Search for Lepton Number Violating $B^0 \to \tau\ell$ Decays at the Belle Experiment with the Hadronic Full Event Interpretation

Nathalie Eberlein

München 2024

Search for Lepton Number Violating $B^0 \to \tau\ell$ Decays at the Belle Experiment with the Hadronic Full Event Interpretation

Nathalie Eberlein

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Nathalie Eberlein
aus Stuttgart

München, den 27.05.2024

Erstgutachter: Prof. Dr. Thomas Kuhr
Zweitgutachter: Prof. Dr. Otmar Biebel
Tag der mündlichen Prüfung: 12.07.2024

Zusammenfassung

Diese Arbeit präsentiert die Suche nach Leptonenzahl-verletzenden Zerfällen $B^0 \to \tau\ell$ mit $\ell = e, \mu$ am Belle Experiment. Im Standardmodell ist der Zerfall $B^0 \to \tau\ell$ stark unterdrückt und kann ausschließlich über Neutrinooszillationen auftreten. Sein winziges erwartetes Verzweigungverhältnis in der Größenordnung von $\mathcal{O}(10^{-50})$ liegt weit unterhalb der Sensitivität heutiger und zukünftiger Experimente. Das macht ihn zu einem idealen Kandidaten für die Suche nach Neuer Physikmodelle, wie zum Beispiel Leptoquarkmodelle. Diese sagen Verzweigungsverhältnisse in der Größenordnung von $\mathcal{O}(10^{-9})$ voraus.

Der Vorteil der Suche am Belle Experiment liegt in seiner sauberen Umgebung der $\Upsilon(4S)$-Zerfälle. Die Elektronen und Positronen kollidieren mit Schwerpunktsenergie von 10,58 GeV auf der $\Upsilon(4S)$-Resonanz, einem gebundenen $b\bar{b}$ Zustand. In nahezu allen Fällen zerfällt das $\Upsilon(4S)$ ausschließlich in zwei B Mesonen. Eines der beiden B-Mesonen (signal B) wird aus einem leichten Lepton und einem τ kombiniert, das in einem von sechs exklusiven Zerfallskanälen rekonstruiert wird. Der monoenergetische Impuls des leichten Leptons im Ruhesystem des signal Bs liefert ein klares Signal für Signalereignisse. Jedoch, verhindern die Neutrinos im Endzustand des τ-Zerfalls, die direkte Bestimmung des signal B Impluses aus seinen Zerfallsprodukten.

Da ausschließlich zwei B-Mesonen im Ereignis vorhanden sind, wird die Gesamtenergie der Elektron-Positron-Kollision gleichmäßig auf diese beiden verteilt, und sie sind im Schwerpunktsystem rückwärts zueinander orientiert. Mit dem vollständigen Wissen über die Kinematik des anderen B-Mesons im Ereignis kann der Impuls des signal B berechnet werden. Die "Full Event Interpretation", basiert auf multivariaten Algorithmen, wird verwendet, um das andere B-Meson in neutrino-freien Zerfallskanälen zu rekonstruieren.

Am Ende wurde die Anzahl der Signalereignisse in den Belle Daten, die einer Luminosität von 711 fb^{-1} entsprechen, in einem Fit des Leptonen-Impulsverteilung im Ruheframe des signal Bs bestimmt. Es wurden keine Signalzerfälle beobachtet, daher wurde obere Grenzwerte für die Verzweigungsverhältnisse von $B^0 \to \tau e$ und $B^0 \to \tau\mu$ bestimmt. Für das $B^0 \to \tau e$ wurde ein neuer bester oberer Grenzwert für das Verzweigungsverhältnis mit $\mathcal{B}(B^0 \to \tau e) < 1.2 \cdot 10^{-5}$ auf dem 90%-Konfidenzniveau ermittelt.

Abstract

This thesis presents the search for lepton-number violating decays $B^0 \to \tau\ell$ with $\ell = e, \mu$ at the Belle experiment. In the Standard Model, the $B^0 \to \tau\ell$ decay is strongly suppressed and can solely occur via neutrino oscillations. Its tiny expected branching ratio at the order $\mathcal{O}(10^{-50})$ is far below the sensitivity of today's and future experiments. The absence of Standard Model contributions makes it an ideal candidate for new physics searches. New physics models, for example models with leptoquarks, predict branching ratios of the order $\mathcal{O}(10^{-9})$.

The advantage of the search at Belle is its clean environment of the $\Upsilon(4S)$ decays. The electrons and positrons collide at a center-of-mass energy of 10.58 GeV at the $\Upsilon(4S)$ resonance, which is a $b\bar{b}$ bound state. In nearly all of the cases, the $\Upsilon(4S)$ decays exclusively into two B mesons and nothing else. One of the two B mesons (signal B) in the event is reconstructed from a light lepton and τ, which is reconstructed in one of six exclusive decay channels. The mono-energetic momentum of the light lepton in the signal B rest frame provides a clear signature for signal events. However, the neutrinos in τ decay final state, prevent the direct determination of the signal B momentum from its decay products.

Since only two B mesons are present in the event, the total energy of the electron-positron collision is equally distributed between these two and they are back-to-back in the center-of-mass frame. With the full knowledge of the kinematics of the other B meson in the event, the momentum of the signal B can be calculated. The Full Event Interpretation, which is based on multivariate algorithms, is employed to reconstruct the other B meson in neutrino-free decay channels.

In the end, the number of signal events in the Belle data corresponding to a luminosity of 711 fb^{-1} was extracted in a fit of the lepton momentum distribution in the signal B rest frame. No signal decays were observed, therefore upper limits on the branching ratios of $B^0 \to \tau e$ and $B^0 \to \tau\mu$ were determined. For the $B^0 \to \tau e$ channel a new best upper limit $\mathcal{B}(B^0 \to \tau e) < 1.2 \cdot 10^{-5}$ at 90% was found.

Unless otherwise declared, ℓ denotes either electrons or muons; natural units are used throughout this thesis, and charge conjugation is implied if not otherwise stated.

Contents

Chapter 1

Theory

This chapter introduces the Standard Model of Particle Physics with its three fundamental interactions with a special focus on the electroweak sector. Additionally, a short overview of the physics of B mesons is provided. Afterward, the physics phenomena beyond the Standard model are discussed with a special emphasis on lepton number violating processes. In the end, the previous analyses with the most stringent upper limits on the branching ratio of $B^0 \to \tau\ell$ decays are described.

1.1 Standard Model of Particle Physics

Most of the information provided in this chapter is based on the books "An Introduction To Quantum Field Theory" by M. Peskin and D. V. Schroeder, "Concepts of Elementary Particle Physics" [1] by M. Peskin and the book "The BABAR Physics Book" [2].

The Standard Model of Particle Physics in its current form was mainly developed in the 1970s and explains the three fundamental forces and the elementary matter particles in our universe to our current understanding.

Elementary matter particles carry a spin of 1/2 and are characterized by their mass, electric charge, and additional quantum numbers, which define the types of interactions they participate in. Each fermionic matter particle has an anti-matter particle with the opposite quantum numbers. There are two different groups of elementary particles: leptons and quarks. The leptons consist of three charged leptons and their respective neutral partners, the lepton neutrinos. Quarks have either an electric charge of 2/3 for up-type quarks or -1/3 for down-type quarks.

Both the leptons and quarks are divided into three generations. The particles of the first generation are the lightest and make up the matter we observe in our universe. The particles of the second and third generations are heavier and consequently decay into lighter first-

generation particles. The first quark generation consists of the up and down quark, the second of the charm and strange quark, and the third of the top and bottom quark. The six different types of quarks are called flavors. The electron and electron-neutrino comprise the first generation of leptons, the muon and muon-neutrino the second generation, and tau and tau-neutrino the third generation.

Originally, all matter particles were assumed to be massless, which contradicted the experimental measurements. With the discovery of the Higgs boson in 2012, the theory of the Higgs mechanism [3; 4; 5] by was verified, which also allows elementary matter particles to gain mass by coupling to the Higgs doublet. The origin of the neutrino masses remains unclear, and they are assumed to be massless in the Standard Model.

The Standard Model is a quantum field theory of the gauge group $SU(3) \otimes SU(2) \otimes U(1)$, describing the strong, weak, and electromagnetic interactions. These interactions are mediated by their respective spin-1 gauge bosons: the gluons (strong interaction), the $W^{\pm}$ and Z^0 (weak interaction), and the photon (electromagnetic force).
Fig. 1.1 provides an overview of the Standard Model particles.

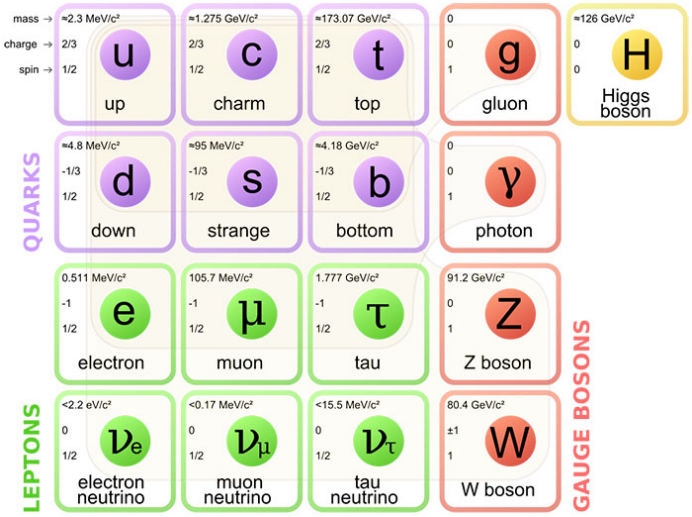

Figure 1.1: Particles of the Standard Model taken from [6]

The fundamental concept of gauge theory is the invariance of the Lagrangian, which describes the dynamics of a system under the gauge transformations. The transformations introduce additional terms in the Lagrangian, that define the interactions of the fermion fields with the mediator fields (covariant derivative D_μ). A particle is considered a singlet of a gauge transformation if its field remains unchanged under the gauge transformation, meaning it does not couple to the corresponding mediator fields.

1.1.1 Strong Interaction

The strong interaction is represented by the non-abelian $SU(3)_C$ symmetry group, where C denotes the associated conserved quantum number "color". The three associated colors are red, blue, and green, and their respective anti-colors. Eight massless gluons, carrying color and anti-color, mediate the strong interactions by coupling to gluons and quarks (color-triplets) with the strong coupling constant g_s. Leptons do not participate in strong interactions and are color singlets. The strong interaction conserves the flavor of quarks.

In experiments, however, no free quarks and gluons are observed. Instead, the strong force binds quarks into color singlets known as hadrons. Hadrons can be divided into mesons consisting of a valence quark-antiquark pair with an integer spin, and baryons composed of three valence quarks or anti-quarks with odd half-integer spins (e.g., 1/2 or 3/2).

At high energies, the otherwise bound quarks and gluons behave as nearly free particles, which is known as the asymptotic freedom of the strong coupling constant. At low energies, the quarks and gluons are confined into hadrons. The behavior of the strong coupling constant complicates predictions in the low-energy regime since perturbation theory is no longer applicable.

1.1.2 Electroweak Interaction

The electroweak theory [7; 8; 9] unifies the electromagnetic and weak interactions under the non-abelian gauge symmetry group $SU(2)_L \otimes U(1)_Y$. This symmetry group introduces the quantum numbers weak isospin T_3 and the hypercharge Y, which are related to the electric charge Q by $Q = \frac{1}{2}Y + T_3$.

Three massless W bosons associated with $SU(2)_L$ and a massless B boson associated with $U(1)_Y$ mediate the electroweak interaction. The field strength tensors of the B field denoted by B_μ and the W field with its components a = 1,..,3 are represented by W^a_μ are given by:

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g\epsilon^{abc} W^b_\mu W^c_\nu \end{aligned} \tag{1.1}$$

The associated covariant derivative is defined as:

$$\begin{aligned} D_\mu &= \partial_\mu - igW_\mu - ig'\tfrac{1}{2}YB_\mu \\ W_\mu &= \tfrac{1}{2}W^a_\mu \sigma_a \end{aligned} \tag{1.2}$$

where g represents the coupling constant of $SU(2)_L$, g' denotes the coupling constant of $U(1)_Y$ and σ_a denotes the three Pauli matrices.
The charged weak interaction only couples to left-handed fermions or right-handed anti-fermions. Right-handed fermions and left-handed anti-fermions are singlets of the charged

weak interaction. Tab. 1.1 provides an overview of the weak isospin and hypercharge for the left-handed and right-handed fermions.

Left-handed fermions			
	Q	(T, T_3)	Y
$\nu_{eL}, \nu_{\mu L}, \nu_{\tau L}$	0	$(1/2, +1/2)$	-1
e_L, μ_L, τ_L	-1	$(1/2, -1/2)$	-1
u_L, c_L, t_L	$+2/3$	$(1/2, +1/2)$	$+1/3$
d_L, s_L, b_L	$-1/3$	$(1/2, -1/2)$	$+1/3$

Right-handed fermions			
	Q	(T, T_3)	Y
$\nu_{eR}, \nu_{\mu R}, \nu_{\tau R}$	Not part of the SM		
e_R, μ_R, τ_R	-1	$(0, 0)$	-2
u_R, c_R, t_R	$+2/3$	$(0, 0)$	$+4/3$
d_R, s_R, b_R	$-1/3$	$(0, 0)$	$-2/3$

Table 1.1: Quantum numbers of left-handed and right-handed fermions.

In experiments, the weak interaction was observed to be short-ranged, which could be caused by a small coupling constant. However, unifying the weak interactions with the infinitely ranged electromagnetic interactions posed a challenge under the assumption of an extremely small coupling constant for the weak interaction. The observation of the massive weak bosons explained that the origin of the short-range interactions is caused by the heavy mass of the bosons and not necessarily by a small coupling constant. The observed masses introduced a new challenge since a non-abelian gauge theory requires massless vector bosons to ensure renormalizability.

The solution was found in the work of Nambu [10] and Goldstone [11], who described the concept of spontaneous symmetry breaking (SSB) and the emergence of Goldstone bosons. The original symmetry of the electroweak interaction is broken $SU(2)_L \otimes U(1)_Y \xrightarrow{\text{SSB}} U(1)_Q$ and the three bosons of the weak interaction acquire a mass, while the fourth boson, photon, remains massless.

1.1.3 Spontaneous Symmetry Breaking

If the choice of a vacuum state does not conserve the symmetry of a symmetry group, it is called spontaneous symmetry breaking. In the simple case of a global $U(1)$ symmetry group, the Lagrangian of a scalar complex field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$ is described by:

$$\begin{aligned} \mathcal{L} &= \partial_\mu \phi^\dagger \partial^\mu \phi - V(\phi) \\ \text{with } V(\phi) &= \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \end{aligned} \tag{1.3}$$

The potential $V(\phi)$ is invariant under $U(1)$ transformations:

$$\phi(x) \to \phi'(x) \equiv e^{i\theta} \phi(x) \tag{1.4}$$

Fig. 1.2 depicts the potential of the scalar field with $\mu^2 > 0$ on the left, which corresponds to a massive particle with mass μ. The potential with $\mu^2 < 0$ shown on the right contains

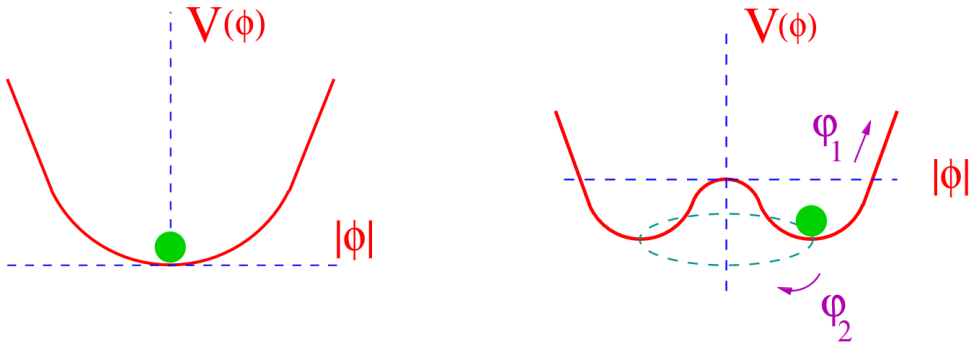

Figure 1.2: Shape of the scalar potential $V(\phi)$ for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). The potential on the right contains a continuous set of degenerate vacuum states connected through a massless field excitation φ_2. The figures are taken from [12, p. 15]

an infinite number of non-zero vacuum states with the same energy. Choosing one specific vacuum state hides the original symmetry of the potential. The vacuum expectation value v of the scalar field is the minimum of the potential $\langle\phi_1\rangle_{min} = \sqrt{-\mu^2/\lambda} \equiv v$.

The scalar field ϕ can be expressed with respect to the vacuum ground state:

$$\begin{aligned} \phi &= \langle\phi\rangle_{min} + \varphi \\ \phi_1 &= v + \varphi_1 \\ \phi_2 &= \varphi_2. \end{aligned} \tag{1.5}$$

The Lagrangian $\mathcal{L}$, see Eq. 1.3, transforms under $U(1)$ to $\mathcal{L}'$ and is described as:

$$\mathcal{L} \Rightarrow \mathcal{L}' = \frac{1}{2}\partial_\mu\varphi_1\partial^\mu\varphi_1 - \frac{1}{2}\mu^2\varphi_1^2 - \frac{1}{2}\partial_\mu\varphi_2\partial^\mu\varphi_2 + \text{self-interaction}. \tag{1.6}$$

The field φ_1 is a scalar field with mass $m^2_{\varphi_1} = -2\mu^2$, while the field φ_2 corresponds to a massless excitation since the Lagrangian contains no mass term associated with φ_2. The massless excitations represent a spin-0 Goldstone boson, which appears due to the broken symmetry in the vacuum state.

1.1.4 Higgs Mechanism

The Higgs mechanism explains the symmetry breaking of $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$. Since three weak bosons must acquire a mass, three Goldstone bosons are required. Therefore, the scalar complex field ϕ is replaced by a $SU(2)_L$ doublet of complex scalar fields with the hypercharge $Y = 1$ and four degrees of freedom.
The $SU(2)$ doublet is represented by :

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(\pi^1 + i\pi^2) \\ \frac{1}{\sqrt{2}}(v + h + i\pi^3) \end{pmatrix} \xrightarrow{\text{vaccuum state}} \phi_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \tag{1.7}$$

and the three Goldstone boson fields π^1, π^2 and π^3 as well as the physical Higgs boson field h. The vacuum state must be zero in the ϕ^+ component to prevent couplings to the photon field and ensures the photon remains massless.

The Lagrangian of the electroweak gauge bosons and the Higgs field is defined as:

$$\begin{aligned} \mathcal{L}_{\text{gauge}} &= (D_\mu \phi)^\dagger D^\mu \phi - V\left(\phi^\dagger \phi\right) - \tfrac{1}{4} W^a_{\mu\nu} W_a^{\mu\nu} - \tfrac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ &\text{with } V\left(\phi^\dagger \phi\right) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \end{aligned} \tag{1.8}$$

The invariance of local gauge transformation results in a coupling between the massless Goldstone bosons to the weak bosons W^a_μ fields. The Goldstone bosons are then absorbed by the W bosons giving them a mass. In the end, only the physical field h out of the four initial fields of the $SU(2)$ doublet ϕ remains.

The Lagrangian describing the mass terms of gauge bosons and the Higgs is given by:

$$\begin{aligned} \mathcal{L}_{\text{gauge, mass}} = &-\tfrac{1}{2} m^2 h^2 + \tfrac{1}{8} v^2 g^2 \left(W^1_\mu W^{1,\mu} + W^2_\mu W^{2,\mu}\right) \\ &+\tfrac{1}{8} v^2 \left(g^2 W^3_\mu W^{3,\mu} + g'^2 B_\mu B^\mu - 2 g g' W^3_\mu B^\mu\right). \end{aligned} \tag{1.9}$$

The experimentally observed charged W bosons $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp i W^2)$ are a linear combinations of the charged W fields W^1 and W^2 with a acquired mass $M_W = \frac{1}{2} v g$. The Z boson and photon are mixtures of the W^3_μ and B_μ fields, which are related via the Weinberg angle θ_w:

$$\begin{aligned} Z_\mu &= \cos(\theta_w) W^3_\mu - \sin(\theta_w) B_\mu \\ A_\mu &= \sin(\theta_w) W^3_\mu + \cos(\theta_w) B_\mu \end{aligned} \tag{1.10}$$

This results in a massless photon and a massive Z boson with $M_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}$, which is in agreement with the experimental observations.

The physical field h represents the Higgs boson with the mass $M_h^2 = 2 v^2 \lambda$ with an electrical charge $Q = 0$ and $T_3 = -1/2$. In 2013, the Nobel Prize in Physics was awarded to F. Englert and P. Higgs for the theoretical explanation of the Higgs mechanism [13], following the experimental confirmation of the Higgs boson's existence by the ATLAS [14] and CMS [15] experiment.

1.1.5 Fermion Masses

The coupling, known as Yukawa coupling, between the fermion fields and the Higgs doublet ϕ generates the fermion masses. If the neutrinos are assumed to be massless, the Lagrangian for the Yukawa couplings in the (weak) interaction basis I is described by:

$$\mathcal{L}^I_{\text{Yuk}} = Y^e_{ij}\bar{L}_{Li}\phi E_{Rj} + Y^d_{ij}\bar{Q}_{Li}\phi D_{Rj} + Y^u_{ij}\bar{Q}_{Li}\phi^c U_{Rj} + \text{ h.c.} \qquad \forall i,j \in [1,2,3]. \tag{1.11}$$

L_{Li} and Q_{Li} denote the $SU(2)$ doublets of the leptons and quarks in generation i, respectively.

$$L_{Li} = \begin{pmatrix} \nu_\ell \\ E_i \end{pmatrix}, \qquad Q_{Li} = \begin{pmatrix} U_i \\ D_i \end{pmatrix} \tag{1.12}$$

E_{Rj}, U_{Rj}, and D_{Rj} represent the $SU(2)$ singlets for the right-handed charged lepton, up-type and down-type quark of generation j.

The Yukawa couplings denoted as Y^e_{ij} for the charged leptons and Y^u_{ij}/Y^d_{ij} for the up-type/down-type quarks, connect the vacuum expectation value v of the Higgs doublet to the mass matrix:

$$M_{ij} = \frac{v}{\sqrt{2}} \cdot Y_{ij}. \tag{1.13}$$

The masses for the up-type quarks are generated through their coupling to the charge conjugate of the Higgs doublet:

$$\phi^c = i\sigma_2\phi^\dagger = \begin{pmatrix} \bar{\phi}^0 \\ -\bar{\phi}^+ \end{pmatrix} \to \frac{1}{\sqrt{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix} \tag{1.14}$$

To express the Lagrangian in the mass basis (denoted by small letters) instead of $SU(2)$-interaction basis (denoted by capital letters), the fields are transformed by unitary matrices:

$$\begin{array}{cc} \text{leptons} & \text{quarks} \\ \ell_{Li} = (U_{\ell L})_{ij}\, L_{Lj} & q_{Li} = (V_{qL})_{ij}\, Q_{Lj} \\ e_{Ri} = (U_{eR})_{ij}\, E_{Rj} & u_{Ri} = (V_{uR})_{ij}\, U_{Rj} \\ & d_{Ri} = (V_{dR})_{ij}\, D_{Rj}. \end{array} \tag{1.15}$$

A mass matrix can be diagonalized via:

$$M_f^{\mathbf{diag}} = V_{fL}^\dagger M_f V_{fR}. \tag{1.16}$$

The charged lepton mass matrix is diagonalized by appropriately choosing the two matrices $U_{\ell L}$ and U_{eR}. The analogous procedure is applied to the up-quark mass matrix.

$$M_u^{\mathbf{diag}} = V_{u_L}^\dagger M_u V_{u_R} \tag{1.17}$$

Since the mass matrices of the up- and down-type quarks differ, a simultaneous diagonalization of these two mass matrices is not possible:

$$\begin{gathered} M_d^{\mathbf{diag}} = V_{d_L}^\dagger M_d V_{d_R} \\ V_{d_L}^\dagger \neq V_{u_L}^\dagger \end{gathered} \tag{1.18}$$

This leads to flavor mixing in the charged weak interaction described by the CKM matrix.

1.1.6 CKM Matrix

The Lagrangian describing flavor-changing interaction of the charged W bosons and the quarks is given by:

$$\mathcal{L}_{W^\pm}^q = \frac{g}{\sqrt{2}} \bar{u}_L^i \gamma^\mu \left(V_{u_L}^\dagger V_{d_L} \right)^{ij} d_L^j W_\mu^+ + \text{ h.c.} \tag{1.19}$$

in the mass basis. The matrix denoted by $V = V_{u_L}^\dagger V_{d_L}$ is the CKM matrix, named after its main contributors Cabibbo, Kobayashi, and Maskawa [16; 17]. Since the CKM matrix is not the identity matrix, the W boson can couple between the different quark generations. This is the only known flavor-changing interaction at tree level in the Standard Model.

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}$$

The CKM matrix contains the sine (s) and cosine (c) values of the three real parameters: the angles $\theta_{12}(12)$, $\theta_{13}(13)$ and $\theta_{23}(23)$, which represent the mixing between the generation i and j. It also includes one imaginary parameter δ, which is the source of CP (joint charge and parity symmetry) violation in the Standard Model.

The relations:

$$\sum_k^{d,s,b} V_{ik}^* V_{jk} = \delta_{ij} \qquad \forall\ i, j \text{ in } [u, c, t] \tag{1.20}$$

ensure the unitarity of the CKM matrix. Precisely measuring the elements of the CKM matrix is an important test of the Standard Model, and observed deviations can potentially indicate new physics phenomena.
The current values according to the PDG [18] are:

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}$$

The small values of the off-diagonal matrix elements represent the lower probability of flavor-changing interactions between different quark generations compared to interactions within the same generation.

1.1.7 Lepton Number Conservation and Lepton Universality

Two fundamental principles of the Standard Model are lepton number conservation and lepton universality. Lepton number conservation implies that each process conserves the individual electron, muon, and tau lepton number. Lepton universality means that the interaction strengths of the different lepton flavors are identical, apart from differences caused by their masses. Observations deviating from those principles, e.g. the anomalies in semi-leptonic B meson decays, hint at possible new physics contributions.

1.1.8 Physics of B mesons

The b quark is the heaviest quark that is bound within color-neutral hadrons since the heavier top quark decays before the formation of hadrons occurs. The B mesons, which contain an anti-b quark and either an u for (B^+) or d quark (B^0) and their charge-conjugated partners, are the central focus of study at Belle. They can only decay via the flavor-changing charged weak interaction, which is suppressed by the small off-diagonal elements in the CKM matrix, resulting in a relatively long B meson lifetime.

The physics of B mesons is governed by interactions occurring at different energy scales. The weak decay can be simplified to the b quark decaying into a lighter quark, while the accompanying light quark in the B meson is just a spectator, which does not interact with the b quark or its decay products. The weak process is in the perturbative energy regime of the Standard Model. In addition to the weak interaction, the strong interaction plays an important role for the dynamics inside of the B meson. The confinement in QCD prevents the color-charged b quark from being seen as an isolated, freely propagating quark and instead binds it via soft gluons into a color-neutral hadron.

To precisely predict the decay rate of B mesons processes, all contributing virtual states need to be calculated, which proves to be difficult. Particularly because the low energetic, strong interactions in hadrons are in the non-perturbative scheme of QCD. The simple weak decay of the B meson is dominated by the complexity of the strong interactions inside of the hadron. Fig. 1.3 illustrates the different interactions in a B meson decay.

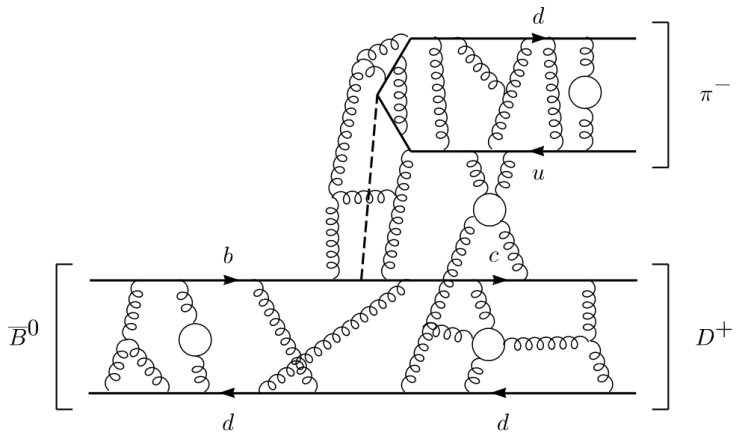

Figure 1.3: Interactions taking place in a B meson decay taken from [19, p. 3].

