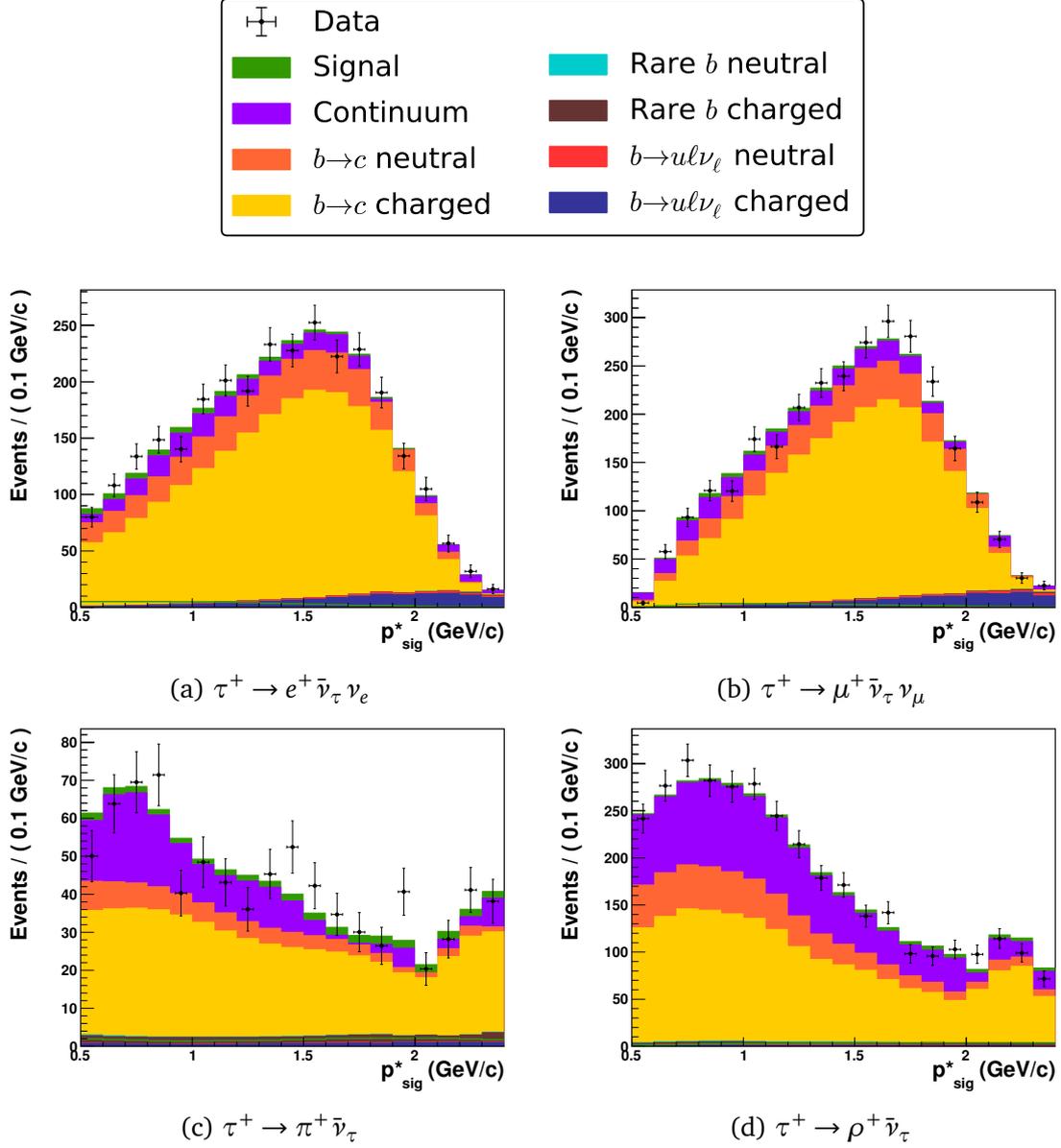


Figure 7.6: Result of the fit on Sample 4 in p_{sig}^* .

7.4 Stability test

In addition to the test using fully simulated events which only deliver a limited amount of data, a test using pseudo experiments is performed. Therefore data samples are generated following the same probability distributions, which are used in the fit. The branching fraction can be set to any desired value. These data samples then are fitted with same PDF which has been used for their generation. So this test cannot verify if the PDFs properly describe the real data distributions, but it can reveal numerical instabilities or biases originating from the fit procedure itself.

The branching fraction is set to multiple values between 0 and 3×10^{-4} , and 500 pseudo-experiments are performed for each value. The distribution of the resulting branching fractions is fitted with a Gaussian distribution and the mean and its uncertainty compared to the true value. The results are illustrated in Figure 7.7 and show that the fit procedure shows no inherent bias over the full range of examined branching fractions.

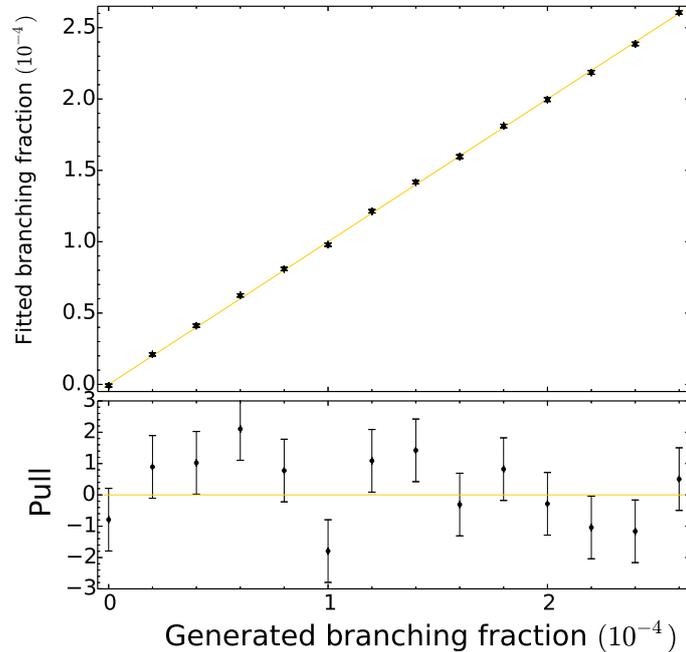


Figure 7.7: Fitted vs. generated branching fraction, determined using pseudo-experiments. The yellow line corresponds to identical generated and fitted values. The lower plot shows the pulls $((\mathcal{B}_{\text{fit}} - \mathcal{B}_{\text{true}})/\sigma_{\mathcal{B}_{\text{fit}}})$.

7.5 Results on data

The fit was performed on the real data set, resulting in a branching fraction of

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (1.25 \pm 0.28) \times 10^{-4}.$$

The log file of the fit can be found in Appendix C. The resulting plots are given in Figures 7.8 to 7.10. The correlation matrix is given in Table 7.2. The signal yields and branching fractions, obtained from fits for each of the τ decay modes separately are given in Table 7.3. The χ^2 value of the hypotheses that the four branching fractions are consistent is 4.75 for 3 degrees of freedom, which results in a probability of 19% to get this or a larger χ^2 value by chance.

	$N_{\text{bkg},e}$	$N_{\text{bkg},\mu}$	$N_{\text{bkg},\pi}$	$N_{\text{bkg},\rho}$
$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$	-0.22	-0.17	-0.26	-0.23
$N_{\text{bkg},e}$		0.04	0.06	0.05
$N_{\text{bkg},\mu}$			0.04	0.04
$N_{\text{bkg},\pi}$				0.06

Table 7.2: Correlation matrix of the fit on the real data sample.

Decay Mode	N_{sig}	$\mathcal{B}(10^{-4})$
$\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$	13 ± 21	0.34 ± 0.55
$\tau^+ \rightarrow e^+ \bar{\nu}_\tau \nu_e$	47 ± 25	0.90 ± 0.47
$\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$	57 ± 21	1.82 ± 0.68
$\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$	119 ± 33	2.16 ± 0.60
Combined	222 ± 50	1.25 ± 0.28

Table 7.3: Signal yields and branching fractions, obtained from fits for the τ decay modes separately and combined.

7. Branching fraction extraction

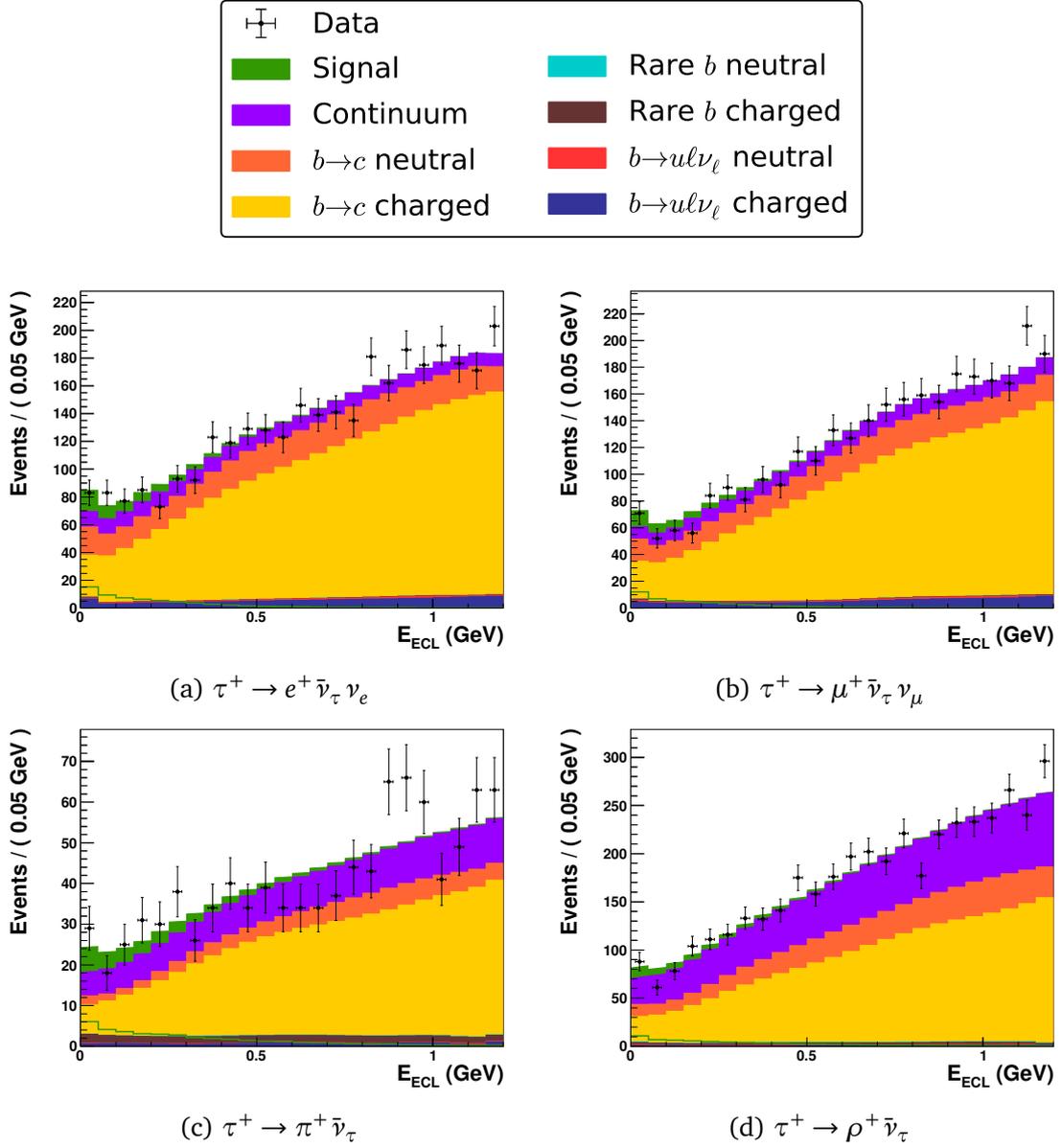
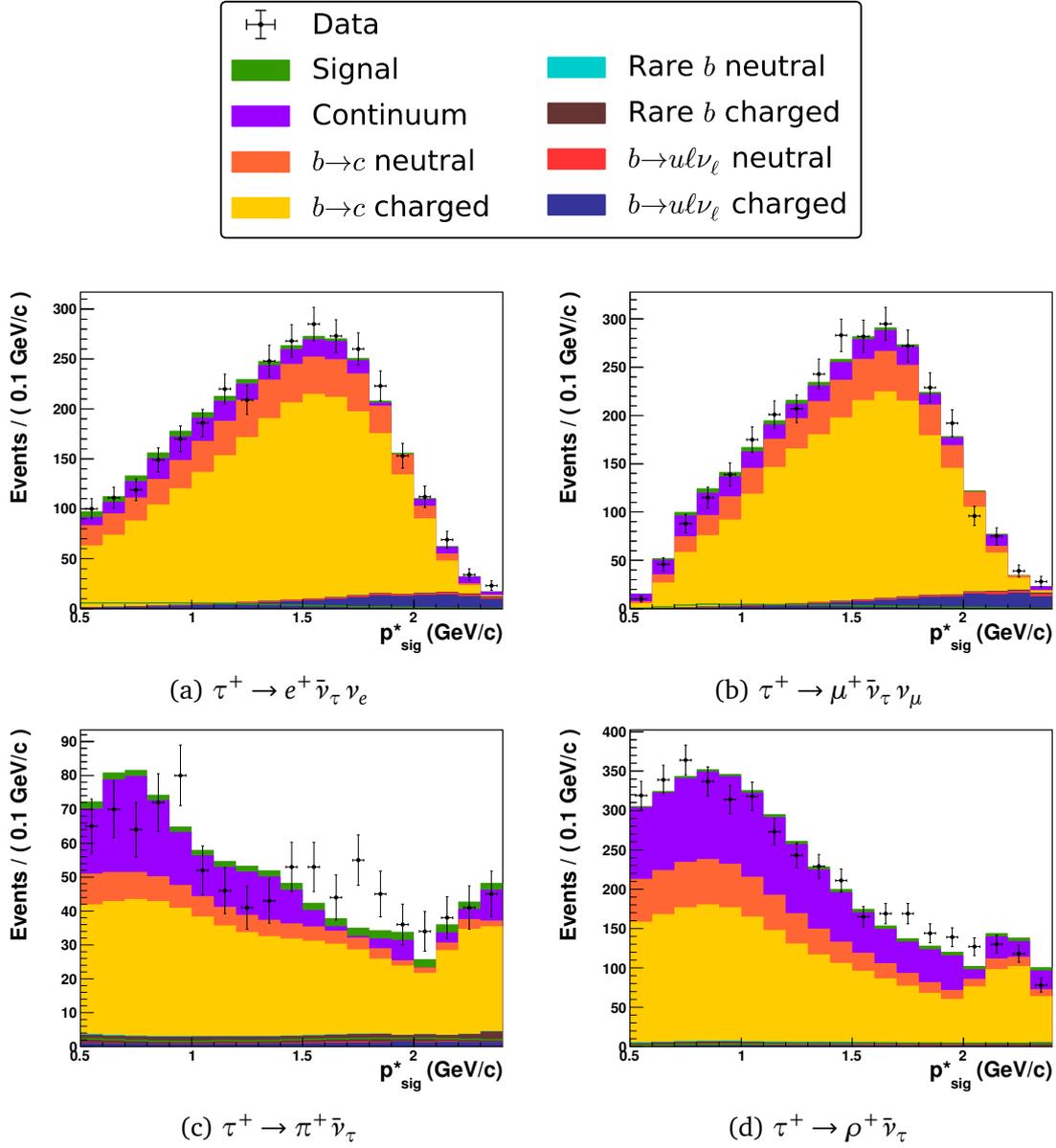


Figure 7.8: Result of the fit on the real data sample in E_{ECL} .

Figure 7.9: Result of the fit on the real data sample in p_{sig}^* .

7. Branching fraction extraction

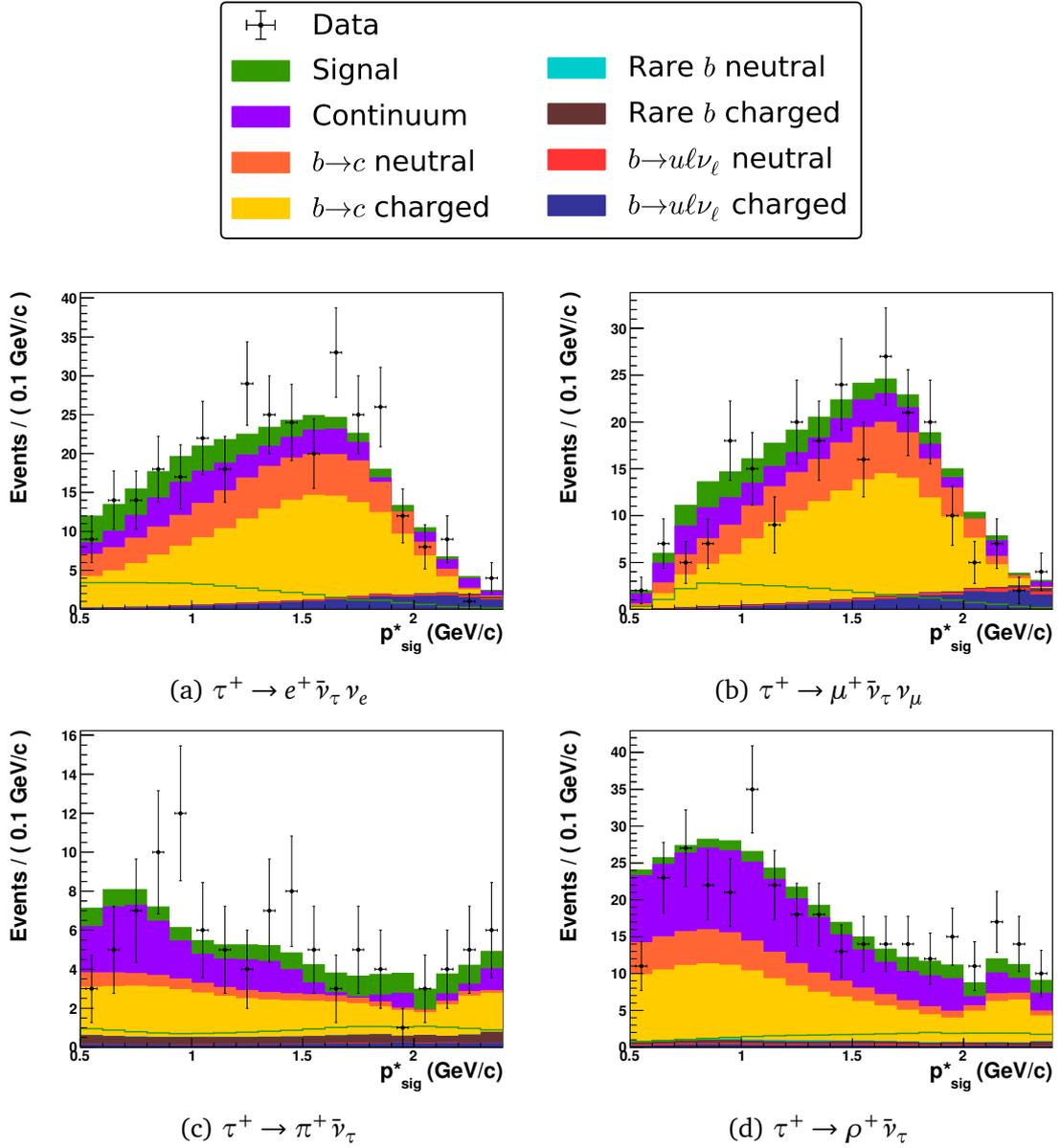


Figure 7.10: Result of the fit on the real data sample in p_{sig}^* , projected in the region $E_{\text{ECL}} < 0.2$ GeV.

7.6 Systematic uncertainties

Potential differences between simulation and real data or insufficient knowledge about certain aspects of the data description are possible sources of systematic errors. Below, all individual points, which could introduce a systematic deviation of the nominal fit result, are listed.

All uncertainties related to the efficiency of the semileptonic tagging, including the tracking efficiency, the unknown intermediate branching fractions, and the selection efficiency are incorporated in the uncertainty of tag-side reconstruction efficiency correction. Therefore, the remaining individual uncertainties only are considered for the signal-side reconstruction.

Efficiency of tracking: The efficiency of the reconstruction of charged tracks has been studied internally by the Belle collaboration using partially reconstructed D^{*+} decays [30]. For this study the decay chain $D^{*+} \rightarrow D^0 \pi^+$, $D^0 \rightarrow K_S \pi^+ \pi^-$, and $K_S \rightarrow \pi^+ \pi^-$ has been reconstructed, where one pion from the K_S decay was explicitly allowed to not be reconstructed. This provided a very clean data sample which was used to compare the efficiency of the reconstruction of charged tracks in MC and data. It has been found to be consistent within 0.35%. Therefore, for the one charged track in all signal decay modes, 0.35% is taken as systematic error.

Efficiency of π^0 reconstruction: Similar to the reconstruction of charged tracks, the π^0 reconstruction efficiency might also be different in simulated and real data. This uncertainty has been determined in an internal study of the Belle collaboration examining the decays of η mesons to multiple neutral pions [31]. The decays $\eta \rightarrow 3\pi^0$ and $\eta \rightarrow \pi^+ \pi^- \pi^0$ have been reconstructed. The reconstruction efficiency of neutral pions then can be compared by forming the double-ratio of these two decay channels and the reconstructed number of events in data and MC. The systematic error has been found to be 4%. Since the ρ^+ signal decay makes up 26% of the total signal, 1.05% is assigned as systematic error.

Efficiency of particle identification: The systematic uncertainty of the efficiency of the particle identification of pions was evaluated according to internal studies of the Belle collaboration [32]. They were performed by reconstructing a clean sample of $D^{*+} \rightarrow D^0 \pi^+$, $D^0 \rightarrow K^- \pi^+$ decays, where the pion and kaon can be identified from their charge. The uncertainty on the selection efficiency of the particle identification was found to be 0.5%. Also the uncertainty on the selection based on the lepton identification is investigated. Therefore the uncertainty on the probability that a lepton is identified as pion is estimated according to the studies by the Belle collaboration [33]. Then also the uncertainty on the probability, that such a lepton is discarded is determined but found to be negligible.

Number of $B\bar{B}$ pairs: The number of produced $B\bar{B}$ pairs is known with an uncertainty of 1.37%.

7. Branching fraction extraction

Histogram PDF shapes: Both the signal component and many of the background components in the fit are described by histogram PDFs. Their shape is affected by statistical fluctuations in the underlying data sample. To take this into account, the histogram content (before the smoothing) is varied in each bin, following a Poisson distribution with the initial value as mean. Then the nominal fit is performed. This procedure is repeated 1000 times and the resulting distribution of fitted branching fractions is fitted with a Gaussian distribution. The width of the distribution is taken as systematic error. This results in a systematic error of 8.5%.

Branching fractions of the τ : The branching fractions of the τ lepton itself have uncertainties, which also may have an impact on the result. To evaluate this, the branching fractions of the dominant τ decay channels in the simulated signal, including the cross feed from $\tau^- \rightarrow \pi^- \pi^0 \pi^0$ are varied by weighting the events and repeating the fits. The resulting individual uncertainties are summed in quadrature which leads to a total error of 0.2%.

Signal reconstruction efficiency: Due to the limited MC statistics, the reconstruction efficiency of signal is known only with an uncertainty. The efficiencies and the errors are shown in Table 5.4. These errors are propagated to the result and lead to an uncertainty of 0.6%.

Continuum description: The shape and the normalization of the continuum background is extracted using the limited dataset of off-resonance data. Therefore the uncertainty in this description is taken into account by varying its parameters by one standard deviation and repeating the fit. The deviations from the nominal fit are given in Table 7.4. The quadratic sum of them is 14.1%.

Also an alternative description for the continuum background is tested. It is fitted with a quadratic instead of a linear function. The deviation from the nominal fit result is negligible compared to the other uncertainty related to the continuum description.

Signal Decay	Slope		Normalization	
	Negative (10^{-6})	Positive (10^{-6})	Negative (10^{-6})	Positive (10^{-6})
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	6.8	7.2	1.7	1.7
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	7.1	7.3	3.2	3.1
$\tau^+ \rightarrow \pi^+ \nu_\tau$	6.7	7.9	0.4	0.5
$\tau^+ \rightarrow \rho^+ \nu_\tau$	10.2	11.2	0.4	0.4

Table 7.4: Deviations from the nominal fit results for variations of the continuum description.

Charged track veto: To estimate the uncertainty of the impact of the charged track veto on the signal efficiency, the double-tagged sample of $B^+ \rightarrow D^0 \pi^+$ decays which only contains signal events is used. It is selected using a charged track veto with a much

narrower cut of $dr < 2$ cm and $dz < 5$ cm. Then the fraction of events with additional charged tracks is compared between data and Monte Carlo. This fraction is directly related to relative efficiency of the nominal charged track veto. The fractions in data and MC are compatible within the uncertainty, so the uncertainty of 1.9% is taken as systematic error.

Background branching fractions: The error arising from the uncertainties of some branching fractions of processes contributing to the background are estimated by varying the branching fractions by the uncertainties and repeating the fit. For the decays $D^0 \rightarrow K_L K_L$ and $D^0 \rightarrow K_L K_L K_L$, where the branching fraction has not been measured, the branching fractions of the decays to K_S have been taken but half of the actual value is taken as uncertainty to account for possible differences.

The variation is done by weighting the background events accordingly. The list of decay processes and the resulting differences of the nominal fit result are shown in Table 7.5. The quadratic sum is an uncertainty of 3.1%.

Process	Negative deviation (10^{-6})	Positive deviation (10^{-6})
$D^0 \rightarrow K_L K_L$	2.03	2.07
$D^0 \rightarrow K_L K_L K_L$	3.00	3.13
$D^0 \rightarrow K_L \pi^0$	0.24	0.23
$D^0 \rightarrow K^- e^+ \nu_e$	0.21	0.21
$D^0 \rightarrow K^- \mu^+ \nu_\mu$	0.53	0.49
$D^+ \rightarrow K_L e^+ \nu_e$	0.13	0.13
$D^+ \rightarrow K_L \mu^+ \nu_\mu$	0.33	0.34

Table 7.5: List of dominant background processes and the errors arising from the uncertainties on their branching fraction.

Tag-side reconstruction efficiency correction: The tag-side reconstruction efficiency is only known with some uncertainty and investigated in Section 6.1. Both the statistical and the systematic uncertainty are taken into account. The systematic error consists of the uncertainty on the $B^+ \rightarrow D^0 \pi^+$ and $B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$ branching fractions and the influence of the ΔM selection on the correction factor for the $B^+ \rightarrow D^{*0} \ell^+ \nu_\ell$ double-tagged samples. The selection is varied in the range where the results are consistent with the $B^+ \rightarrow D^0 \pi^+$ double-tagged sample.

The efficiency corrections are varied and the deviations from the nominal fit result taken as systematic errors. Since these errors are correlated for different signal decay channels, the deviations are added linearly and taken as systematic error. The total uncertainty is 12.6%.

Best candidate selection: To estimate the influence of the best candidate selection, a sample is prepared where the best candidate selection is not applied. This sample is fitted assuming the nominal efficiencies and PDF shapes. The result is divided by the mean multiplicity of 1.07 and compared to the nominal fit result. The difference of 0.4% is taken as systematic error.

7. Branching fraction extraction

Total error: The full list of systematic errors can be found in Table 7.6.

Source	Relative uncertainty (%)
Histogram PDF shapes	8.5
Continuum description	14.1
Signal reconstruction efficiency	0.6
Background branching fractions	3.1
Tag-side reconstruction efficiency correction	12.6
Branching fractions of the τ	0.2
Best candidate selection	0.4
Efficiency of tracking	0.4
Efficiency of π^0 reconstruction	1.1
Efficiency of particle identification	0.5
Charged track veto	1.9
Number of $B\bar{B}$ pairs	1.4
Total	21.2

Table 7.6: Full list of systematic errors.

The final result is given by

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (1.25 \pm 0.28 \pm 0.27) \times 10^{-4},$$

where the first error is statistical and the second systematic.

7.7 Significance of the result

The significance of the result denotes the confidence to exclude the hypotheses of no $B^+ \rightarrow \tau^+ \nu_\tau$ decays. To evaluate this significance, a likelihood ratio test is performed. According to Wilk's Theorem [34], the ratio of the logarithmic likelihood values of two hypotheses asymptotically follows a χ^2 distribution. The difference in free parameters of the two hypothesis is the number of degrees of freedom in the χ^2 distribution. By calculating the quantile of the χ^2 distribution the probability to measure the present values or some value further away from the null hypotheses if the null hypothesis is true can be determined. This probability can be converted to units of standard deviations of the corresponding normal distribution, covering the same probability (σ). In the case, where the difference of free parameters of the two hypotheses is one, the according calculation simplifies to

$$\Sigma = \sqrt{2 \ln(L_{\max}/L_0)},$$

where L_{\max} and L_0 are the values of the likelihood functions when the signal yield is allowed to vary or set to zero, respectively.

To take into account the systematic error, the likelihood curve of the fit is numerically convoluted with a Gaussian distribution with a width corresponding to the systematic error. The branching fraction depends on the efficiency, the number of $B\bar{B}$ pairs, and the signal yield. To get the significance in falsifying the hypothesis of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$ being zero, the total systematic uncertainty needs to be taken into account. But to falsify the hypotheses of no $B^+ \rightarrow \tau^+ \nu_\tau$ decays, only the systematic uncertainty related to the signal yield is of interest, while the components related to the efficiency and $N_{B\bar{B}}$ do not impact the confidence. Therefore, the convolution is done for the total uncertainty and only for the components related to the signal yield. These components are the the histogram PDF shapes, the continuum description, and the background branching fractions.

The likelihood curve is obtained by fixing the branching fraction to 1600 values between -10^{-4} and 3.5×10^{-4} and maximizing the likelihood at each of them. The difference of the logarithmic likelihood curves and the corresponding minima ($\Delta\mathcal{L}$) before and after the convolution are shown in Figure 7.11. The obtained significance is 3.8σ . Additional numbers, including the corresponding probability, are given in Table 7.7.

Error	$-2 \ln\left(\frac{L_{\max}}{L_0}\right)$	Probability	in σ
Statistical	22.6	$2.0 \cdot 10^{-6}$	4.8
Statistical and systematic (signal yield)	16.7	$4.4 \cdot 10^{-5}$	4.1
Statistical and systematic	14.2	$1.6 \cdot 10^{-4}$	3.8

Table 7.7: Probabilities and statistical significances of the result.

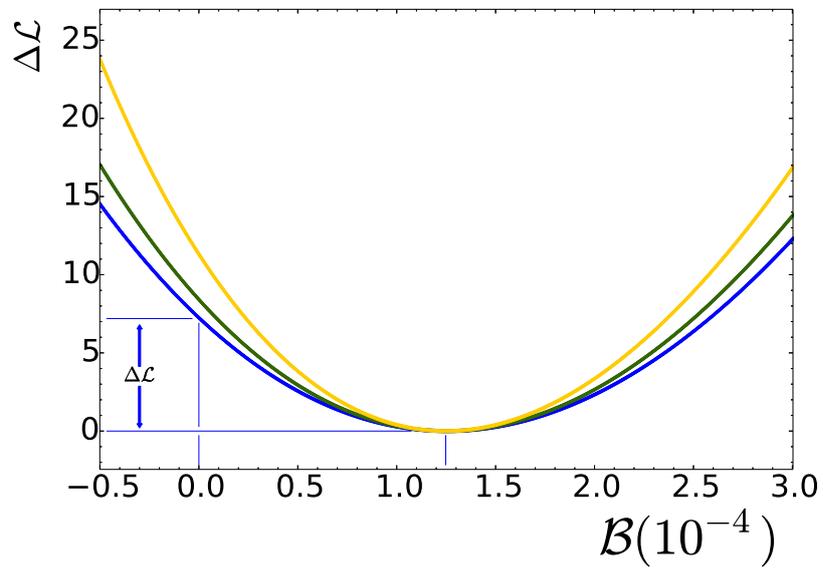


Figure 7.11: Curves of $\Delta\mathcal{L}$ before (orange) and after the convolution with the systematic uncertainty (blue) and only the systematic uncertainty related to the signal yield (green).

8. A track-finding algorithm for the Belle II central drift chamber

In this chapter, the development of a track-finding algorithm for the Belle II central drift chamber (CDC) is described. It is inspired by an algorithm for the ATLAS detector described in Reference [35] but does not rely on any previously established implementation of a track-finding technique. It works in the context of the Belle II software framework and is part of the overall tracking effort within the Belle II collaboration.

The aim of the development was the implementation of a working version of the track finding, given the computational time constraints, and the estimation of its potential capabilities.

Section 8.1 gives a very brief introduction to the Belle II experiment and its CDC, Section 8.2 explains the developed algorithm, and Section 8.3 shows an estimation of its performance. Within Section 8.2, the underlying formalism is explained in Sections 8.2.1 and 8.2.2, while the original works of this thesis, their application and implementation in an efficient way, is described in Sections 8.2.3 to 8.2.5.

8.1 The Belle II experiment

Belle II is the successor to the Belle experiment. It is currently being constructed at the SuperKEKB accelerator which is the replacement for the KEKB accelerator, and it is planned to start collecting data in 2017. The goals of the upgrade from Belle to Belle II are a significantly higher luminosity and considerably better measurement capabilities. This will be achieved by a new detector concept which includes new and redesigned detector components and a completely newly developed software for the detector and analysis of the data. A detailed description of the detector can be found in Reference [36]. The intended measurements and the underlying physics are described in Reference [37].

Part of the detector concept is the new CDC. Its main purpose is the measurement of charged particles. The ionization of the gas in the CDC is measured by so-called sense

8. A track-finding algorithm for the Belle II central drift chamber

wires. Therefore, an electric field is induced by a large number of field wires which accelerate and multiply the free charge carriers, whose charge is measured in the sense wires. The time between the collision and its measurement is directly related to the distance between track and wire.

The design of the Belle II CDC is similar to the CDC in the Belle detector, but the parameters are different. The main parameters are given in Table 8.1. A section of the CDC is illustrated in Figure 8.1.

The wires in the CDC are organized in concentric layers around the interaction region. Each layer consists of 160 to 384 wires. The layers are combined in superlayers. There are two different kinds of superlayers depending on their angle to the beam line. A superlayer with wires parallel to the beam is called axial layer. The other kind of superlayers, stereo layers, consist of wires with a small angle of about 70 mrad to the beam direction.

A magnetic field of 1.5 T is induced by a solenoid magnet surrounding the CDC. Its field lines are parallel to the beam axis and therefore the trajectories of charged particles are changed to circles or circular arcs in the plane orthogonal to the beam. This allows for the determination of the transverse component of their momentum from the curvature of the track.

The actual information delivered by the CDC which is related to the detection of charged particles consists of the activated wires and the corresponding time difference between the collision and the measurement. By assuming a relation between the time and the covered distance this information can be translated to circles around wires, denoted as drift circles or as hits. This information is illustrated in Figure 8.2.

	Belle	Belle II
Inner radius (mm)	77	160
Outer radius (mm)	880	1130
Number of layers	50	56
Number of sense wires	8400	14336

Table 8.1: Main parameters of the Belle and Belle II CDC.

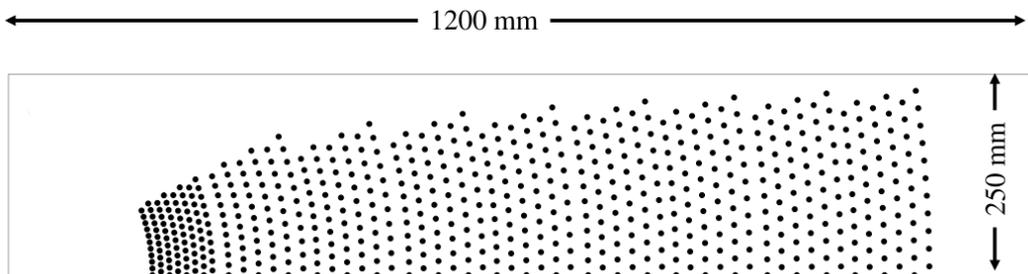


Figure 8.1: Section of the CDC. Taken from Reference [36].

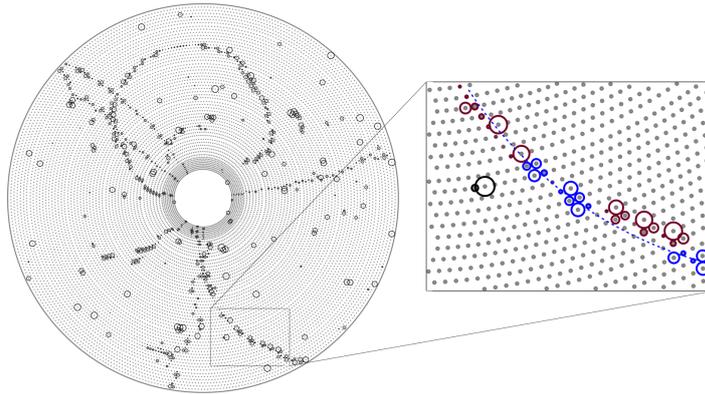


Figure 8.2: Illustration of the information delivered by the CDC. The enlarged area shows drift circles from axial (stereo) layers in solid, blue (red) circles and the tangential trajectory of the corresponding particle in a dashed, blue line.

8.2 Track finding

The data delivered by the CDC does not include information on which drift circle is induced by which particle. Therefore, the combination of drift circles to a track needs to be performed by an algorithm. Such a method is called *pattern recognition* since it exploits the fact that a track in the CDC follows a certain pattern: because of the magnetic field, it forms a circle tangential to the drift circles. This is illustrated in the enlarged area in Figure 8.2.

The task must be fulfilled under computational time constraints. The considerations and assessments of the Belle II collaboration lead to the condition that the processing of a single event, thus the recognition of about 10 tracks from some hundred drift circles, must not exceed 50 milliseconds using a current CPU. While staying within this constraint, the algorithm shall maximize the efficiency in finding tracks.

A coordinate system is established: the direction of the beam is taken as z -axis, the horizontal direction as x -axis, and the vertical direction as y -axis. Only the xy -plane is considered for the track finding and in a first step only the axial wires are taken into account since the position of the stereo wires in the xy -plane is ambiguous due to their angle with respect to the beam direction. The hits originating from stereo wires are afterwards assigned to the track candidates.

The track-finding algorithm consists of these steps:

1. The drift circles are transformed to the inverted plane.
2. The drift circles are further transformed into curves in the Legendre space.
3. A voting procedure in the Legendre space is performed and clusters are identified.
4. The stereo hits are assigned to the track candidate.
5. The properties of the particle are estimated.

Each of these steps is described in the following.