Effective field theory provides a useful approach for separating the physics of different interactions into single-energy-scale problems. It fully describes the physics occurring below a chosen energy threshold μ and reduces the number of degrees of freedom of the

physics above the threshold by integrating out the quantum fluctuations of particles with masses $M > \mu$. The effective Lagrangian is only valid to a cutoff energy, which corresponds to the mass of the removed particle and can exhibit approximate symmetries, which are not necessarily symmetries of the Lagrangian of the complete underlying theory.

Heavy Quark Effective Theory

One example of an effective field theory is the Heavy Quark Effective Theory (HQET), which is discussed in more detail in [19], [18, chapter 16] and [2, chapter 2]. It describes the interactions of a single heavy quark with the other light constituents in the hadrons. The energy threshold μ is defined such that $\Lambda_{QCD} \ll \mu \ll m_Q$, which is illustrated in Fig. 1.4.

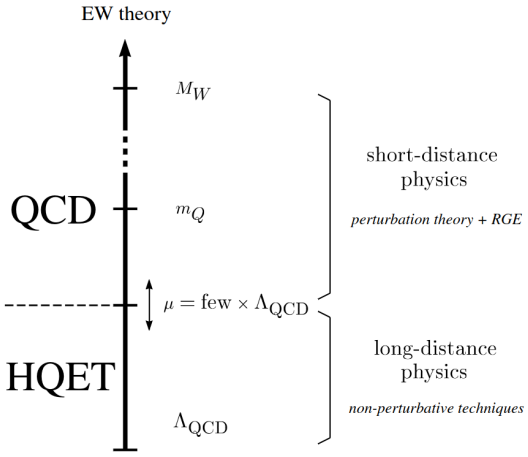

Figure 1.4: Energy scales of interactions in HQET taken from [19, p. 18].

The short-distance (high-energy) interactions, charged weak interactions and QCD interactions above the mass of heavy c or b quarks (m_Q) can be calculated using perturbation theory and renormalization group techniques, while the long-distance (low-energy) interactions rely on nonperturbative techniques.

In the energy regime of the HQET, the heavy b or c quark carries the same momentum of the order few Λ_{QCD} as the so-called light constituents: gluons and light quarks. Thus, the momentum transfers between them and the heavy quark are too low to resolve the flavor or spin of the heavy quark $1/m_Q \gg 1/\Lambda_{QCD}$. Instead, they only observe the electric and chromoelectric fields of the heavy quark. In the limit $m_Q \to \infty$, the properties of the hadrons containing a b or a c quark are related by the flavor and spin symmetry of the effective Lagrangian for hadrons with the same velocity at the leading order. The uncertainties of the symmetry-breaking effects can be studied in a systematic way.

1.2 Physics beyond the Standard Model

Despite the remarkable success of the Standard Model in explaining the three interactions, it leaves several unanswered questions. A selection of these unresolved questions is listed here:

1. Does a grand unified theory of all three interactions exist, similar to the unification of the electromagnetic and weak interaction? The symmetry of the unified theory would be broken down to our known Standard Model symmetry at short distances.

2. What explains the observed asymmetry between matter and antimatter in our universe? According to the inflationary universe theory, matter and antimatter were in equilibrium at the beginning of the universe. The CP violation phase in the Standard Model is orders of magnitude too small to explain the observed asymmetry of approximately 10^{-10} [20].

3. Neutrinos are assumed to be massless in the Standard Model, which was contradicted by the observation of neutrino oscillations [21]. Consequently, it proves that the lepton flavor is not a conserved quantity of nature. Studying neutrinos improves our insight into the CP violation in the lepton sector, which might help us understand the observed matter-antimatter asymmetry.

Many new physics models exist, offering possible solutions to the phenomena that the Standard Model can not explain.

1.2.1 Lepton Flavor Violation in the Neutrino Sector

Neutrinos in the Standard Model are assumed to be massless Dirac particles, which implies the conservation of the three lepton numbers L_e, L_μ, L_τ. The idea of neutrino oscillation was first proposed by Pontecorvo [22] even before the discovery of the ν_μ and ν_τ. Similar to the oscillation of the K^0 [23], he discussed the notion of $\nu \leftrightarrow \overline{\nu}$ oscillations leading to a $\Delta L = 2$ violation of the lepton number. In 1962, following the direct discovery of the ν_μ neutrino, Maki, Nakagawa, and Sakata [24] proposed the concept of oscillating neutrinos due to their masses in analogy to the quark sector. Neutrino oscillations occur if the weak flavor eigenstate $(\nu_e, \nu_\mu, \nu_\tau)$ differ from the mass eigenstates (ν_1, ν_2, ν_3) and if their masses are distinct.

The low interaction rate of neutrinos complicated the first conclusive observation of neutrino oscillations. Finally, in 1998, significant evidence of neutrino oscillations was observed at the Super-Kamiokande experiment [21] in a deficit of measured atmospheric muon neutrinos. This discovery was awarded with the Nobel Prize in 2015 [25].

The Superkamiokande experiment measured the muon and electron neutrino flux generated from interactions in the atmosphere. Cosmic rays interact with atomic nuclei in the atmosphere and create, for example, pions. These pions preferentially decay into muons and muon neutrinos:

$$\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}\,(\bar{\nu}_{\mu})$$

The created muons, with typically energies E $<$ 1 GeV, decay further before reaching the earth's surface.

$$\mu^{\pm} \rightarrow e^{\pm} + \nu_{e}\,(\bar{\nu}_{e}) + \bar{\nu}_{\mu}\,(\nu_{\mu})$$

These interactions result in an expected muon neutrino flux, approximately twice the size of the electron neutrino flux. The flavor of the neutrino was determined in a 50,000t water Cherenkov detector by detecting either an electron or a muon created in charged weak interactions of the incoming neutrino with nuclei. In addition to the flavor of the neutrinos, their original direction was reconstructed as the zenith angle Θ. Fig. 1.5 shows the measured neutrino flux for the electron and muon neutrino events.

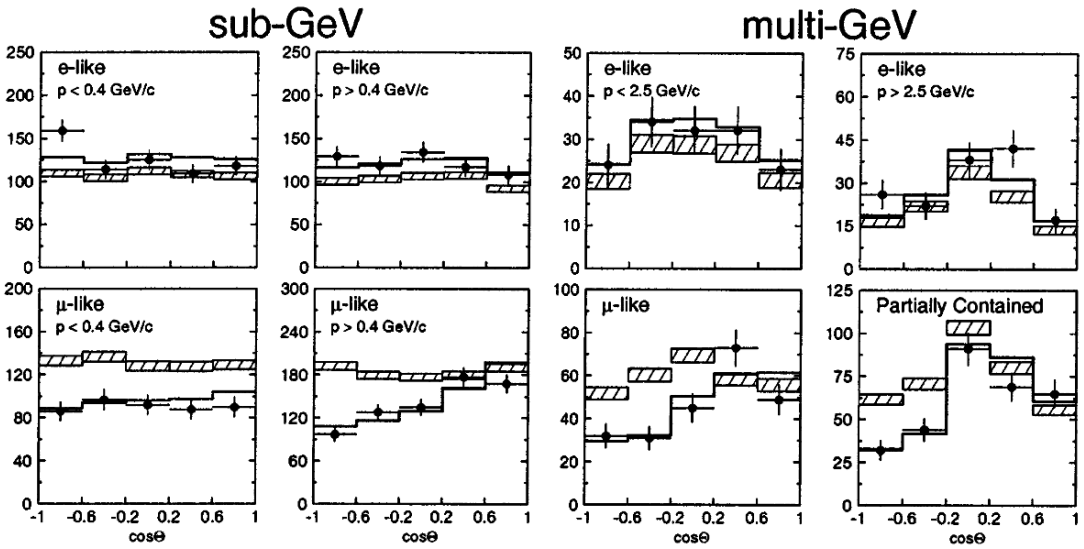

Figure 1.5: The zenith angular distributions for sub- and multi-GeV e-like and μ-like atmospheric neutrino events observed in Super-Kamiokande [21]. Values of $\cos(\Theta) < 0$ represent the events where the neutrino traversed the earth before being detected, and $\cos(\Theta) > 0$ events where the neutrino reached the detector directly from the atmosphere. The Monte Carlo expectation assumes no neutrino oscillation and is represented by the hatched region, while the black points denote the measured data. The black line is the best-fit result for $\nu_{\mu} \leftrightarrow \nu_{\tau}$ oscillation. The partially contained events denote the events where part of the Cherenkov light was detected in the outer detector instead of being fully contained in the inner detector.

In the case of no neutrino oscillations, the number of neutrinos is expected to be approximately independent of the origin direction because the neutrinos interact only very weakly with the matter in the earth. The result showed a significant deviation from the expected flux of muon neutrinos traversing the earth before reaching the detector. Since the electron neutrino flux agrees with the expected distribution, it indicates that these μ-neutrinos most likely oscillated into τ-neutrinos instead which was proven later in 2013 [26].

Neutrino oscillations occur if their interaction eigenstates (I) differ from their mass eigenstates (M) and their mass difference $m_i - m_j \neq 0$. In honor of its main contributors Pontecorvo, Maki, Nakagawa, and Sakata, the neutrino mass mixing matrix is known as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix}_I = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}_M \tag{1.21}$$

The PMNS matrix has the same structure as the CKM matrix, with possible additional phases α_1 and α_2, if neutrinos are their own anti-particles, so-called Majorana particles.

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} e^{i\alpha_1/2} & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.22}$$

The probability of detecting a neutrino of a specific flavor changes as it propagates through space. Since the mass differences between the neutrinos are small, neutrino oscillations can only be observed over large distances. The probability, that a neutrino of flavor α oscillated into a neutrino of flavor β over the distance L is given by:

$$\begin{aligned} P_{\nu_\alpha \to \nu_\beta} = \delta_{\alpha\beta} &- 4 \sum_{j>i} \mathrm{Re}\left(U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j}\right) \sin^2\left(\frac{\Delta m^2_{ij} L}{4E}\right) \\ &+ 2 \sum_{j>i} \mathrm{Im}\left(U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta j}\right) \sin\left(\frac{\Delta m^2_{ij} L}{2E}\right). \end{aligned} \tag{1.23}$$

It depends on the energy E of the neutrino and the squared mass difference $\Delta m^2_{ij} = m^2_i - m^2_j$ of the neutrino mass eigenstates. By measuring the difference between the probability of neutrino and antineutrino oscillation, the CP asymmetry can be calculated.

The angles of the PMNS matrix are studied using solar, atmospheric, and reactor neutrinos. There are two possible ordering for the neutrino masses: the normal $m_1 < m_2 < m_3$ and the inverted ordering $m_3 < m_1 < m_2$. The squared mass difference and the values of PMNS matrix elements are obtained from a global fit of neutrino oscillation parameters [27].

These are the results of the fit assuming normal ordering:

$$\begin{aligned} m_2^2 - m_1^2 &= 7.50^{+0.22}_{-0.20} \times 10^{-5}\text{eV}^2 & |m_3^2 - m_1^2| &= 2.55^{+0.02}_{-0.03} \times 10^{-3}\text{eV}^2 \\ \theta_{12} &= 34.3^\circ \pm 1.0^\circ & \theta_{23} &= 49.26^\circ \pm 0.79^\circ \\ \theta_{13} &= 8.53^{\circ\,+0.13^\circ}_{\quad -0.12^\circ} & \delta &= 194^{\circ\,+24^\circ}_{\quad -22^\circ} \end{aligned} \tag{1.24}$$

Since neutrino mixing between flavors is less suppressed than for quarks, the PMNS matrix exhibits larger off-diagonal elements than the CKM matrix.
The question about the nature of the mechanism giving the neutrinos their masses and their absolute values remains. Answers may be found in the neutrinoless double-β decays studied at KATRIN [28], which would indicate that neutrinos are Majorana particles, and in cosmology observations, which provide an upper limit on the sum of all three neutrino masses.

1.2.2 Lepton Flavor Violation in the Charged Lepton Sector

The observed lepton flavor violation in the neutrino sector proved that lepton flavor conservation is not a symmetry of nature and can also be extended to the charged leptons. If the neutrinos gain their mass by the same Yukawa coupling to the Higgs as the other fermions, short-range charged lepton flavor violating decays remain strongly suppressed due to the unitarity constraints of the PMNS matrix. The expected branching ratios for charged lepton flavor-violating decays are at the order of 10^{-50} [29], which is far below the sensitivity of today's experiments.

Flavor-changing neutral currents at tree level are not permitted in the Standard Model because both the Z bosons and the photon conserve flavor. Therefore, the lowest-order process for flavor-changing neutral currents is only possible via the exchange of two W bosons. Fig. 1.6 shows the lowest order box diagram of the Standard Model, including neutrino masses.

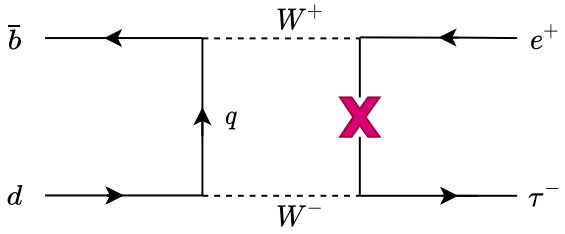

Figure 1.6: Sketch of the lowest order box diagram of the $B^0 \to \tau e$ decay in the Standard Model with the extension of massive neutrinos. The q quark in the loop represents all up-type quarks, where the contribution from the top quark dominates due to its heavy mass. The X denotes the oscillation between an electron-neutrino and a τ-neutrino.

Discovering charged lepton flavor-violating decays is a clear sign of new physics contributions, and even null results can help constrain their parameter space. Possible contributions can arise from decays involving leptoquarks with expected branching ratios at the order of 10^{-9} [30] for a vector leptoquark with a lower mass of 86 TeV.

Fig. 1.7 shows the tree-level diagram of the $B^0 \to \tau\ell$ decay via the exchange of a leptoquark.

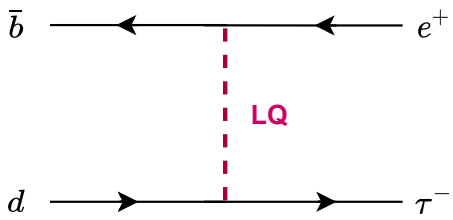

Figure 1.7: Sketch of the tree-level diagram of the $B^0 \to \tau e$ decay via the exchange of a leptoquark (LQ).

In recent years, leptoquark models have gained significant interest. They provide possible explanations for the observed deviation from the Standard Model prediction for the lepton universality in B meson decays and the anomalous magnetic moment of the muon.

1.2.3 Leptoquarks

Leptoquarks are mediator bosons that unify the lepton and color charges, allowing them to couple to individual quarks and leptons. They are often associated with grand unified theories and thus serve as tentative indicators of the unification of the three forces. The concept of leptoquark was first proposed by Pati and Salam in 1974 [31]. Because no leptoquarks have been observed so far, the generation of their masses is still unclear; one possible mechanism is spontaneous symmetry breaking to the Standard Model symmetry group analogous to the Higgs mechanism in the Standard Model.

The coupling of the leptoquarks to Standard Model particles strongly depends on the assumed leptoquark masses of each model, which tends to be quickly outdated with new experimental measurements. Therefore, only an overview of their general properties is discussed instead of specific leptoquark models. A more detailed description can be found in [32]. The leptoquark masses are expected to be in the order of TeV due to the lack of observation in the dedicated leptoquark searches at CERN. Leptoquarks can be either vector or scalar bosons and potentially mediate proton decay. Determining the upper limits on the proton lifetime severely constrains the parameter space of leptoquark models.

Tab. 1.2 lists the general properties of various scalar and vector leptoquark types with respect to the Standard model symmetry group. F denotes the fermion number, defined as $F = 3B + L$, where B represents the baryon and L represents the lepton number.

$(SU(3)_C, SU(2)_L, U(1)_{Y/2})$	Spin	Coupling to fermions	Symbol	Fermion number F
$(\overline{\mathbf{3}}, \mathbf{3}, 1/3)$	0	LL	S_3	-2
$(\mathbf{3}, \mathbf{2}, 7/6)$	0	RL, LR	R_2	0
$(\mathbf{3}, \mathbf{2}, 1/6)$	0	$RL, \overline{LR}$	$\tilde{R}_2$	0
$(\overline{\mathbf{3}}, \mathbf{1}, 4/3)$	0	RR	$\tilde{S}_1$	-2
$(\overline{\mathbf{3}}, \mathbf{1}, 1/3)$	0	$LL, RR, \overline{RR}$	S_1	-2
$(\overline{\mathbf{3}}, \mathbf{1}, -2/3)$	0	$\overline{RR}$	$\bar{S}_1$	-2
$(\mathbf{3}, \mathbf{3}, 2/3)$	1	LL	U_3	0
$(\overline{\mathbf{3}}, \mathbf{2}, 5/6)$	1	RL, LR	V_2	-2
$(\overline{\mathbf{3}}, \mathbf{2}, -1/6)$	1	$RL, \overline{LR}$	$\tilde{V}_2$	-2
$(\mathbf{3}, \mathbf{1}, 5/3)$	1	RR	$\tilde{U}_1$	0
$(\mathbf{3}, \mathbf{1}, 2/3)$	1	$LL, RR, \overline{RR}$	U_1	0
$(\mathbf{3}, \mathbf{1}, -1/3)$	1	$\overline{RR}$	$\bar{U}_1$	0

Table 1.2: List of leptoquark multiples and their Standard Model representation modified from [32, p. 4]. L denotes the left-handed $SU(2)_L$ doublets, and R the right-handed $SU(2)_L$ singlet with the respective charge conjugates $\overline{L}$ and $\overline{R}$. The definition for the $Q = T_3 + Y/2$.

Quarks are triplets of the $SU(3)_C$ gauge symmetry, while leptons are singlets. To ensure the conservation of the gauge symmetry, leptoquarks that couple to a quark and lepton must be triplets of $SU(3)_C$, which also allows leptoquark couplings to diquark-pairs, but not to a lepton pair. Depending on their coupling to the left- or right-handed fermions, they can be singlets, doublets, or triplets under $SU(2)_L$ transformations. The last value denotes the 1/2 leptoquark hypercharge, where the electric charge $Q = T_3 + \frac{Y}{2}$. Leptoquarks can carry electric charges of $\pm\frac{5}{3}$, $\pm\frac{4}{3}$, $\pm\frac{2}{3}$, and $\pm\frac{1}{3}$.

1.3 Previous Analyses

The lepton flavor violating decay $B^0 \to \tau\ell$ has been investigated in several studies without the observation of any signal events. Therefore, upper limits on the signal branching ratio of $B^0 \to \tau e$ and $B^0 \to \tau\mu$ were determined, and the most stringent upper limits are given in Tab. 1.3.

	UL $\mathcal{B}$ at 90% CL	Experiment
$B^0 \to \tau e$	$< 1.6 \times 10^{-5}$	Belle(2021) [33]
$B^0 \to \tau\mu$	$< 1.2 \times 10^{-5}$	LHCb(2019) [34]

Table 1.3: Best upper limits on the branching ratio of $B^0 \to \tau\ell$ decays at 90% confidence level(CL).

The Belle analysis benefits from a clean experimental environment because it operates just above the threshold of the $\Upsilon(4S)$ resonance, which almost always decays exclusively into two B mesons. The energy of the initial collision is well known, which simplifies the study of decays with neutrinos in the final state. One of the two B mesons is reconstructed in exclusive decay modes without neutrinos in the final state, employing neural nets. The full knowledge of the B meson kinematics helps to constraint the kinematics of the signal B meson. The invariant mass of the inclusively reconstructed τ is calculated by utilizing the two-body decay of the signal B. In the end, the upper limit of signal events is extracted in a fit of the invariant τ mass, and the corresponding upper limit on the branching ratio is determined. The study was performed on the full 711 fb^{-1} Belle data set.

The best upper limit for the branching ratio of $B^0 \to \tau\mu$ decays was determined on 3 fb^{-1} of LHCb data collected at CMS energy of 7 and 8 TeV in proton-proton collisions. Final states with neutrinos present a challenge, because the colliding partons (gluons and quarks) carry only an arbitrary fraction of the total proton momentum and thus the initial energy of the collision is unknown. Additionally, the spectator partons of the protons produce large QCD background rates in the detector. Taking advantage of the long lifetime and high momentum of the B meson, its displaced decay vertex with respect to the initial collision point serves as a good discriminator against these background contributions. The τ candidates are reconstructed in the $\tau^- \to \pi^-\pi^-\pi^+\nu$ channel, which primarily proceeds through the resonances a_1 and ρ. This helps in the reconstruction of the B meson mass, even with the undetected neutrino in the final state. The upper limit was determined from a fit of the reconstructed B meson mass.

Chapter 2

Experimental Methods

This chapter provides an overview of the relevant experimental methods in this analysis. It introduces the fundamental concepts necessary for extracting the number of signal events from measured data and for determining the upper limit of branching ratios. The second part presents multivariate algorithms, which are used to classify data into signal and background categories.

2.1 Fitting Tools

The signal extraction usually relies on prior knowledge of the expected distributions in data, which is typically blinded in the signal region to prevent biasing the selection and fit procedure. Therefore, Monte Carlo events are studied to determine the expected composition of various processes in the measured data and to identify a variable distribution with distinct characteristics for signal events.

The signal and background shapes of a distribution are described by probability density functions (PDFs), which depend on the parameters $\vec{\theta}$. The PDFs of all contributing processes are combined into a model PDF, which is then fitted to the distribution of the measured data. During the fitting process, statistical methods such as maximum likelihood are used to adjust the model parameter values $\vec{\theta}$ for optimal agreement between the model and data $\vec{x}$. For the upper limit determination of the $B^0 \to \tau\ell$ branching ratio, hypothesis tests are performed.

Further information on the statistical methods can be found in the book "Statistical Data Analysis" [35] and the lecture series "Practical Statistics for Particle Physicists" [36].

2.1.1 Maximum Likelihood Technique

The maximum likelihood technique is a common tool for determining the best values of model parameters $\vec{\theta}$, where $\vec{\theta} = (\theta_1, ..., \theta_M)^T$, based on a measured set of N independent values $\vec{x} = (x_1, ..., x_N)^T$ of the fit variable, also known as the measured data. The likelihood function $\mathcal{L}(\vec{\theta})$ expresses the probability of observing the data $\vec{x}$ for specific values of the parameters $\vec{\theta}$.

$$\mathcal{L}(\vec{\theta}) = \prod_{i=1}^{N} f(x_i; \vec{\theta}) \tag{2.1}$$

Here, $f(x_i; \vec{\theta})$ denotes the probability density function for a measured data point x_i given the parameter values $\vec{\theta}$. The values of $\vec{\theta}$, that maximize the likelihood function are called maximum likelihood estimators $\hat{\vec{\theta}}$ and are calculated by solving:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = 0, \qquad \forall i \in [1, .., M]. \tag{2.2}$$

Instead of maximizing the likelihood function, the log-likelihood function $\log\mathcal{L}(\vec{\theta})$ can be calculated. This approach can reduce the computational cost and is numerically more stable.

$$\log\mathcal{L}(\vec{\theta}) = \sum_{i=1}^{N} \log(f(x_i; \vec{\theta})) \tag{2.3}$$

Since the data has a finite size and can randomly fluctuate, the maximum likelihood estimators $\hat{\vec{\theta}}$ do not necessarily coincide with the true values of the parameters. Therefore, uncertainties on the maximum likelihood estimators are determined to obtain intervals that contain the true values of the parameters with a chosen probability, typically 68%. The uncertainty σ on the estimators depends on the covariance matrix $\mathbf{V}$:

$$\sigma^2_{\hat{\theta}_i} = \mathbf{V}_{ii}(\hat{\theta}). \tag{2.4}$$

For a large number of data points and unbiased likelihood estimators, the variance of the estimator is defined as:

$$\mathbf{V}_{ij}(\hat{\theta})^{-1} = - \left. \frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \right|_{\vec{\theta}=\hat{\vec{\theta}}}. \tag{2.5}$$

2.1.2 Extended Maximum Likelihood

In case the number of measured data points itself is Poissonian distributed around the true value ν, the probability of observing N data points can be included in the maximum likelihood function. This results in the extended maximum likelihood function:

$$\mathcal{L}(\nu, \vec{\theta}) = \frac{\nu^N}{N!} e^{-\nu} \prod_{i=1}^{N} f(x_i; \vec{\theta}). \tag{2.6}$$

If ν depends on the parameters $\vec{\theta}$, the log-likelihood function additionally constrains the likelihood estimators with the number of measurements N. The log-likelihood function, neglecting the constant terms, is defined as:

$$\log\mathcal{L}(\nu, \vec{\theta}) = N \cdot \log(\nu(\vec{\theta})) - \nu(\vec{\theta}) + \sum_{i=1}^{N} \log(f(x_i; \vec{\theta})) \tag{2.7}$$

If ν is independent of the parameters $\vec{\theta}$, the log-likelihood function simplifies to Eq. 2.3.

2.1.3 Binned Maximum Likelihood Fit

For large data sets, calculating the log-likelihood function becomes too computationally expensive. Therefore, the data is binned into a histogram with B bins, where each bin i contains n_i entries that are distributed around the true value ν_i. The expectation value ν_i is defined as:

$$\nu_i(\vec{\theta}) = N \int_{x_{i,\min}}^{x_{i,\max}} f(x; \vec{\theta}) dx \quad \xrightarrow{N \to \nu(\vec{\theta})} \quad \nu_i(\nu, \theta) = \nu \int_{x_{i,\min}}^{x_{i,\max}} f(x; \theta) dx. \tag{2.8}$$

Here, $x_{i,\min}$ and $x_{i,\max}$ denote the lower and upper bin boundary of bin i, respectively. The corresponding likelihood functions is defined as:

$$\mathcal{L}(\vec{\theta}) = N! \prod_{i=1}^{B} \frac{1}{n_i!} \cdot \left(\frac{\nu_i(\vec{\theta})}{N} \right)^{n_i} \quad \xrightarrow{N \to \nu(\vec{\theta})} \quad \mathcal{L}(\nu, \vec{\theta}) = \frac{\nu^N e^{-\nu}}{N!} N! \prod_{i=1}^{B} \frac{1}{n_i!} \cdot \left(\frac{\nu_i(\nu, \vec{\theta})}{\nu} \right)^{n_i} \tag{2.9}$$

$$= \prod_{i=1}^{B} \frac{e^{\nu_i}}{n_i!} \cdot \nu_i(\nu, \vec{\theta})^{n_i}.$$

$$\Rightarrow \log\mathcal{L}(\nu, \vec{\theta})) = -\nu + \sum_{i=1}^{B} n_i \ \log(\nu_i(\nu, \vec{\theta})) \tag{2.10}$$

2.1.4 Hypothesis Test

A hypothesis test quantifies the agreement between the measured data $\vec{x}$ and the prediction of a hypothesis. The hypothesis assumed to describe the data well is called the null hypothesis H_0, and its agreement with the data is tested against an alternative hypothesis H_1. Therefore, a test statistic $t(\vec{x})$ is constructed, which contains all relevant information of the measured data to distinguish between the two hypotheses. A PDF assigns a probability for observing a specific value of the test statistic under the respective hypothesis. Fig. 2.1 illustrates the PDFs for the two hypotheses of a test statistic t.

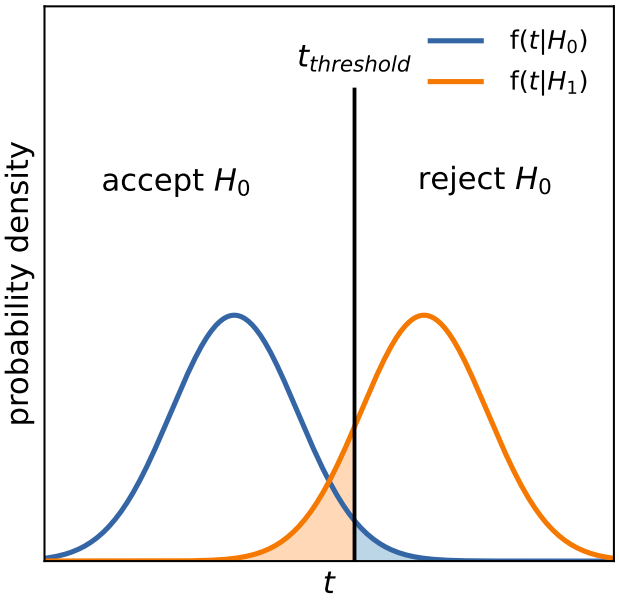

Figure 2.1: Probability density functions f of the null hypothesis H_0 and the alternative hypothesis H_1 for the test statistic t. The black line denotes the threshold, which determines whether the H_0 is rejected or accepted in favor of H_1. The blue area represents the significance α, and the orange area represents β. Figure modified from [35, p. 47].

If the observed test statistic value t_{obs} is below the threshold, $t_{threshold}$, the null hypothesis is accepted; for values larger than $t_{threshold}$, the null hypothesis is rejected in favor of the alternative hypothesis H_1. The probability α defines the significance of the test, representing the probability of rejecting the null hypothesis H_0 even though H_0 is true:

$$\alpha = \int_{t_{threshold}}^{\infty} f(t|H_0)dt. \tag{2.11}$$

Likewise, the probability β represents the likelihood of not rejecting H_0 in favor of H_1 even though H_1 is true:

$$\beta = \int_{-\infty}^{t_{threshold}} f(t|H_1)dt. \tag{2.12}$$

$1 - \beta$ determines the probability of correctly rejecting H_0 when H_1 is true and thus gives the power of a test to discriminate against H_1. The significance and power of the test are chosen before the test is performed.

The p-value expresses the probability of observing data equally or more incompatible with the null hypothesis, assuming the null hypothesis H_0 is true.

$$p = \int_{t_{obs}}^{\infty} f(t|H_0)dt \tag{2.13}$$

Fig. 2.2 shows the p-value for the observed test statistic value t_{obs} under the assumption of hypothesis H_0.

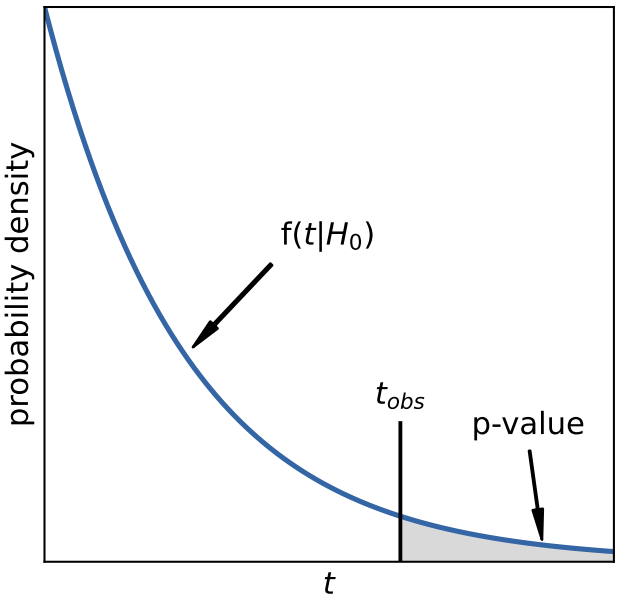

Figure 2.2: Probability density function of the null hypothesis H_0 and the corresponding p-value(grey area) for the observed value t_{obs} of the test statistic.

If the p $\leq \alpha$, the null hypothesis is rejected in favor of H_1; otherwise, there is no sufficient evidence against H_0 and H_0 is accepted.

2.1.5 Upper Limit Searches

In upper-limit searches on branching ratios, hypothesis tests are employed, where the null hypothesis $H_0 = b + s_i$ assumes that the data contains signal (s_i) and background (b) events, and the alternative hypothesis assumes that the data only contains background events $H_1 = b$. Multiple null hypotheses with varied numbers of signal events s_i are tested against the background-only hypothesis.

The null hypothesis with the highest number of signal events s_i that is not rejected in favor of the background-only hypothesis determines the upper limit on the signal events in data and, consequently, the upper limit on the branching ratio of the signal decay.

Typically, 90% or 95% ($1 - \alpha$) confidence levels (CL) are chosen for upper limits. They quantify how often the true value of the branching ratio would fall within the upper limit result if the same measurement were repeated multiple times. A commonly used test statistic in upper-limit searches relies on the ratio of profile likelihoods.

Profile Likelihood

The profile likelihood separates the parameters $\vec{\theta}$ of the PDF into a parameter of interest μ, which is the number of signal events in upper limit searches, and all other parameters, the so-called nuisance parameters $\vec{\lambda}$. The (unconditional) likelihood function $\mathcal{L}$ is defined as:

$$\mathcal{L}(\vec{\theta}) = \sum_{i=1}^{N} f\left(x_i; \vec{\theta}\right). \tag{2.14}$$

Instead of maximizing the likelihood simultaneously for all parameters, the conditional likelihood functions $\mathcal{L}_\mu(\vec{\lambda})$ is optimized for a fixed value of μ with respect to the nuisance parameters $\vec{\lambda}$:

$$\mathcal{L}_\mu(\vec{\lambda}) = \sum_{i=1}^{N} f(x_i; \mu, \vec{\lambda}). \tag{2.15}$$

The estimators of the nuisance parameters for a specific μ value are defined as:

$$\hat{\vec{\lambda}}_\mu = \arg\max_{\vec{\lambda}} \mathcal{L}_\mu(\vec{\lambda}) \tag{2.16}$$

This procedure is repeated n times for different parameter of interest values, resulting in n log-likelihood functions to find the value of μ, which agrees the most with the observed data. The estimator $\hat{\mu}$ is determined by finding the value of μ with the maximal $\mathcal{L}_\mu(\hat{\vec{\lambda}})$.

$$\hat{\mu} = \arg\max_{\mu} \mathcal{L}_\mu\left(\hat{\vec{\lambda}}_\mu\right) = \arg\max_{\mu} \mathcal{L}\left(\mu, \hat{\vec{\lambda}}\right) \tag{2.17}$$

The profile likelihood ratio used to determine the test statistic for upper limits is defined as:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\vec{\lambda}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\lambda}})}. \tag{2.18}$$

The likelihood ratio ranges between 0 and 1, where 1 represents a good agreement between the observed data with the hypothesized value of μ.

CL_s Method

The CL_s method is a modified version of the standard frequentist hypothesis test that prevents the discovery or exclusion of signals in low-sensitivity measurements. Rather than using the p-value of the null hypothesis $p_{s+b} \leq \alpha$ as a criterion to accept the $H_0(= s + b)$ hypothesis in favor of the background-only hypothesis $H_1(= b)$, a CL_s value is used.

$$CL_s = \frac{p_{s+b}}{1 - p_b} \tag{2.19}$$

with the p-values:

$$\begin{aligned} p_{s+b} &= P\left(t \geq t_{obs} \mid s + b\right) = \int_{t_{obs}}^{\infty} f(t \mid s + b)dt \\ 1 - p_b &= P\left(t \geq t_{obs} \mid b\right) = 1 - \int_{-\infty}^{t_{obs}} f(t \mid b)dt. \end{aligned} \tag{2.20}$$

The denominator in the CL_s formula varies between zero and one and thus leads to more conservative values than the p_{s+b} approach. The upper limit is determined by requiring $CL_s \leq \alpha$.

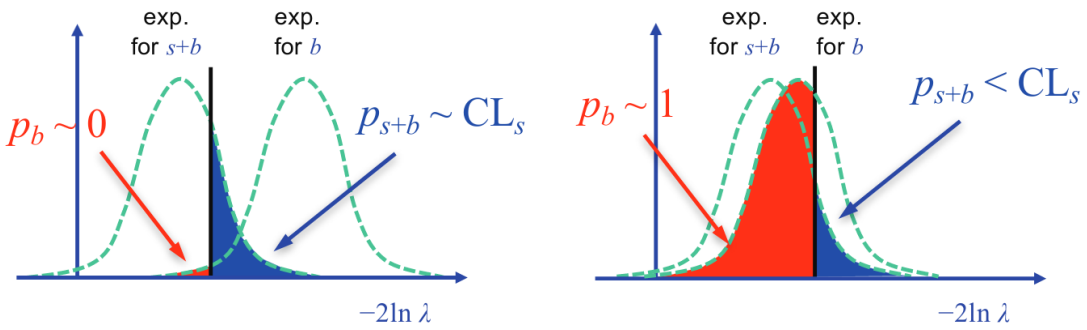

Figure 2.3: Comparison of the probability density functions for the signal + background (s + b) hypothesis and the background (b) only hypothesis for the test statistic t = − 2lnλ. The graphic is taken from [36, p. 248].

Fig. 2.3 shows the probability density functions for the two hypotheses. The overlapping distributions of the probability density functions in the right plot imply a low sensitivity of the measurement. In that case, the CL_s method won't satisfy the condition $CL_s \leq \alpha$ and therefore won't reject the H_0 hypothesis for a specific μ.

Test Statistic for Upper Limits

Numerous possible test statistics exist to calculate the upper limit for the parameter of interest μ, which corresponds to the number of signal events. The test statistic $\tilde{q}$ implemented in this analysis protects against negative signal events ($\mu < 0$) and is defined as:

$$\tilde{q}(\mu) = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{\mathcal{L}(\vec{x}|\mu,\hat{\hat{\vec{\lambda}}}_\mu(\mu))}{\mathcal{L}(\vec{x}|0,\hat{\hat{\vec{\lambda}}}_\mu(0))} & \hat{\mu} < 0, \\ -2\ln\frac{\mathcal{L}(\vec{x}|\mu,\hat{\hat{\vec{\lambda}}}_\mu(\mu))}{\mathcal{L}(\vec{x}|(\hat{\mu},\hat{\vec{\lambda}})} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{2.21}$$

The PDFs of the two hypotheses for the test statistic can be either generated using toy experiments with a high computational cost or determined using Wilk's theorem [37] and Wald's approximations [38].

The $\tilde{q}(\mu)$ with the asymptotic approximation is defined as:

$$\tilde{q}(\mu) = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0, \\ \frac{(\mu-\hat{\mu})^2}{\sigma^2} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{2.22}$$

where the estimator $\hat{\mu}$ is Gaussian distributed with a mean μ' and a standard deviation σ. Instead of generating multiple toy data sets to determine the expected upper limit and its uncertainty bands, Asimov data can be used. It serves as the perfect representative by replacing all parameters with their expected values. This approach saves computational resources while still providing a reliable estimation of the expected upper limit and its uncertainty due to statistical fluctuations in data.

2.2 Multivariate Algorithms

Multivariate classification models are implemented in the analysis to classify data into signal and background categories based on their properties, the so-called features. They are trained on data points with known classes, the training data set, and then applied to

data points that need to be classified. In the Belle II software, a speed-optimized multivariate classification algorithm called fastBDT [39] is integrated, which relies on stochastic gradient-boosted decision trees.

2.2.1 Decision Tree

A decision tree classifies data via multiple consecutive selection criteria starting at the root node, where the total data set is split into subsets. At each internal node, the subset is further divided based on a feature's value. The best selection value of a feature is calculated via the cumulative probability of the signal and background class, and the threshold with the highest separation power is chosen for each node. In the end, the various leaf nodes predict the class of data subset based on its features. Fig. 2.4 illustrates the structure of a decision tree.

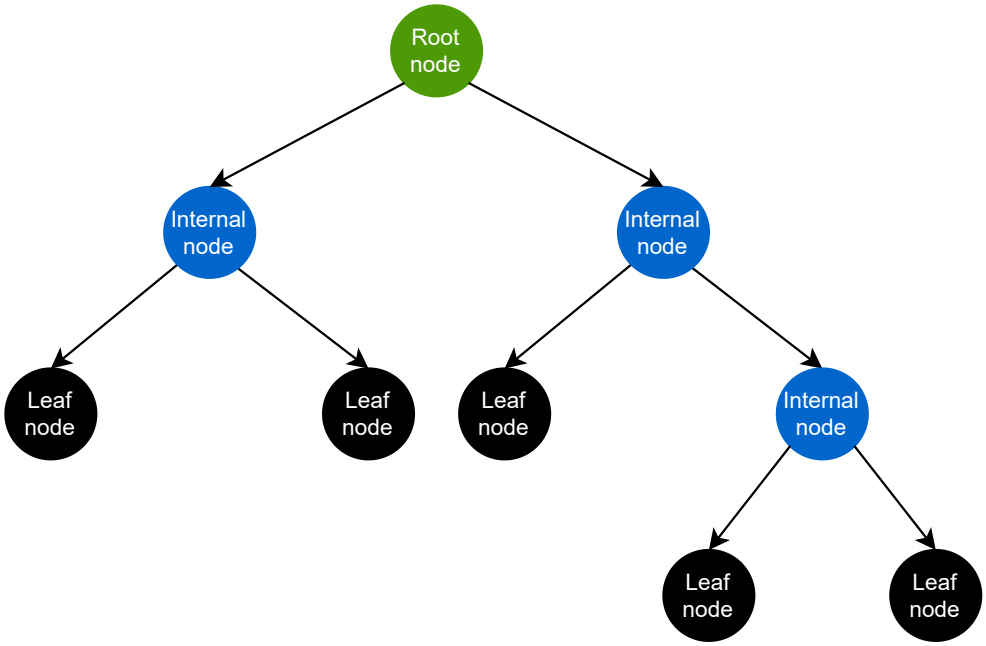

Figure 2.4: Structure of a decision tree.

The order of the consecutive nodes is determined based on the highest class separation. A decision tree is defined by its depth, which is the path with the highest number of nodes. Trees with many nodes can be highly dependent on statistical fluctuations in the training data set, which leads to overfitting. Therefore, pruning is applied, which removes nodes that do not provide important separation power for the final prediction.

2.2.2 Stochastic Gradient-boosted Decision Trees

Stochastic gradient-boosted decision trees are a sequential combination of pruned decision trees. Each decision tree provides a low separation power between the classes to prevent overfitting. However, by combining multiple decision trees, a high separation power can be reached.

Because the trees are sequentially connected, each tree can try to correct the mistakes in the previous tree's prediction during its training. This is accomplished with a gradient descent of a loss function that reweights the original data set and assigns a higher weight to the previously wrongly classified data points. Fig. 2.5 shows the working principle of the gradient-boosted decision trees.

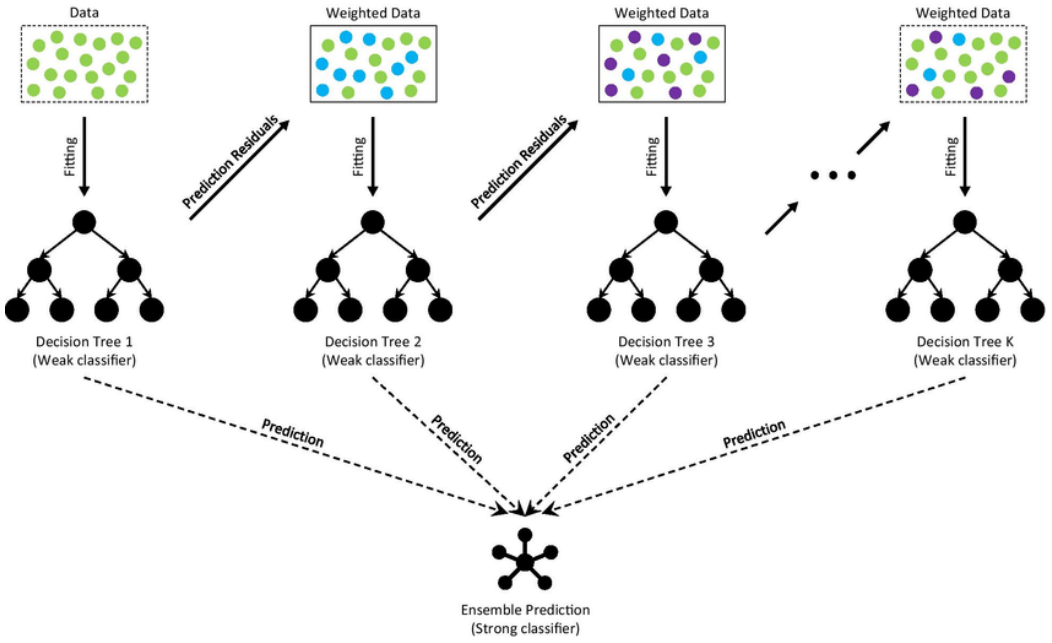

Figure 2.5: Working principle of boosted decision trees, taken from [40, p. 4]. The initial data, shown in green, is reweighted based on the correctness of the prediction by decision tree 1. This reweighting and fitting is repeated for all trees until a final prediction is made from all individual predictions.

The learning rate defines the influence of each tree on the final prediction. In the case of a small learning rate, more trees might be necessary to reach a certainty accuracy, but the advantage is that the final prediction is less prone to overfitting. Instead of using the full training data for each tree, a randomly drawn subset is used, which also helps against overfitting. The fraction of the subset compared to the total training data set is called the sub-sampling rate.

The importance of a feature is determined by the decrease in the overall performance if it is left out of the calculation of the final prediction. Compared to a single decision tree, stochastic gradient-boosted decision trees increase the robustness of the model against overfitting.

2.2.3 Full Event Interpretation

The Full Event Interpretation [41; 42], called FEI, is a multivariate classification algorithm that reconstructs B meson decays in approximately 10,000 exclusive decay modes. It is based on the idea of the previous reconstruction mechanism at Belle, the Full Reconstruction, which relies on neural nets. Starting from detector information like clusters, tracks, and displaced vertices and intermediate particles f.e. D mesons, and finally the B mesons are reconstructed. Each reconstructed B meson is assigned a signal probability, which predicts how likely the B meson decay was correctly reconstructed.

Typical cases for the application of the FEI are searches of B meson decays with neutrinos in the final state. Most of the time, the $\Upsilon(4S)$ resonance decays into two B mesons without any additional particles. When studying a specific B meson decay $\mathrm{B_{sig}}$, the FEI algorithm can be used to reconstruct the other B meson ($\mathrm{B_{tag}}$) in the event. The gained knowledge of the kinematics of the $\mathrm{B_{tag}}$ constrains the kinematics of the $\mathrm{B_{sig}}$, even if the final state of the $\mathrm{B_{sig}}$ contains neutrinos, because the initial energy of the e^+e^- collision is well-known.

The performance of the FEI training can be expressed with the following variables:

- **Tag-side efficiency**

$$\epsilon = \frac{\text{events with a correctly reconstructed } \mathrm{B_{tag}}}{\text{all events}} \tag{2.23}$$

- **Purity**

$$P = \frac{\text{events with a correctly reconstructed } \mathrm{B_{tag}}}{\text{events with a reconstructed } \mathrm{B_{tag}}} \tag{2.24}$$

A correctly reconstructed $\mathrm{B_{tag}}$ is either identified via Monte Carlo matching, which checks if the reconstructed decay chain matches the generated decay chain, or from a fit of the beam constraint mass $\mathrm{M_{bc}}$, which peaks at the nominal mass of the B meson for correctly reconstructed B mesons. The beam constraint mass $\mathrm{M_{bc}}$ is defined as:

$$\mathrm{M_{bc}} = \sqrt{\mathrm{E_{beam}^2} - \vec{p}_{\mathrm{B}}^{\,2}}, \tag{2.25}$$

where $\vec{p}_B$ represents the three-momentum of the $\mathrm{B_{tag}}$ in the center-of-mass frame (CMS), and the beam energy in the CMS frame replaces the reconstructed $\mathrm{B_{tag}}$ energy.

The FEI is trained to reconstruct two categories of decays:

- **Hadronic FEI** only includes B meson decays without neutrinos in the final state. This results in well-known kinematics of the B_{tag} with a high purity but comes with the cost of a tiny tag-side efficiency because the typical branching fractions are at the order of magnitude $\mathcal{O}(10^{-3})$.
- **Semileptonic FEI** has the advantage of a higher branching fraction, which leads to higher efficiency. The resulting data set has a lower purity, and the kinematics of the B_{tag} are not as well known due to undetected neutrinos.

Mechanism of the FEI

The FEI operates with a hierarchical approach consisting of six steps, which are shown in Fig. 2.6. In each step, classifiers predict the correctness of the reconstruction of each particle and only select the most likely candidates for the next step.

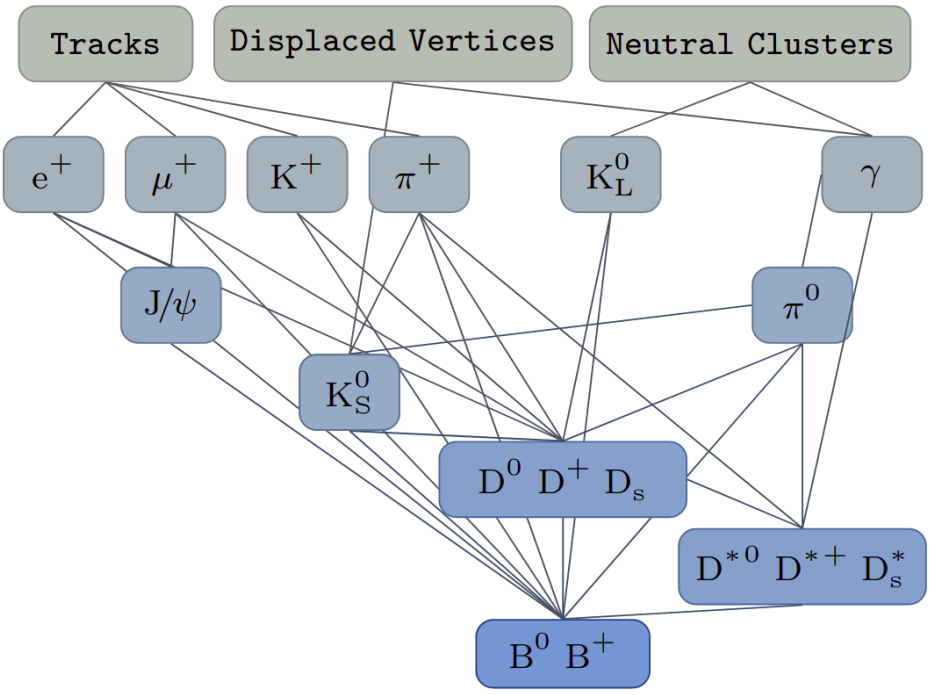

Figure 2.6: Schematic overview of FEI taken from [41, p. 3]. The grey boxes represent the detector information, and the blue boxes represent the reconstructed particles. The lines connect the detector information or the decay products to their corresponding particle.

In the first step, the algorithm trains classifiers to assign the correct particle type to detector information, i.e. tracks and clusters. Neutral particles are created from neutral clusters, oppositely charged tracks, and displaced vertices. Charged final state particles are matched to tracks assuming different particle hypotheses. To reduce the complexity due to the vast possible combinations, the charged final state particles must originate from the point of the e^+e^- collision.

In the next steps, the final state particles are combined to form intermediate particle f.e J/ψ, and a vertex fit is performed. The prediction of each step is the so-called signal probability, which combines information from all classifiers in a single value. In all further steps, the signal probability, kinematics, and displaced vertices are input values for finding the correct mother particle. In each reconstruction step, only ten to twenty candidates with the highest signal probabilities are selected. Intermediate particles are discarded if the vertex fit fails. A usual step includes reconstructing a particle, applying selections, fitting the vertex, determining the multivariate classifiers, and then applying selections again.

In the end, multiple B candidates with different signal probabilities are reconstructed. For example, a B^0 can be formed from D^- and π^+ or with an additional π^0. Depending on the kinematics of the daughter particles and the vertex fit, these two B mesons have a different signal probability. The signal probability does not only contain information from the last reconstruction step but takes the complete decay chain into account.

Chapter 3

Belle Experiment

This chapter outlines the main research topics of the Belle experiment and the setup of the Belle detector. A more thorough description of the Belle experiment and detector is given in the technical design report [43], the book "The Physics of the B Factories" [44] and the report "Physics achievements from the Belle experiment" [45].

The primary goals of the Belle experiment are precisely measuring the Standard Model parameters and studying τ leptons, hadrons containing c quarks, and B mesons with special emphasis on the CP violation in the neutral B meson system [46]. Therefore, the electron-positron ring collider KEKB, located in Tsukuba, Japan, operated mainly with a center-of-mass energy of 10.58 GeV, which is just above the threshold of the $\Upsilon(4S)$ resonance, a $b\bar{b}$ bound state. The $\Upsilon(4S)$ resonance decays exclusively in two B mesons with a probability of 96% at 95% confidence level [18], providing a clean environment.

When two neutral B mesons are created in a $\Upsilon(4S)$ decay, they are in an entangled quantum state necessary for the time-dependent CP measurement. The CP violation is determined by measuring the distance between the decay vertices of the B mesons along the positron beam direction. Because most of the available energy is consumed in creating the two B mesons, they are nearly at rest in the center-of-mass (CMS) system. By using asymmetric beam energies of 8.0 GeV (electrons) and 3.5 GeV (positrons), a Lorentz boost $\beta\gamma = 0.425$ is reached, allowing the measurement of the vertices distance with a resolution of approximately 100 μm.

3.1 Belle Detector

The Belle Detector measures the signatures of the stable decay particles originating from the initial electron-positron collision. Starting with these stable particles, the complete decay chain can be reconstructed. Therefore, it needs to satisfy multiple conditions.

The first requirement is a good 4π coverage of the solid angle around the collision point of the electron and the positron, the so-called interaction point (IP), to detect as many stable particles as possible. This is realized by a cylindrical detector with a 90% solid angle coverage.

Additionally, a high detection efficiency of these particles, even for low-momentum particles, is required, with excellent momentum and energy resolutions over a wide momentum range. For photons, a precise energy and position measurement is necessary to reconstruct neutral pions. Furthermore, a highly efficient particle identification for charged particles is crucial, with a special focus on separating pions and kaons at various momenta.

One of the main challenges is the presence of various beam-induced background processes. The most dominant one is Touscheck scattering, where two particles in the same bunch interact via Coulomb scattering. Other typical beam-induced background processes include beam-gas scattering, where beam particles interact with residual gas molecules in the beam pipe; synchrotron radiation, which is the emission of photons due to the acceleration of particles in the ring; as well as radiative Bhabha and two-photon processes. The background rates are reduced by colliding the electron and positron beam with a crossing angle of 22 mrad. Overall, the detector must be highly radiation-hard and capable of operating with high background levels.

The high data volume relies on an efficient and fast trigger system and data acquisition to identify the events of interest. The trigger system consists of multiple hardware triggers and a software trigger. The hardware triggers provide information from individual subdetectors and are often redundant to maintain high efficiency even under varying beam conditions. They are processed in parallel to ensure a fast information flow to the global decision logic, which determines whether to keep or discard an event.

3.1.1 Beam Pipe

The detector surrounds the beryllium beam pipe, and most of its components are encompassed in a homogeneous magnetic field of 1.5 T to determine the charge and momentum of charged particles. The beam pipe is constructed as a double-walled cylinder with cooling between the inner and outer radius, as shown in Fig. 3.1. The advantage of beryllium is its atomic number of 4 because it minimizes the multiple scattering of the particles and thus

improves the vertex resolution of composite particles, such as the B, D, and τ, decaying near the IP. The initially 20 mm inner radius was reduced to 15 mm [47] when the adjacent vertex detector was replaced.

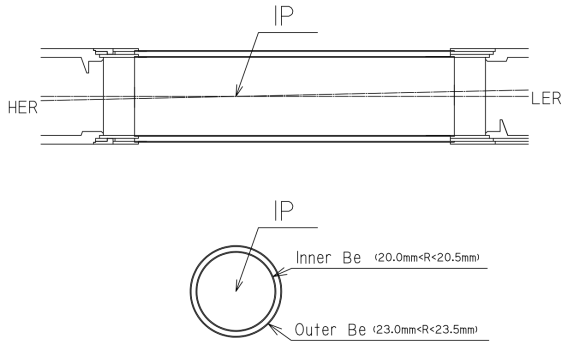

Figure 3.1: Double-walled beryllium beam pipe before the vertex detector was replaced with a newer one. IP denotes the interaction point, HER the high-energy electron ring and LER the low-energy positron ring. The figure is taken from [43, p. 126].

Fig. 3.2 illustrates the Belle Detector with its components described in the following sections, starting from the innermost subdetector.

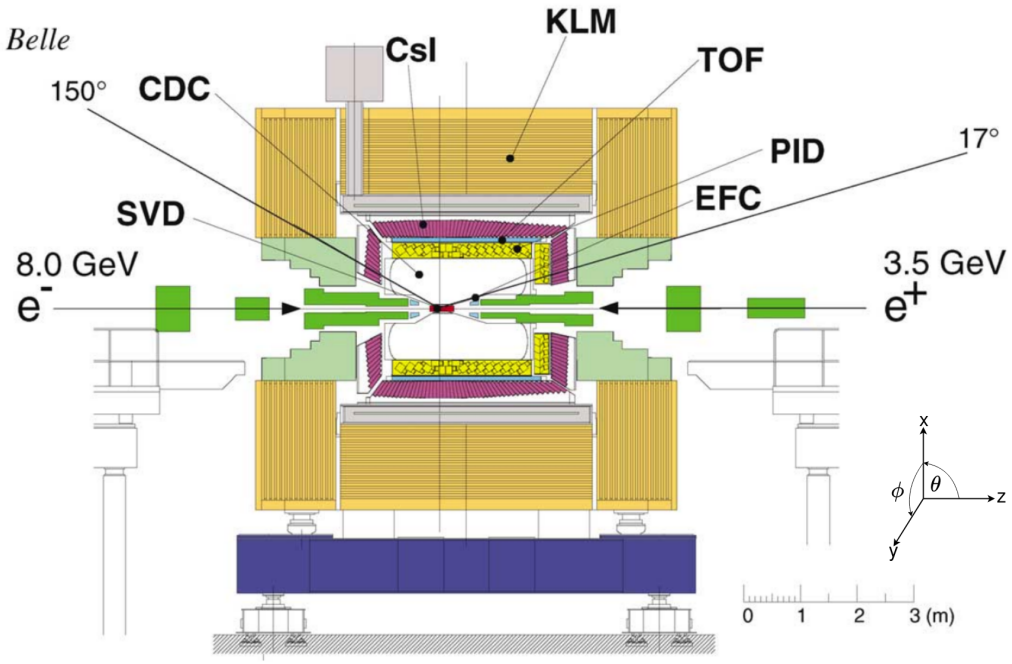

Figure 3.2: Side view of the Belle Detector, modified from [43, p. 124].