8.2.1 Inversion transformation

Before the actual track finding, a coordinate transformation, denoted as inversion, is performed. It is more feasible to search for tracks in the inverted plane as they are transformed to straight lines. The inversion is defined as

$$\begin{aligned}x' &= \frac{2x}{x^2 + y^2} \\y' &= \frac{2y}{x^2 + y^2}.\end{aligned}$$

This non-linear transformation translates circles with the center (x_0, y_0) and radius R fulfilling

$$(x - x_0)^2 + (y - y_0)^2 = R_0^2$$

to circles with

$$\left(x' - \frac{2x_0}{x_0^2 + y_0^2 - R_0^2}\right)^2 + \left(y' - \frac{2y_0}{x_0^2 + y_0^2 - R_0^2}\right)^2 = \left(\frac{2R_0}{x_0^2 + y_0^2 - R_0^2}\right)^2.$$

Consequently, the parameters of the circle in the inverted plane are given by

$$\begin{aligned}x'_0 &= \frac{2x_0}{x_0^2 + y_0^2 - R_0^2} \\y'_0 &= \frac{2y_0}{x_0^2 + y_0^2 - R_0^2} \\R'_0 &= \frac{2R_0}{x_0^2 + y_0^2 - R_0^2}.\end{aligned}\tag{8.1}$$

This translation is not valid for a circle passing through the origin at $(x, y) = (0, 0)$, as $x_0^2 + y_0^2 - R_0^2 = 0$. With this constraint, a circle is translated to

$$x'x_0 + y'y_0 = 1,$$

which is the equation for a straight line.

The transformation is illustrated in Figure 8.3. Consequently, the search for straight lines implicitly assumes that all tracks are circular and pass through the origin.

8.2.2 Legendre transformation

Furthermore, each drift circle in the inverted plane is translated to the Legendre space. The defining property of this space in the context of this work is the fact that it allows to parametrize all possible tangents to a circle as a set of two functions. Thus, a curve in the Legendre space corresponds to a collection of lines in the original coordinate system. The information in this section is taken from Reference [35], where a more detailed

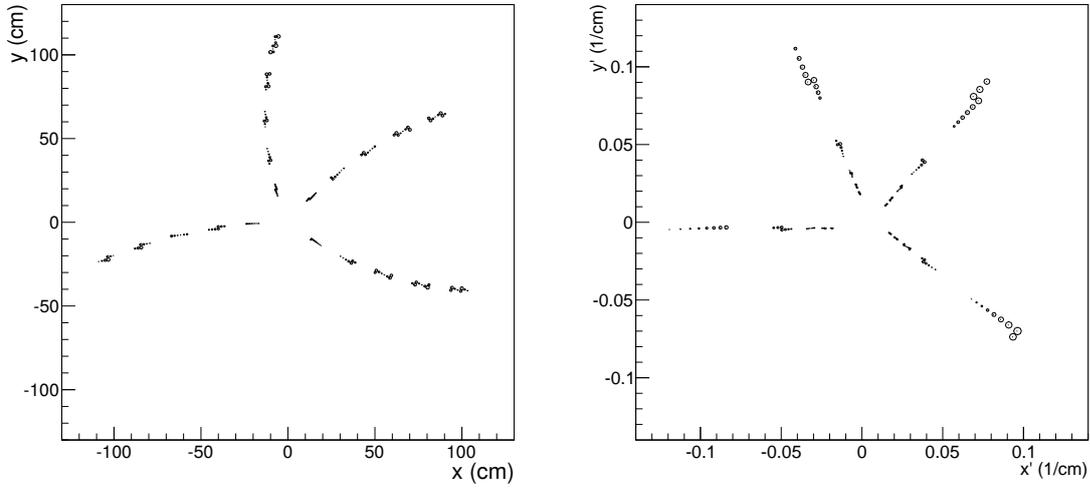


Figure 8.3: Illustration of the inversion transformation. The left and right plots show the drift circles of four tracks before and after the inversion, respectively.

description of the Legendre transformation and its application in track finding can be found. The parameters of the Legendre space are denoted as r and θ , where r is the distance of the line to the point $(0, 0)$ and θ the angle between the line and the x' -axis. All possible tangents to a drift circle are given by the two curves

$$\begin{aligned} r_{(1)} &= x'_0 \cos \theta + y'_0 \sin \theta + R'_0 \\ r_{(2)} &= x'_0 \cos \theta + y'_0 \sin \theta - R'_0, \end{aligned} \quad (8.2)$$

where x'_0 , y'_0 , and R'_0 are the parameters of the circle in inverted space given by Equation 8.1. The relation between a point in the Legendre space and the corresponding tangent to a circle is illustrated in Figure 8.4. Furthermore, the left plot in Figure 8.5 shows the two curves representing all possible tangents to a single drift circle. Since a track candidate needs to be a tangent to all contributing drift circles there is one point in the Legendre space in which the curves from all contributing drift circles intersect. The right plot of Figure 8.5 shows such a collection of curves for four tracks where the four intersections can be seen clearly.

A given point in the Legendre space represents a straight line in the inverted plane and therefore a circle through the origin in normal space. The relation to the parameters of the circle is given by

$$\begin{aligned} R &= 1/|r| \\ \varphi &= \text{sign}(r) \times \theta, \end{aligned} \quad (8.3)$$

where R is the radius of the circle and φ the angle between the circle and x -axis at the origin.

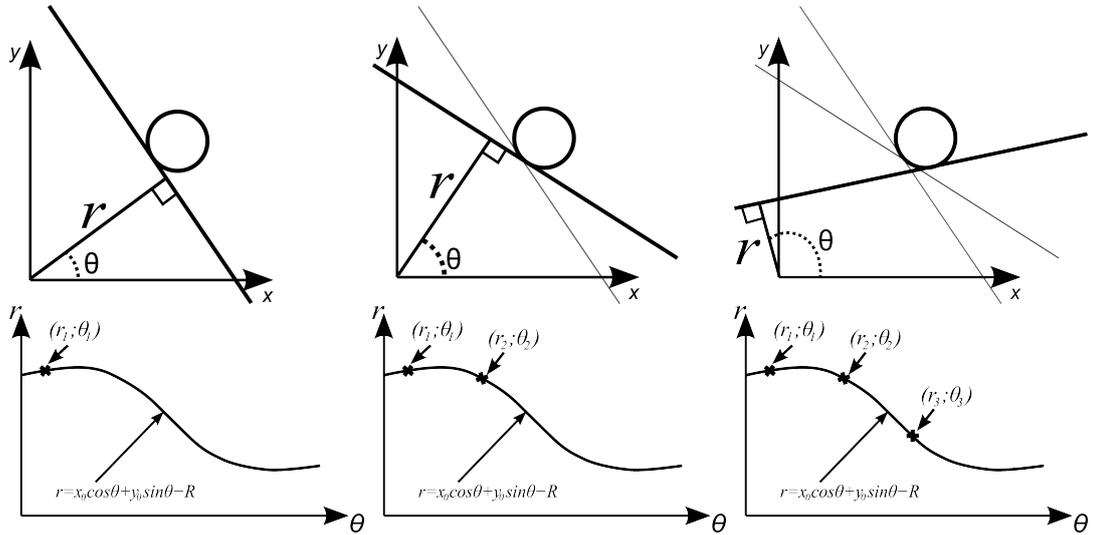


Figure 8.4: Illustration of the relation between a tangent to a circle (upper plots) and the corresponding point in the Legendre space (lower plots) [38].

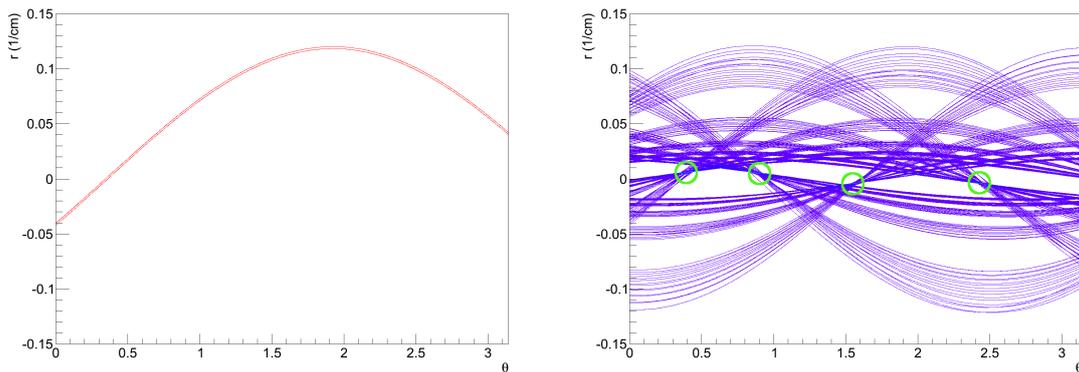


Figure 8.5: Illustration of the translation of circles to the Legendre space for a single drift circle (left) and for four tracks (right).

8.2.3 Voting and peak finding

The problem of finding the intersection of a large number of sinusoidal curves cannot be solved analytically within the given computational constraints. Therefore the θr -plane is divided into bins and taken as a two-dimensional histogram. The ranges of θ and r are given by the potential properties of the measured particles. The values of θ between 0 and π and values of r between -0.15 and 0.15 cm^{-1} are considered. This corresponds to a full coverage in φ and a minimal transverse momentum of $30 \text{ MeV}/c$. Particles with a smaller transverse momentum do not reach the CDC but curl within the inner parts of the detector.

The curves given in Equation 8.2 are used to perform a voting procedure. For each drift circle each value of θ is taken, the equation is used to calculate the according values of r , and the corresponding entries in the histogram are increased by one. This results in a two-dimensional distribution of integer values with the highest values corresponding to the most likely tangents to the drift circles. Therefore this process is often denoted as *voting*.

The next step is to find the correct values of r and θ . Each track candidate is related to a peak in the two-dimensional histogram. The identification of peaks and clusters is a non-trivial task as the requirement to find as many tracks as possible stands in tension to the danger of finding a single track multiple times. A very good solution for that problem was found in an iterative ansatz. The voting is performed, the track candidate with the highest number of contributing hits in the CDC is searched for, the corresponding hits are removed, and it is started over. Since this involves repeated voting and peak-finding, a very fast implementation of both of them is required.

For the voting, this is achieved by using a lookup table for the calculation of $\sin \theta$ and $\cos \theta$. By binning the θ -dimension beforehand all occurring values are known and the results of the sin and cos calculation can be determined once and repeatedly for each voting procedure.

To identify the entry with the highest content in a two-dimensional histogram might seem like a trivial task, but the trivial approach, just testing all bins, is far too slow. Therefore, an algorithm similar to a Fast Hough finder [39] is applied. For this algorithm, the θr -plane is divided in only 2×2 bins. The voting is performed and bins with a number of contributing hits lower than a certain threshold are not treated further. This is repeated for each of the kept bins, which are further divided in 2×2 bins, until the binning is fine enough. By neglecting a significant part of the parameter space, much computation time is saved. It can safely be neglected as only the bin with the highest number of entries is of interest. The algorithm is illustrated in Figure 8.6. The consecutive character of the computation makes it possible to further optimize the procedure. As soon as an appropriate track candidate is found, the number of contributing hits can be used as new threshold for the ongoing voting procedure.

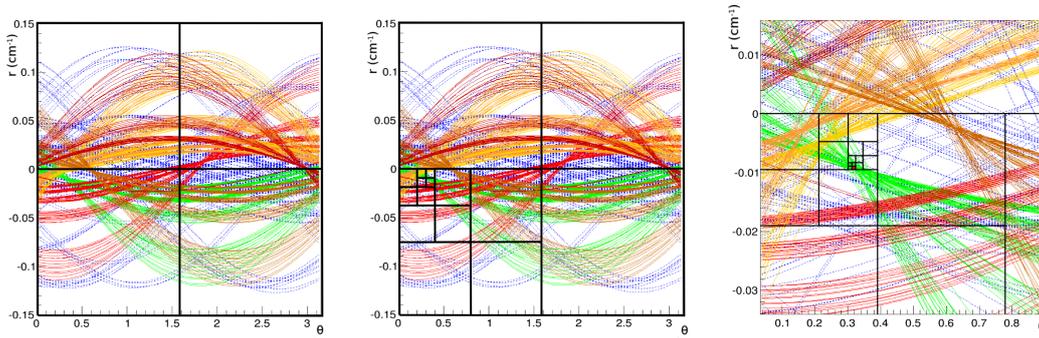
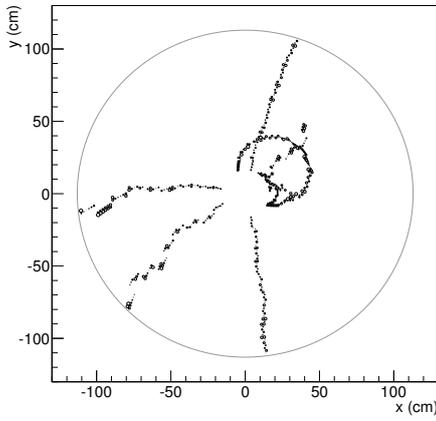


Figure 8.6: Step 1, 5, and 10 of a single instance of the peak finder [38].

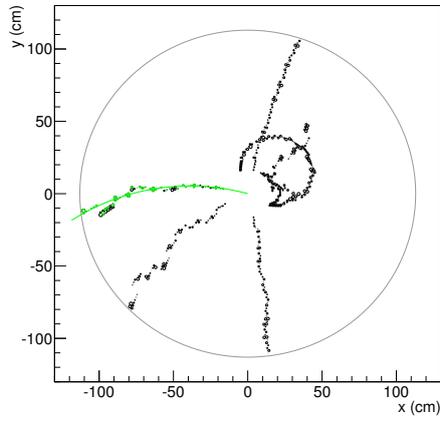
The overall procedure for the track finding in an event is organized as follows:

1. An initial threshold for the voting procedure is taken.
2. The track candidate is searched for.
3. If a track candidate is found, the contributing hits are removed from the overall list of hits.
4. If no candidate is found for the given threshold it is tested if it is too low to continue. If yes, the peak-finding is finished.
5. The threshold is scaled by a certain factor < 1 and Step 2 is repeated.

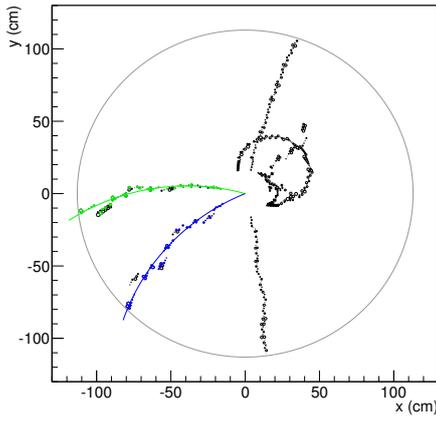
This iterative procedure is illustrated in Figure 8.7. The initial threshold for the peak finding is chosen to be 30. In steps 2 to 7 a track candidate is found and the new threshold is given by the number of hits contributing to the new track candidate multiplied by the scaling factor of 0.75. Then, in steps 8 and 9 no track candidate is found and the threshold is scaled down by 0.75. The peak finding and the adjustment of the threshold is continued until no new track candidate with at least 10 contributing hits is found.



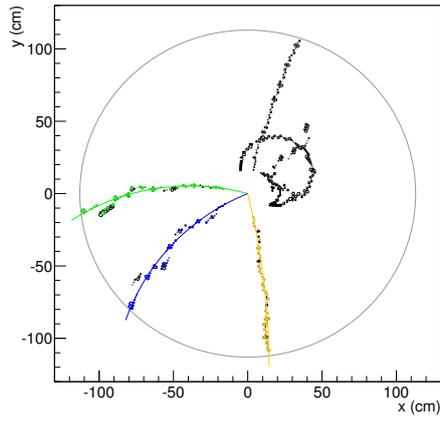
(1) Threshold: 30 / $N_{\text{hits}}(\text{Track})$: -



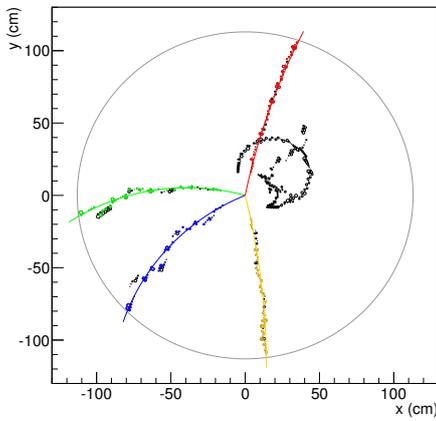
(2) Threshold: 30 / $N_{\text{hits}}(\text{Track})$: 40



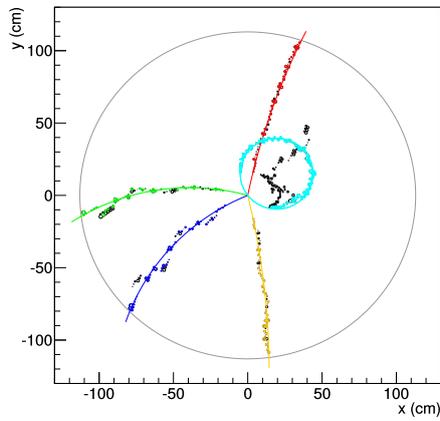
(3) Threshold: 30 / $N_{\text{hits}}(\text{Track})$: 39



(4) Threshold: 30 / $N_{\text{hits}}(\text{Track})$: 36

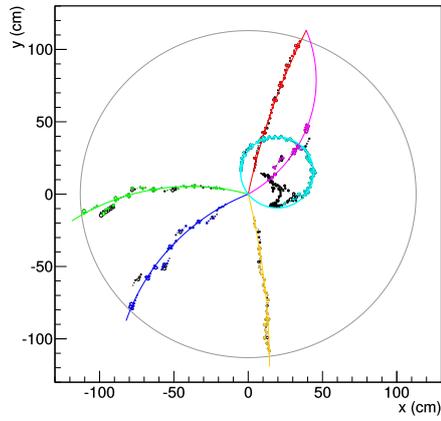


(5) Threshold: 27 / $N_{\text{hits}}(\text{Track})$: 35

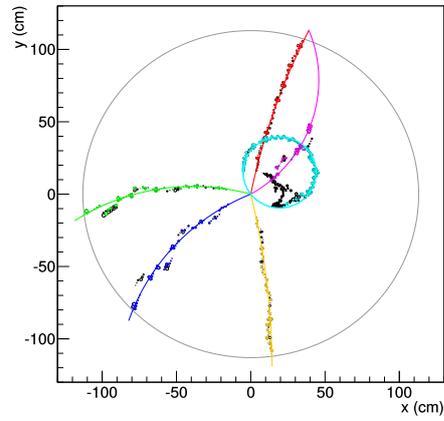


(6) Threshold: 27 / $N_{\text{hits}}(\text{Track})$: 33

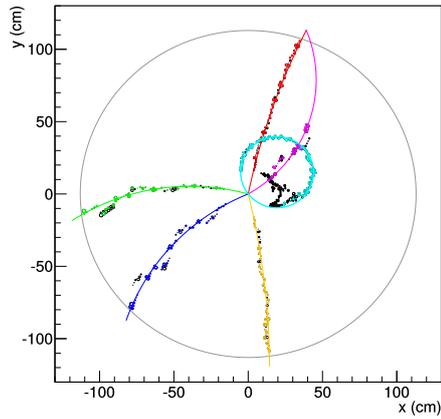
8. A track-finding algorithm for the Belle II central drift chamber



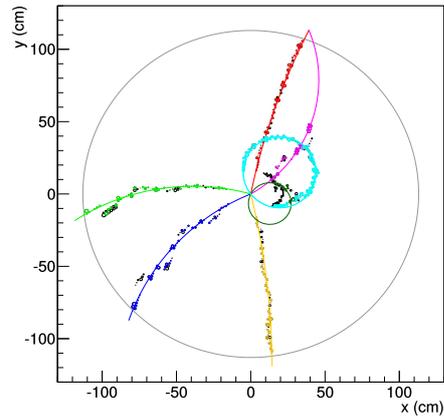
(7) Threshold: 25 / $N_{\text{hits}}(\text{Track})$: 25



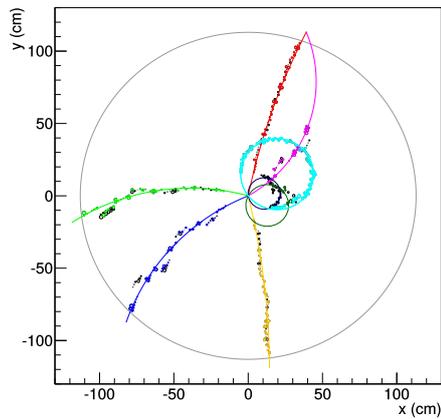
(8) Threshold: 19 / $N_{\text{hits}}(\text{Track})$: -



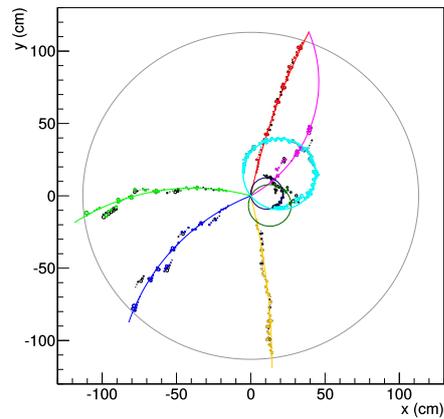
(9) Threshold: 15 / $N_{\text{hits}}(\text{Track})$: -



(10) Threshold: 10 / $N_{\text{hits}}(\text{Track})$: 14



(11) Threshold: 10 / $N_{\text{hits}}(\text{Track})$: 11



(12) Threshold: 10 / $N_{\text{hits}}(\text{Track})$: -

Figure 8.7: Illustration of the track-finding procedure.

8.2.4 Stereo-hit assignment

The contributing axial hits to a track candidate are given by the track-finding procedure itself, but the stereo hits need to be assigned afterwards. Their xy -position is not unambiguous as it depends on the z -component of the momentum of the particle. The position is given by

$$\vec{x}_{\text{hit}} = \vec{x}_{\text{wire}_{\text{begin}}} + t(\vec{x}_{\text{wire}_{\text{end}}} - \vec{x}_{\text{wire}_{\text{begin}}}), \quad (8.4)$$

where t is a number between 0 and 1 and $\vec{x}_{\text{wire}_{\text{begin/end}}}$ are the xy -positions of the corresponding wire at its minimal and maximal z -position. A stereo hit can be assigned to a track candidate if its xy -position can be varied, by altering t in Equation 8.4, so that the distance of its drift circle and the track candidate can become compatible with 0 within the measurement resolution.

The distance of two circles is given by

$$d = \left| \sqrt{(x_{\text{track}} - x_{\text{hit}})^2 + (y_{\text{track}} - y_{\text{hit}})^2} - (r_{\text{track}} \pm r_{\text{hit}}) \right|. \quad (8.5)$$

Inserting Equation 8.4 in Equation 8.5, the equation $d = 0$ can be solved analytically and can have 0, 1, or 2 solutions for t . Solutions with values of $t < 0$ or $t > 1$ are discarded. This is done for each combination of track candidates and stereo hits near enough to potentially intersect with each other. If at least one suitable solution is found, the hit is assigned to the track candidate. If a hit could be matched to multiple track candidates it is assigned to the one with the highest number of contributing axial hits.

8.2.5 Parameter estimation

Besides the contributing hits, more information about the track candidates is required as starting value for the following parameter estimation of the tracks which uses a more elaborate model for their shape.

The x - and y -components of the momentum are given by Equation 8.3, where the transverse momentum of the particle and the radius of its track are directly related by the strength of the magnetic field.

The z -component of the momentum can only be deduced from the assigned stereo hits. With the assumption that the tracks originate from the point (0,0), in principle one stereo hit would be enough to give the z -component of the momentum as its z -position is known after the assignment to the track candidate. It can be calculated with

$$p_z = \tan(\rho)p_t$$

$$\rho = \text{atan2}(z, R),$$

where R and z are the radial and z -component of the position of the stereo hit. But there is limited experimental resolution and there are wrongly assigned hits and thus a single hit is not enough. Therefore the median of all assigned hits is taken. The median is chosen over the mean as it is faster to calculate and most of all robust against outliers which often occur for randomly assigned hits.

The charge can be determined from the curvature of the track, but only if the track does not form a full circle in CDC. So the charge of curling tracks cannot be determined easily within the context of this track finder. If the shape of the track is a circular arc, as it is the case for particles with a momentum high enough, the curvature can be determined from the individual hits. This is done by comparing the φ -angle of the center of track circle and the center of the drift circle. This is illustrated in Figure 8.8. The difference $\varphi_{\text{track}} - \varphi_{\text{hit}}$ is constrained to be in the range $[0, 2\pi]$ and if it is larger than π the curvature is related to a positively charged particle and vice versa. This is performed for each axial hit of a track candidate and the majority of the charge assumptions determines the charge of the track candidate.

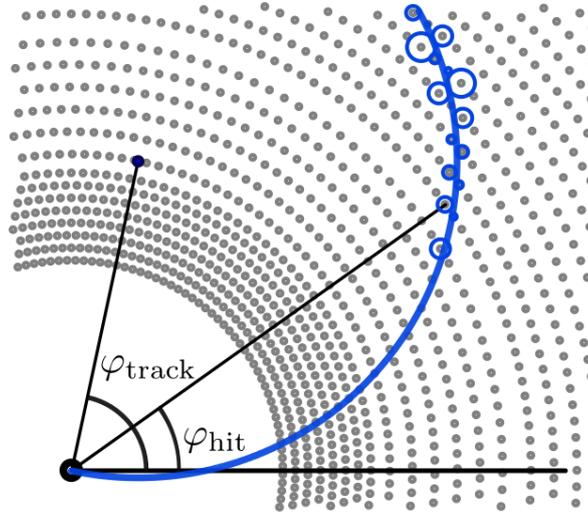


Figure 8.8: Illustration of the φ -angles used to determine the curvature of a track candidate.

8.3 Performance

This algorithm has some free parameters which can be varied to optimize the performance:

- The minimal number of hits required to form a track candidate.
- The maximal depth of the peak-finding algorithm, i.e., the number of bins in the θr -plane.
- The initial threshold for the voting procedure.
- The scaling of the threshold after each voting procedure.

As benchmark observables for an optimization the computing time per event and the efficiency in reconstructing tracks from simulated $B^+ \rightarrow \bar{D}^0(\rightarrow K^+ \pi^-)\pi^+$ decays were chosen. This decay was taken as the momentum of the resulting particles is considerably higher than the average particle at Belle II and no long-lived particles occur which could produce tracks not passing through the origin. Therefore their tracks are easier

to reconstruct, which limits the general validity but is a very good starting point for the development. To count a $B^+ \rightarrow \bar{D}^0 \pi^+$ decay as reconstructed all of its tracks need to be found. As there was no appropriate simulation available at the time of the tests, they were performed without the contribution of background hits from sources like interaction between the electron/positron beams and the gas in the beam pipe.

The maximal depth of the peak-finding algorithm has only a minimal effect on the two benchmark variables. It was found that a depth of 10, thus $2^{10} \times 2^{10}$ bins in θ and r , is sufficient to distinguish individual tracks and give a good enough estimation of the parameters of the track.

The optimization was performed on a single core of a Intel Xeon X5650 CPU and the time per event was measured using the statistics tool of the Belle II software framework. In the following, the CPU time denotes the time required per event without the initialization of the lookup table and the memory allocation which are only performed once at the start of the software.

The computing time can be optimized by varying all parameters. As the initial threshold and its scaling after each voting procedure cannot have an impact on the efficiency, they are two-dimensionally optimized with respect to the computing time per event. As minimal number of hits 10 was chosen. The result can be seen in Figure 8.9. It also shows the impact on the efficiency as a cross-check which confirms that the efficiency is flat. An initial threshold of 48 and a scaling factor of 0.75 were found as parameters to achieve a minimal computing time of 48 ms per event. Of course, these values may change with another number of minimal required hits or further performance optimization.

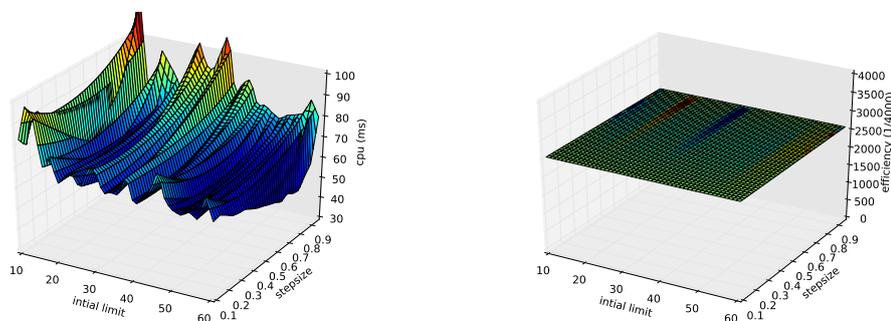


Figure 8.9: Impact of the initial threshold and its scaling on the consumed CPU time (left) and the efficiency (right).

The relation between the minimal number of hits and the efficiency and the computing time is shown in Figure 8.10. The benchmarks for the efficiency are the track-finding algorithm developed for the Belle experiment called Trasan and an algorithm which uses Monte Carlo (MC) information. The MC based method also requires the track candidates to consist of a minimal number of hits in the CDC and therefore cannot reach an efficiency of 1 but represents the maximally achievable efficiency for a track-finding algorithm. The Legendre track finder is considerably faster than Trasan for the full range of minimal hits. For the lower values of minimal hits also the efficiency is better.

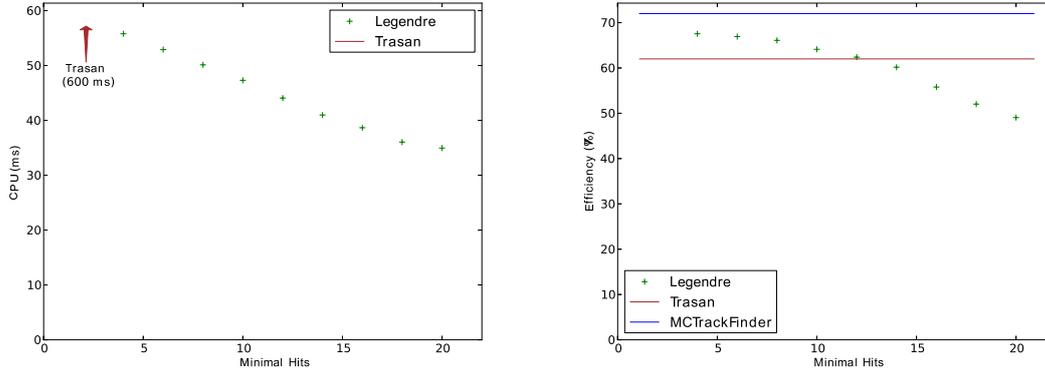


Figure 8.10: Relation between the minimal number of hits and the CPU time per event (left) and the efficiency in completely finding $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\pi^+$ decays (right).

8.4 Result

The presented status of the track-finding algorithm represents the first phase of a continuous development. The achieved performance is very promising as it was possible to surpass the previously employed method in the reconstruction of the benchmark events both in speed and efficiency with the newly developed track finder. Consequently, the work on it has been picked up and is still ongoing. The current development concentrates on improving certain aspects of the algorithm, adding mechanisms to it like the merging of track candidates, and a wider range of tests taking into account background hits and more general decay topologies.

The restriction on tracks originating from the interaction point (0, 0) prevents the track finder from being the sole algorithm of this kind at Belle II, but it is especially fast and efficient for high-momentum tracks and therefore an excellent choice to be applied before a second, slower track-finding algorithm.

9. Conclusion

In summary, the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ decays has been measured to be

$$\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (1.25 \pm 0.28 \pm 0.27) \times 10^{-4},$$

where the first error is statistical and the second systematic. The number of reconstructed $B^+ \rightarrow \tau^+ \nu_\tau$ decays was found to be 222 ± 50 . The measurement was performed with the semileptonic tagging method and used the full Belle data sample of $770 \times 10^6 B\bar{B}$ pairs which was recorded by the Belle detector at the $\Upsilon(4S)$ resonance at the asymmetric-energy KEKB- e^+e^- collider. The result supersedes the previous Belle measurement with semileptonic tagging of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau) = (1.54_{-0.37-0.31}^{+0.38+0.28}) \times 10^{-4}$, which was performed with an only slightly smaller data set containing $660 \times 10^6 B\bar{B}$ pairs. The large increase in statistical sensitivity could be achieved by performing a completely new analysis and implementing several improvements. An improved semileptonic tagging method was employed which reconstructs a large number of decay channels and uses multivariate selection to achieve a high reconstruction efficiency. An additional τ decay channel, $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$, was reconstructed which contributes with 0.25 to the total branching fraction of the τ lepton. The selection was optimized and includes a dedicated suppression of background containing converted photons and a multivariate suppression of continuum $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) background. Furthermore, a second variable, the momentum of the decay product of the τ , was included in the final fit.

A large part of the analysis consists of the validation of different aspects of the measurements. The efficiency of the tagging algorithm was calibrated with double-tagged events. This requires independent measurements of branching fractions including the necessary reconstruction and selection. The decay $B^+ \rightarrow D^0 \pi^+$ turned out to be especially suitable for such a calibration since it is fully reconstructible and suffers from only low uncertainties related to its reconstruction efficiency. But the low number of reconstructed signal events required the calibration using other, semileptonic decay channels. In the study of the background simulation the non- b background was found to be problematic and the off-resonance data sample was taken to describe this background component

in the final fit. Also other tests, like the comparison of various variable distributions in different sideband samples or the investigation of the E_{ECL} composition, were performed. A comparison of the result with other measurements and the expectation is shown in Figure 9.1. It displays that the result is compatible both with the previous measurements and the expectation from the fit on the CKM triangle. It is closer to the Standard Model expectation than the previous result. Therefore the difference between the Standard Model expectation and the measurements which used to be a prominent issue in flavor particle physics was even reduced further.

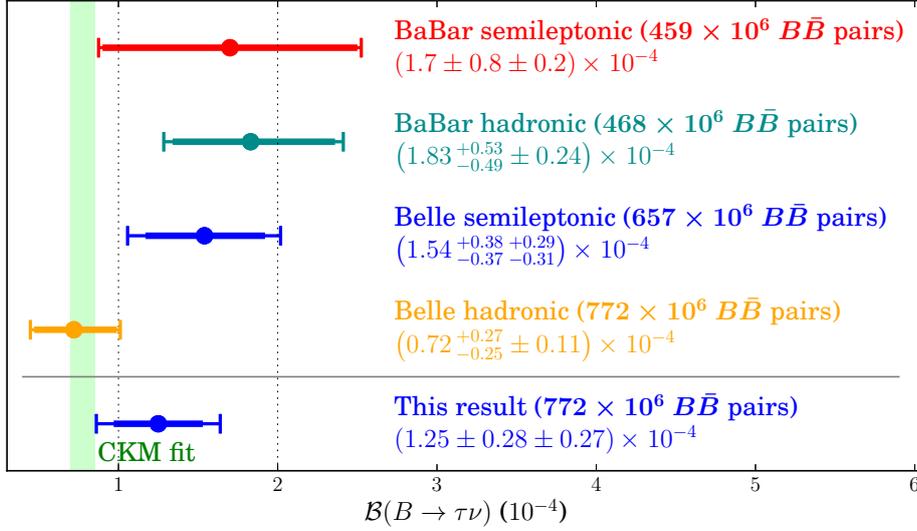


Figure 9.1: Comparison of the result of this thesis with previous measurements of $\mathcal{B}(B^+ \rightarrow \tau^+ \nu_\tau)$.

The systematic uncertainty of the result mainly consists of components which are extracted from MC data, control samples, or sidebands. So future measurements at the upcoming Belle II experiment will be able to further reduce the systematic error just using the larger data samples. The Belle II collaboration plans to record a data sample of 50 ab^{-1} and wants to achieve a total relative uncertainty of 3% on the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ decays. To reach this goal, also the non-scaling systematic uncertainties need to be reduced. This is only possible with a better knowledge of the dominant background components, a very good understanding of the tagging efficiency, and a good background reduction. When this is achieved, the decay $B^+ \rightarrow \tau^+ \nu_\tau$ can be used to drastically reduce the parameter space of various models of physics beyond the Standard Model.

This result was first publicly shown at the International Workshop on the CKM Unitarity Triangle (CKM 2014) in the presentation “New physics searches in leptonic decays of B mesons” and multiple times afterwards.

Additionally, a track-finding algorithm for the Belle II central drift chamber is described. The track finder needs to fulfill strong computational time constraints while reaching a maximal efficiency. Therefore, it was developed using various optimization mechanisms, e.g. the employment of a peak-finding algorithm similar to a fast Hough transform and the

iterative search for peaks in a $2D$ -plane instead of a clustering algorithm. Its performance was determined with simulated $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\pi^+$ decays and was found to be very promising. For this special decay, the newly developed track finder was able to exceed the method employed at the Belle experiment both in speed and efficiency. This led to the continuation of the development. In this way, the described work represents the foundation of what will most likely be a significant part of the Belle II tracking software.

Bibliography

- [1] W. Greiner and B. Müller, *Gauge theory of weak interactions : with 74 worked examples and problems*, 2., rev. ed. (Springer, Berlin, 1996).
- [2] J. Charles *et al.* (CKMfitter Group), Eur. Phys. J. C **41**, 1–131 (2005), updated results and plots available at: <http://ckmfitter.in2p3.fr>.
- [3] K. Olive *et al.* (Particle Data Group), “The Review of Particle Physics”, Chin. Phys. C **38**, 090001 (2014).
- [4] S. Aoki *et al.*, “Review of lattice results concerning low energy particle physics”, (2013) arXiv:1310.8555 [hep-lat].
- [5] A. Crivellin, C. Greub, and A. Kokulu, “Explaining $B \rightarrow D\tau\nu$, $B \rightarrow D^*\tau\nu$ and $B \rightarrow \tau\nu$ in a 2HDM of type III”, Phys.Rev. **D86**, 054014 (2012) arXiv:1206.2634 [hep-ph].
- [6] A. Bevan *et al.* (BaBar Collaboration, Belle Collaboration), “The Physics of the B Factories”, (2014) arXiv:1406.6311 [hep-ex].
- [7] J. P. Lees *et al.* (The BABAR Collaboration), “Measurement of an excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ decays and implications for charged Higgs bosons”, Phys. Rev. D **88**, 072012 (2013).
- [8] D. Buskulic *et al.* (ALEPH Collaboration), “Measurement of the $b \rightarrow \tau^-\bar{\nu}_\tau X$ branching ratio and an upper limit on $B^- \rightarrow \tau^-\bar{\nu}_\tau$ ”, Physics Letters B **343**, 444–452 (1995).
- [9] M. Artuso *et al.* (CLEO Collaboration), “Search for $B \rightarrow \ell\bar{\nu}_\ell$ ”, Phys. Rev. Lett. **75**, 785–789 (1995).
- [10] B. Aubert *et al.* (BABAR Collaboration), “Search for $B^+ \rightarrow \ell^+\nu_\ell$ recoiling against $B^- \rightarrow D^0\ell^-\bar{\nu}_\ell X$ ”, Phys. Rev. D **81**, 051101 (2010).
- [11] J. P. Lees *et al.* (BABAR Collaboration), “Evidence of $B^+ \rightarrow \tau^+\nu$ decays with hadronic B tags”, Phys. Rev. D **88**, 031102 (2013).
- [12] K. Hara *et al.* (Belle Collaboration), “Evidence for $B^- \rightarrow \tau^-\bar{\nu}_\tau$ with a semileptonic tagging method”, Phys. Rev. D **82**, 071101 (2010).
- [13] K. Hara *et al.* (Belle Collaboration), “Evidence for $B^- \rightarrow \tau^-\bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle”, Phys. Rev. Lett. **110**, 131801 (2013).