The setup of the Belle detector is chosen in a manner that:

- z-direction is opposite to the direction of the positron beam
- y-direction is in the horizontal plane, pointing toward the middle of the accelerator
- x-direction is oriented upwards for the right-handed coordinate system
- ϕ (azimuthal) direction lies in the x-y plane
- θ direction is the polar angle between the z-direction and ϕ plane
- radius r is defined as $r = \sqrt{x^2 + y^2}$
- homogeneous magnetic field points in the z-direction
- barrel (BR) region covers $30° < \theta < 130°$
- forward (FWD) region covers $17° < \theta < 30°$
- backward (BWD) region covers $130° < \theta < 150°$

3.1.2 Silicon Vertex Detector

The innermost component of the detector, a Silicon Vertex Detector (SVD), is positioned around the beam pipe. It detects hits from charged particles and serves as the only source of information for low-momenta particles that do not reach the main tracking component. Its primary goal is to provide precise information for the vertex reconstruction of the B, D mesons and τ leptons, whose decay vertices inside the beam pipe are determined by extrapolating tracks from the tracking detectors.

The first SVD, known as the SVD1, consisted of three layers of double-sided silicon strip detectors. After four years of operation, it was replaced by a newer, better-performing version, the SVD2. The SVD2 contained four layers of double-sided silicon strip detectors [48] with an increased coverage from $23° < \theta < 140°$ to $17° < \theta < 150°$ and exhibited a higher radiation hardness.

A sensor of the SVD has long p-side strips, oriented parallel to the z-direction, which are embedded into an n-doped bulk with n-side strips rotated in r-ϕ direction. Multiple sensors, depending on the layer, are combined into a ladder. The ladders are arranged cylindrically around the IP in layers with 8, 10, and 14 ladders for SVD1 and 6, 12, 18, and 18 ladders for SVD2. The neighboring ladders overlap in the ϕ plane to ensure total coverage and easier calibration of the alignment.

Fig. 3.3 presents the SVD1 setup in the transverse plane and the working principle of a silicon strip module. When a charged particle traverses the depleted n-bulk of an SVD sensor, it creates electron-hole pairs. The electrons drift toward the n-side, while the holes drift to the p-side due to a voltage difference of 75 V between the readout nodes.

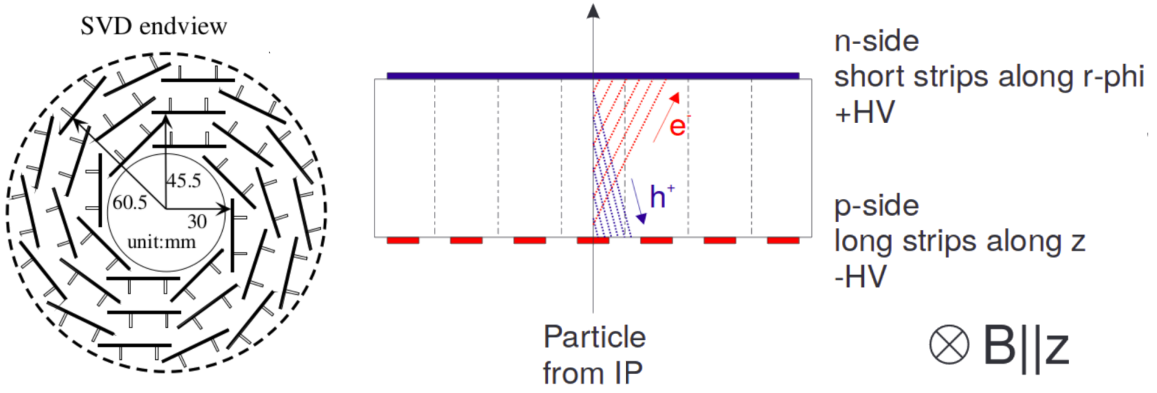

Figure 3.3: Setup of SVD1 shown in the transverse plane on the left, taken from [43, p. 134]. The working principle of a silicon strip detector is illustrated on the right and taken from [49, p. 144].

3.1.3 Central Drift Chamber

The main purposes of the Central Drift Chamber (CDC) are to reconstruct charged tracks; identify charged particles based on their energy loss in a gas volume (Bethe-Bloch)—especially for low-momentum particles that do not reach the other particle identification detectors; and provide a trigger signal. The central challenge is mitigating the multiple scattering processes to ensure a good momentum resolution while still obtaining a high energy loss resolution.

The CDC is a He-C_2H_6 gaseous ionization detector containing 8400 drift cells. If a charged particle traverses the volume, it ionizes the gas in the drift cell, creating electron-ion pairs. The electrons drift towards the anode (sense wires) in an electric field applied by the field wires. In the amplification region close to the sense wires, the electrons are further accelerated, resulting in additional ionization of the gas and an amplified current. The strength of the current depends on the number of primary generated electrons, which are related to the energy and momentum of the traversing charged particle.

Three-dimensional tracking is achieved with axial wires aligned along the z-direction and stereo wires skewed by a small angle with respect to the axial wires. To resolve the left-right ambiguities, the stereo wires are oriented with positive and negative angles with respect

to the axial wire. Due to the higher occupancy in the drift cells closer to the beam pipe, they are smaller than the average size of 15.5 - 17.0 mm. Fig. 3.4 shows the drift cells and the general setup of the CDC before the installation of the SVD2.

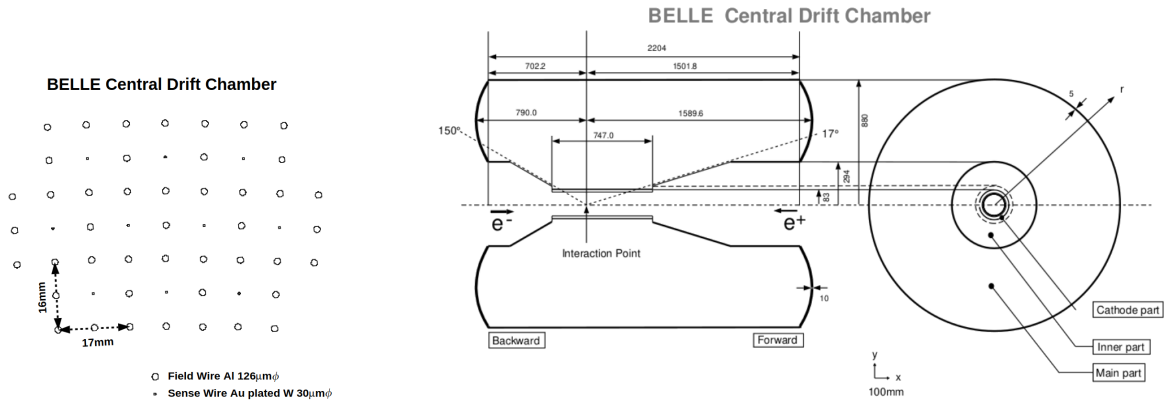

Figure 3.4: Central drift chamber setup for the Belle Detector with the installed SVD1. The figures are taken from [43, p. 144 - 145].

Before the installation of the SVD2, the CDC contained cathode strips as the innermost part measuring the z-direction of traversing charged particles. These cathode strips were replaced by smaller drift cells to provide enough space for the new SVD2. Due to the Lorentz boost and the resulting higher number of particles in the forward direction, the CDC, like the SVD, is asymmetric in the z-direction.

3.1.4 Particle Identification System

Two subdetectors are solely used to provide particle identification of charged particles via the detection of Cherenkov photons in a dielectric medium. Their primary focus is a good pion and kaon separation for momenta between 0.5 GeV and 3.5 GeV.

The inner particle identification detector is an Aerogel Cherenkov Counter (ACC, denoted as PID Fig. 3.2) and is used to distinguish pions and kaons with momenta between 1.2 and 3.5 GeV. The refraction index of the dielectric medium is chosen so that a pion traversing the volume emits Cherenkov photons, while a kaon has a velocity below the speed of light in the medium and does not emit Cherenkov photons. It only returns a yes or no answer for the detection of Cherenkov photons and covers the region of $17° < \theta < 127°$. Fig. 3.5 shows an ACC module installed in the barrel and forward region.

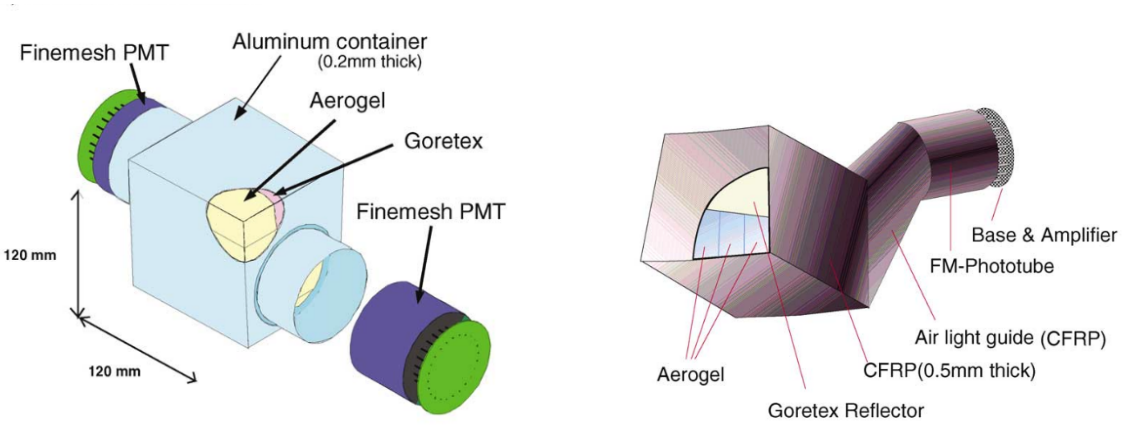

Figure 3.5: Aerogel Cherenkov Counter installed in the barrel (left) and the forward endcap region (right). These figures are taken from [43, p. 157]

The second system is the Time of Flight (TOF) detector with plastic scintillators as the dielectric medium. At a radial distance of 1.2 m from the IP and with a time resolution of 100 ps, it identifies particles with momenta up to 1.2 GeV and thus complements the ACC pion kaon separation. Its acceptance region covers the polar angles $33° < \theta < 121°$.

A TOF module consists of two trapezoidal-shaped counters and one thin trigger scintillation counter and is read out by photo multiplier tubes (PMT) on both ends of the module. Combined with the beam collision time, the time of flight of charged particles is determined, and with information on the other subdetectors, the mass can be calculated. In addition to the particle identification, the TOF provides an important input for the trigger decision. The working principle of the TOF system is illustrated in Fig. 3.6.

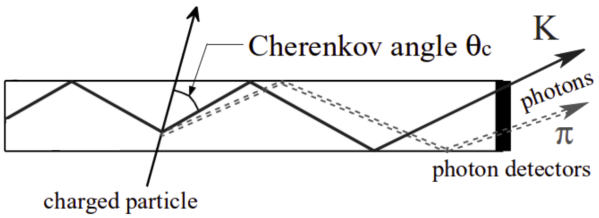

Figure 3.6: Cherenkov radiation for kaons and pions in a dielectric medium, taken from [49, p. 122].

A charged particle passing through the dielectric medium with a velocity greater than the speed of light in this medium emits Cherenkov radiation. The emission angle, Cherenkov angle, depends on the velocity v of the charged particle and the refraction index n of the medium: $\cos(\theta_c) = 1/(n \cdot v)$. The photons propagate under total reflection to the readout

system while preserving the information regarding the Cherenkov angle. If a kaon and a pion have the same momentum, the velocity of the pion is higher due to its lower mass. This results in a larger Cherenkov angle. Therefore, the photons of the pion travel a shorter distance and arrive faster at the photon detector than the photons of the kaon if they traverse the dielectric medium at the distance to the readout system.

3.1.5 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECL, denoted as CsI Fig. 3.2) is primarily designed to measure the energy of electromagnetic showers created by electrons and photons and to provide crucial information for the electron identification. Because a large portion of the created particles are neutral pions that subsequently decay into two photons, the ECL must efficiently measure electromagnetic showers of photons with energies between a few tens MeV to 4 GeV as well as their angular distribution. Therefore, it is installed as a highly segmented system consisting of scintillating thallium-doped cesium iodide (CsI(Tl)) crystals that have a high light output and a short radiation length. The ECL also provides trigger information and is used to determine the luminosity.

The barrel section spans 3 m in z-direction with an inner radius of 1250 mm. Including the circular endcaps, an angular coverage between 17° to 150° is reached with 8736 crystals. These crystals have an average length of 300 mm (equivalent to 16.2 radiation lengths) to minimize energy leakage. Additionally, they are slightly tilted with respect to the IP to prevent particles from hitting the gap between crystals and to provide angular resolution. Fig. 3.7 shows the layout of the ECL.

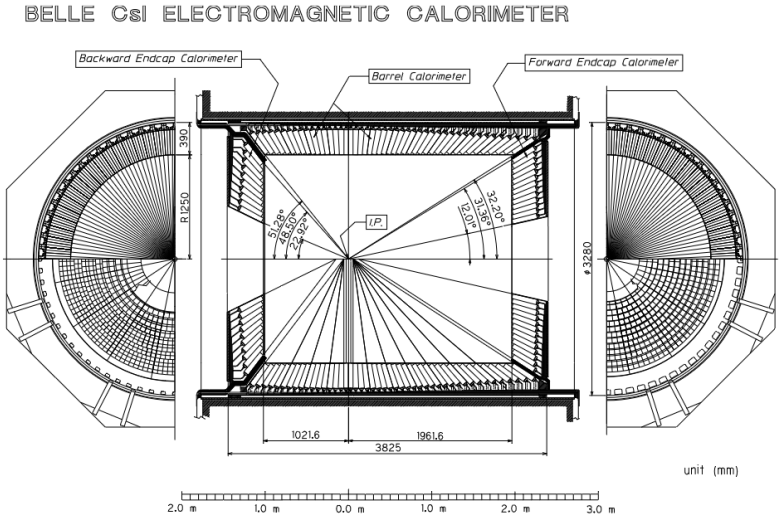

Figure 3.7: Configuration of the electromagnetic calorimeter taken from [43, p. 175].

Photons and electrons with sufficient energy passing through the crystal interact with the detector material and produce electromagnetic showers. High-energetic photons can convert into electron-positron pairs, which subsequently can emit Bremsstrahlung photons or excite the scintillation detector material. These processes continue repeatedly until all the energy is absorbed. The scintillating material de-excites by emitting low-energetic photons, which photodiodes detect at the end of each crystal. The number of detected photons is proportional to the energy of the initial photon or electron.

To determine whether a shower belongs to an initial photon or electron, the presence of a track leading to the shower is checked. If a track is detected, the shower is classified as an electron shower; otherwise, it is a photon shower. Hadrons can also shower in the ECL. However, the additional strong interaction with the nuclei results in distinct shower shapes compared to purely electromagnetic showers.

The extreme forward calorimeter (EFC) operates on the same principle as the ECL, using bismuth germanate oxide crystals as scintillators. It is positioned in the forward (6.4° to 11.5°) and the backward (163.3° to 171.2°) regions of the detector. The EFC serves two purposes: it acts as a shield for the CDC against the beam-induced background and provides information on the luminosity during data taking (online luminosity). The luminosity is determined using Bhabha events due to their high cross-section and clear signature.

3.1.6 K_L and μ Detector

The outermost subdetector is the K_L and μ detector (KLM). As the name suggests, its primary aim is identifying K_L and muons with high efficiency. In addition, it serves as magnetic reflux of the magnetic field and provides information to the trigger system. Its working principle relies on double gap resistive plate counters (RPCs) as active material that alternate with 4.7 cm thick iron plates.

A RPC consists of two parallel glass plates surrounding a thin gas volume. A high-voltage field is applied between the two glass plates. When a charged particle passes through, it ionizes the gas. The ions are accelerated toward the cathode, and the electrons toward the anode. Due to the high electric field, the electrons further ionize the gas, creating an avalanche of electrons. That leads to a temporarily reduced electric field in the nearest electrode, measured by the external pick-up strips on the plate. The pick-up strips on the ground plate are arranged orthogonally around a pair of RPCs. The layout of the RPC system is illustrated in Fig. 3.8.

K_L usually deposit their entire energy in hadronic showers, mainly in the iron plates due to its high atomic number. Although the KLM does not provide energy information, the shower position can be used to determine the original direction of the K_L by extrapolating

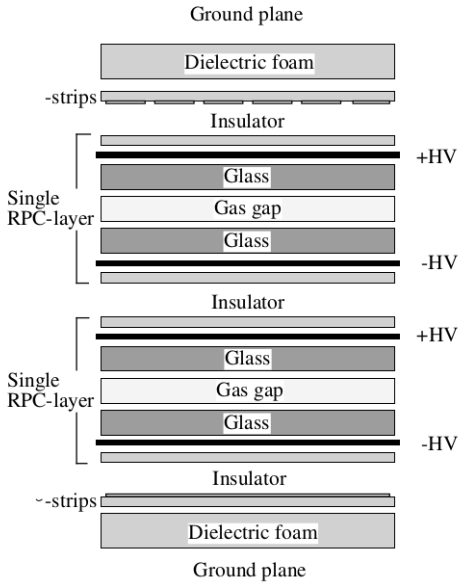

Figure 3.8: Schematic sketch of the RPC system taken from [43, p. 195].

to the IP. Because muons are minimal ionizing particles, they pass through the detector with small deflections. The barrel region of the KLM consists of fifteen active material layers and fourteen iron plate layers (adding up to a total of 3.9 interaction lengths) and the endcaps of fourteen active layers and fourteen iron plate layers with total polar coverage from 20° to 155°.

3.2 Event Types at Belle

The most dominant interactions at Belle are QED processes, whose cross sections overshadow the approximately 1/50 factor smaller cross section of $B\overline{B}$ events. To mitigate most of the QED contributions while retaining 99.1% of all $B\overline{B}$ events, a selection based on the track multiplicity, energy in the ECL, and the total visible energy in the events is applied. For a detailed description of the individual criteria, refer to the internal Belle Note "HadronB"[50].

Chapter 4

Data Sets

This chapter presents the Belle data and Monte Carlo data sets studied in this analysis and the implemented Monte Carlo corrections to improve the data - Monte Carlo agreement.

The Belle experiment collected data between 1999 and 2010 at various center-of-mass (CMS) energies [45]. Most data were recorded at the $\Upsilon(4S)$ resonance, slightly above the production threshold of two B mesons. An overview of the data sets at the various resonances is given in Tab. 4.1. Two silicon vertex detectors were used for the data set taken on the $\Upsilon(4S)$ resonance, where the SVD1 was replaced by the SVD2. For each on-resonance data set, a small off-resonance data set was collected to study the continuum events.

Resonance	On-resonance luminosity [fb^{-1}]	Off-resonance luminosity [fb^{-1}]
$\Upsilon(1S)$	5.7	1.8
$\Upsilon(2S)$	24.9	1.7
$\Upsilon(3S)$	2.9	0.25
$\Upsilon(4S)$ SVD1	140.0	15.6
$\Upsilon(4S)$ SVD2	571.0	73.8
$\Upsilon(5S)$	121.4	1.7
Scan	–	27.6

Table 4.1: Summary of the luminosity integrated by Belle, broken down by CM energy. The Tab. is taken from [45, p. 7].

This analysis studies the Belle data, referred to as real data, recorded at the center of mass energy of the $\Upsilon(4S)$ resonance at 10.58 GeV with a luminosity of 711 fb^{-1}. Due to varying the beam parameters and background conditions over the data-taking periods, the real data is divided into experiments with experiment-specific conditions. To ensure an unbiased final result, the analysis is conducted and optimized using Monte Carlo events, which require that they precisely represent the real data.

4.1 Monte Carlo Data

The Monte Carlo events contain information on the underlying generated events, which allows the identification of signal and background events after the detector simulation. The generation is composed of two stages.

In the first stage of the Monte Carlo generation, the complete decay chain of Monte Carlo particles created in a e^+e^- collision is generated according to properties defined in a decay file. Since no single generator accurately models the complete decay chain, various generators are employed. EVTGEN [51] is used for B meson decays; PYTHIA [52] simulates the fragmentation and hadronization of quarks; and PHOTOS [53] is used to include the QED final state radiations. In the second stage, the interactions of the Monte Carlo particles with the detector and its responses are implemented with GEANT3 [54; 55].

Monte Carlo matching connects a detected particle to its generated Monte Carlo particle. In the case of final-state particles, the detected particle is a "true" particle if the matched Monte Carlo particle is of the same type. A decaying particle is correctly reconstructed if all its decay products match their respective Monte Carlo particles. The Monte Carlo data sets are generated for each experiment of the real data under the experiment-specific condition and the corresponding luminosity.

4.2 Signal Monte Carlo

Because the signal decay is lepton number violating, it is not part of the officially produced Belle Monte Carlo data, which only contains Standard Model decays. Therefore, 10 million signal events are generated, where one B^0 meson decays in the signal channel $B^0 \to \tau^+\ell^-$ with $\ell = e, \mu$, with the phase-space model PHSP of EVTGEN. The subsequent τ decays are modeled based on the conditions defined in the official Belle decay file and rely on either the τ-specific models of EVTGEN or the general phase space models of PHSP and PYTHIA. The second B meson in the events is generated in an electroweak $b \to c$ transition, following the official Belle decay file. Mixing between the two neutral B mesons and QED radiative corrections are included.

Tab. 4.2 lists the reconstructed τ decays of interest and their corresponding branching ratios used in the generation according to the official Belle decay file and the updated values from Particle Data Group (PDG) [18].

τ Decay mode	Belle old $\mathcal{B}$ ($\times 10^{-2}$)	PDG new $\mathcal{B}$ ($\times 10^{-2}$)
$e^- \bar{\nu}_e \nu_\tau$	17.79	17.82 ± 0.04
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.32	17.39 ± 0.04
$\pi^- \nu_\tau$	11.00	10.82 ± 0.05
$\pi^- \pi^0 \nu_\tau$	25.30	25.49 ± 0.09
$\pi^- \pi^0 \pi^0 \nu_\tau$	8.70	9.31 ± 0.05
$\pi^- \pi^- \pi^+ \nu_\tau$	7.53	9.26 ± 0.10
total	87.64	90.09 ± 0.16

Table 4.2: Branching ratios of the reconstructed τ decays according to the Belle decay file and the PDG [18].

To correct the branching ratios in signal events, each event is assigned a weight with respect to the generated τ decay branching ratio and the updated branching ratio. For the specific reconstructed τ decay, an event weight w_{rec} is applied:

$$w_{\text{rec}} = \frac{\mathcal{B}_{\tau,\text{PDG}}}{\mathcal{B}_{\tau,\text{Belle}}}. \tag{4.1}$$

For all other types of generated τ decays (nrec), a correction weight w_{nrec} is determined relative to the sum of all signal τ decays i.

$$w_{\text{nrec}} = \frac{1 - \sum_i \mathcal{B}_{\tau_i,\text{PDG}}}{1 - \sum_i \mathcal{B}_{\tau_i,\text{Belle}}} \tag{4.2}$$

4.3 Background Monte Carlo

A good description of the background events is essential to determine the expected background rate and composition in the real data. Therefore, various Monte Carlo data sets are investigated. The most important ones are events with Standard Model B meson decays and continuum events, where the initial $e^+e^- \to q\bar{q}$ with $q = u, d, s, c$. Since Belle was foremost intended for studying B physics, the applied initial selection of the Hadron B skim, see section 3.2, rejects most of the low-multiplicity events. Their contribution after the reconstruction of the signal channel was found to be negligible.

Generic Belle Monte Carlo

Events of the most probable e^+e^- interactions were generated in the official Belle Monte Carlo campaign. These data sets were produced with multiple times the luminosity of the real data set, where one stream corresponds to the online luminosity of the full real data set. The large number of Monte Carlo events decreases the statistical uncertainty of the final result. One of the available streams (stream 6), which was only generated for B meson events, was used in the training of the FEI and was therefore not included in the analysis to avoid biasing the final result.

Continuum

Continuum events are light-quark events, where the e^+e^- directly decays into a light-quark pair $e^+e^- \to q\bar{q}$ with $q = u, d, s, c$. Their contribution to the total cross-section is approximately three times larger than $\Upsilon(4S)$ events. Six streams of continuum events were generated, which are separated into uds ($ee \to u\bar{u}/d\bar{d}/s\bar{s}$) and charm ($ee \to c\bar{c}$) data sets.

B Meson with $b \to c$ Transition

If a $\Upsilon(4S)$ particle is created, it decays with a probability of over 96% into two B mesons. These B mesons most likely decay through electroweak $b \to c$ transitions because $b \to u$ transitions are CKM suppressed, and $b \to d, s$ transitions do not occur at tree level and are also CKM suppressed.

In the official Belle Monte Carlo campaign, the $\Upsilon(4S)$ decays were generated under the assumption of isospin symmetry, and both of the B mesons decay via $b \to c$ transitions. These events are referred to as B($b \to c$) events in the further analysis. The total data set of B($b \to c$) events consists of ten streams. Tab. 4.3 shows the generated branching ratios for the B meson Belle data set and the PDG result assuming $\mathcal{B}(\Upsilon(4S) \to B\bar{B}) = 100\%$.

Decay mode	Belle $\mathcal{B}$ ($\times 10^{-2}$)	PDG $\mathcal{B}$ ($\times 10^{-2}$)
$\Upsilon(4S) \to B^+ B^-$	49.948	51.4 ± 0.6
$\Upsilon(4S) \to B^0 \bar{B}^0$	49.947	48.6 ± 0.6
$\Upsilon(4S) \to$ other	0.105	–

Table 4.3: Branching ratios of $\Upsilon(4S)$ relative to the total $\Upsilon(4S)$ branching ratio as defined in the official Belle decay file and the updated values from the PDG, assuming a 100% decay into B mesons.

Additional Belle Monte Carlo

The generic Belle Monte Carlo data set excludes B meson decays via the $b \to d, s$ and $b \to u\ell\nu$ transition. Because they are crucial background processes in searches of rare or forbidden Standard Model B meson decays, they were generated in a separate Monte Carlo campaign, known as the SpecialMC2010 campaign. In these events, one B meson decays via the $b \to d, s$ or $b \to u\ell\nu$ transition, while the other B meson decays through the more probable $b \to c$ transition.

B Meson with $b \to d, s$ Transition

This data set contains events where one of the two B mesons decays via $b \to d, s$ transition. Due to their tiny branching ratios, see Tab. 4.4, data sets with fifty times the expected luminosity of the real data were generated. Events of this type are referred to as B $\to$ rare events in the further analysis.

$\mathcal{B}(\mathrm{B}^+ \to \text{rare})$	$\mathcal{B}(\mathrm{B}^0 \to \text{rare})$
$3.27945 \cdot 10^{-3}$	$4.05703 \cdot 10^{-3}$

Table 4.4: Branching ratios of rare B meson decays used in the SpecialMC2010 campaign.

B Meson with $b \to u\ell\nu$ Transition

The SpecialMC2010 data set also contains events where one B meson decays via a $b \to u\ell\nu$ transition. While the data accurately models the kinematic distribution of the lepton momentum over a wide range, it exhibits deviations at the endpoint of the leptonic momentum region. Since this region is significant for distinguishing signal and background events, a new data set with updated branching ratios and theory predictions is used instead.
The branching ratios for the inclusive (resonant and non-resonant) decays of B mesons are defined as:

$$\mathcal{B}\left(\mathrm{B}^+ \to X_\mathrm{u}^0 \ell\nu\right) = \tau_{+0} \frac{1 + f_{+0}}{1 + \tau_{+0} f_{+0}} \mathcal{B}\left(\mathrm{B} \to X_\mathrm{u} \ell\nu\right)_{\Upsilon(4\mathrm{S})} = (2.2 \pm 0.3) \cdot 10^{-3}, \tag{4.3}$$

$$\mathcal{B}\left(\mathrm{B}^0 \to X_\mathrm{u}^+ \ell\nu\right) = \frac{1 + f_{+0}}{1 + \tau_{+0} f_{+0}} \mathcal{B}\left(\mathrm{B} \to X_\mathrm{u} \ell\nu\right)_{\Upsilon(4\mathrm{S})} = (2.0 \pm 0.3) \cdot 10^{-3}, \tag{4.4}$$

where $\mathcal{B}\left(\mathrm{B} \to X_\mathrm{u} \ell\nu\right)_{\Upsilon(4\mathrm{S})}$ denotes the combined branching ratio of neutral and charged B mesons.

$$\mathcal{B}\left(\mathrm{B} \to X_\mathrm{u} \ell\nu\right)_{\Upsilon(4\mathrm{S})} = (2.13 \pm 0.31) \cdot 10^{-3} \tag{4.5}$$

$\tau_{+0} = \frac{\tau_{\mathrm{B}^+}}{\tau_{\mathrm{B}^0}} = 1.076 \pm 0.004$ is the lifetime ratio and $f_{+0} = \frac{\Gamma(\Upsilon(4\mathrm{S}) \to \mathrm{B}^+\mathrm{B}^-)}{\Gamma(\Upsilon(4\mathrm{S}) \to \mathrm{B}^0\bar{\mathrm{B}}^0)} = 1.058 \pm 0.007$ the fraction of the $\Upsilon(4\mathrm{S})$ decay rates into charged and neutral B meson pairs.