- [14] D. Besson and T. Skwarnicki, “Upsilon Spectroscopy: Transitions in the Bottomonium System”, *Annual Review of Nuclear and Particle Science* **43**, 333–378 (1993).
- [15] A. Abashian *et al.*, “The Belle detector”, *Nucl. Instrum. Methods in Phys. Res., Sect. A* **479**, 117–232 (2002).
- [16] J. Brodzicka *et al.*, “Physics achievements from the Belle experiment”, *Prog. Theor. Exp. Phys.* **2012** (2012), detector section.
- [17] E. Nakano, “Belle PID”, *Nucl. Instrum. Methods in Phys. Res., Sect. A* **494**, 402–408 (2002), Proceedings of the 8th International Conference on Instrumentation for Colliding Beam Physics.
- [18] D. J. Lange, “The EvtGen particle decay simulation package”, *Nucl. Instrum. Methods in Phys. Res., Sect. A* **462**, 152–155 (2001), BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.
- [19] R. Brun *et al.*, *GEANT 3: user’s guide Geant 3.10, Geant 3.11; rev. version* (CERN, Geneva, 1987).
- [20] T. Sjöstrand, “High-energy-physics event generation with PYTHIA 5.7 and JETSET 7.4”, *Comput. Phys. Commun.* **82**, 74–89 (1994).
- [21] M. Feindt *et al.*, “A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories”, *Nucl. Instrum. Methods Phys. Res., Sect. A* **654**, 432–440 (2011).
- [22] A. Sibidanov *et al.* (Belle Collaboration), “Study of exclusive $B \rightarrow X_u \ell \nu$ decays and extraction of $|V_{ub}|$ using full reconstruction tagging at the Belle experiment”, *Phys. Rev. D* **88**, 032005 (2013).
- [23] O. Lutz *et al.* (Belle Collaboration), “Search for $B \rightarrow h^{(*)} \nu \bar{\nu}$ with the full Belle $\Upsilon(4S)$ data sample”, *Phys. Rev. D* **87**, 111103 (2013).
- [24] K. Kirchgessner, “Semileptonische Markierungsseiten-Rekonstruktion”, MA thesis (2012), <http://ekp-invenio.physik.uni-karlsruhe.de/record/48181>.
- [25] M. Röhrken *et al.* (Belle Collaboration), “Measurements of branching fractions and time-dependent CP violating asymmetries in $B^0 \rightarrow D^{(*)\pm} D^\mp$ decays”, *Phys. Rev. D* **85**, 091106 (2012).
- [26] M. Prim *et al.* (Belle Collaboration), “Angular analysis of $B^0 \rightarrow \phi K^*$ decays and search for CP violation at Belle”, *Phys. Rev. D* **88**, 072004 (2013).
- [27] M. Feindt and U. Kerzel, “The NeuroBayes neural network package”, *Nucl. Instrum. Methods Phys. Res., Sect. A* **559**, 190–194 (2006).
- [28] D. M. Asner *et al.* (CLEO Collaboration), “Search for exclusive charmless hadronic B decays”, *Phys. Rev. D* **53**, 1039–1050 (1996).
- [29] V. Blobel, *Smoothing of Poisson distributed data*, <http://www.desy.de/~blobel/splft.f>.
- [30] B. Bhuyan, “High P_T Tracking Efficiency Using Partially Reconstructed D^* Decays”, BelleNote 1165 (internal).

- [31] S. W. Lin, “ π^0 systematics Using Inclusive η ”, BelleNote 645 (internal).
- [32] S. Nishida, “Study of Kaon and Pion Identification Using Inclusive D^* Sample”, BelleNote 779 (internal).
- [33] L. Hinz, “Lepton ID efficiency correction and systematic error”, BelleNote 954 (internal).
- [34] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, English, The Annals of Mathematical Statistics **9**, pages (1938).
- [35] T. Alexopoulos *et al.*, “Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers”, Nucl. Instrum. Methods in Phys. Res., Sect. A **592**, 456–462 (2008).
- [36] T. Abe *et al.* (Belle II collaboration), “Belle II Technical Design Report”, (2010) arXiv:1011.0352 [physics.ins-det].
- [37] T. Aushev, W. Bartel, and A. Bondar, “Physics at Super B Factory”, (2010) arXiv:1002.5012 [hep-ex].
- [38] Illustrative plots generously provided by Viktor Trusov.
- [39] H. Li, M. A. Lavin, and R. J. L. Master, “Fast Hough transform: A hierarchical approach”, Computer Vision, Graphics, and Image Processing **36**, 139–161 (1986).

Appendix

A Data / MC comparisons

A.1 E_{ECL} sideband

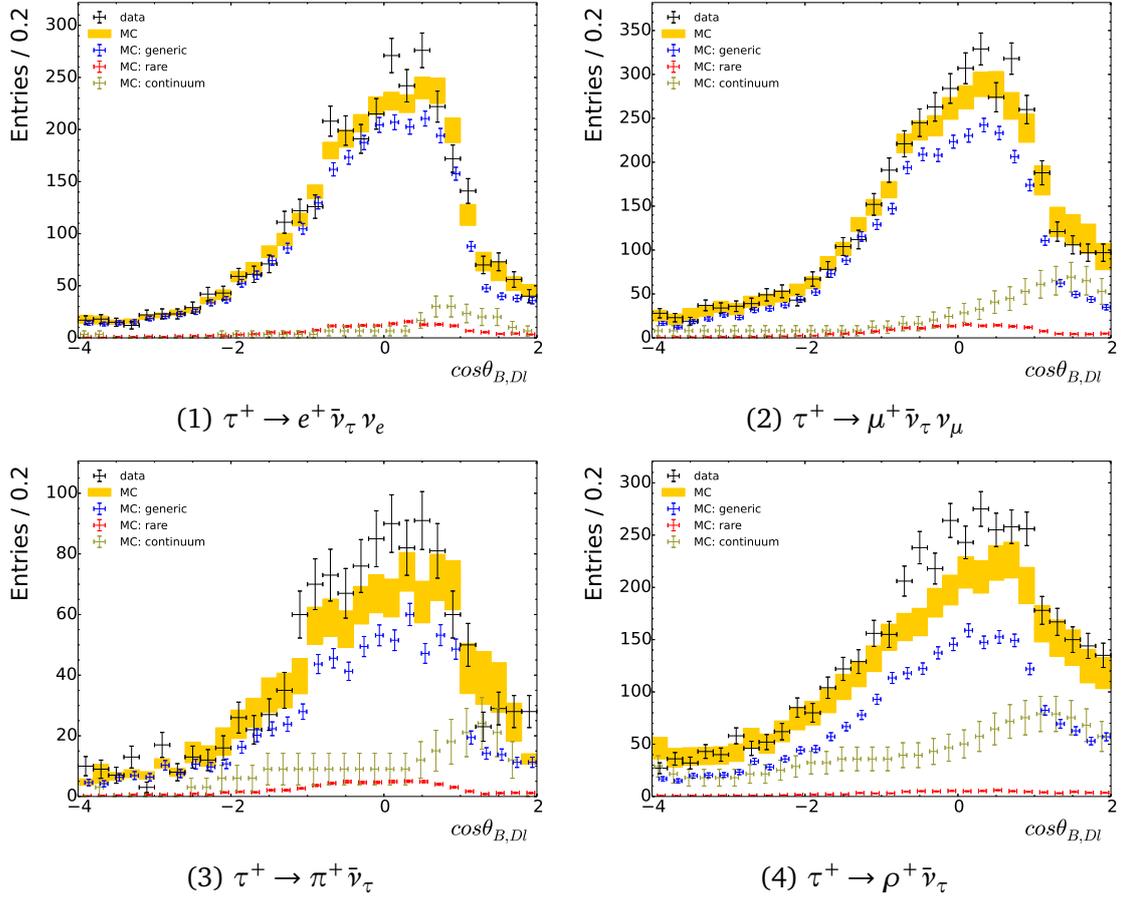


Figure A.1: Comparison of the $\cos \theta_{B,D^{(*)}\ell}$ distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

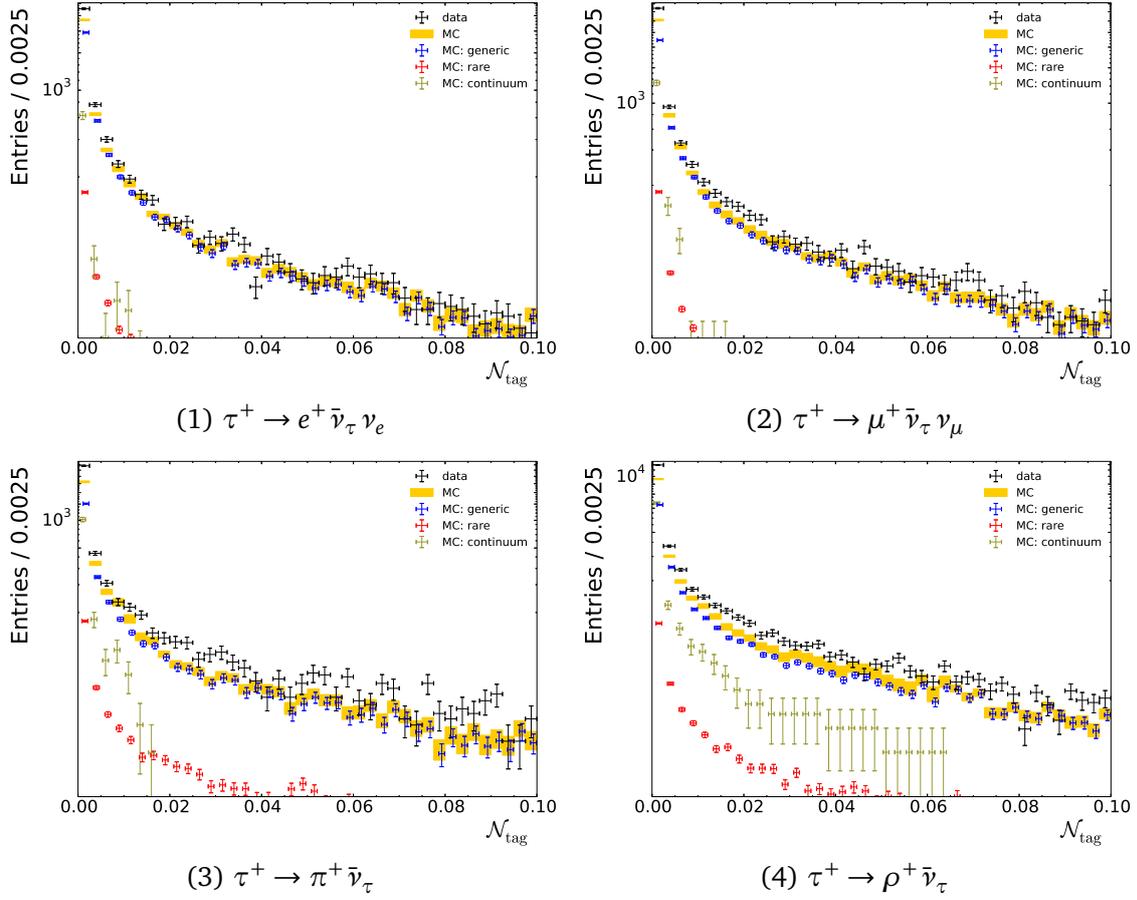


Figure A.2: Comparison of the \mathcal{N}_{tag} distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

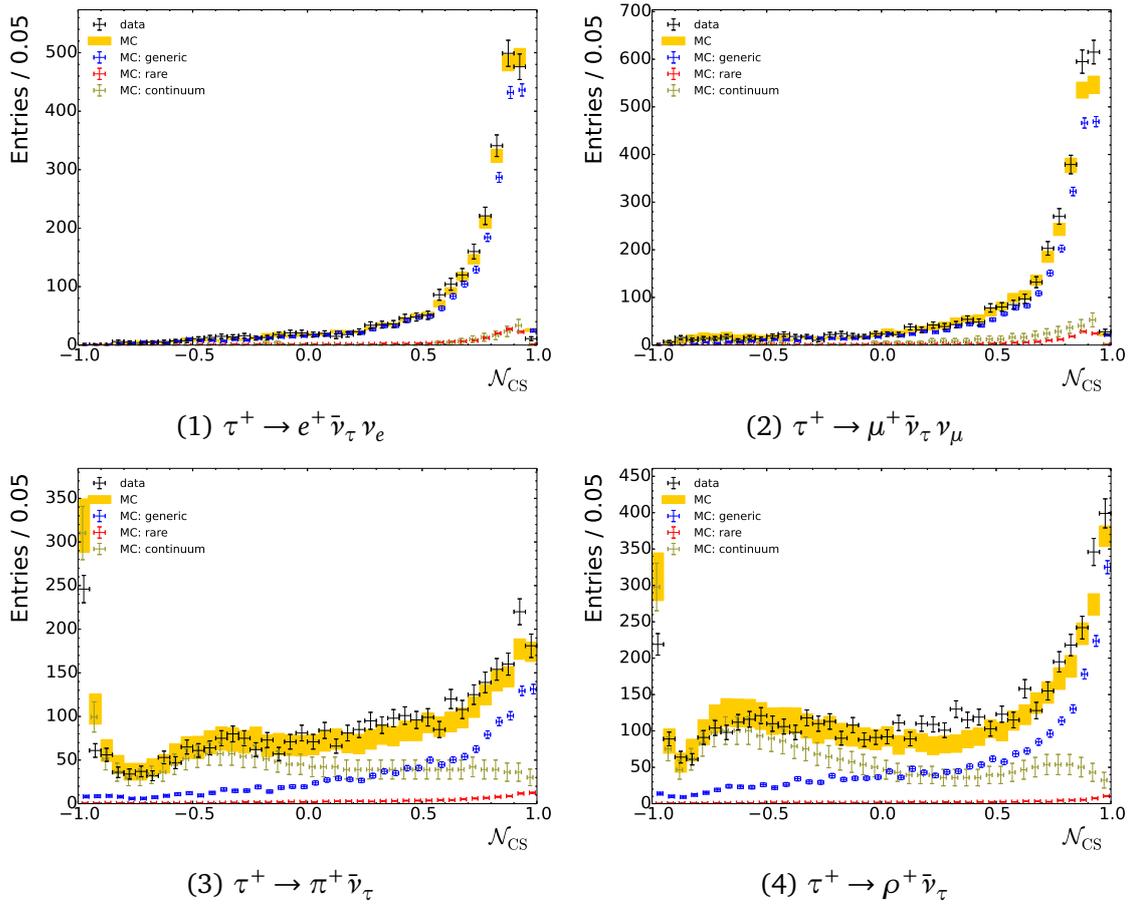


Figure A.3: Comparison of the \mathcal{N}_{CS} distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

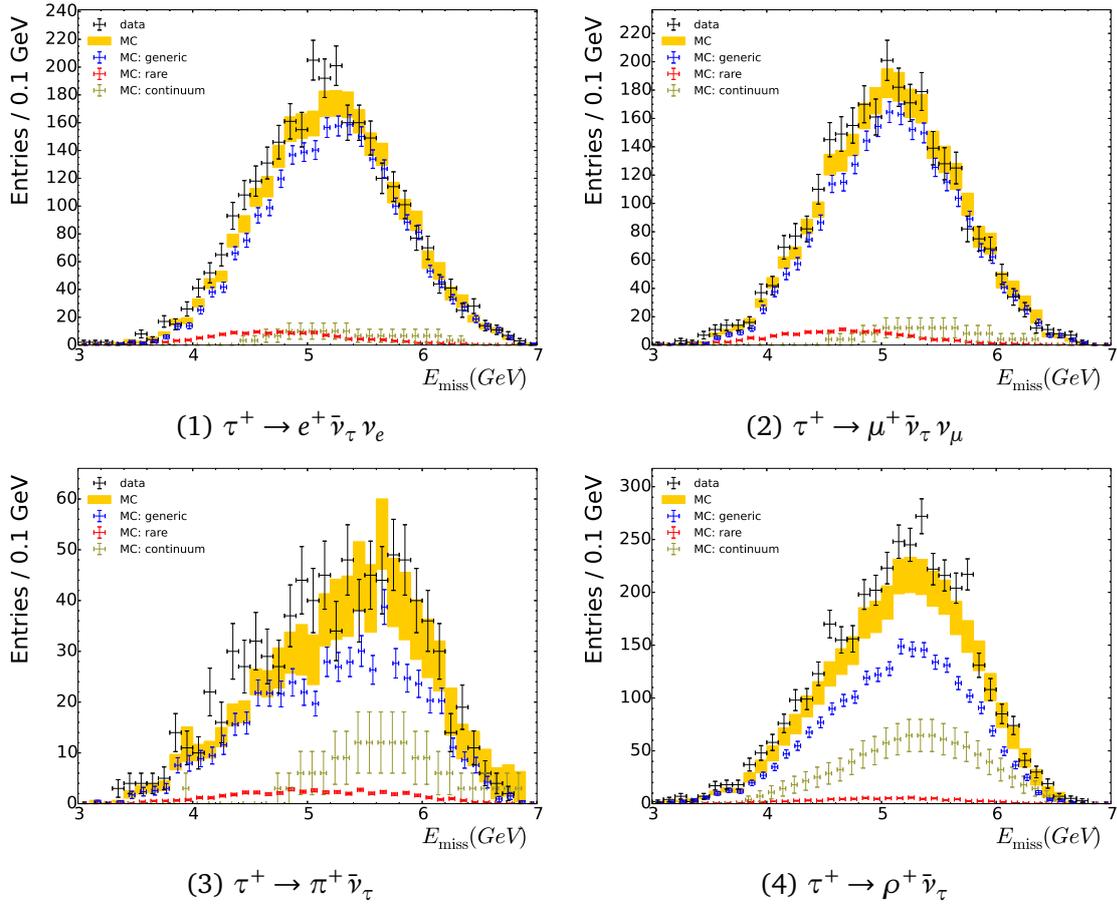


Figure A.4: Comparison of the missing energy distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

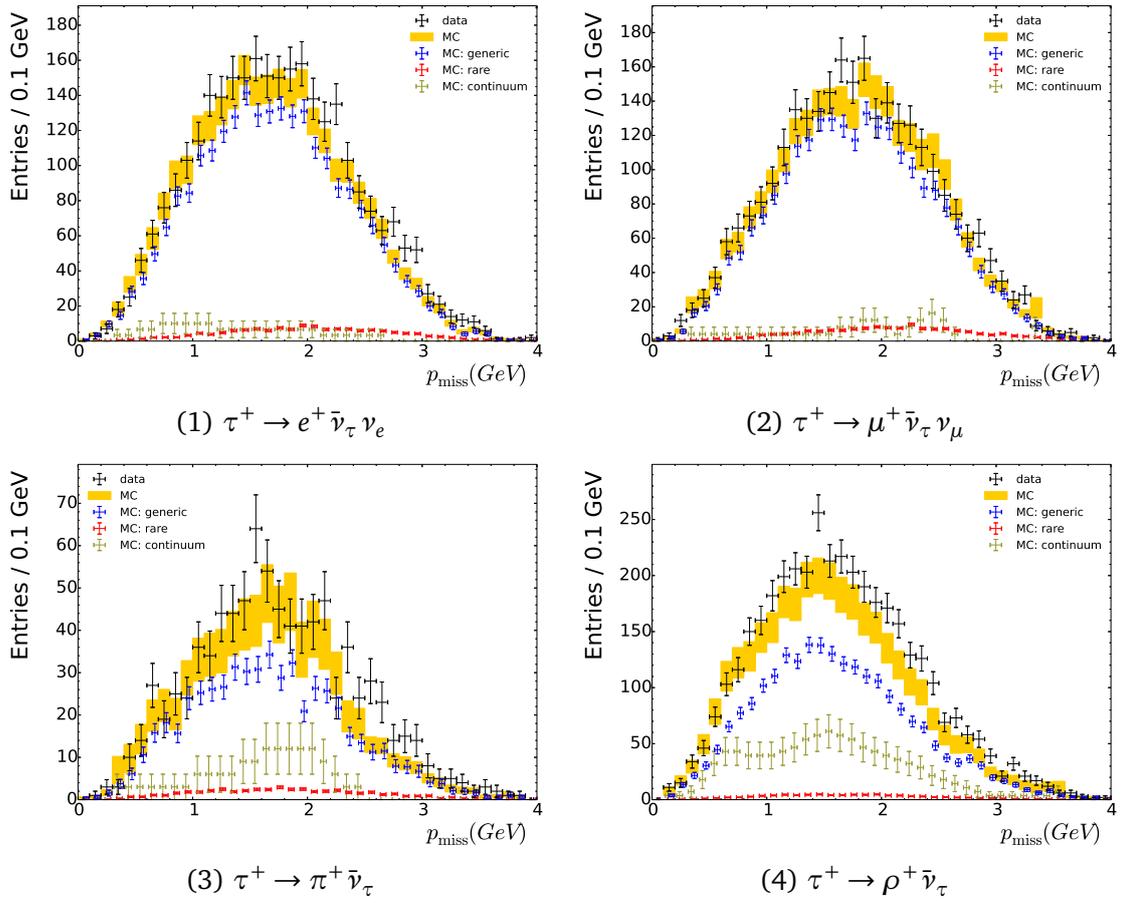


Figure A.5: Comparison of the missing momentum distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

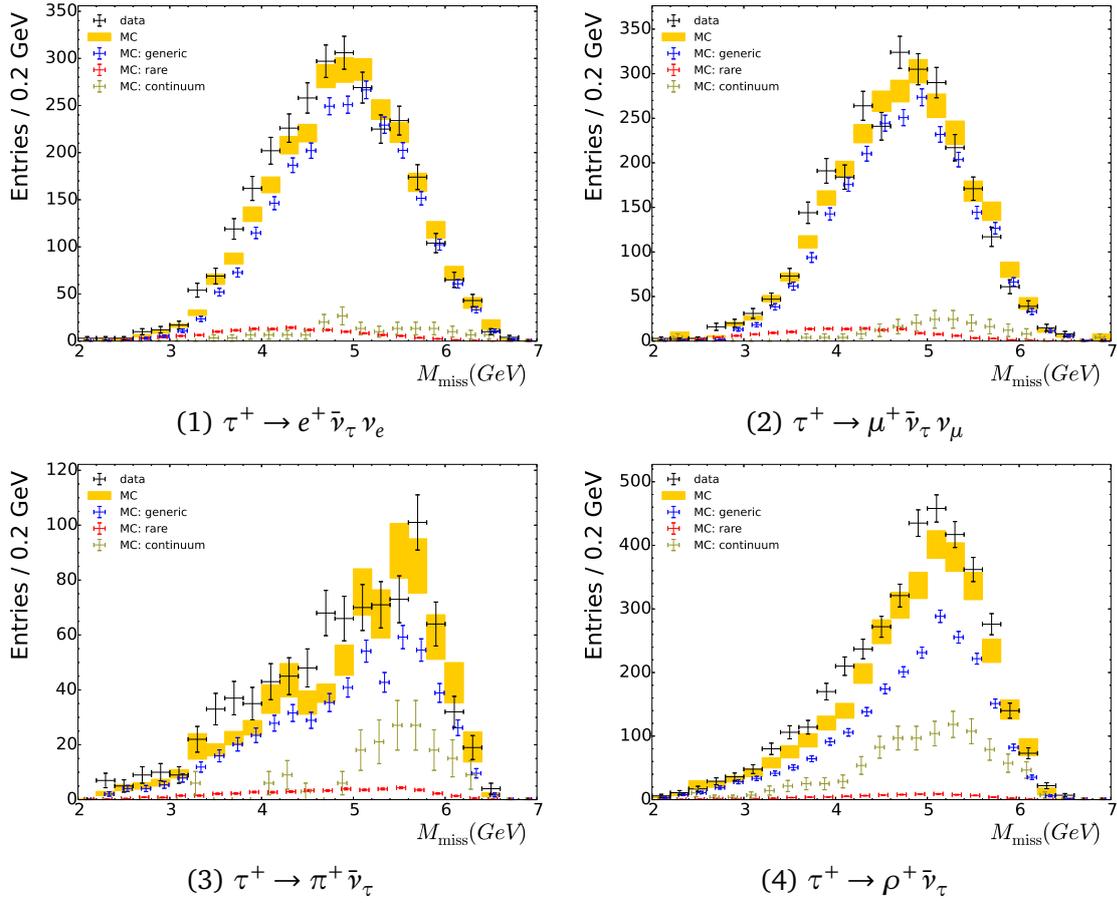


Figure A.6: Comparison of the missing mass distribution of simulated and real data in the E_{ECL} sideband, separated by the τ decay channels.

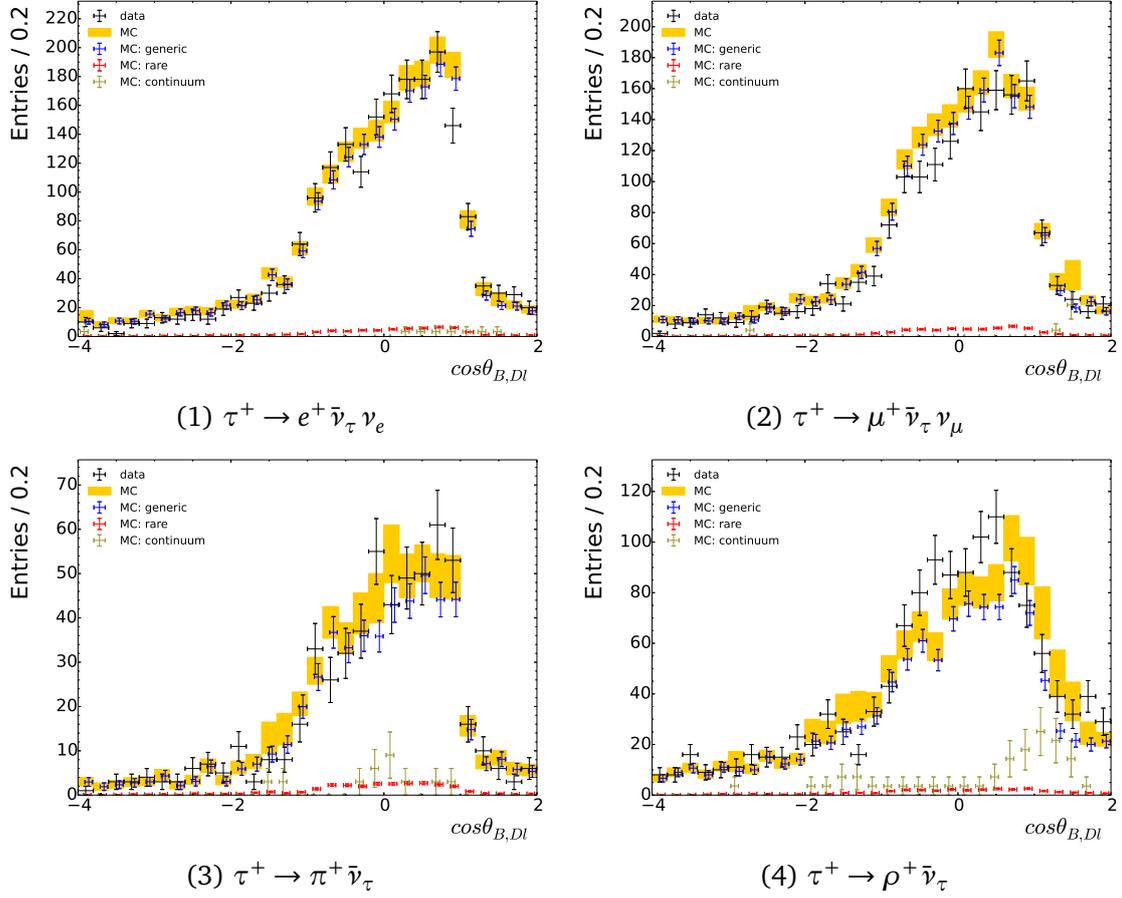
A.2 B^0 tagged sample

Figure A.7: Comparison of the $\cos \theta_{B,D(*)\ell}$ distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

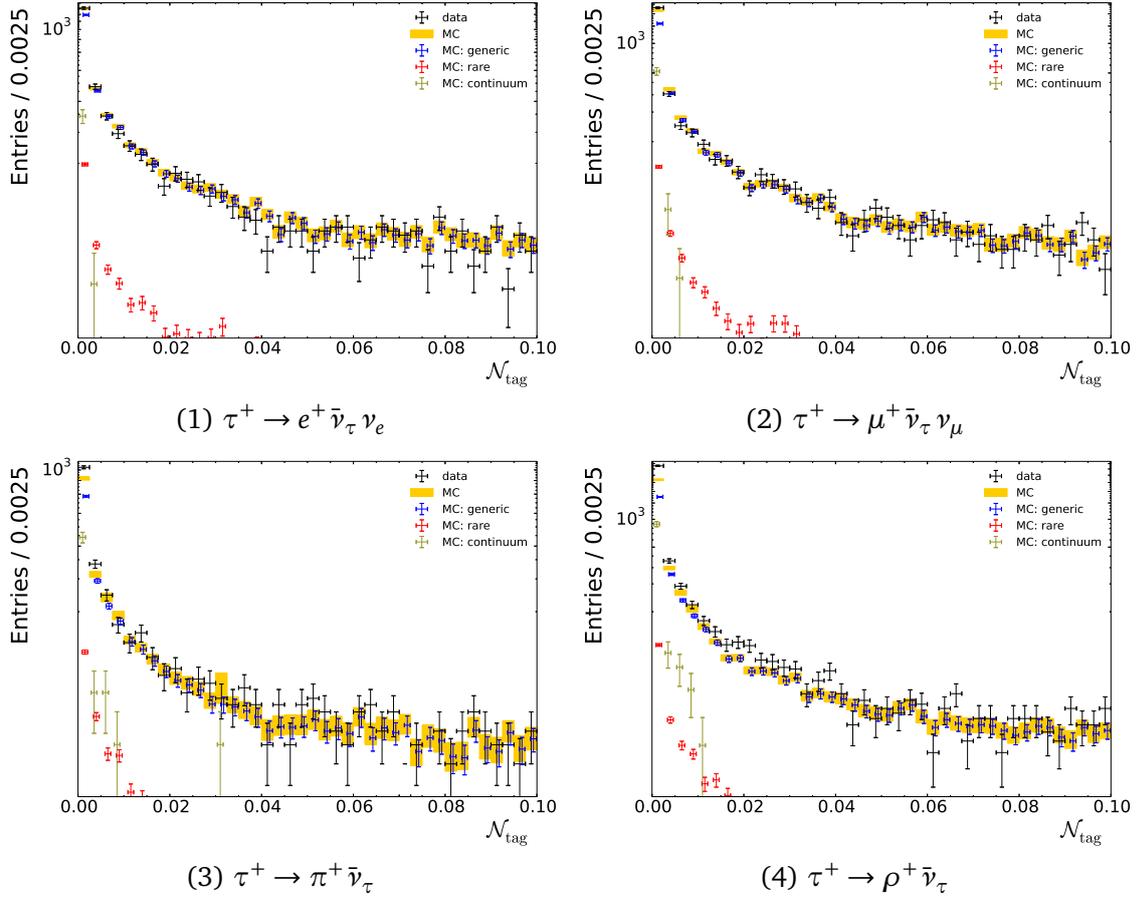


Figure A.8: Comparison of the \mathcal{N}_{tag} distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

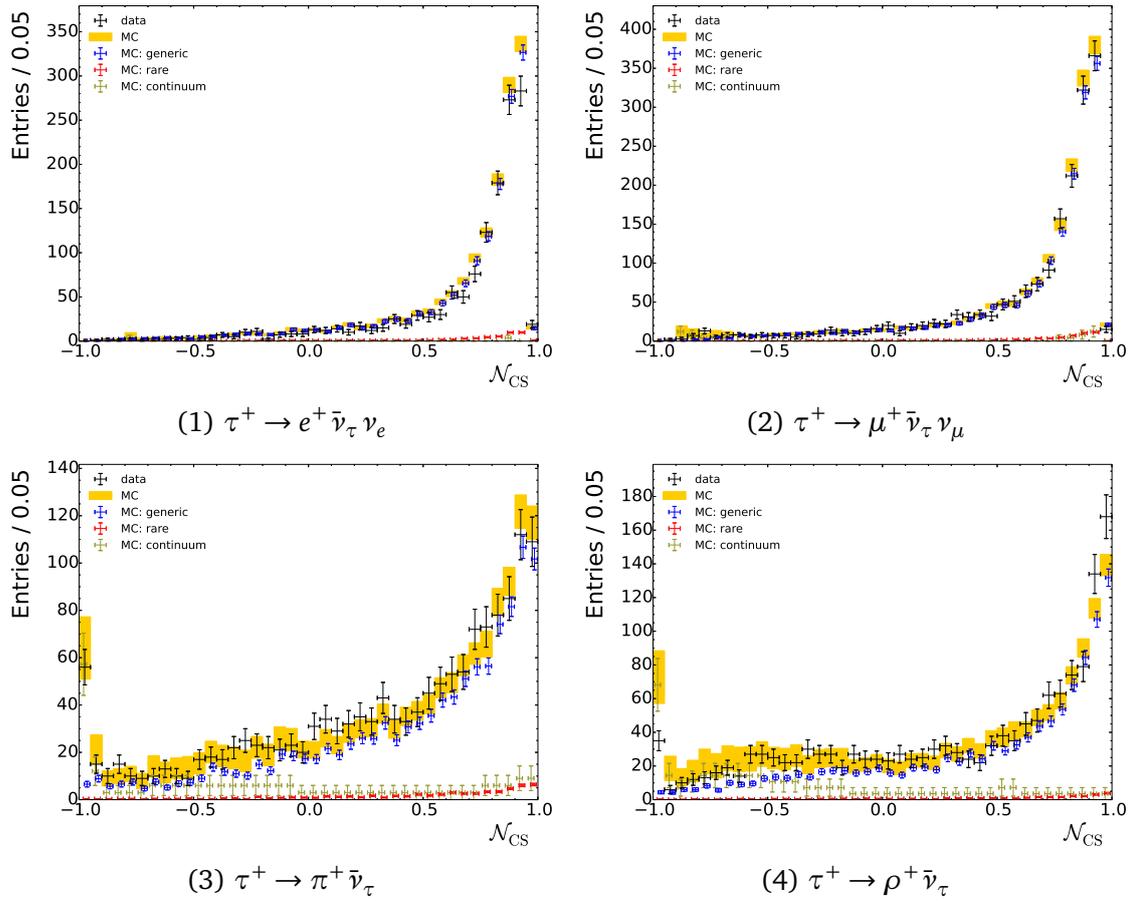


Figure A.9: Comparison of the \mathcal{N}_{CS} distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

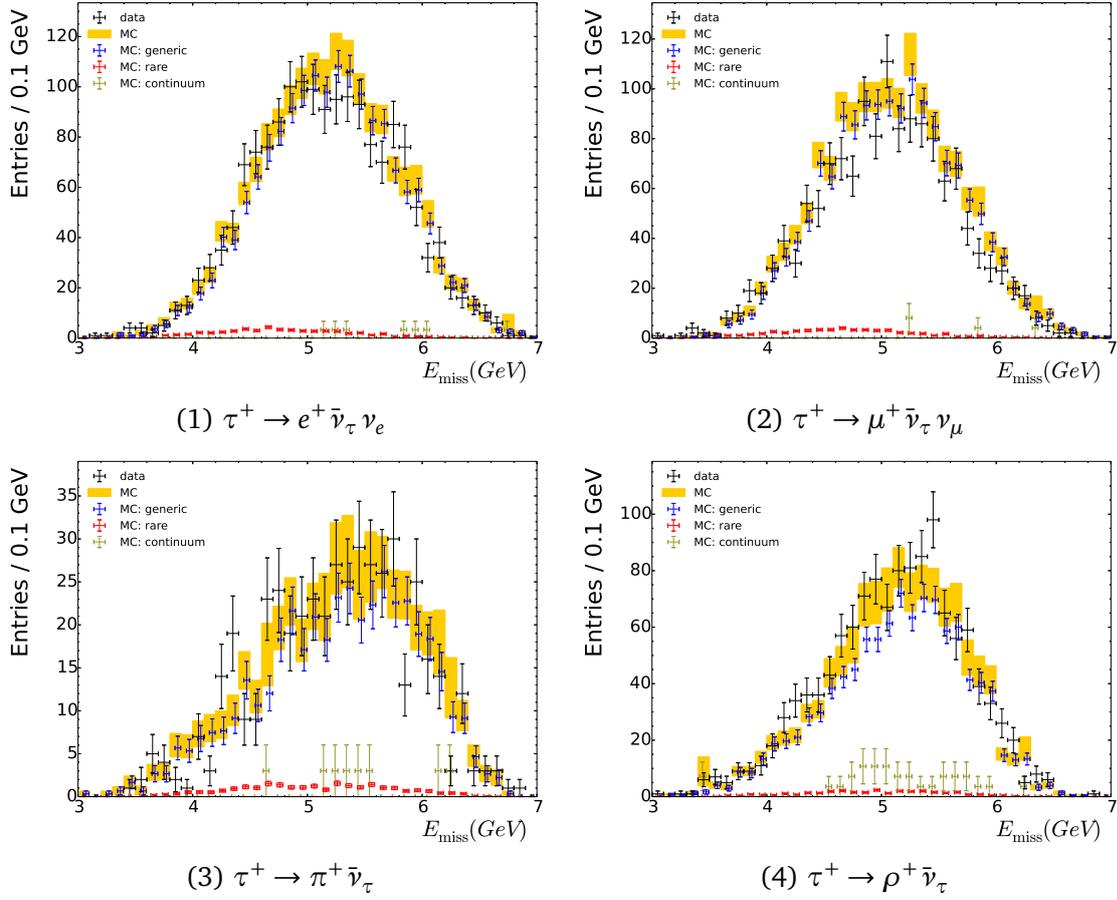


Figure A.10: Comparison of the missing energy distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