Theory models correctly predict the triple differential decay rate, which depends on the invariant mass of the hadronic final states m_X, the energy of the lepton in the B rest frame E_ℓ^B and the momentum transfer q^2 from the B meson to the leptonic system:

$$\frac{d\Gamma}{dm_X\, dE_\ell{}^B\, dq^2}. \tag{4.6}$$

However, these theory models can only model non-resonant decays for hadronic final states with $m_X \geq 2\ m_\pi$ hadronic final states. The only available generator in the Belle software for such non-resonant decays is EvtVUB, which relies on the De Fazio and Neubert (DFN) model [56]. The non-resonant Monte Carlo was generated with the inclusive branching ratio and later reweighted when adding the resonant decays. The missing resonant decays were generated with updated form factors and branching ratios provided by Heavy Flavor Averaging Group (HFLAV) [57] and PDG 2019 [58].

Tab. 4.5 compares the branching ratios of the $B(b \to u\ell\nu)$ decays in the old (specialMC2010) and the new Monte Carlo data. Additional resonances were included in the generation of the old Monte Carlo data. The branching ratios of these higher resonances are combined with the non-resonant X_u in the "X_u + higher resonances: component. The listed branching ratios are identical for electrons or muons.

B meson	Decay mode	old $\mathcal{B}$ ($\times 10^{-3}$)	new $\mathcal{B}$ ($\times 10^{-3}$)	(new-old)/new
Neutral	π^+	0.136	0.150	90.7%
	ρ^+	0.277	0.294	94.2%
	X_u^+ + higher resonances	1.296	1.606	80.7%
	total	1.709	2.050	83.4%
Charged	π^0	0.073	0.078	93.6%
	ρ^0	0.149	0.158	94.3%
	η	0.084	0.039	215.4%
	η'	0.033	0.023	143.5%
	ω	0.115	0.119	96.6%
	X_u^0 + higher resonances	1.381	1.789	77.2%
	total	1.835	2.206	83.2%

Table 4.5: Branching ratios for the old (SpecialMC2010) and new Belle Monte Carlo data sets. The new branching ratios are obtained from the 2018 results of HFLAV and the 2019 results of the PDG.

Hybrid Model

In the Hybrid Model, events with non-resonant and resonant decays are combined according to the triple differential decay rate. The resonant and non-resonant Monte Carlo events are binned in the following variables:

- m_X: [0.0, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0, 3.5]
- E_ℓ: [0.0, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0, 3.5]
- q^2: [0.0, 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, 20.0, 25.0]

For each bin i of the triple differential decay rate, the number of inclusive Monte Carlo events I_i is defined as the sum of the resonant R_i and non-resonant N_i contributions.

$$I_i = R_i + N_i \tag{4.7}$$

Because the non-resonant contribution was generated according to the total inclusive decay rate, it needs to be reweighted by a weight ω_i:

$$N_i = \omega_i \cdot I_i \tag{4.8}$$

to conserve the total inclusive branching ratio when adding the resonant decays. The Hybrid Model data set contains events ten times the expected luminosity. Events where one B meson decays in a $b \to u\ell\nu$ transition are referred to as B($b \to u\ell\nu$) events.

4.4 Monte Carlo Data Set Overview

An overview of the studied signal and background Monte Carlo data is given in Tab. 4.6.

Monte Carlo Type	Number of events	Number of streams
charm	900,306,306	6
uds	1,447,406,660	6
$B^0(b \to c)$	377,426,369	9 (10)
$B^+(b \to c)$	377,426,374	9 (10)
$B^0(b \to u\ell\nu)$	1,543,162	10
$B^+(b \to u\ell\nu)$	1,702,108	10
$B^0 \to$ rare	3,130,327	50
$B^+ \to$ rare	2,530,361	50
signal	10,000,000	–

Table 4.6: Number of generated events of the Monte Carlo background events expected in data with luminosity of 711 fb^{-1} and the number of available streams. For $B(b \to c)$ Monte Carlo events, the number of streams in brackets represents the total number of available streams, including the stream used for the FEI training. The signal events are listed with the total number of generated signal events.

4.5 Monte Carlo Corrections

Multiple corrections are applied to the Monte Carlo events to improve the data - Monte Carlo agreement and assign uncertainties introduced by these corrections. While the simulation of the detector attempts to mimic the detector responses as precisely as possible, deviations are unavoidable. These deviations result in differences in particle identification, particle finding efficiencies, and the performance of multivariate algorithms, which were trained on Monte Carlo data.

Additionally, measurements of branching ratios and form factors have been updated since the Monte Carlo generation. However, their uncertainties were neglected during the Monte Carlo generation. They need to be included to ensure the accurate determination of the uncertainty of the final result.

For each Monte Carlo event i, the so-called event weight w_i is determined as the product of the individual correction factors.

4.5.1 Luminosity

The generic Monte Carlo data were generated based on the online luminosity, which is less precise than the later determined offline luminosity. Additionally, some of the Monte Carlo files were lost. For a good representation of the real data, the generic Monte Carlo data

are weighted according to the number of N_{BB} in real data to match the expected rate of the individual background processes. The scaling factors for the charged and neutral B meson events are calculated as follows:

$$C_{B^0} = \frac{N_{BB,data}}{N_{BB,MC}} \cdot f_{00}, \tag{4.9}$$

$$C_{B^+} = \frac{N_{BB,data}}{N_{BB,MC}} \cdot f_{+-}. \tag{4.10}$$

The branching ratios of the $\Upsilon(4S) \to B\bar{B}$ relative to the total branching ratio of the $\Upsilon(4S)$ into neutral B mesons are $f_{00} = 0.486$ and $f_{+-} = 0.514$ for charged B meson pairs assuming a 100% decay of the $\Upsilon(4S)$ resonance into B mesons. The total number of B meson pair events in real data $N_{BB,data}$ is $(771.581 \pm 10.566) \cdot 10^6$ [59] and corresponds to the offline luminosity $\mathcal{L} = 711$ fb^{-1}.

The scaling factor for continuum events is calculated based on the detector efficiency corrected cross sections of the continuum and B meson events. These cross sections are $\sigma_c = 1.30$ nb and $\sigma_{uds} = 2.09$ nb [2] for the continuum events at $\Upsilon(4S)$ resonance and $\sigma_B = 1.09$ nb [60] for B meson events.
For light-quark pairs, when $q = c$

$$C_c = \frac{\sigma_c}{\sigma_B} \cdot \frac{N_{BB,data}}{N_{c,MC}}. \tag{4.11}$$

For light-quark pairs, when $q = u, d, s$:

$$C_{uds} = \frac{\sigma_{uds}}{\sigma_B} \cdot \frac{N_{BB,data}}{N_{uds,MC}}. \tag{4.12}$$

The additional background Monte Carlo data were generated according to the offline luminosity and are thus not reweighted.

4.5.2 Uncertainties of Correlated Parameters

Uncertainties on the corrections due to correlated parameters can be determined as uncorrelated variations by rotating the parameter space. For a set of correlated parameters $\vec{x} = (x_1,...,x_n)^T$, the covariance matrix $\mathbf{C}$ has non-zero off-diagonal elements.

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & ... & \rho_{1n}\sigma_1\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & ... & \sigma_n^2 \end{pmatrix} \tag{4.13}$$

where ρ_{ij} denotes the correlation between the parameters i and j and ranges between -1 to 1, and σ_i denotes the uncertainty of parameter x_i.

By rotating the parameter space and thus diagonalizing the covariance matrix, the correlated parameters $\vec{x}$ become the uncorrelated parameters $\vec{y} = (y_1, ..., y_n)^T$.

$$\vec{y} = \vec{x}\ \mathbf{U}^{-1}, \tag{4.14}$$

where $\mathbf{U} = (\vec{u_1},..,\vec{u_n})$ consists of the n normalized orthogonal eigenvectors $\vec{u_i}$ belonging to the eigenvalues λ_i of $\mathbf{C}$. The diagonalized covariance matrix $\mathbf{V}$ is defined by:

$$\mathbf{V} = \mathbf{U}^{-1}\mathbf{C}\mathbf{U}. \tag{4.15}$$

The variation of the parameters y_i in the orthogonal space is then determined for the n different eigenvalues λ_i by adding or subtracting the corresponding uncertainty of λ_i to the nominal parameters.

$$\vec{y}_{\pm i} = \vec{y} \pm \sqrt{\lambda_i}. \tag{4.16}$$

In the last step, the varied parameters are transformed back into the original space via:

$$\vec{x}_{\pm i} = \vec{y}_{\pm i}\ \mathbf{U}. \tag{4.17}$$

4.5.3 Form Factor $B \to D^{(*)}\ell\nu$

In the generation of Monte Carlo $B \to D^{(*)}\ell\nu$ decays, the form factor parameterization as proposed by Caprini, Lellouch and Neubert [61] (CLN) was used. It relies on the heavy-quark symmetry to reduce the number of free parameters. Recent experimental results yield a better agreement between observation and theory for the form factor parameterization of Boyd, Grinstein, and Lebed [62; 63] (BGL). Therefore, each B meson decaying in a $B \to D^{(*)}\ell\nu$ channel is reweighted according to the differential decay rate from the CLN to the BGL parametrization with the eFFORT package [64] assuming $|V_{cb}| = 40.83 \cdot 10^{-3}$. If both B mesons decay in a $D^{(*)}\ell\nu$ channel, the event is reweighted for both B mesons.

The differential decay rate of $B \to D^* \ell\nu$ depends on the following variables, which are shown in Fig. 4.1:

- θ_ℓ: angle between the lepton and the direction opposite the B meson in W rest frame.
- θ_v: angle between the D meson and the direction opposite the B meson in the D^* rest frame.
- χ: tilting angle between the two decay planes spanned by the W – ℓ and D^*– D systems in the B meson rest frame.
- $q^2 = (p_B - p_{D^*})^2$ four-momentum transfer squared.

For $B \to D\ell\nu$ decays, the differential decay rate only depends on q^2.

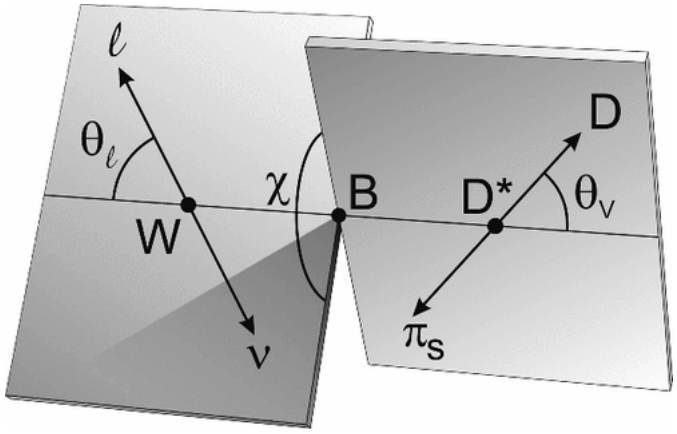

Figure 4.1: Definition of the angles θ_ℓ, θ_v , and χ for the $B \to D^* \ell \nu_\ell$ decay, mediated by a vector boson W; π_s refers to the low momentum pion from the decay $D^* \to D^0 \pi_s$ taken from [65, p. 5].

The correction factor C_{BGL} for each B meson is applied according to the following formula:

$$C_{BGL} = \frac{d\Gamma_{BGL}(\theta_\ell, \theta_v, \chi, q^2)}{d\Gamma_{CLN}(\theta_\ell, \theta_v, \chi, q^2)} \cdot \frac{\Gamma_{CLN}}{\Gamma_{BGL}}, \tag{4.18}$$

where Γ denotes the total and $d\Gamma(\theta_\ell, \theta_v, \chi, q^2)$ the differential decay rate.

The BGL parametrization of the $B \to D\ell\nu$ depends on five correlated parameters [66] and for the $B \to D^*\ell\nu$ decays on six correlated parameters [67]. The variations of the correlated parameters are determined as described in the previous section 4.5.2. In appendix A, the BGL parameters with their uncertainties and the correlation matrices are provided.

4.5.4 Branching Ratio $B \to D^{(**)}\ell\nu$

Additionally, the branching ratios of the generated $B \to D^{(**)}\ell\nu$ decays, where D^{**} denotes the four excited D states: D_1, D_1', D_0^* and D_2^*, are updated. They are calculated as isospin averages from the results of HFLAV [57] and PDG [68]; further details on the calculation can be found in [69] and [70].

The following Tab. 4.7 presents the branching ratios in the Belle Monte Carlo generation and the updated values.

B Type	Decay mode	old $\mathcal{B}$ ($\times 10^{-2}$)	new $\mathcal{B}$ ($\times 10^{-2}$)
Neutral	$B^0 \to D\ell\nu$	2.31	2.24 ± 0.07
	$B^0 \to D^*\ell\nu$	5.33	5.11 ± 0.11
	$B^0 \to D_1\ell\nu$	0.74	0.62 ± 0.10
	$B^0 \to D_0^*\ell\nu$	0.84	0.39 ± 0.07
	$B^0 \to D_1'\ell\nu$	0.74	0.39 ± 0.08
	$B^0 \to D_2^*\ell\nu$	0.36	0.27 ± 0.03
Charged	$B^+ \to D\ell\nu$	2.31	2.41 ± 0.07
	$B^+ \to D^*\ell\nu$	5.79	5.50 ± 0.11
	$B^+ \to D_1\ell\nu$	0.81	0.66 ± 0.11
	$B^+ \to D_0^*\ell\nu$	0.91	0.42 ± 0.08
	$B^+ \to D_1'\ell\nu$	0.81	0.42 ± 0.90
	$B^+ \to D_2^*\ell\nu$	0.39	0.29 ± 0.03

Table 4.7: Isospin-averaged branching ratios of $B \to X_c\ell\nu$ decays used in the Belle Monte Carlo generation and the updated ones obtained from HLFAV(2018) and PDG(2020) results.

The correction factor $C_{\mathcal{B},i}$ for each B meson decaying in the decay mode i is defined by:

$$C_{\mathcal{B}_i} = \frac{\text{new } \mathcal{B}_i}{\text{old } \mathcal{B}_i}. \tag{4.19}$$

4.5.5 Hybrid Model

The uncertainties of the hybrid model arise due to uncertainties on the inclusive branching ratio, uncertainties of the (DFN) model parameters in the generation of the non-resonant $B \to X_u\ell\nu$ decays, and from choosing a particular model out of all possible models for the generation of the non-resonant decays.

Branching Ratio

The non-resonant $B \to X_u \ell \nu$ decays are generated with the inclusive branching ratio and are later reweighted when adding the resonant decays. When generating the hybrid weights, the branching ratios of the inclusive decay modes are varied within one standard deviation uncertainty.

DFN Parameters

The non-resonant Monte Carlo events were simulated with the EVTgenVUB generator using the nominal values of the parameters m_b^{1S} and a. They depend on the HQET parameters $\overline{\Lambda}$ and λ_1, which are obtained from fits of radiative B meson decays using the Kagan-Neubert scheme [71]. The EVTgenVUB parameters are defined as:

$$m_b^{1S} = m_B - \overline{\Lambda} \quad \text{and} \quad a = \frac{3 \cdot \overline{\Lambda}^2}{-\lambda_1} - 1, \tag{4.20}$$

where m_B denotes the nominal mass of the B meson and the HQET parameters are $\overline{\Lambda} = 0.621 \pm 0.041$ and $\lambda_1 = -0.497^{+0.072}_{-0.086}$ with the correlation $\rho(\overline{\Lambda}, \lambda_1) = -0.17$.

Since the HQET parameters are correlated, the parameters m_b^{1S} and a are also correlated. The values of m_b^{1S} and a for the nominal values and the two uncorrelated variations within the uncertainty of the HQET parameters are presented in Tab. 4.8.

	nominal	variation 1 up	variation 1 down	variation 2 up	variation 2 down
m_b^{1S}	4.659	4.699	4.619	4.650	4.668
a	1.328	1.021	1.660	1.042	1.734

Table 4.8: Parameters for Monte Carlo generation obtained with the DFN model and the two variations within one σ uncertainty of the heavy quark parameters.

The two variations are determined by diagonalizing the covariance matrix. For the final uncertainty on the parameters m_b^{1S} and a, the mass of the B meson $m_B = 5.27966 \pm 0.00012$ GeV is neglected, and the larger uncertainty of λ_1 is chosen when calculating the covariance matrix to avoid underestimating the overall uncertainty.

Model Choice

Choosing the DFN model rather than any of the other models to generate the Monte Carlo data set introduces uncertainties due to the model choice. It is calculated by comparing the fit variable distribution for the default model (DFN) with the distribution of the reweighted events according to the BLNP [72] model, which serves as the representative for all other models.

4.5.6 Efficiency Corrections

Because the detector simulation does not perfectly match the real detector, particle identification and the particle finding efficiencies differ between Monte Carlo and real data. This also extends to discrepancies in the performance of multivariate algorithms, like the FEI, which is trained on the simulated particles. For the final result of the branching ratio, a precise knowledge of the signal efficiency and its uncertainty is essential. Since it is determined using signal Monte Carlo data, all efficiency corrections and their uncertainties must be applied to calculate the correct efficiency for the signal events in real data. Moreover, correcting the background of Monte Carlo also improves the overall agreement between data and Monte Carlo.

Track Finding Efficiency

One correction that needs to be applied is the track finding corrections for charged particles. A study, presented in the internal Belle Note [73], of $D^* \to \pi D^0$ decay with subsequent $D^0 \to \pi\pi K_s^0$ and $K_s^0 \to \pi^-\pi^+$ decays, revealed that the statistical uncertainty on the track finding correction for tracks with transverse momentum $p_t > 0.2$ GeV in the laboratory frame (lab frame), exceeds the correction itself. Instead of adjusting the efficiency of each track, a systematic uncertainty of 0.0035 per track is applied.

Low Momentum Particle Finding Efficiency

An additional correction for finding neutral and charged particles with a transverse momentum $p_t < 0.2$ GeV is applied. The efficiency of low momentum pions was investigated in $B^0 \to D^*\pi^+$ and a $B^+ \to \overline{D^{*0}}\pi^+$ decays, described in further detail in the internal Belle Note 1176 [74]. Correction factors for six low-momentum bins are calculated by comparing the number of signal and background events in Monte Carlo and real data obtained from fits.

Due to the large branching ratio uncertainties of the $B \to D^*\pi$ decays, the correction factors of the low momenta bins are normalized to the high momentum bin with $p_t > 0.2$ GeV, resulting in a correlation of the correction factors between the individual the momentum bins. The total correction factors uncertainty σ_{C_i} consists of an uncorrelated and a 100% correlated uncertainty for each bin i.

$$\sigma_{C_i}^2 = \sigma_{\text{uncorr}}^2 + \sigma_{\text{corr}}^2 \,, \forall i \in \{1, \ldots, 6\} \tag{4.21}$$

Particle Identification

In the reconstruction of an event, the charged final state particles are assigned a certain particle type, which depends on the information obtained from the subdetectors. Discrepancies in the response of the simulated detector and the real detector can result in particle identification differences. These are typically investigated in well-known and clean channels.

Individual correction factors for the different particle types are determined in so-called particle identification bins (pid bins) depending on the polar angle θ and the momentum in the lab frame. Furthermore, the correction factors are separated into particles, which are correctly identified, and particles imitating a different particle type, known as fakes.

Lepton Identification

The lepton identification gives a probability in the form of a likelihood ratio to distinguishing either electrons or muons from pions and kaons. The lepton identification is defined as

$$\ell_{\mathrm{ID}} = \frac{\mathcal{L}_\ell}{\mathcal{L}_\ell + \mathcal{L}_\pi + \mathcal{L}_K} \tag{4.22}$$

where $\mathcal{L}_\ell$ represents the likelihood for either electrons or muons, $\mathcal{L}_\pi$ represents the likelihood for pions, and $\mathcal{L}_K$ the likelihood for kaons.

The correction factors for true electrons and muons were obtained in a study of two-photon events $e^-e^+ \to e^-e^+\ell^-\ell^+$ with $\ell = e, \mu$. This process benefits from the large data size, resulting in low statistical uncertainties on the correction. Separate correction factors are calculated depending on the particle type, experiment (data-taking period), θ angle, momentum in the lab frame, and the lepton identification selection.

Each correction factor includes a statistical uncertainty, which is independent across all pid bins, and a systematic uncertainty, which is assumed to be 100% correlated among the bins. The systematic uncertainty also contains an additional uncertainty due to the different efficiency ratios between real data and Monte Carlo in two-photon events and $B \to XJ/\psi$, where X includes all additional particles directly originating from the B meson. More information is provided in the internal Belle Notes 777 [75] and 954 [76]. True pions and kaons mimicking electrons or muons, fake rate corrections are studied D^* decays in the analysis of semi-inclusive semi-leptonic $\mathrm{B} \to \mathrm{D}^0/\mathrm{D}^+/\mathrm{D}_s\ell\nu X$ decays, described in the internal Belle Note [77].

Pion & Kaon Identification

The pion and kaon identification gives a probability in the form of a likelihood ratio to distinguish pions from kaons and vice versa:

$$\pi_{\mathrm{ID}} = \frac{\mathcal{L}_\pi}{\mathcal{L}_\pi + \mathcal{L}_K} \quad \text{and} \quad K_{\mathrm{ID}} = \frac{\mathcal{L}_K}{\mathcal{L}_\pi + \mathcal{L}_K} \tag{4.23}$$

Efficiency corrections for pion vs. kaon identification are obtained with an D^* sample, in which the D^{*+} decays into a $\mathrm{D}^0\pi^+$ and the D^0 further decays into a K^- and π^+. This analysis, documented in the internal Belle Note 779 [78], studies the discrepancies between Monte Carlo and data in the particle identification for true kaons and true pions and Monte Carlo kaons being misidentified as pion candidates and vice versa. The correction factors depend on the momentum in the lab frame and the θ angle of the particles.

Neutral Pion

The efficiency ratio for finding neutral pion in real data compared to Monte Carlo data was analyzed by measuring the branching ratios of $\tau^- \to \pi^-\pi^0\nu$ decays in $ee \to \tau\tau$ events; see the internal Belle Note 1224 [79]. One τ was identified by a lepton in its decay, and the other τ was reconstructed in the $\tau^- \to \pi^-\pi^0\nu$ channel. Additionally, the efficiency ratio was determined for double $\tau^- \to \pi^-\pi^0\nu$ events. The resulting correction factor is given with a statistical (second term) and systematic uncertainty (third term):

$$\mathrm{C}_{\pi^0} = 0.957 \pm 0.008 \pm 0.023 \tag{4.24}$$

Full Event Interpretation

The Full Event Interpretation (FEI) employs multivariate algorithms that are trained on Monte Carlo events. To determine the correct efficiency of the FEI and its uncertainty for signal events in real data, the FEI is calibrated by comparing its efficiency on real data and Monte Carlo data in a well-known and nearly background-free channel. Assuming independence between the tag-side and signal-side reconstruction, the overall efficiency is given by $\epsilon = \epsilon_{\mathrm{B_{tag}}} \cdot \epsilon_{\mathrm{B_{sig}}}$. In the calibration, the $\mathrm{B_{tag}}$ is reconstructed analogously to the tag side of the signal events in $\mathrm{B}^0 \to \tau\ell$ decays, and the signal side is reconstructed in the $\mathrm{B}^0 \to \mathrm{D}^{*-}(\to \mathrm{D}^0(\to K^+\pi^-)\pi^-)\ell^+\nu$ mode. A detailed description of the calibration is given in chapter 6. The resulting calibration factors are separately determined for the $\mathrm{B}^0 \to \mathrm{D}^{*-}e^+\nu$ and $\mathrm{B}^0 \to \mathrm{D}^{*-}\mu^+\nu$ channels.

$$\mathrm{C}_{\mathrm{FEI},e} = 0.7796 \pm 0.0415$$
$$\mathrm{C}_{\mathrm{FEI},\mu} = 0.7871 \pm 0.0421$$

Chapter 5

Reconstruction of $B^0 \to \tau\ell$ Events

This chapter presents the reconstruction of the $\Upsilon(4S)$ resonance for signal events, where the B_{tag} is reconstructed with the hadronic FEI and B_{sig} is reconstructed in six exclusive τ decay modes. Afterward, the different background rates are studied and various selections are applied to reduce their contributions.

5.1 Analysis Strategy

My analysis follows the strategy of the previous Babar study from 2008 [80]. A sketch of an exemplary signal event is illustrated in Fig. 5.1.

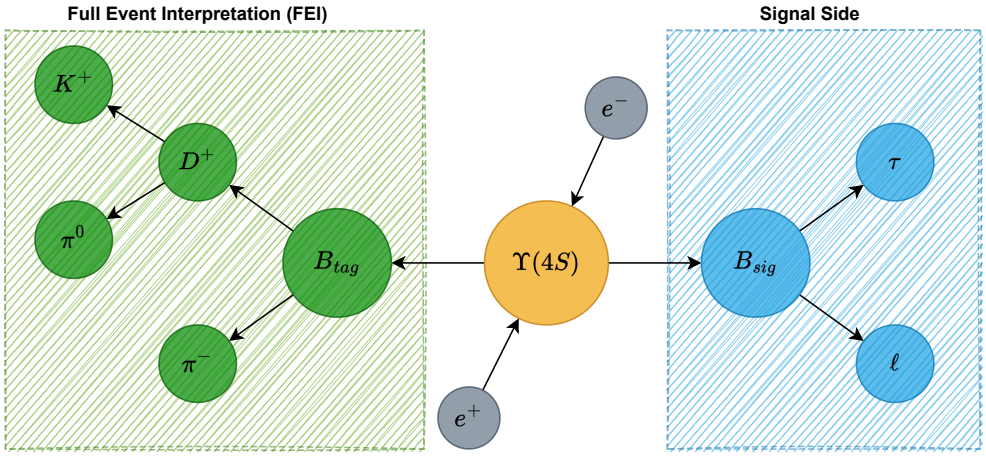

Figure 5.1: Sketch of a signal event with a B (B_{sig}) decaying in a signal channel and a hadronic decay of the other B (B_{tag}) in the event.

A signal event contains the two B mesons, which originate from a $\Upsilon(4S)$ decay. The signal B meson, referred to as B_{sig}, decays into a light lepton and a τ. Due to the large mass of the B meson, the light lepton carries a high momentum of 2.34 GeV in the B_{sig} rest frame. It is mono-energetic due to the two-body decay of the B_{sig} and serves as a clear signature for signal decays. However, the undetected neutrino(s) in the τ decays prevent the direct determination of the B_{sig} momentum from its decay products.

Exploiting the clean environment of the $\Upsilon(4S)$ decay into mostly two B mesons and nothing else allows constraining the kinematics of B_{sig} based on the knowledge of the other B meson's kinematics. Therefore, the Full Event Interpretation is employed to reconstruct the other B meson in the event, referred to as B_{tag}, in decay modes with neutrino-free final states. The B_{sig} rest frame can be calculated via momentum conservation in the CMS frame.

$$\vec{p}_{B_{sig}} = -\vec{p}_{B_{tag}} \tag{5.1}$$

By reconstructing both B mesons, no particle remains unassigned after the $\Upsilon(4S)$ reconstruction for correctly reconstructed signal decays.

The complete analysis is implemented and optimized on Monte Carlo data while the real data is blinded in the signal region of the lepton momentum. Taking advantage of the new Belle II software [81; 82], the reconstruction is implemented with the conversion tool b2bii [83], which allows the application of the FEI algorithm developed for the Belle II Experiment on Belle data.

5.2 Reconstruction of $B^0 \to \tau\ell$ Events

This section describes the reconstruction of the $\Upsilon(4S)$ candidate starting from the final state particles.

5.2.1 Event Selection

Before the reconstruction of the $\Upsilon(4S)$ an initial event selection is applied to reduce the number of background events.

1. An event must contain fewer than 18 tracks originating from the primary collision, which is ensured by selections on the minimal radial distance (dr) and the minimal distance in z-direction (dz) between the track and the IP: $dr < 2$ cm and $|dz| < 4$ cm.

2. Furthermore, an event must contain fewer than 18 photons, each with an energy requirement of E > 100/50/150 MeV in the forward (FWD)/barrel (BR)/backward (BWD) region of the detector.

3. The most crucial criterion requires at least one particle per event with a momentum $p_{\ell,\mathrm{CMS}} > 0.5$ GeV and a high lepton identification value for either an electron (eID > 0.9) or a muon (muID > 0.9). This selection discards approximately 60% of B meson events and even more so for continuum events while maintaining nearly 90% of the signal events.