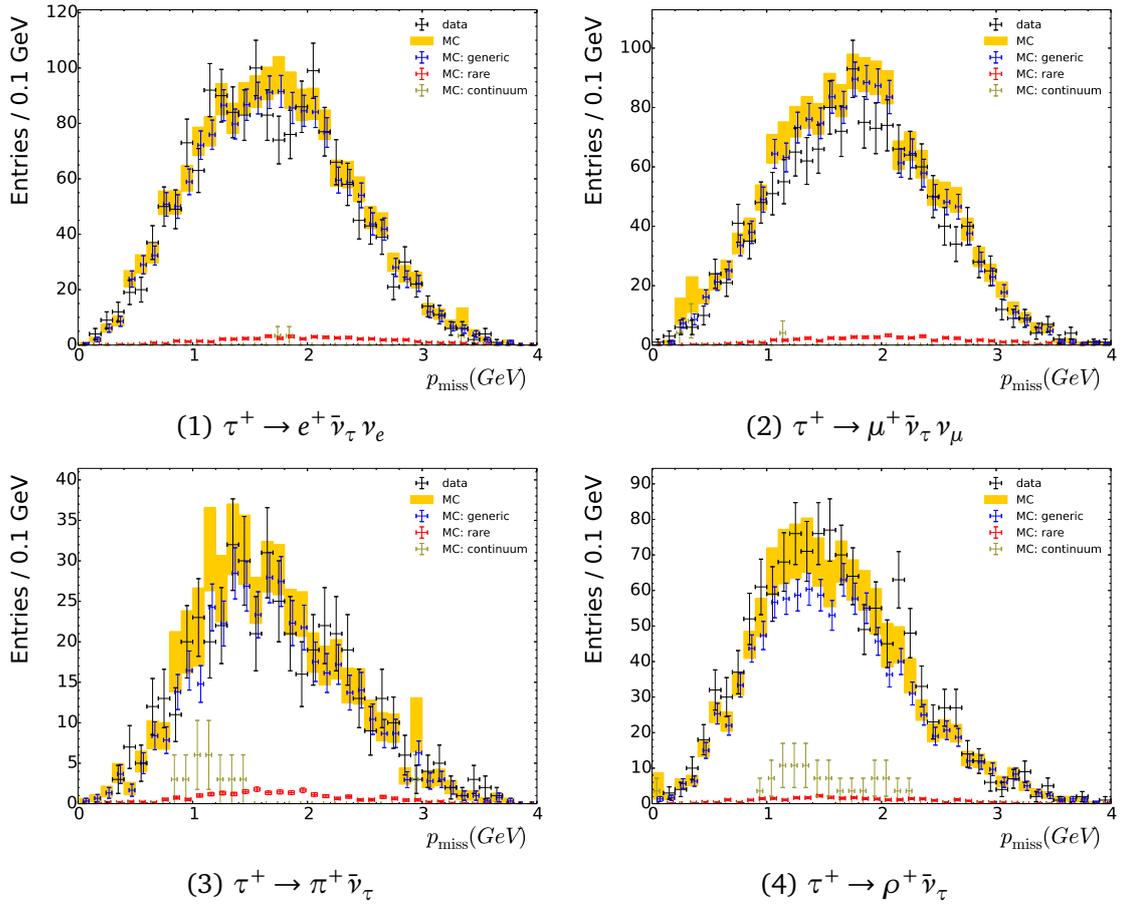


Figure A.11: Comparison of the missing momentum distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

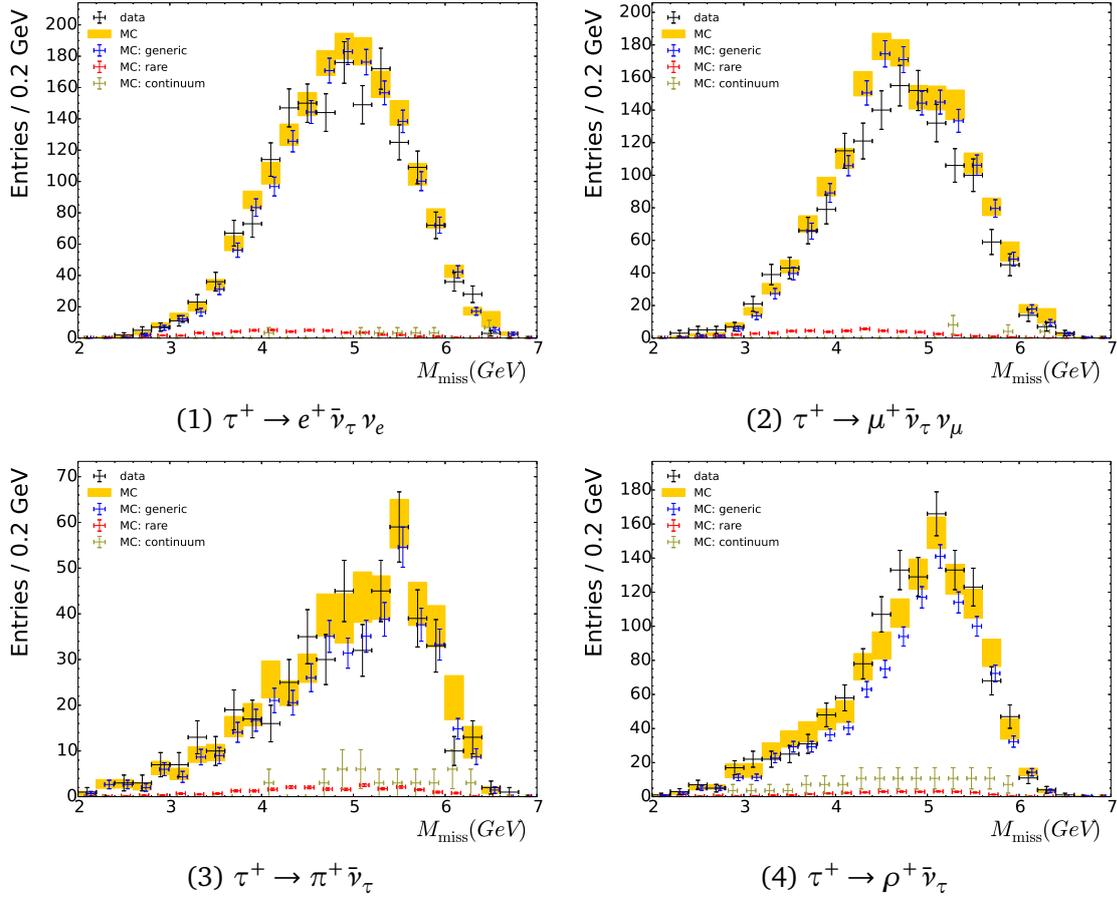


Figure A.12: Comparison of the missing mass distribution of simulated and real data in the B^0 tagged sample, separated by the τ decay channels.

B PDF shapes

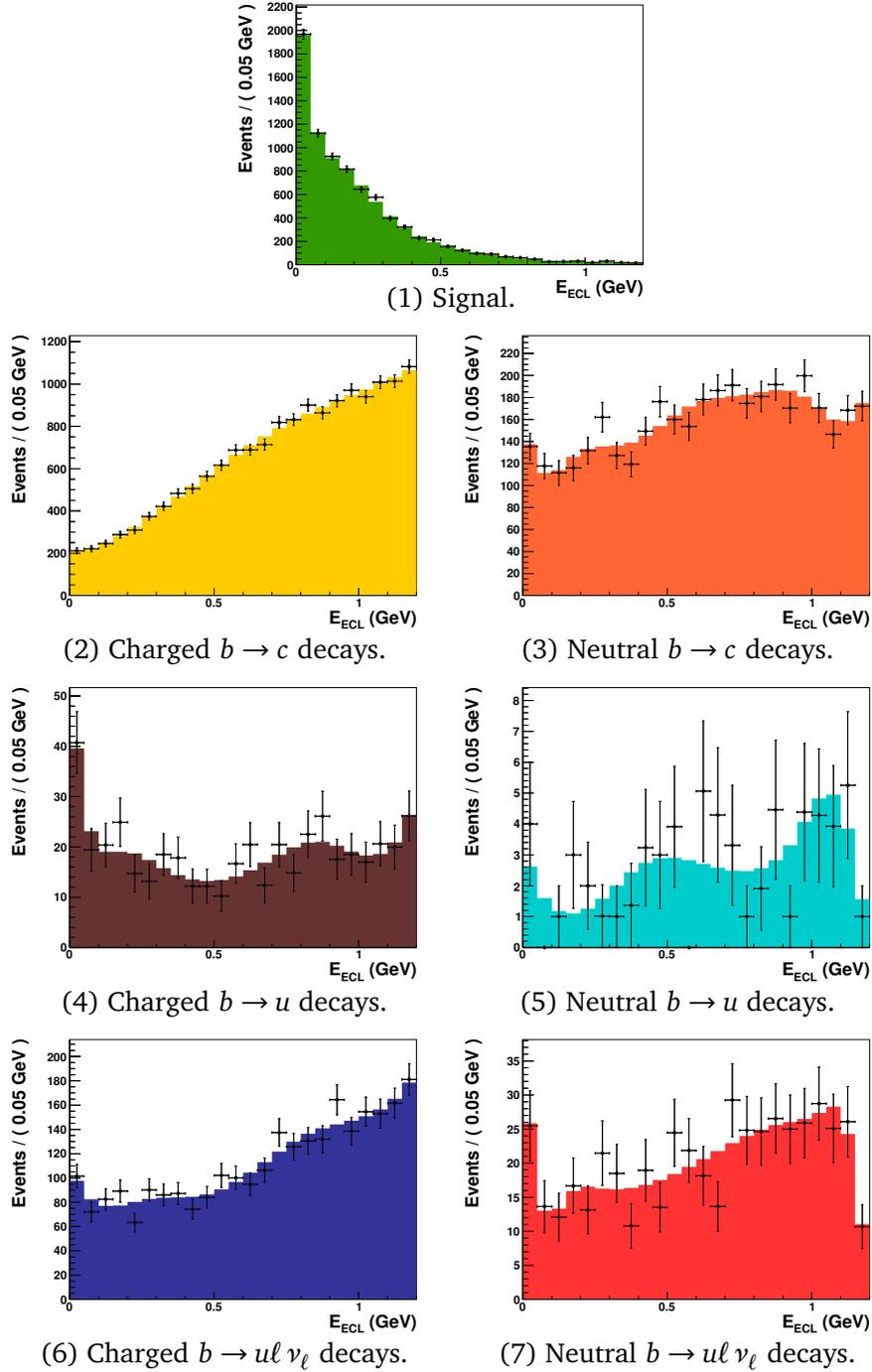


Figure B.13: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$ in E_{ECL} . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled by the available amount of simulated data.

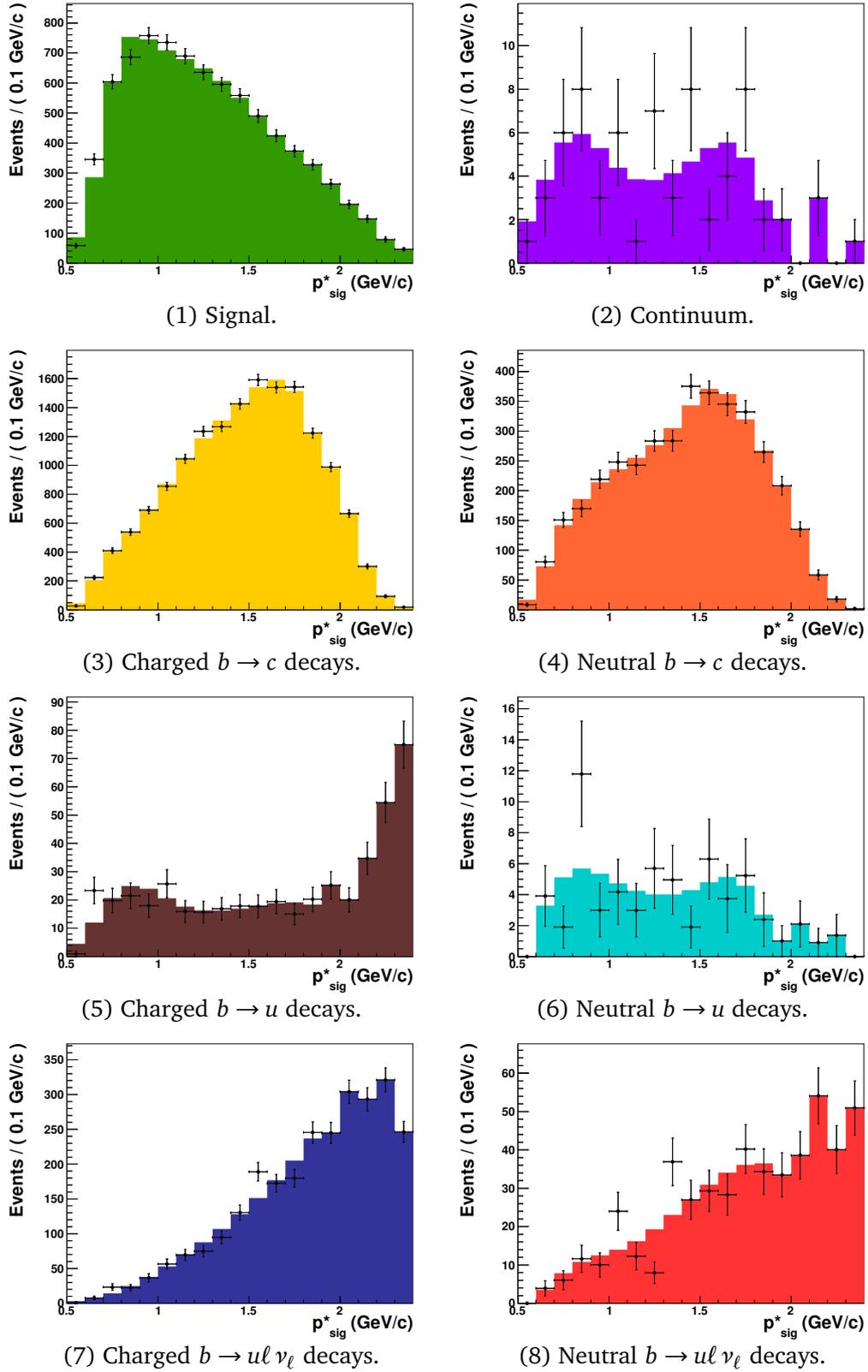


Figure B.14: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow \mu^+ \bar{\nu}_\tau \nu_\mu$ in p_{sig}^* . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled by the available amount of simulated data.

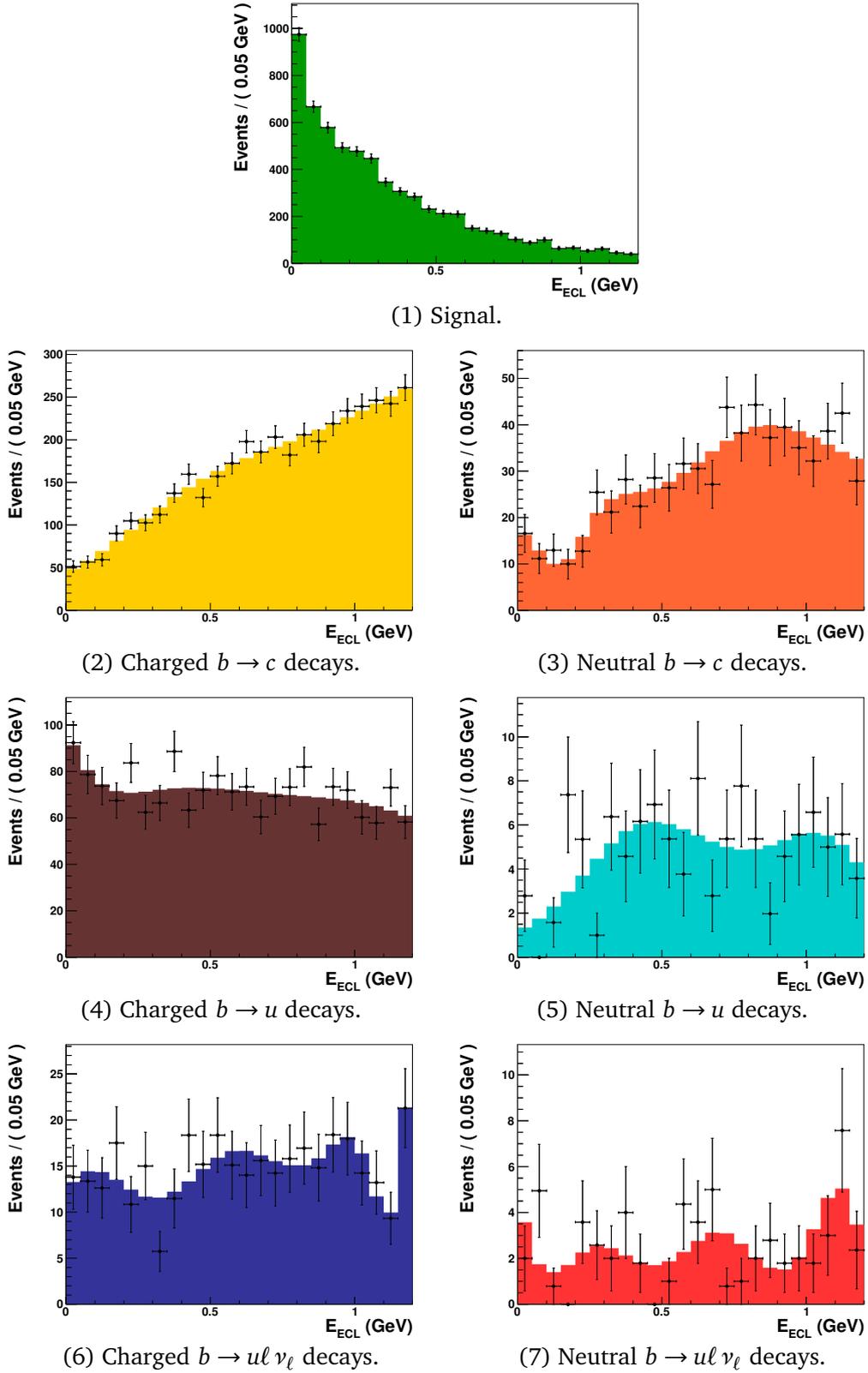


Figure B.15: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ in E_{ECL} . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled by the available amount of simulated data.

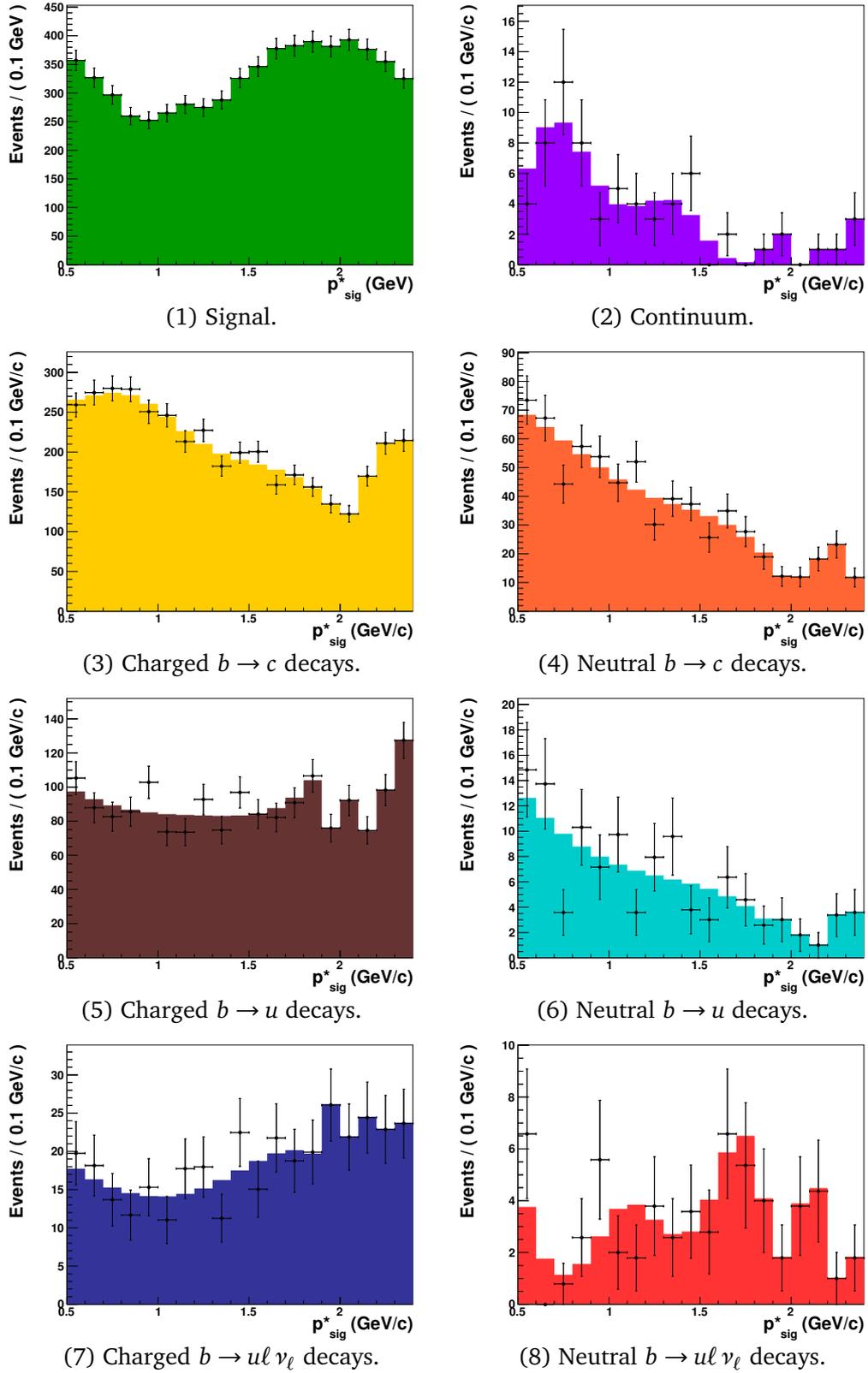


Figure B.16: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow \pi^+ \bar{\nu}_\tau$ in p_{sig}^* . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled by the available amount of simulated data.