Fig. 5.2 compares the light lepton momentum distribution in the B_{sig} rest frame and the CMS frame for signal events.

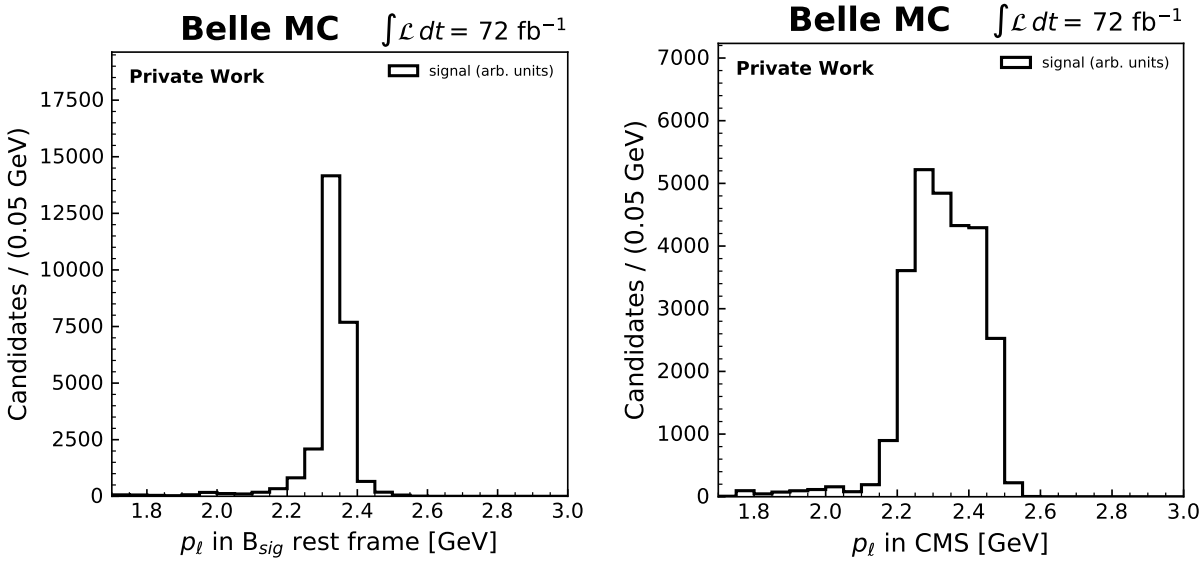

Figure 5.2: Lepton momentum distribution in the B_{sig} rest frame (left) and the CMS frame (right) for correctly reconstructed B_{sig} candidates.

Since the B meson only has an average momentum of 0.3 GeV in the CMS frame, the momentum of the light lepton is smeared out by the B meson momentum and is mostly still above 2.0 GeV.

An overview of the efficiency for the individual conditions and their combination are provided in Tab. 5.1, which are determined separately for signal and the main background events.

	$N_{tracks} < 18$	$N_\gamma < 18$	$p_{\ell,CMS} > 0.5$ GeV	all criteria
ϵ signal	100.00%	99.85%	89.32%	89.19%
ϵ charm	100.00%	99.12%	20.87%	20.77%
ϵ uds	100.00%	99.59%	10.44%	10.41%
ϵ B^0 $(b \to c)$	99.89%	96.91%	41.98%	41.48%
ϵ B^+ $(b \to c)$	99.93%	96.22%	42.65%	42.03%

Table 5.1: Efficiency of the initial selection for signals and the main background events.

5.2.2 Reconstruction of the Tag Side

After the initial event selection, B_{tag} candidates are reconstructed using the hadronic FEI. Specific B_{tag} variables are calculated, which removes events where neither of the two B mesons decays into a neutrino-free final state. The following variables are defined in the CMS frame:

ΔE is defined as the energy difference between the reconstructed B and the beam energy:

$$\Delta E = E_B - E_{beam}. \tag{5.2}$$

Since the $\Upsilon(4S)$ resonance almost always decays into two B mesons and nothing else, the initial energy ($E = 2 \cdot E_{beam}$ in the CMS frame) is distributed equally between them in the CMS frame. For correctly reconstructed B_{tag} decays $|\Delta E| = 0$ and the distribution is only smeared out by the finite detector resolution. A selection of $|\Delta E| < 0.2$ GeV is implemented.

The beam constraint mass M_{bc} is calculated via:

$$M_{bc} = \sqrt{E_{beam}^2 - \vec{p}_B^{\,2}}. \tag{5.3}$$

where, $\vec{p}_B$ represents the three-momentum of the B_{tag}, and the beam energy replaces the reconstructed B_{tag} energy. For correctly reconstructed B_{tag} decays, this yields the nominal B meson mass and improves the resolution due to the more precisely measured beam energy in comparison to the reconstructed B_{tag} energy. Requiring the reconstructed B_{tag} to satisfy $M_{bc} > 5.27$ GeV mitigates continuum events and wrongly reconstructed B meson candidates.

The signal probability $\mathcal{P}_{B_{tag}}$ is the classifier provided by the FEI, indicating the probability that a B_{tag} is correctly reconstructed. The FEI returns up to 20 possible B meson candidates per event, and the selection $\mathcal{P}_{B_{tag}} > 0.001$ discards those that are likely wrongly reconstructed. The distributions of the $\mathcal{P}_{B_{tag}}$, M_{bc}, and ΔE of the B_{tag} candidates in signal events are shown in Fig. 5.3.

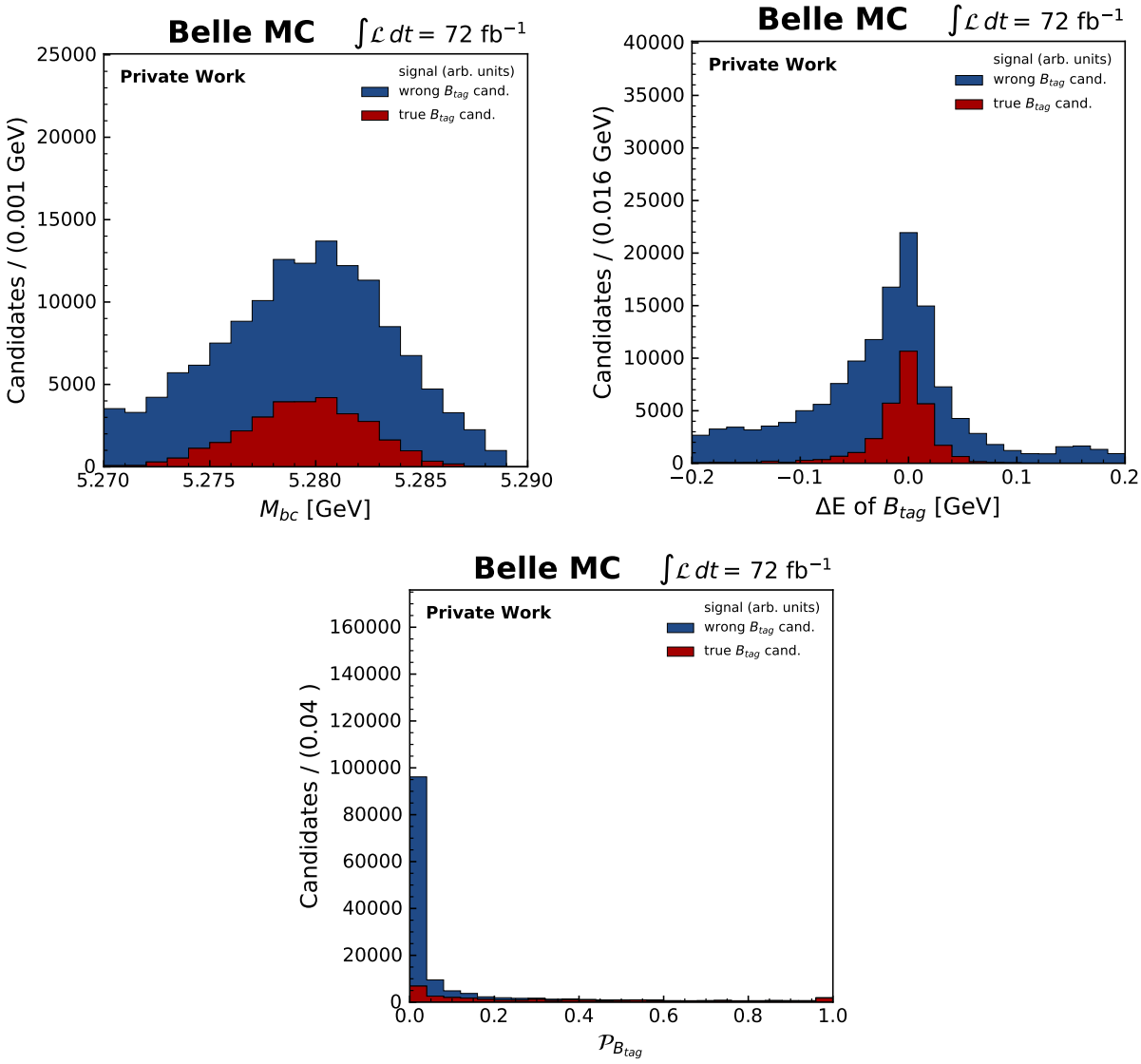

Figure 5.3: Distributions of B_{tag} variables in signal events with correctly reconstructed B_{tag} candidates shown in red and wrongly reconstructed B_{tag} candidates shown in blue.

5.2.3 Reconstruction of the Signal Side

For the signal side reconstructions, all particles not associated with the B_{tag} are considered. Since the Bremsstrahlung emitted from electrons is not necessarily negligible, a Bremsstrahlungs correction is implemented, which also slightly improves the overall resolution of the lepton momentum distribution.

The initial electron momentum $\vec{p}_{e,\text{corr}}$ is calculated by adding the momenta of photons $\vec{p}_{\gamma_{\text{brems}}}$ within a cone with an opening angle $\theta = 0.05$ rad around the measured electron momentum $\vec{p}_e$. Each of these Bremsstrahlungs photons must fulfill an energy selection of E > 100/50/150 MeV based on the FWD/BR/BWD region of the detector.

$$\vec{p}_{e,\text{corr}} = \vec{p}_e + \sum_i \vec{p}_{\gamma^i_{\text{brems}}} \tag{5.4}$$

Fig. 5.4 illustrates the procedure of the Bremsstrahlung correction. Only 6% of all electrons are Bremsstrahlung-corrected.

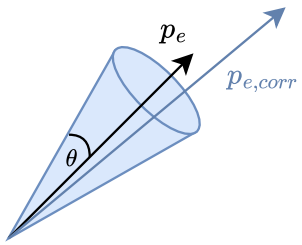

Figure 5.4: Bremsstrahlungscorrection for electrons. The original electron momentum p_e is Bremsstrahlungs corrected by adding all the momenta of photons within a cone with opening angle θ to obtain the Bremsstrahlungs corrected momentum of the electron.

$$\vec{p}_{e,\text{corr}} = \vec{p}_e + \sum_i \vec{p}_{\gamma^i_{\text{brems}}} \tag{5.5}$$

To increase the purity of the charged final state particle – electrons, muons, and pions– a selection is applied. They are required to originate from the initial collision: $dr < 2$cm and $|dz| < 4$ cm. Additionally, individual selections on the particle identification need to be fulfilled. Tab. 5.2 lists the selection for the different final state particles.

Neutral pions are reconstructed from two photons, where each photon must satisfy an energy selection of $E_\gamma > 0.05$ GeV and their combined invariant mass is required to be distributed around the nominal neutral pion mass $m_{\pi^0} = 0.135$ GeV.

	Selection Criteria
electron	eID > 0.9
muon	muID > 0.9
pion	muID ≤ 0.9 and eID ≤ 0.9
neutral pion	$E_\gamma > 0.05$ GeV and 0.12 GeV $< M_{\gamma\gamma} <$ 0.15 GeV

Table 5.2: Selection criteria for charged and neutral final state particles.

The purity for a final state particle is defined as the fraction of true particles of the hypothesized particle type relative to all particles identified as particles of the hypothesized type. Fig. 5.5 shows the composition of Monte Carlo particles for the electron, muon, and charged pions, where charged pions are referred to as π in the further analysis.

Figure 5.5: Composition of the Monte Carlo particles for the different hypothesized final state particles in signal events.

The purity for electrons candidates is high, with less than 2% of the assumed electrons being other particles mimicking the electron-like signature in the detector; these particles are called fakes. They are matched to true pions, kaons, and muons, and in some cases can't be assigned to any Monte Carlo particles, which are contained in the "other" component. Most of the muons are true muons; the leading contribution of fakes are true pions because of their similar mass. The charged pions show the lowest purity of all charged final state particles with 78%. Their main contributions of fakes originate from true kaons with 11% followed by 5% muons. Instead of further tightening the selection for charged pions, the current selection is kept to prevent a loss in efficiency.

The purity for neutral pions is only 28% in signal events. This is mainly caused by the difficulty of matching detected energy clusters to Monte Carlo photons. If an energy cluster can not be matched to its associated Monte Carlo photon, the neutral pion is not correctly reconstructed.

τ Reconstruction

Using the final state particles, τ candidates are reconstructed in six exclusive decay modes, which are given in Tab. 5.3. Overall, nearly 90% of all τ decay modes are included in the reconstruction.

τ decay mode	$\mathcal{B}$ (%)
$\tau^- \to e^- \nu_\tau \bar{\nu}_e$	17.82
$\tau^- \to \mu^- \nu_\tau \bar{\nu}_\mu$	17.39
$\tau^- \to \pi^- \nu_\tau$	10.82
$\tau^- \to \pi^- \pi^0 \nu_\tau$	25.49
$\tau^- \to \pi^- \pi^- \pi^+ \nu_\tau$	9.02
$\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$	9.29

Table 5.3: Reconstructed τ decay modes and their respective branching ratio in the signal events.

If the final state contains multiple pions, it is dominated by resonant decays. The $\tau \to \pi\pi^0\nu_\tau$ decays mostly proceed via the ρ resonance and the $\tau \to \pi\pi\pi\nu_\tau$ decays via the a_1 resonance. In the case of resonant decays, the invariant mass spectrum of the final state pions exhibits peaks at the ρ and a_1 mass. With an invariant mass selection on the hadronic daughters for these decay modes, the number of combinatorial background candidates can be reduced.

Furthermore, a selection on the cosine of the angle (θ) between the hadronic τ daughter(s) and the τ in the B_{sig} rest frame is used to discriminate against the combinatorial background candidates. Decays with multiple pions in the final state mainly occur via spin-1 ρ or a_1 resonances. This results in a peaking $\cos(\theta)$ distribution at 1, where θ denotes the angle between the three-momenta of the resonance and the τ in the B_{sig} rest frame. For the pion final state, $\cos(\theta)$ is distributed between -1 and 1, neglecting the finite detector resolution. The $\cos(\theta)$ is determined as:

$$\cos(\theta_{\tau,\mathrm{had}}) = \frac{2 \cdot \mathrm{E}_\tau \cdot \mathrm{E}_{\tau,\mathrm{had}} - m_\tau^2 - m_{\tau,\mathrm{had}}^2}{2 \cdot |\vec{p}_\tau| \cdot |\vec{p}_{\tau,\mathrm{had}}|}. \tag{5.6}$$

All energies and momenta are calculated in the B_{sig} rest frame, where the τ kinematics are determined by the four-momentum conservation of the two body B_{sig} decay. The masses are $m_\tau = 1.777$ GeV, $m_\pi = 0.140$ GeV, $m_\rho = 0.775$ GeV and $m_{a_1} = 1.245$ GeV [18]. Tab. 5.4 lists the additional selections for the hadronic τ decays.

τ decay mode	Selection Criteria	
$\tau \to \pi\nu$	–	$-1.00 < \cos(\theta_{\tau,\pi}) < 1.30$
$\tau \to \pi\pi^0\nu$	$0.6\ \mathrm{GeV} < \mathrm{M_{inv}} < 1.3\ \mathrm{GeV}$	$0.50 < \cos(\theta_{\tau,\rho}) < 1.30$
$\tau \to \pi\pi\pi\nu/\pi\pi^0\pi^0\nu$	$0.9\ \mathrm{GeV} < \mathrm{M_{inv}} < 1.8\ \mathrm{GeV}$	$0.45 < \cos(\theta_{\tau,a_1}) < 1.30$

Table 5.4: Selection criteria for hadronic τ decays.

Fig. 5.6 shows the angular distribution between the hadronic τ daughters and the τ in B_{sig} rest frame in signal events. The fraction of wrongly reconstructed B_{sig} is especially high in the $\tau \to \pi\nu$ decay mode because the other three hadronic decay modes are also reconstructed in this decay mode and seen as down-feed. For τ decay modes with π^0 in the final state, the purity is low because of the low efficiency in correctly identifying π^0. In addition, pions from the B_{tag} decay can also contribute to the combinatorial background.

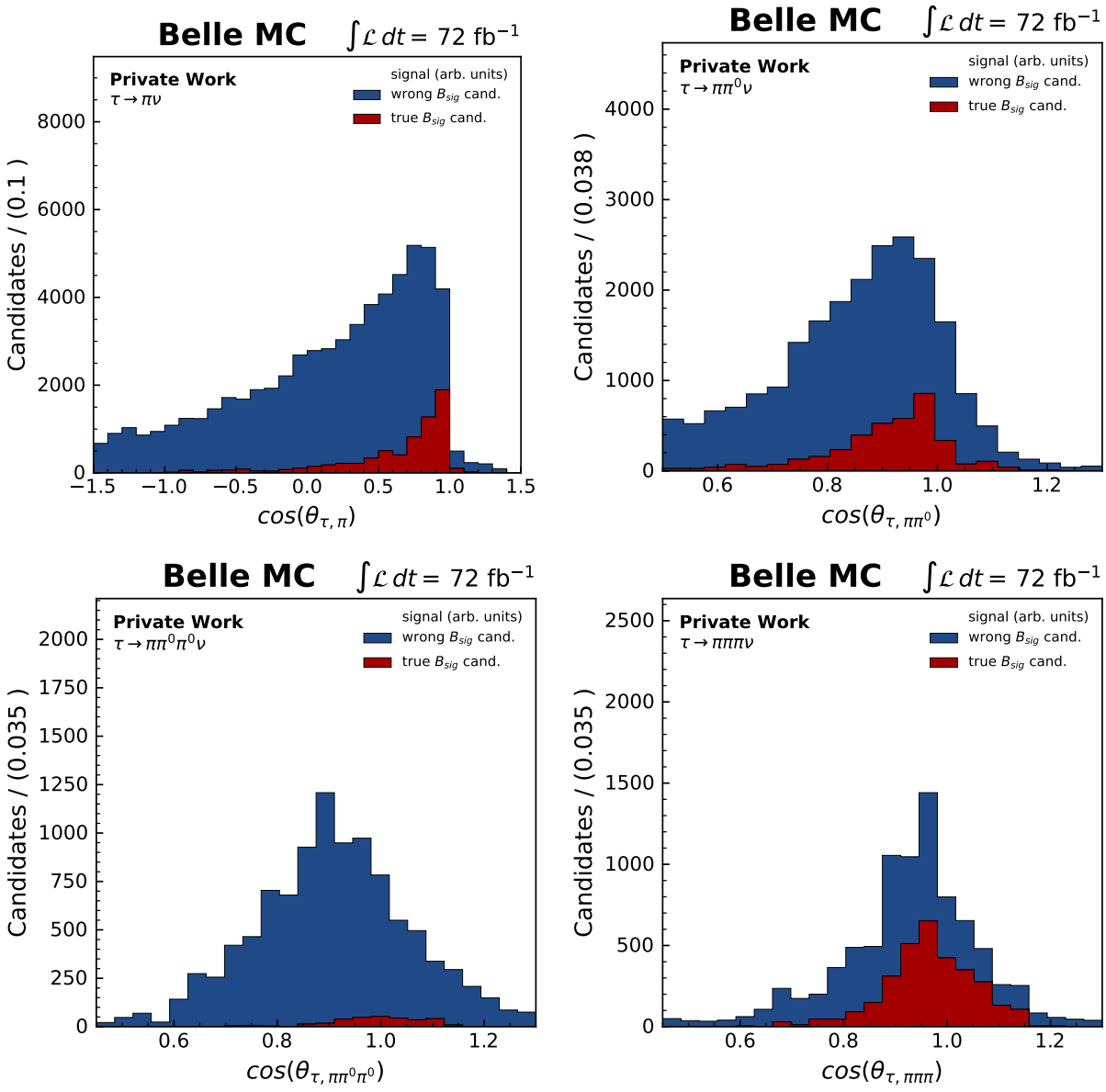

Figure 5.6: Angular distribution of the hadronic τ daughters and the τ in the B_{sig} rest frame for the four hadronic τ decay modes in signal events. The blue component represents the B_{sig} candidates, which are wrongly reconstructed, and the red component denotes the B_{sig} candidates, which are correctly reconstructed.

Reconstruction of B_{sig} and $\Upsilon(4S)$

A B_{sig} is reconstructed from a τ and light lepton. Afterward, it is combined with a B_{tag} to complete the reconstruction of the $\Upsilon(4S)$ resonance. A minimum requirement on the lepton momentum in the B_{sig} rest frame of $p_\ell \geq 1.0$ GeV decreases the number of wrongly reconstructed $\Upsilon(4S)$ candidates. Because of the neutrinos on the signal side, the total momentum of the B_{sig} cannot be determined. Instead, the two-body decay of the $\Upsilon(4S)$ and the full knowledge of the B_{tag} kinematics due to the reconstruction with the hadronic FEI are utilized.

In the CMS frame, the four-momentum of the B_{sig} is defined as:

$$p_{B_{sig}} \equiv \left(E_{B_{sig}}, \vec{p}_{B_{sig}}\right) = \left(E_{B_{tag}}, -\vec{p}_{B_{tag}}\right) = \left(E_{beam}, -\frac{\vec{p}_{B_{tag}}}{|\vec{p}_{B_{tag}}|}\sqrt{E_{beam}^2 - m_B^2}\right). \tag{5.7}$$

The beam energy is replaced with the energy of the B_{tag} because of its higher measurement precision. Additionally, only the direction of the B_{tag} three-momentum is used, and its magnitude is determined from the beam energy and the nominal B meson mass m_B. Fig. 5.7 shows the lepton momentum distribution in the B_{sig} rest frame for signal and background Monte Carlo events. The momentum distribution for signal events is not mono-energetic. Instead, it is smeared out by the finite detector resolution, with a small tail towards lower energies due to the unrecovered Bremsstrahlung-photons of the electrons.

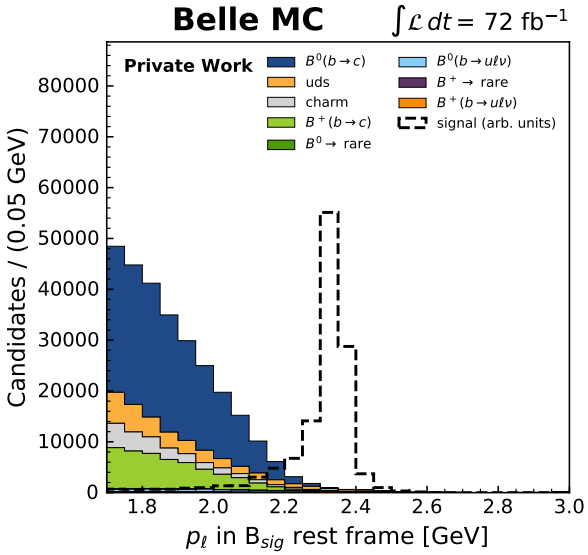

Figure 5.7: Lepton momentum distribution in the B_{sig} rest frame for signal (dashed black line) and background events on a subset of the total Monte Carlo data set corresponding to the luminosity of 72 fb^{-1}.

5.3 Study of Background Contributions

This section presents the individual steps implemented to reduce the background contributions. A continuum suppression is applied to minimize light quark background events. The large number of wrongly reconstructed signal events caused by self-cross feed is reduced with the best candidate selection, which chooses only one $\Upsilon(4S)$ candidate per event. Afterward, the various background Monte Carlo events are examined, and a J/ψ veto for events with leptonic τ decays is introduced. Moreover, fastBDTs are trained to mitigate the dominant B meson background events in the signal region of the lepton momentum.

5.3.1 Multiplicity

The multiplicity expresses the number of reconstructed particle candidates per event. Tab. 5.5 lists the average multiplicities of the B_{sig}, B_{tag}, $\Upsilon(4S)$ candidates. A particle is only used once in the reconstruction of the $\Upsilon(4S)$, but it can be assigned at different positions in the decay chain. For example, in a final state with two electrons, e_1 and e_2, electron e_1 can be assigned as the light lepton origination from the B_{sig} and electron e_2 is identified as lepton from a τ decay or vice versa, resulting in two B_{sig} candidates. Hadronic τ decays with multiple pions in the final state are the primary source for the high B_{sig} multiplicity. They contribute as down-feed in the $\tau \to \pi\nu$ decay mode and, for the $\tau \to \pi\pi^0\pi^0\nu$ decays, additionally in the $\tau \to \pi\pi^0\nu$ decay mode.

The B_{tag} multiplicity is lower because the FEI already applies selections to reduce the number of particles. The multiplicity of $\Upsilon(4S)$ candidates averages between five and eight, which is lower than the B_{sig} multiplicity due to the additional selection on the light lepton in the B_{sig} rest frame. Overall, the highest average multiplicity is observed in continuum events due to their abundance of pions. Before the multiplicity is reduced a large fraction of continuum events are discarded with a continuum suppression.

MC Event Type	B_{sig}	B_{tag}	$\Upsilon(4S)$
signal	26.11	2.24	4.35
charm	76.26	1.59	8.12
uds	67.93	1.50	7.99
$B^0(b \to c)$	52.80	2.44	7.62
$B^+(b \to c)$	62.68	1.70	6.45
$B^0 \to$ rare	34.69	1.62	5.42
$B^+ \to$ rare	42.56	1.52	5.59
$B^0(b \to u\ell\nu)$	41.44	2.38	6.38
$B^+(b \to u\ell\nu)$	47.97	1.70	5.35

Table 5.5: Average multiplicity for B_{sig}, B_{tag}, and $\Upsilon(4S)$ candidates in Monte Carlo events.

5.3.2 Continuum Suppression

By exploiting the different event topologies of continuum and B meson events, the continuum contributions can be reduced while retaining most of the signal events. Fig. 5.8 illustrates the typical event shape of continuum and B meson events in the CMS Frame. The decay products of the continuum events are distributed back-to-back, since most of the initial energy is available for the hadron momentum. In contrast to the continuum events, the creation of the two B mesons consumes almost all of the initial energy, leading to two B mesons nearly at rest in the CMS frame, resulting in isotropically distributed decay products.

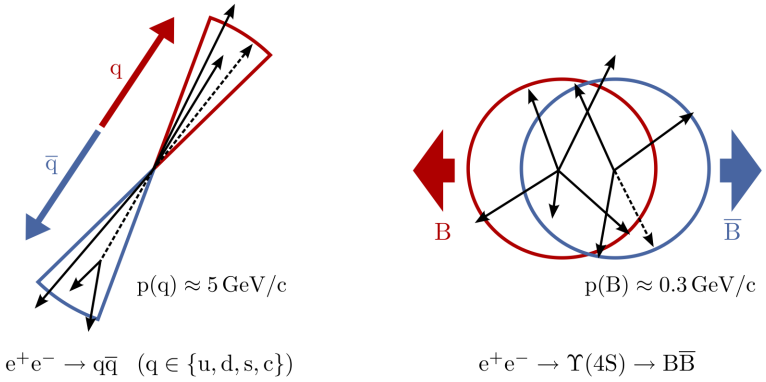

Figure 5.8: Event topology of continuum events (left) and B meson events (right) taken from [84, p. 91].

The continuum suppression (cs) was trained with fastBDTs by Felix Metzner as part of his analysis [85]. Initially, the FEI calibration especially performed for his study, was supposed to be implemented in this study as well. Therefore, the same continuum suppression with the same selection on the continuum suppression classifier output was applied to reduce the possible sources of differences for the B_{tag} candidates. In the final result of the calibration study, a correlation between the signal side and the tag side was observed. The composition of B_{tag} candidates in the semi-inclusive calibration events differed from the composition of B_{tag} candidates in Felix's study. Therefore, FEI calibration is also not used in this analysis.

Nevertheless, the continuum suppression is kept due to time constraints and because only a small decrease in signal efficiency was observed. It employs B_{tag} variables and variables of all the other particles in the event to distinguish continuum from B meson events. Particles not associated with the B_{tag} must fulfill the following selections: Photons must satisfy the $E > 100/50/150$ MeV selection in the FWD/BR/BKW detector region. And charged particles are assumed to be pions and must originate from an IP with impact parameters: $dr < 2$ cm, $|dz| < 4$ cm.