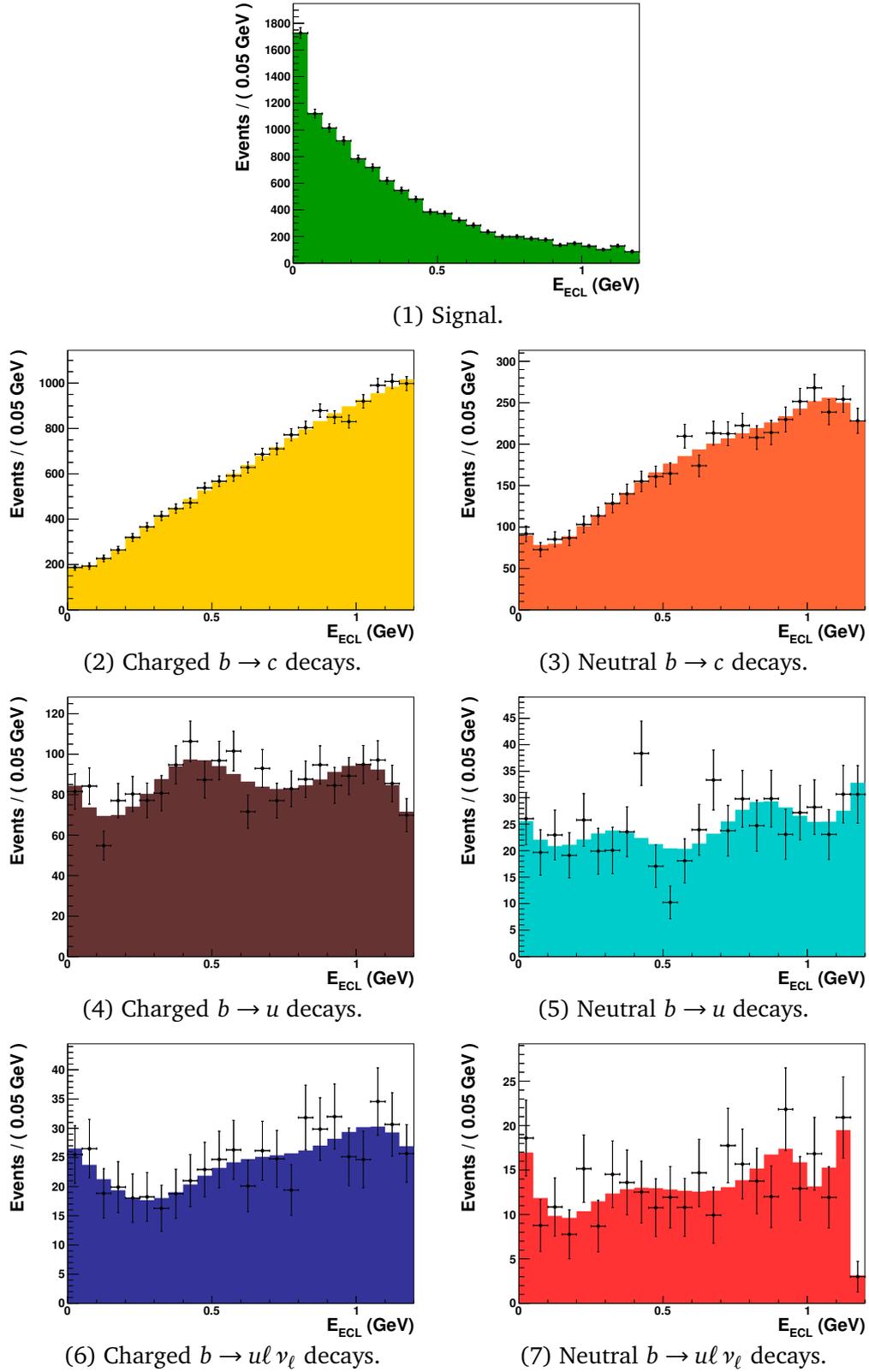
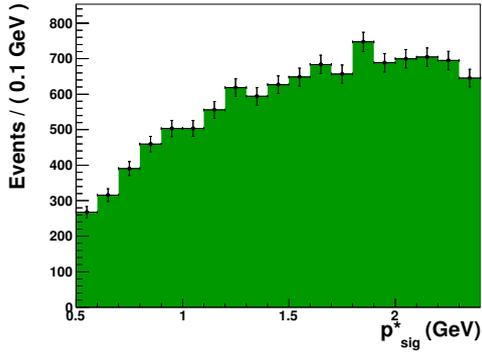
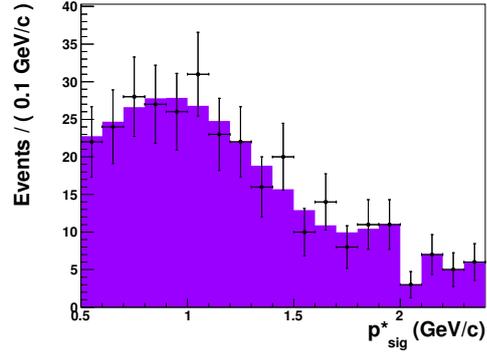


Figure B.17: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$ in E_{ECL} . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled by the available amount of simulated data.



(1) Signal.



(2) Continuum.

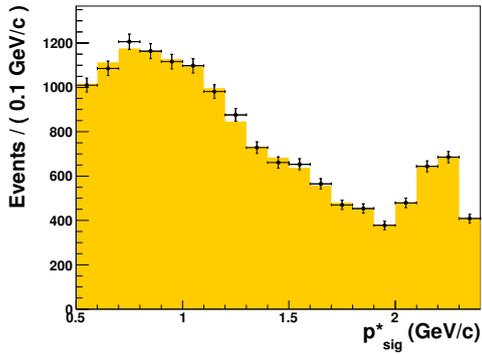
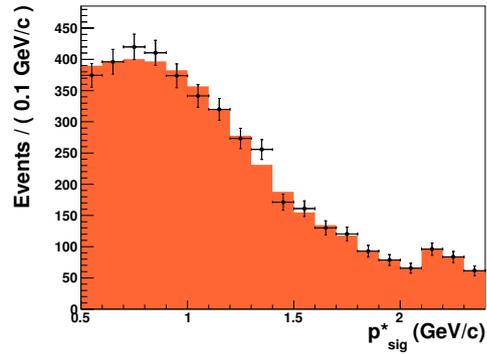
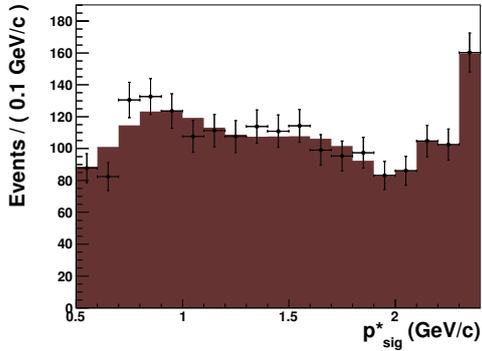
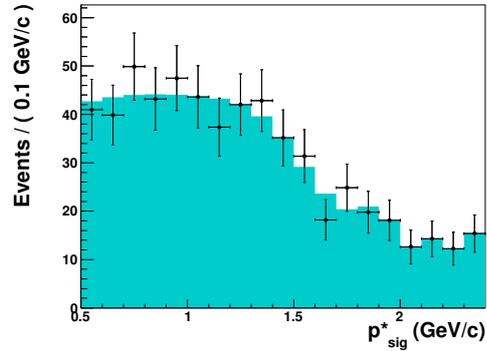
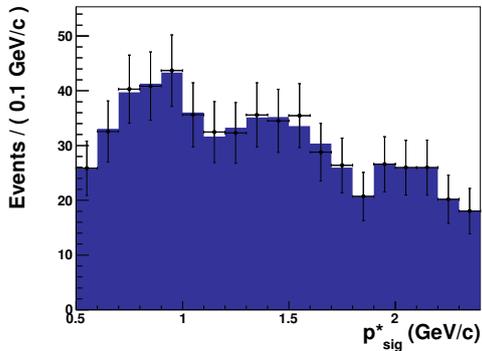
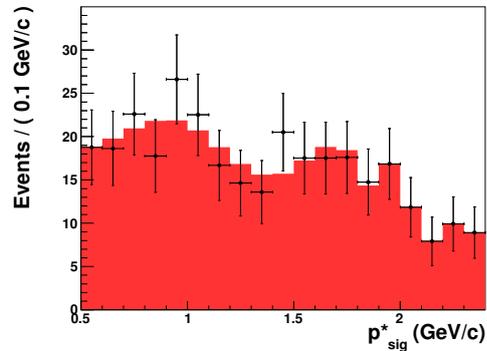
(3) Charged $b \rightarrow c$ decays.(4) Neutral $b \rightarrow c$ decays.(5) Charged $b \rightarrow u$ decays.(6) Neutral $b \rightarrow u$ decays.(7) Charged $b \rightarrow ul \nu_\ell$ decays.(8) Neutral $b \rightarrow ul \nu_\ell$ decays.

Figure B.18: Individual PDFs for different components for the signal decay $\tau^+ \rightarrow \rho^+ \bar{\nu}_\tau$ in p_{sig}^* . The black markers show the simulated data distribution and the solid shape the PDF after the smoothing. The number of events is scaled by the available amount of simulated data.

C Log of the data fit

```

[#1] INFO:Minization -- RooMinuit::optimizeConst: deactivating const optimization
[#1] INFO:Minization -- p.d.f. provides expected number of events, including extended term in likelihood.
[#1] INFO:Minization -- RooMinuit::optimizeConst: activating const optimization
RooAbsTestStatistic::initSimMode: creating slave calculator #0 for state Mu (3015 dataset entries)
RooAbsTestStatistic::initSimMode: creating slave calculator #1 for state E (3212 dataset entries)
RooAbsTestStatistic::initSimMode: creating slave calculator #2 for state Pi (977 dataset entries)
RooAbsTestStatistic::initSimMode: creating slave calculator #3 for state Rho (4186 dataset entries)
[#1] INFO:Fitting -- RooAbsTestStatistic::initSimMode: created 4 slave calculators.
[#1] INFO:Minization -- The following expressions have been identified as constant and will be precalculated and cached: (Pdf_signal_Mu,Pdf_backgroun
[#1] INFO:Minization -- The following expressions have been identified as constant and will be precalculated and cached: (Pdf_signal_E,Pdf_backgroun
[#1] INFO:Minization -- The following expressions have been identified as constant and will be precalculated and cached: (Pdf_signal_Pi,Pdf_backgroun
[#1] INFO:Minization -- The following expressions have been identified as constant and will be precalculated and cached: (Pdf_signal_Rho,Pdf_backgroun
*****
** 13 **MIGRAD          2500          1
*****
FIRST CALL TO USER FUNCTION AT NEW START POINT, WITH IFLAG=4.
START MIGRAD MINIMIZATION. STRATEGY 1. CONVERGENCE WHEN EDM .LT. 1.00e-03
FCN=-57110.6 FROM MIGRAD STATUS=INITIATE 12 CALLS 13 TOTAL
EDM= unknown STRATEGY= 1 NO ERROR MATRIX
EXT PARAMETER          CURRENT GUESS          STEP          FIRST
NO.  NAME  VALUE  ERROR  SIZE  DERIVATIVE
1  BR_B2taunu  2.16439e-04  5.96203e-05  1.34389e-01  5.27991e+01
2  nBkg_E  3.16499e+03  6.10001e+01  8.35710e-03  1.25702e+02
3  nBkg_Mu  3.00181e+03  5.87443e+01  8.22416e-03  1.43331e+02
4  nBkg_Pi  9.20315e+02  3.63284e+01  8.67111e-03  3.93497e+01
5  nBkg_Rho  4.06656e+03  7.09687e+01  8.81684e-03  3.21606e-02
ERR DEF= 0.5
MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-57116.7 FROM MIGRAD STATUS=CONVERGED 89 CALLS 90 TOTAL
EDM=4.4519e-06 STRATEGY= 1 ERROR MATRIX ACCURATE
EXT PARAMETER          STEP          FIRST
NO.  NAME  VALUE  ERROR  SIZE  DERIVATIVE
1  BR_B2taunu  1.24772e-04  2.80358e-05  9.41217e-03  -2.97736e-02
2  nBkg_E  3.14844e+03  5.80429e+01  1.28206e-03  -7.57857e-02
3  nBkg_Mu  2.96927e+03  5.56599e+01  1.27241e-03  -2.58724e-02
4  nBkg_Pi  9.36481e+02  3.22472e+01  1.21892e-03  -1.57809e-01
5  nBkg_Rho  4.11358e+03  6.63704e+01  1.32051e-03  7.62799e-02
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.  NDIM= 86  NPAR= 5  ERR DEF=0.5
 7.870e-10 -3.587e-04 -2.618e-04 -2.293e-04 -4.204e-04
-3.587e-04  3.369e+03  1.193e+02  1.045e+02  1.916e+02
-2.618e-04  1.193e+02  3.098e+03  7.630e+01  1.399e+02
-2.293e-04  1.045e+02  7.630e+01  1.040e+03  1.225e+02
-4.204e-04  1.916e+02  1.399e+02  1.225e+02  4.405e+03
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL  1  2  3  4  5
1  0.41021  1.000 -0.220 -0.168 -0.253 -0.226
2  0.22027  -0.220  1.000  0.037  0.056  0.050
3  0.16769  -0.168  0.037  1.000  0.043  0.038
4  0.25349  -0.253  0.056  0.043  1.000  0.057
5  0.22579  -0.226  0.050  0.038  0.057  1.000
*****
** 18 **HESSE          2500
*****
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=-57116.7 FROM HESSE STATUS=OK 31 CALLS 121 TOTAL
EDM=4.44817e-06 STRATEGY= 1 ERROR MATRIX ACCURATE
EXT PARAMETER          INTERNAL          INTERNAL
NO.  NAME  VALUE  ERROR  STEP SIZE  VALUE
1  BR_B2taunu  1.24772e-04  2.80669e-05  3.76487e-04  -5.60923e-02
2  nBkg_E  3.14844e+03  5.80606e+01  5.12825e-05  -7.54818e-01
3  nBkg_Mu  2.96927e+03  5.56692e+01  5.08964e-05  -7.79710e-01
4  nBkg_Pi  9.36481e+02  3.22641e+01  4.87569e-05  -1.13457e+00
5  nBkg_Rho  4.11358e+03  6.63886e+01  5.28203e-05  -6.29378e-01
ERR DEF= 0.5
EXTERNAL ERROR MATRIX.  NDIM= 86  NPAR= 5  ERR DEF=0.5
 7.888e-10 -3.613e-04 -2.637e-04 -2.314e-04 -4.231e-04
-3.613e-04  3.371e+03  1.208e+02  1.060e+02  1.938e+02
-2.637e-04  1.208e+02  3.099e+03  7.737e+01  1.415e+02
-2.314e-04  1.060e+02  7.737e+01  1.041e+03  1.242e+02
-4.231e-04  1.938e+02  1.415e+02  1.242e+02  4.408e+03
PARAMETER CORRELATION COEFFICIENTS
NO.  GLOBAL  1  2  3  4  5
1  0.41245  1.000 -0.222 -0.169 -0.255 -0.227
2  0.22158  -0.222  1.000  0.037  0.057  0.050
3  0.16865  -0.169  0.037  1.000  0.043  0.038
4  0.25541  -0.255  0.057  0.043  1.000  0.058
5  0.22694  -0.227  0.050  0.038  0.058  1.000
[#1] INFO:Minization -- RooMinuit::optimizeConst: deactivating const optimization

```

Danksagung

Diese Arbeit wäre nicht möglich gewesen ohne die Unterstützung vieler, vieler Personen.

Zuallererst gilt mein Dank natürlich Prof. Michael Feindt, meinem Referent und Doktorvater. Er hat mir Freiraum gewährt, war jederzeit zur Stelle wenn Hilfe oder Unterstützung notwendig war und hat eine tolle Arbeitsgruppe aufgebaut.

Prof. Günter Quast danke ich vielmals für die Übernahme des Korreferats. Sein Engagement für meine Arbeit und die terminliche Flexibilität haben vieles einfacher gemacht.

Ein riesen Dankeschön gilt selbstverständlich der *B*-Gruppe des EKP für die großartige Arbeitsatmosphäre. Besonders möchte ich mich bei Michael Prim, Christian Pulvermacher, Johannes Grygier, Michael Ziegler und Matthias Huschle für den direkten Input zu meiner Arbeit bedanken. Manch einer von ihnen musste auch die ein oder andere penetrante Nachfrage ertragen. Diese Liste muss leider unvollständig bleiben. Jedem Einzelnen zu danken, der mir bei dieser Arbeit geholfen hat würde den Rahmen einer Danksagung sprengen.

I also want to thank the great post-docs in our group. Thomas Kuhr, Anže Zupanc, Martin Heck, and Pablo Goldenzweig contributed a great part to my work; each of them in his own manner and each of them was of major importance.

Das EKP ist ein tolles Institut und die Leute, die es am Laufen halten verdienen definitiv eine Menge Dankbarkeit. Das Admin-Team ist stets zur Stelle, wenn es etwas zu reparieren gibt, egal ob am Wochenende oder nach Feierabend und ich habe noch nie einen kürzeren Dienstweg als den zum Core-Admin des Vertrauens erlebt. Ebenso bedanken möchte ich mich bei Frau Bräunling. Sie als die gute Seele des Instituts zu bezeichnen ist definitiv keine Übertreibung.

Die immerwährende Unterstützung meiner Eltern und der Rückhalt den sie mir geboten haben, haben es mir überhaupt erst möglich gemacht diesen Weg einzuschlagen. Meine Schwester hat dafür gesorgt, dass ich häufig genug über die richtigen Dinge nachgedacht habe und mir immer geholfen, den passenden Blickwinkel auf meine Arbeit zu behalten. Dafür, und für Tausende andere Dinge, bedanke ich mich vielmals bei ihnen und bei meiner ganzen Familie.