The magnetic field in the detector can bend low-momentum charged particles onto a curling path, resulting in multiple tracks in the tracking volume that belong to the same particle. Two tracks are assigned to the same particle if both have a $p_t < 275$ MeV, their transverse momenta difference $|\Delta p_t| < 20$ MeV, and the cosine of the angle between their three-momenta $|\cos(\vec{p_1}, \vec{p_2})| < \cos(5°)$. Of the two tracks, the one with no SVD hits is identified as a curling track called curler. If both tracks have either no SVD hits or SVD hits, the one with the higher impact parameters is marked as the curler.

$$\left(\frac{d_0}{\sigma_{d_0}}\right)^2 + \left(\frac{z_0}{\sigma_{z_0}}\right)^2 \tag{5.8}$$

where d_0 is the minimal distance of the track to the origin in $r - \phi$ and z_0 in the z-direction, and σ denotes the respective uncertainties. The basf2 software does not provide the variable for the uncertainties on the distance (dr, dz) to IP. The difference is negligible since the distances of the curlers to the origin are typically much larger than the distance between the IP and the origin.

The variables of the continuum suppression are present in the following section and a more detailed description is provided in [44].

Thrust Variables and B Meson Direction Variables

The thrust axis $\vec{T}$ of the B_{tag} is the axis for which the projection of the B_{tag} decay particles momenta is maximized.

$$T = \frac{\sum_{i=1}^{N} |\vec{T} \cdot \vec{p_i}|}{\sum_{i=1}^{N} |\vec{p_i}|} \tag{5.9}$$

Due to their jet-like topology, continuum events have thrust values close to 1, while B meson events exhibit a distribution around 0.5, which is explained by their isotropically distributed decay products.

The variable $|cos\theta_{T_B,TO}|$ describes the cosine of the angle between the thrust axis of the B_{tag} and the thrust axis of all other particles denoted by "O". B mesons events demonstrate a uniform distribution because the two axes of the B mesons are randomly distributed. Continuum events are jet-like, and therefore; the angle between the two thrust axes is distributed around 180°.

Additionally, the cosine of the angle between the thrust of the B and the beam axis (= z-axis) ($|\cos\theta_{T_B,z}|$) discriminates continuum and B meson events. Since the ee $\to$ B$\bar{\text{B}}$ proceeds through the spin-1 resonance of the Υ(4S), the distribution is described by a $\sin^2(\theta_{T_B,z})$ function. For the ee $\to f\overline{f}$ processes, f being a spin $\frac{1}{2}$ fermion, follows a $1 + \cos^2(\theta_{T_B,z})$ distribution.

Cleo Cones

Cleo Cones [86], describe the scalar sum of all momenta contained in concentric cones around the B_{tag} thrust axis in 10° intervals, f.e. Cleo Cone 2 covers the region between a cone with opening angle 10° and 20°. In total, nine Cleo Cones are calculated, and a 360° coverage is reached by combining the forward and backward cones. Fig. 5.9 presents the configuration of Cleo Cones around the thrust axis.

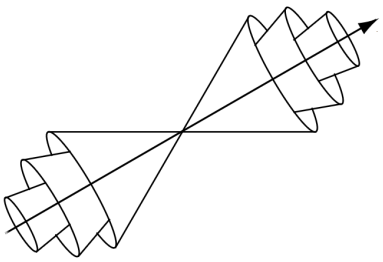

Figure 5.9: Illustration of Cleo Cones modified from [44, p. 111].

Fox-Wolfram Moments

Fox-Wolfram moments [87] quantify the momentum distribution of an event. For N particles, with each a momentum p_i, the Fox-Wolfram moment of order k is defined as:

$$H_k = \sum_{i,j}^{N} |\vec{p_i}||\vec{p_j}| P_k(cos\theta_{ij}) \tag{5.10}$$

where θ_{ij} denotes the angle between the momentum of particle i and j, and P_k is the k-th order Legendre polynomial. Instead of the Fox Wolfram moments, a modified version, the so-called Kakuno Super Fox-Wolfram moments, are used in the continuum suppression. They include additional information on the event topology, f.e. the number of charged particles. A detailed description can be found in [44].

The ratio of the second to zeroth order Fox-Wolfram Moment is defined as:

$$R_2 = \frac{H_2}{H_0}. \tag{5.11}$$

Continuum Suppression Classifier Output

The distribution of the continuum suppression classifier output in Fig. 5.10 demonstrates the separation of continuum and B meson events in Monte Carlo data. Continuum events are distributed close to zero, while most B meson events tend toward higher values close to one.

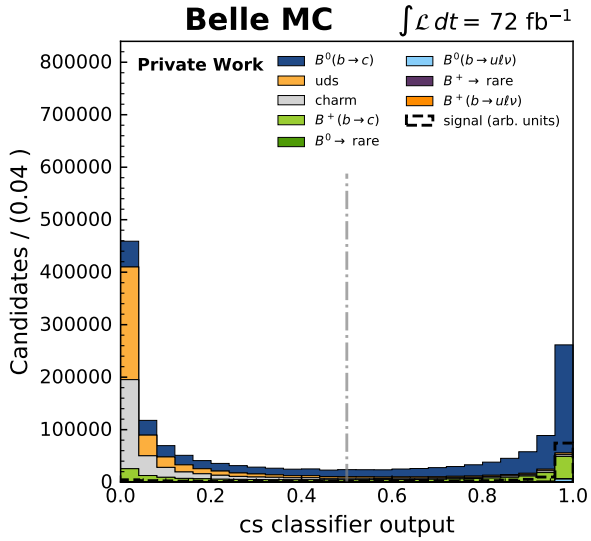

Figure 5.10: Continuum suppression (cs) classifier output for signal and background events corresponding to a luminosity of 72 fb^{-1}. The stacked components represent the different background contributions, and the dashed black line denotes the distribution for signal events. The grey line marks the implemented selection of the classifier output.

A selection of the continuum suppression classifier output > 0.5 is applied to suppress a large fraction of the continuum contributions. Tab. 5.6 compares the efficiencies of the cs classifier output > 0.5 selection for signal and background events.
The efficiencies are calculated once for the loss of events ($\epsilon_{\mathrm{events}}$) and the efficiencies for the loss of $\Upsilon(4S)$ candidates (ϵ_{cand}). These two efficiencies differ due to multiple $\Upsilon(4S)$ candidates per event. Approximately 90% of the continuum events are discarded while keeping 85% of signal events.

MC Event Type	$\epsilon_{\text{cand.}}$(cs)	ϵ_{events}(cs)
signal	81.24%	85.47%
signal true $\Upsilon(4S)$	90.49%	90.61%
charm	9.93%	11.12%
uds	9.19%	9.85%
$B^0(b \to c)$	67.89%	72.31%
B^+ $(b \to c)$	57.63%	62.89%
$B^0 \to$ rare	60.18%	66.28%
$B^+ \to$ rare	52.89%	58.25%
$B^0(b \to u\ell\nu)$	76.65%	82.53%
B^+ $(b \to u\ell\nu)$	70.46%	73.92%

Table 5.6: Efficiencies of the continuum suppression (cs) selection on signal and background Monte Carlo events. The component "signal true $\Upsilon(4S)$" denotes signal events with at least one correctly reconstructed $\Upsilon(4S)$ candidate. The efficiencies for the continuum events are marked in blue.

5.3.3 Best Candidate Selection

After the continuum suppression, a best candidate selection chooses the $\Upsilon(4S)$ candidate that is most likely correctly reconstructed based on B_{tag} and signal-side variables. The following order of the individual steps showed the best efficiency for signal events:

I: Select B_{tag} candidate with the highest signal probability.

II: Choose lepton from B_{sig} with highest lepton momentum in B_{sig} rest frame.

IIIa: For leptonic τ decays, choose the lepton with the second highest momentum in the B_{sig} rest frame.

IIIb: For hadronic τ decays choose the candidate with the smallest energy difference ΔE_τ in τ rest frame. ΔE_τ is defined as:

$$\Delta E_\tau = E_H + |\vec{p}_\nu| - m_\tau. \tag{5.12}$$

with $|\vec{p}_\nu| = |\vec{p}_H|$, since the they are back-to-back in the τ rest frame as shown in Fig. 5.11.

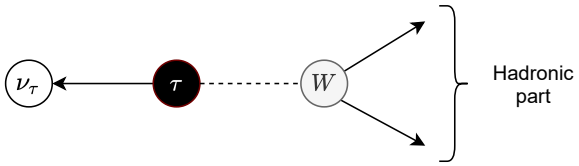

Figure 5.11: Hadronic τ decay in the τ rest frame.

For correctly reconstructed hadronic τ decays ΔE_τ peaks at zero. The $|\Delta E_\tau|$ distribution for correctly reconstructed leptonic and hadronic τ decays in signal events is shown in Fig. 5.12.

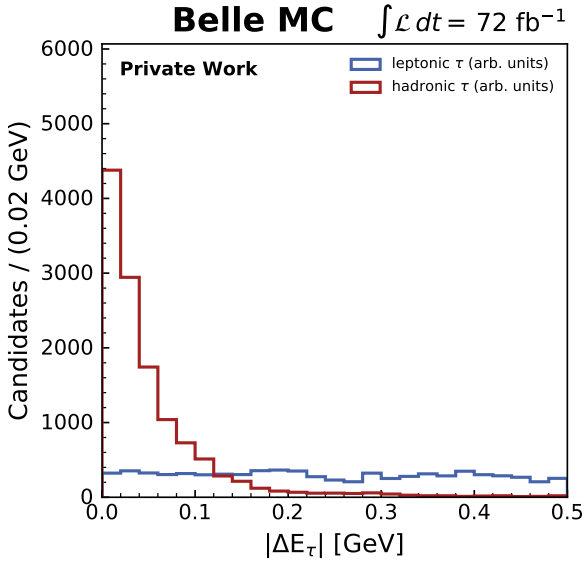

Figure 5.12: ΔE distribution for correctly reconstructed B_{sig} candidates with leptonic τ daughters in blue and hadronic τ daughters in red.

III: If both a leptonic and a hadronic τ candidate are selected in the same event, the hadronic τ candidate is chosen if its $|\Delta E_\tau| < 0.1$ GeV, otherwise the leptonic τ candidate is selected.

As the last step of the best candidate selection, only the events where the light lepton from the B_{sig} passes the selection $p_\ell > 1.5$ GeV in the B_{sig} rest frame are kept.

Reconstructed vs. Generated τ Decay Modes in Signal Events

The impact of the best candidate selection on the purity of the individual τ decays is investigated on a subset of signal events. Fig. 5.13 compares the generated τ decay modes with the reconstructed ones before and after the best candidate selection.

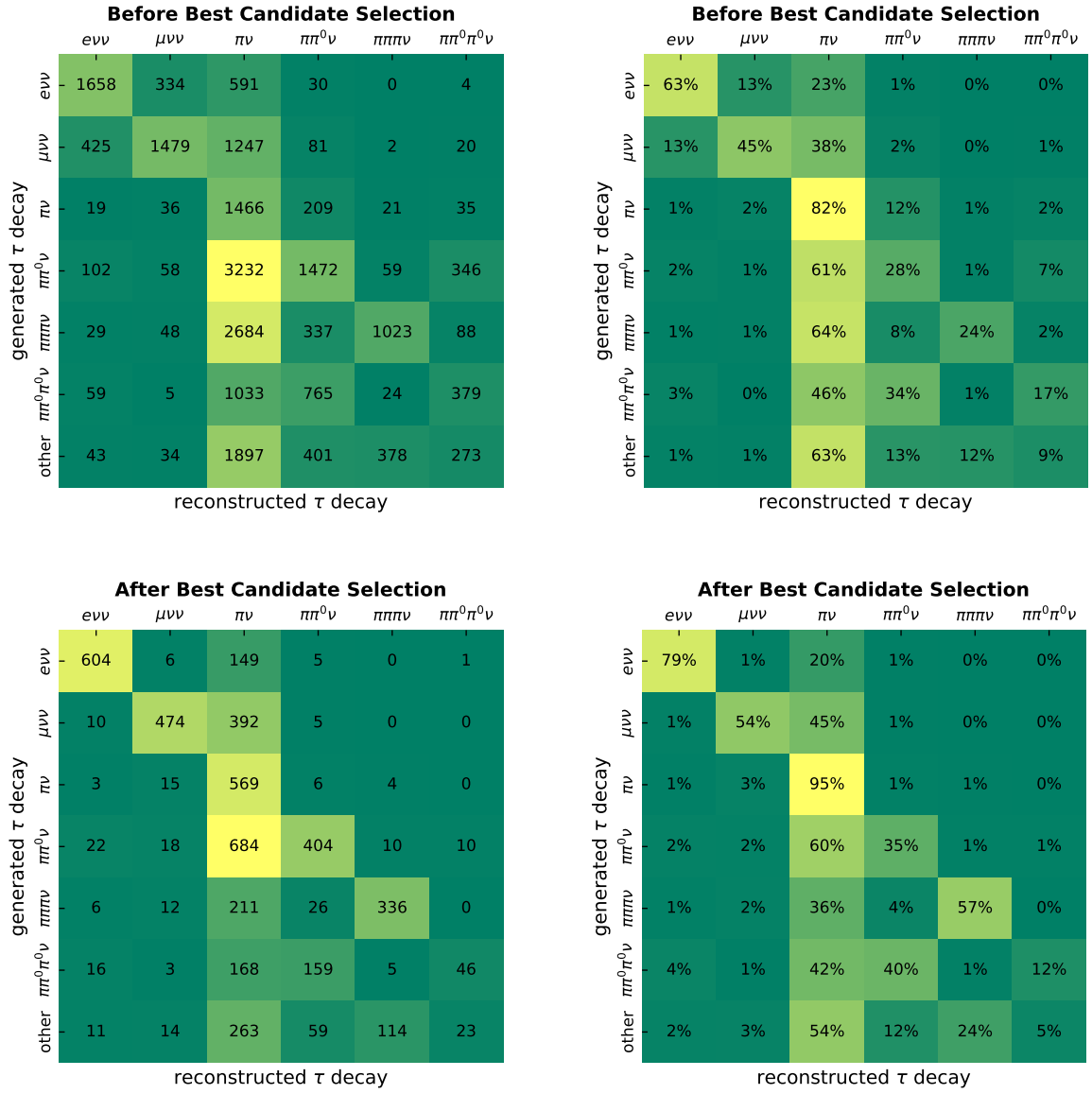

Figure 5.13: Cross-feed matrix between generated and reconstructed τ decays for signal events before (upper plots) and after the best candidate selection (bottom plots). The matrix on the left displays the absolute number of $\Upsilon(4S)$ candidates for each reconstructed τ decay mode per generated τ decay mode. The matrix on the right presents the percentage of $\Upsilon(4S)$ candidates for each reconstructed τ decay mode relative to the total number of $\Upsilon(4S)$ candidates per generated τ decay mode. Since not all generated τ decay modes are reconstructed, they are combined into the "$\tau \rightarrow$ other" component.

Overall, the best candidate selection improves the agreement between the generated and reconstructed decay modes. However, a slight decrease is observed for the generated $\tau \to \pi\pi^0\pi^0\nu$ decays. Due to its small contribution compared to all other decay modes and the difficulty of detecting two π^0, the best candidate selection is left unchanged.

A considerable cross-feed of the $\tau \to \mu\nu\nu$ into the $\tau \to \pi\nu$ mode is visible. Although the true muons are assigned as the τ daughters, they mostly have a muID below 0.2 and are therefore not identified as muons. Since the only requirement for charged pions is a leptonID < 0.9, all those true muons are identified as pions instead. Even a looser selection of the muID value would not reduce this cross-feed.

For hadronic τ decays with multiple pions in the final state, a down-feed into the $\tau \to \pi\nu$ is observed. The best candidate selection reduces this cross-feed, but undetected particles or neutral pions that do not fulfill the neutral pion criteria still appear as down-feed.

Efficiency of Continuum Suppression and Best Candidate Selection

The signal efficiency of correctly reconstructed $\Upsilon(4S)$ candidates for the continuum suppression and the best candidate selection are shown in Tab. 5.7. As the previous study shows, the efficiency decreases with an increasing number of τ daughters.

τ decay mode	ϵ(cs)	ϵ(bcs)	ϵ(cs + bcs)
$\tau \to e\nu\nu$	90.33%	89.11%	80.50%
$\tau \to \mu\nu\nu$	90.97%	90.45%	82.28%
$\tau \to \pi\nu$	96.95%	92.43%	89.61%
$\tau \to \pi\pi^0\nu$	82.91%	79.53%	65.93%
$\tau \to \pi\pi\pi\nu$	92.34%	90.88%	83.92%
$\tau \to \pi\pi^0\pi^0\nu$	64.74%	77.84%	50.39%

Table 5.7: Efficiencies of the continuum suppression (cs) and best candidate selection (bcs) for the different generated τ decays in signal events.

The efficiencies for signal and background events are presented in Tab. 5.8. If a B_{sig} is correctly reconstructed, it is in 94% of all cases chosen by the best candidate selection. The lepton momentum requirement in the best candidate selection removes less than 9% of signal events while discarding between 27% and 80% of the individual background components. The smallest decrease in background events is observed for the $B^0(b \to u\ell\nu)$ events because their leptons also have a high momentum.

MC Event Type	ϵ(bcs)	ϵ(cs + bcs)
signal	91.74%	78.41%
signal true $\Upsilon(4S)$	88.92%	80.57%
signal true B_{tag}	86.51%	75.23%
signal true B_{sig}	94.00%	85.62%
charm	19.94%	2.22%
uds	22.77%	2.24%
$B^0(b \to c)$	42.09%	30.43%
$B^+(b \to c)$	39.53%	24.86%
$B^0 \to$ rare	41.36%	27.41%
$B^+ \to$ rare	39.82%	23.19%
$B^0(b \to u\ell\nu)$	63.01%	52.00%
$B^+(b \to u\ell\nu)$	54.95%	40.62%

Table 5.8: Efficiencies of best candidate selection (bcs) and continuum suppression + best candidate selection (cs + bcs) for Monte Carlo signal and background events.

5.3.4 Study of Background Contributions in the Signal Region

Although a large fraction of background events was already discarded in the previous analysis steps, the background contributions in the signal region of the lepton momentum in the B_{sig} rest frame distribution (2.20 GeV - 2.42 GeV) remain. Fig. 5.14 presents the lepton momentum distribution of Monte Carlo events after the best candidate selection.

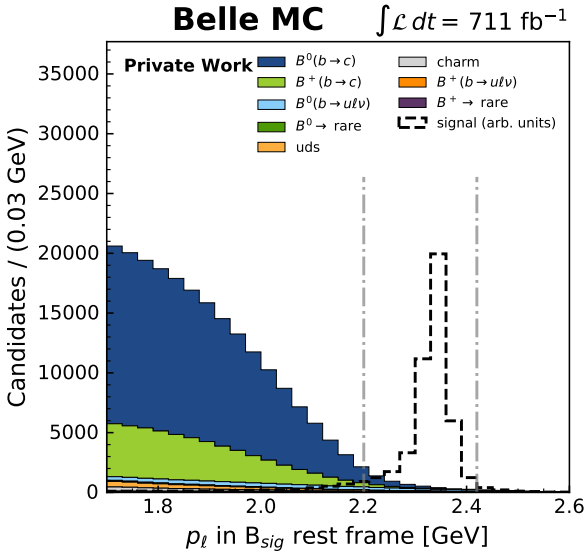

Figure 5.14: Lepton momentum distribution in the B_{sig} rest frame for signal (dashed black line) and background (colored) Monte Carlo events after the best candidate selection. The signal region is denoted by the area between the dashed grey lines.

The $B(b \to c)$ background events in the signal region are studied with the TopoAna tool [88] to identify dominant background decays. It relies on Monte Carlo information of each particle to determine the complete generated $\Upsilon(4S)$ decay chain and individual sub-decays. Typical decay modes of signal events, where one B meson decays through a $b \to c$ transition, are compared with the main decay modes of $B(b \to c)$ background events. Because the B_{tag} is reconstructed with the full hadronic FEI and is assumed to have a negligible correlation to the signal side, the difference in the identified generated decay modes is attributed to the B_{sig} reconstruction.

Fig. 5.15 confirms this assumption for the $B^0(b \to c)$ component, where the normalized distributions of the B_{tag} decay modes for the signal and the $B^0(b \to c)$ events in the signal region show a good agreement.

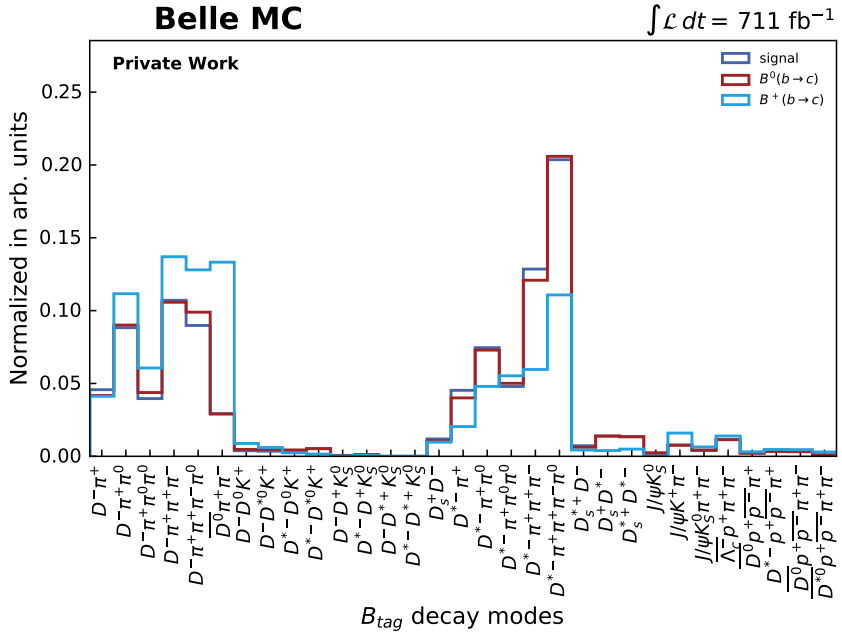

Figure 5.15: Reconstructed decay modes of the B_{tag} with the full hadronic FEI for signal events in dark blue, $B^0(b \to c)$ events in red, and $B^+(b \to c)$ events in light blue. Only the events in the signal region are included and the distributions are normalized.

Because different background processes can contribute to different reconstructed τ decay modes, they are analyzed separately. Tab. 5.9 and 5.10 summarize the most significant insights on the generated decay modes. The percentage is given by the number of events with at least one B meson decaying in a specific channel, and for the $\Upsilon(4S)$ final states, relative to the total number of events.

	rec. τ mode: $\tau \to e\nu\nu$			rec. τ mode: $\tau \to \mu\nu\nu$		
MC information	$B^0(b \to c)$	$B^+(b \to c)$	signal	$B^0(b \to c)$	$B^+(b \to c)$	signal
$B \to D^{(*)}\ell\nu$	31%	60%	1%	29%	22%	0%
$B \to c\bar{c}$ res. + X	63%	67%	2%	59%	40%	2%
ℓ, ν in FS	0%	0%	0%	63%	67%	2%
no ν in FS	58%	30%	0%	62%	66%	0%

Table 5.9: Typical decay modes of the neutral and charged $B(b \to c)$ background and signal events in the leptonic reconstructed τ decay modes $\tau \to e\nu\nu$ and $\tau \to \mu\nu\nu$. The last components, 'no ℓ, ν in FS' and 'no ν in FS', denote the events without any light leptons and neutrinos in the final state (FS) or only without any neutrinos in the final state.

Only 2% of signal events contain a B meson decaying through a $c\bar{c}$ resonance, which can be explained with the $B^0 \to J/\psi(\to \ell\ell) + X$ channels reconstructed in the full hadronic FEI,

where X denotes all other particles. For background events, the percentages vary between 40% and 67%. Assuming most of the leptons from the $B \to J/\psi(\to \ell\ell) + X$ decays are part of the signal side reconstruction, a peak in the invariant mass of the lepton of the B_{sig} and the oppositely charged lepton of the τ decay should be present.

	rec. τ mode: $\tau \to \pi\nu$			rec. τ mode: $\tau \to \pi\pi^0\nu$		
MC information	$B^0(b \to c)$	$B^+(b \to c)$	Signal	$B^0(b \to c)$	$B^+(b \to c)$	Signal
$B \to D^{(*)}\ell\nu$	80%	89%	1%	67%	85%	1%
$B \to c\bar{c}$ res. + X	5%	3%	2%	3%	5%	2%
No ℓ, ν in FS	12%	8%	0%	16%	10%	0%
No ν in FS	17%	11%	0%	19%	15%	0%

	rec. τ mode: $\tau \to \pi\pi\pi\nu$			rec. τ mode: $\tau \to \pi\pi^0\pi^0\nu$		
MC information	$B^0(b \to c)$	$B^+(b \to c)$	Signal	$B^0(b \to c)$	$B^+(b \to c)$	Signal
$B \to D^{(*)}\ell\nu$	75%	88%	1%	76%	80%	0%
$B \to c\bar{c}$ res. + X	10%	4%	1%	12%	10%	2%
No ℓ, ν in FS	4%	0%	0%	6%	10%	0%
No ν in FS	22%	0%	0%	12%	20%	0%

Table 5.10: Typical decay modes of the neutral and charged $B(b \to c)$ background and signal events in the hadronic reconstructed τ decay modes $\tau \to \pi\nu$, $\tau \to \pi\pi^0\nu$, $\tau \to \pi\pi\pi\nu$, and $\tau \to \pi\pi^0\pi^0\nu$. The last components, 'no ℓ, ν in FS' and 'no ν in FS', denote the events without any light leptons and neutrinos in the final state (FS) or only without any neutrinos in the final state.

The main background contributions for hadronic τ decay modes are events with at least one $B \to D^{(*)}\ell\nu$ decay. Since the full hadronic FEI does not reconstruct semi-leptonic B decays, the lepton of those decays is mainly identified as the lepton from the B_{sig} candidate. Between 12% and 66% of the $\Upsilon(4S)$ resonances decay into neutrino-free final states in background events. Variables based on the total energy in the event or the missing momentum of the event can be used to suppress those background contributions. The missing three-momentum of the event is defined as:

$$\vec{p}_{miss} = \vec{p}_{B_{tag}} - \vec{p}_{B_{sig,vis}} \tag{5.13}$$

where $\vec{p}_{B_{tag}}$ represents the three-momentum of the B_{tag} and $\vec{p}_{B_{sig,vis}}$ the detected three-momentum of the B_{sig}. For correctly reconstructed B_{sig} candidates, the $\vec{p}_{B_{sig,vis}}$ is the sum of all signal-side final state particles momenta apart from the undetected neutrino(s).

Comparing the small fraction of decays into the $c\bar{c}$ resonance in hadronic reconstructed τ decays, with the large fraction in the leptonic reconstructed τ decays strengthens the argument of the lepton pair being reconstructed in the signal side.

J/ψ Veto

Based on the findings of the study with TopoAna, the invariant mass of the electron and muon pair of the two signal-side leptons is investigated; see Fig. 5.16.

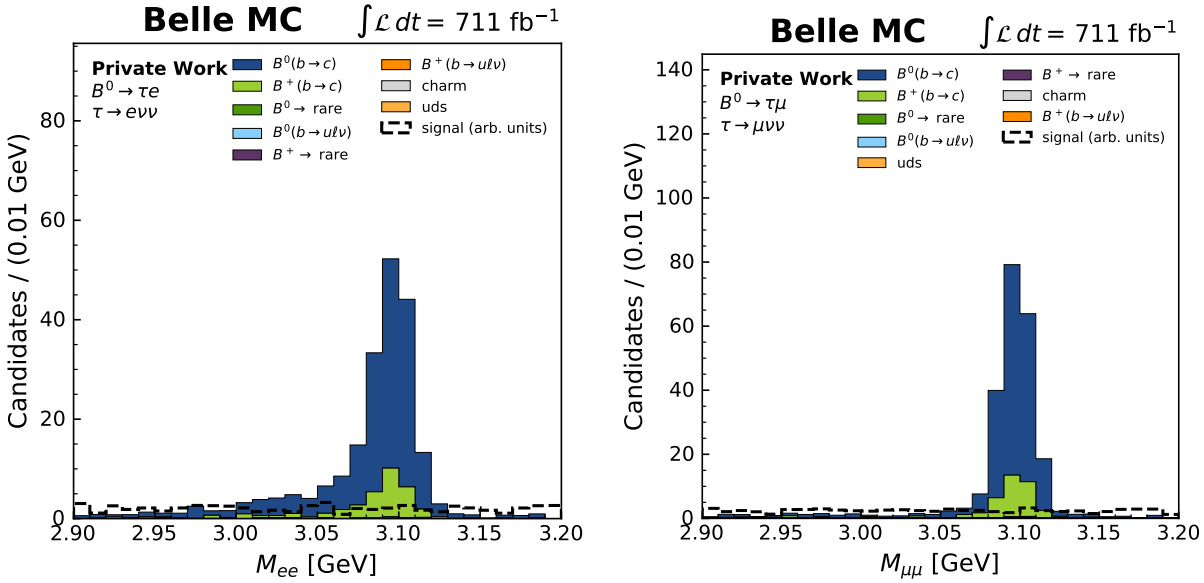

Figure 5.16: Invariant mass of electron pair (left) in the $B^0 \to \tau(\to e\nu\nu)e$ channel and muon pair (right) in the $B^0 \to \tau(\to \mu\nu\nu)\mu$ channel for signal events (dashed black line) and the different background contributions.

An enhancement in the B meson background around the nominal J/ψ mass at 3.1 GeV is visible for the same lepton-type pairs while the signal distribution remains flat. The resonance in the invariant mass spectrum is broader for the electron pair with a tail towards smaller masses, which arises from unrecovered Bremsstrahlung photons.

A J/ψ veto discards events, where the invariant mass of the same type of light lepton pairs falls within the region of invariant J/ψ mass. The selection is given in Tab. 5.11.

	J/ψ
muon pair	3.07 GeV$<$ $M_{\mu\mu}$ $<$ 3.13 GeV
electron pair	3.00 GeV$<$ M_{ee} $<$ 3.12 GeV

Table 5.11: Selections on invariant mass of lepton pairs for the electron- and muon-pair. Lepton-pairs not satisfying the invariant mass selection are discarded.

Since a non-negligible fraction of true muons is identified as pions, the invariant mass spectrum of a muon originating from the B_{sig} and a pion from the τ with a hypothesized muon mass is analyzed. As expected from the previous TopoAna result, only a small peak around the nominal J/ψ mass, shown in Fig. 5.17, is visible. Consequently, no J/ψ veto is applied for the hadronic τ decays to prevent the loss of signal efficiency.

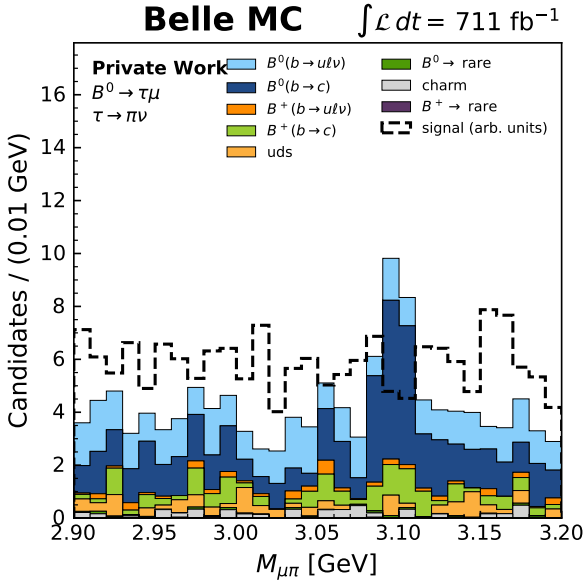

Figure 5.17: Invariant mass of muon and pion assuming the muon mass hypothesis for the pion candidate in the $B^0 \to \tau(\to \pi\nu)\mu$ channel for signal events (dashed black line) and the different background contributions.

5.3.5 Background Suppression

Variables of the signal side and the rest of the event of the $\Upsilon(4S)$ are investigated to further mitigate these background contributions. The rest of the event of the $\Upsilon(4S)$ contains all charged and neutral particles that are not associated with the tag- or signal-side reconstruction. Two selections, referred to as masks, are applied to the neutral and charged particles of the rest of the event; see Tab. 5.12.

	charged particle		neutral particle
roe mask	$dr < 10$ cm $\lvert dz\rvert < 20$ cm	not curler	$E > 100/50/150$ MeV in FWD/BR/BKW
mask 1	✓	✗	✓
mask 2	✓	✓	✓

Table 5.12: Requirement for each charged and neutral particle in the rest of the event for the two different rest of events masks.

Charged particles are assumed to be pions and must fulfill a loose selection on the distance to the interaction point to ensure they originate from the primary collisions. Neutral particles are assumed to be photons and must satisfy an energy selection depending on the detector region.

For correctly reconstructed signal events, the rest of the event is empty because all particles were used in the B_{tag} and B_{sig} reconstruction. However, mistakes in the reconstruction, f. e. if a multi-pion τ decay is reconstructed as a down-feed in the $\tau \to \pi\nu$ channel, can lead to a non-empty rest of the event.

A loose selection is applied on the number tracks < 4, the number of neutral ECL clusters < 4, and the number of neutral pions < 2 in the rest of the event, reducing 51% of background events while still retaining 82% of signal events. Fig. 5.18 shows the distributions of these three rest of the event variables before the selection. The number of neutral pions is independent of the rest of the event mask, and counts the number of neutral pions, satisfying the 0.08 GeV $< M_{\gamma\gamma} <$ 0.18 GeV requirement.

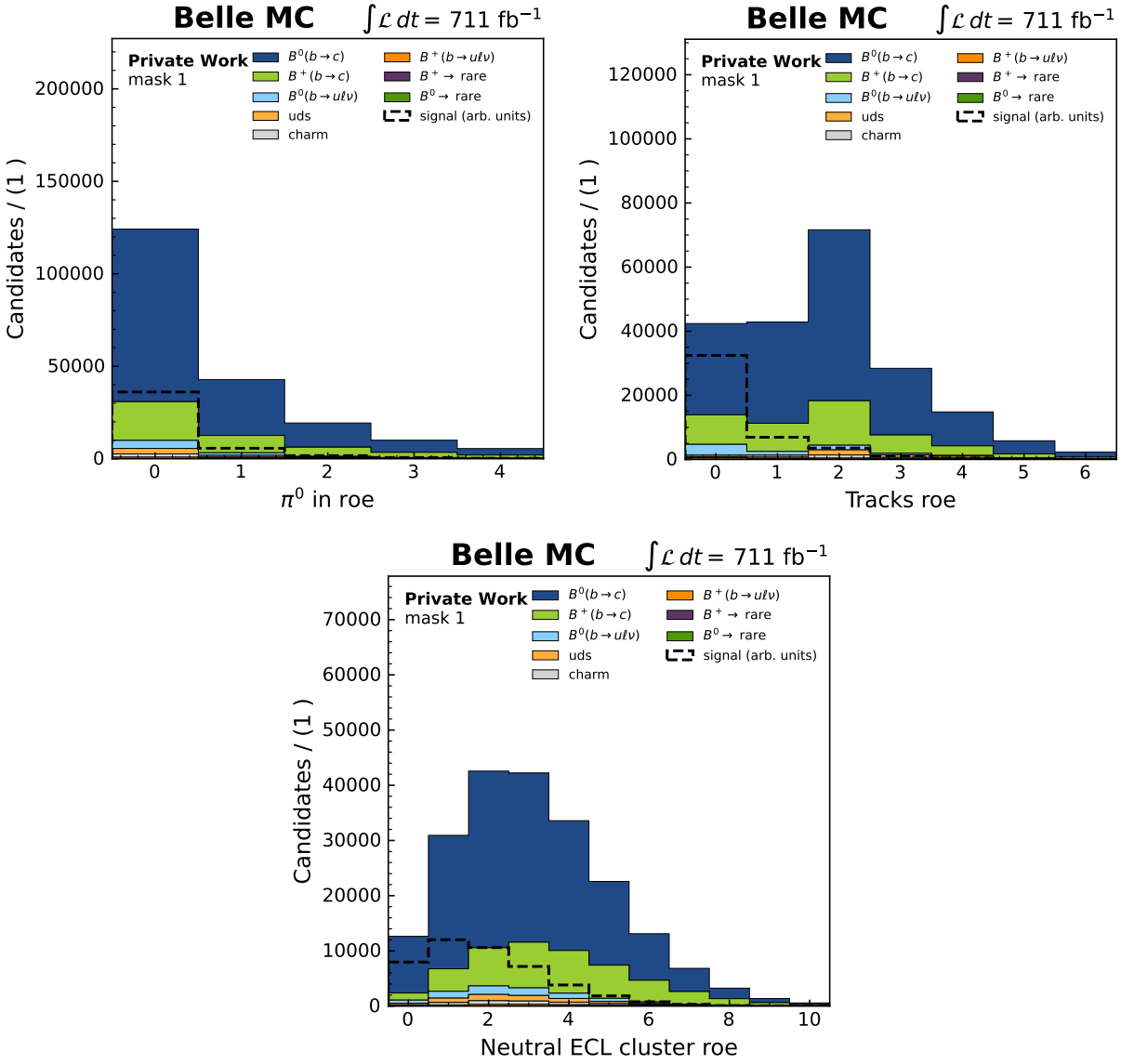

Figure 5.18: Rest-of-the-event (roe) variables for signal and background Monte Carlo events: the number of neutral pions (upper left), the number of tracks (upper right) for mask 1, and the number of neutral ECL clusters (bottom) for mask 1.

Afterward, multiple B meson background suppressions based on fastBDT [39] are trained using signal-side and the rest-of-the-event variables to distinguish signal from background events. The training variables are listed in Tab. 5.13.

	core variables	optional variables
signal variables	visible τ momentum in B_{sig} rest frame ΔE_τ in τ rest frame	$\lvert p_{miss}\rvert$ of the event reconstructed τ decay modes
roe variables	number of tracks transverse momentum total energy neutral energy in ECL	number neutral pions momentum number of all clusters in ECL number of neutral clusters in ECL invariant mass charge

Table 5.13: Variables of the B meson background suppressions divided into the core variables, which are included in all trainings, and the optional variables, which are only included in some of the trainings. Apart from the decay mode identifier of the reconstructed τ decays (blue), either all or none of the optional variables are included in the training variable set.

Four different sets of variables are used in the B meson background suppressions, which is shown in Fig. 5.19. The core variables are included in all trainings, and the optional variables are only included in a subset of trainings.

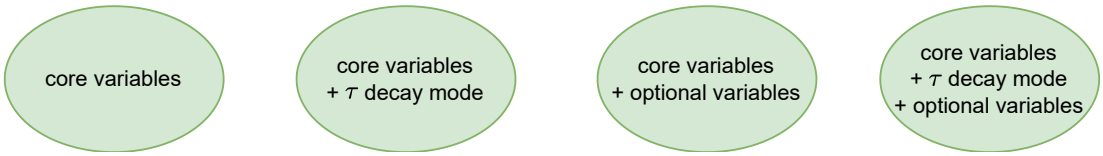

Figure 5.19: The four different variable sets used in the training of the B meson background suppressions.

The B meson background suppression relies on the good representation of the real data events by Monte Carlo events. Discrepancies between Monte Carlo and real data in the training variables can influence the shape of the final fit distribution and signal efficiency. If more variables are included in the B meson background suppression training, the uncertainty in the signal efficiency can increase. Meanwhile, more variables in the suppression training can also improve the separation power between background and signal events.

If the correlation between training variables and lepton momentum is small, the distribution of the training variables can be extrapolated to the signal region to estimate the impact of their Monte Carlo - data discrepancies on the lepton momentum.

The correlation coefficient [35, p. 18] quantifies the linear correlation between the lepton moment in the region [1.7, 2.6] GeV and the training variables in their shown range. The correlation is defined as:

$$\rho = \frac{V_{xy}}{\sigma_x \sigma_y} \tag{5.14}$$

where V_{xy} denotes the covariance between the two variables x and y and σ_x and σ_y their corresponding standard deviations. Fig. 5.20 - 5.21 show the distributions of the four training variables with the largest correlation to the lepton momentum for one example rest of the event mask 1. The requirement in mask 2 that a charged particle in the rest of the event is not identified as a curler, only has a low impact on the training variable distributions and they are similar to those of mask 1.

For correctly reconstructed signal events, no particles are in the rest of the event; therefore, the roe variables peak at low values. The maximal correlation between lepton momentum and training variables is $|\rho| \leq 0.13$. As expected, the signal events have a slightly higher correlation between the training variables and the lepton momentum than the background events.

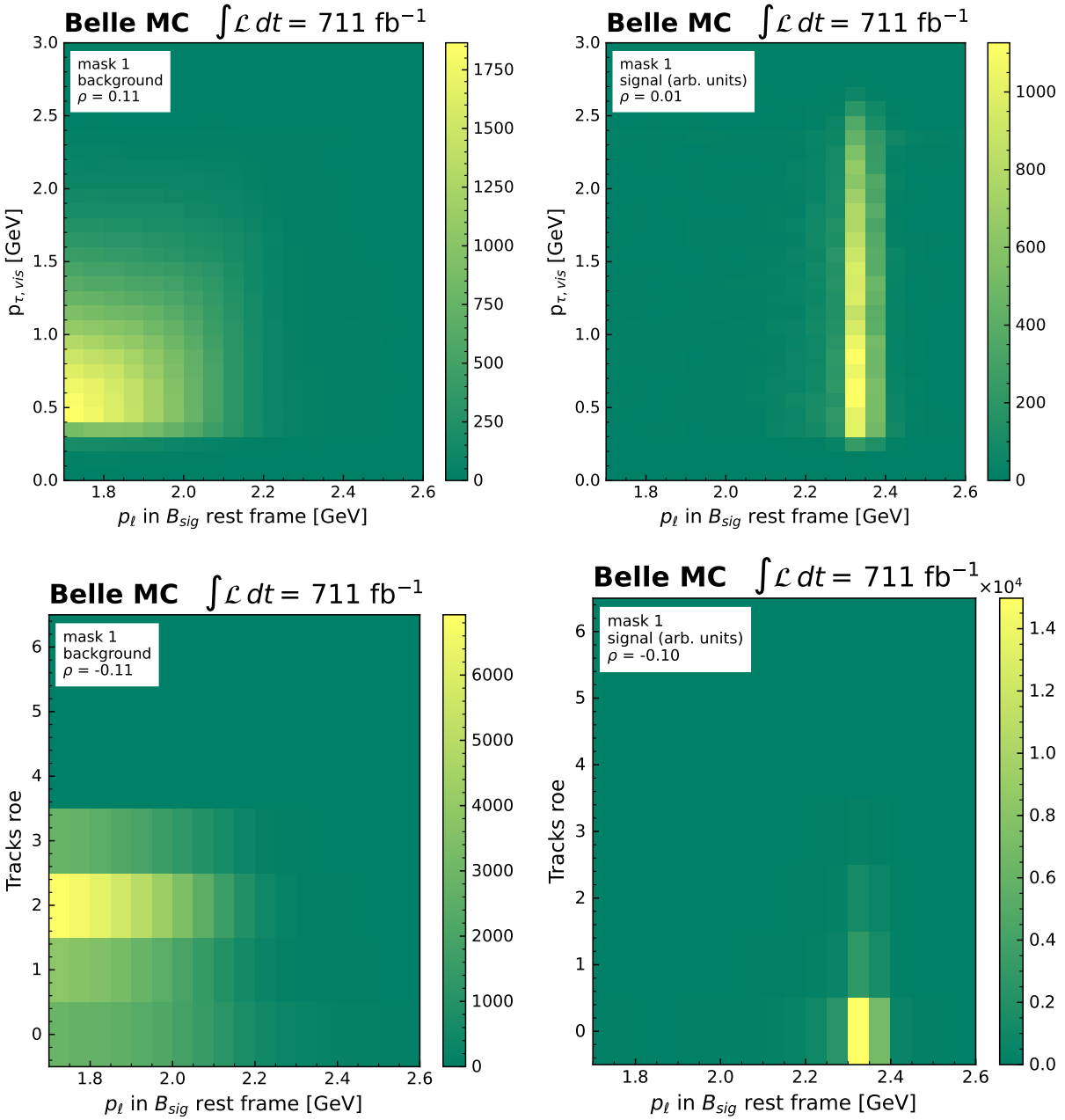

Figure 5.20: Two-dimensional histograms of the lepton momentum in the B_{sig} rest frame and the visible momentum of the τ daughters (top) in the B_{sig} rest frame and the number tracks (bottom) used in the reconstruction of the rest of the event for background Monte Carlo events on the left and signal events on the right.

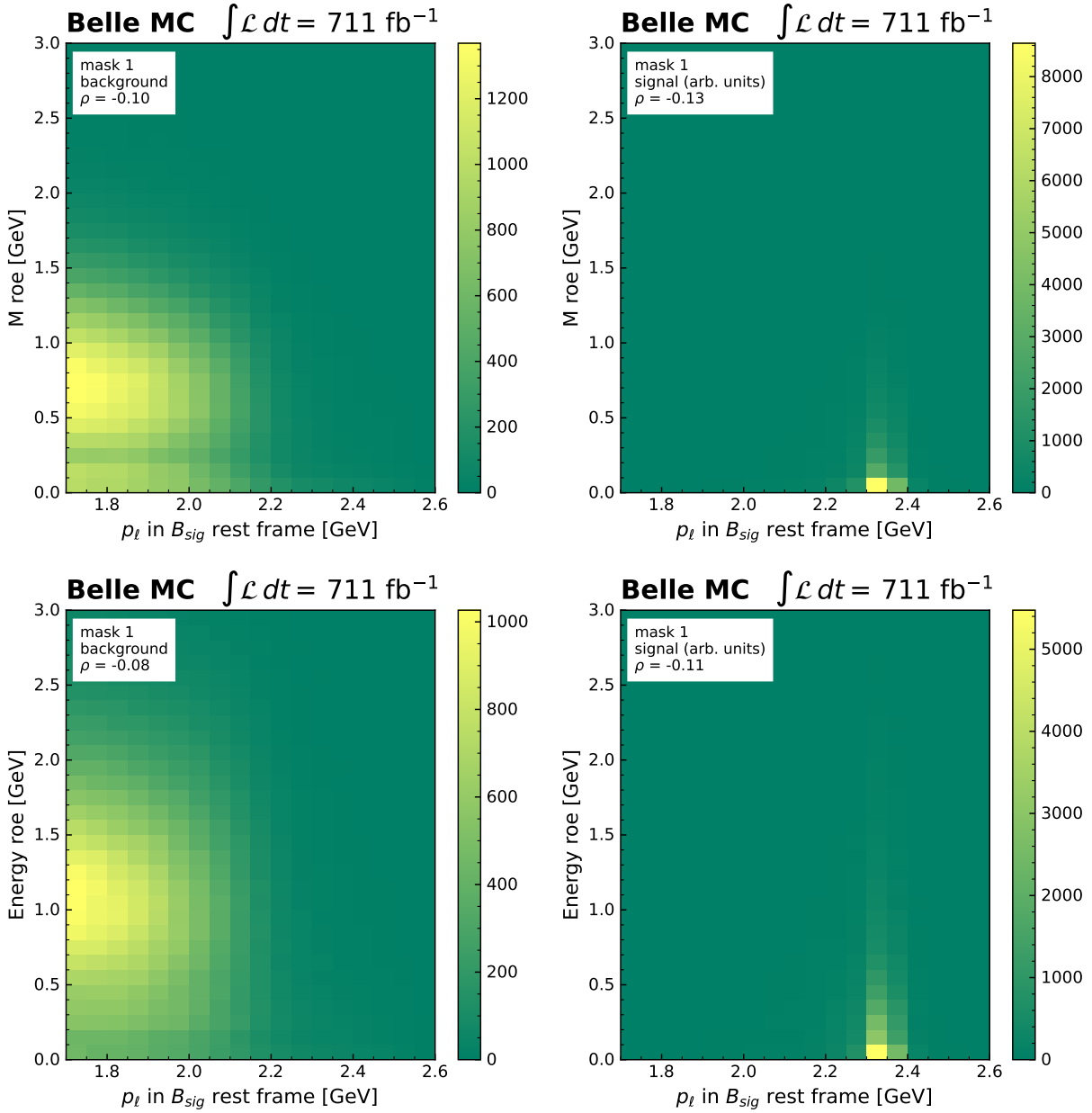

Figure 5.21: Two-dimensional histograms of the lepton momentum in the B_{sig} rest frame and the reconstructed invariant mass (top) and the total energy (bottom) of the rest of the event for background Monte Carlo events on the left and signal events on the right.

B Meson Background Suppression Training and Testing Data Sets

The Monte Carlo data of signal and background events are split into training and testing data for the training of the B meson background suppression. All events in the training data are discarded afterward to avoid overfitting and biasing the final result. Tab. 5.18 lists the signal and background events in the training and testing data. In total, 32 B meson background suppressions are trained, 16 for each rest of the event mask.

MC type	training data p_ℓ reg.		testing data p_ℓ reg.	
	[1.7,3.0] GeV	[2.0,2.5] GeV	[1.7,3.0] GeV	[2.0,2.5] GeV
signal	18361.7	17906.7	18288.0	17809.8
signal true B_{sig}	9139.7	9011.8	9143.1	8999.2
background	108921.2	25003.9	107917.3	24716.9
charm	1.49%	1.73%	1.27%	1.57%
uds	2.14%	3.47%	2.20%	3.48%
$B^0(b \to c)$	76.19%	70.23%	76.17%	70.41%
$B^+(b \to c)$	15.45%	13.38%	15.61%	13.32%
$B^0(b \to u\ell\nu)$	4.01%	9.59%	4.03%	9.62%
$B^+(b \to u\ell\nu)$	0.73%	1.60%	0.72%	1.59%

Table 5.14: Number of signal and background events with background events in the training and test data for roe mask 1 with two different lepton momentum regions for the training. The background events are weighted to their expectation for a luminosity of 711 fb^{-1}. Individual background contributions a given relative to the total number of background events.

Two different signal data sets were used in the supppressions. One version included all signal events, regardless of whether the reconstruction was correct or not, while the other version only consisted of signal events with correctly reconstructed B_{sig}. This approach was chosen because distributions of the rest of the event and signal variables vary for correctly and wrongly reconstructed B_{sig} candidates.

As an example, the distribution of the energy of neutral ECL cluster in the rest of the event for wrongly and correctly reconstructed B_{sig} candidates is shown in Fig. 5.22. Most of the B_{sig} candidates are incorrectly reconstructed because of the τ and not the lepton from the B_{sig}. Since the τ reconstruction does not necessarily impact the lepton momentum, the lepton momentum in the B_{sig} rest frame is similar for correctly and wrongly reconstructed B_{sig} candidates.

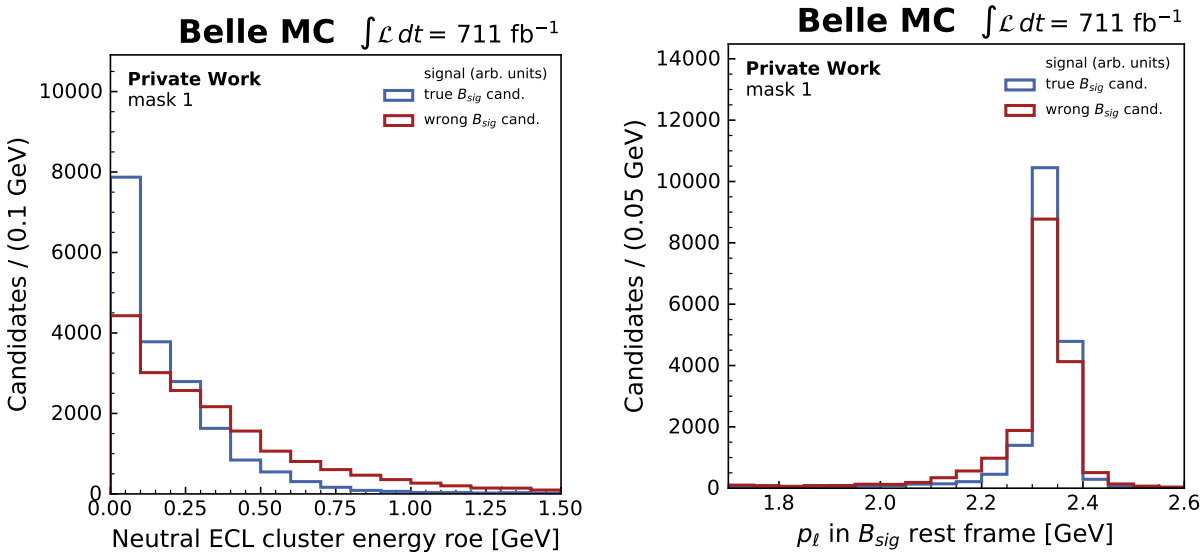

Figure 5.22: The neutral ECL cluster energy of the rest-of-the-event candidate (left) and the lepton momentum in the B_{sig} rest frame (right) for signal events with correctly reconstructed B_{sig} candidates in blue and wrongly reconstructed B_{sig} candidates in red.

Rare B meson decays are excluded from the training data due to their negligible contribution. All background contributions are weighted to their expected number of events for a luminosity of 711 fb^{-1} due to the varying number of streams available for the individual background contributions. The contributions of the different background events are given as fractions relative to the total number of background events.

Two training regions for the lepton momentum in the B_{sig} rest frame [1.7,3.0] GeV and [2.0,2.5] GeV are chosen.

The hyperparameters of the B meson background suppression are given in Tab. 5.15.

Hyperparameter	Value
Number of Trees	200
Depth	3
Learning rate	0.1
Random sub-sample per tree	0.5

Table 5.15: Hyperparameters of the B meson background suppression training.

5.3.6 Determination of Best B Meson Background Suppression Selection

For each of the 32 B meson background suppressions, the best selection on the classifier output is determined from the fit of the lepton momentum distribution in the B_{sig} rest frame in the region [1.7,2.6] GeV, separately for the $B^0 \to \tau e$ and $B^0 \to \tau \mu$ channel. The fit is a simplified version with fewer nuisance parameters than the final fit, which is described in chapter 7.

The expected upper limits on the signal branching ratio obtained on Asimov data with zero signal events are determined for selections on the classifier output > cut value. The cut values vary between 0.100 and 0.990 in steps of 0.005. For each selection, the fitting process starts with generating two background PDFs and one signal PDF. To reduce the time of the fit, only the three yield parameters are included in the model PDF, which still gives a good estimate of the expected upper limit. Fig. 5.23 shows the best upper limits depending on the classifier output selection for the B meson suppression chosen in the final fit.

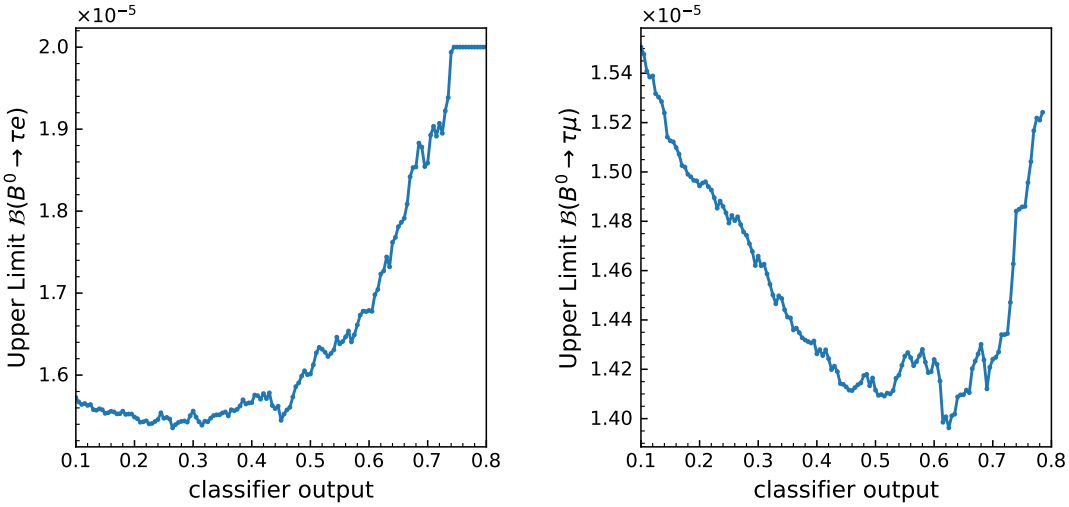

Figure 5.23: Expected upper limits on the signal branching ratio for various selections on the B meson background suppression classifier output. Only the suppressions chosen in the final fit for the $B^0 \to \tau e$ (left) and $B^0 \to \tau \mu$ (right) are shown.

The plateau in the upper limit on the branching ratio distribution for the $B^0 \to \tau e$ channels is maximal value of 0.3 for the signal strength in the upper limit scan. Afterward, the systematic uncertainties due to the limited size of the Monte Carlo data used for the PDF generation are included in the fit as nuisance parameters. Starting from the selection on classifier output with the best upper limit on the branching ratio, the fit is repeated. If the

fit converges with the additional systematic uncertainties, the selection is chosen; otherwise, the same procedure is repeated with the next best selection until the fit converges.

Importance of the Training Variables in the B Meson Background Suppression

The importance of each variable for the B meson background suppression used in the final fit is given in Tab. 5.16 and 5.17. The most important variable has an importance of 100 and the least important has an importance of 0.

Training Variable	Importance
number of tracks roe	100
ΔE_τ in τ rest frame	92
energy of roe	85
visible τ momentum in B_{sig} rest frame	43
τ decay modes	30
transverse momentum of roe	3
energy of neutrals in ECL of roe	2
number of π^0 in roe	0

Table 5.16: Importance of the signal-side and rest-of-the-event (roe) training variables in the B meson background suppression used in the final fit of the $B^0 \to \tau e$ channel.

Training Variable	Importance
energy of roe	100
ΔE_τ in τ rest frame	50
visible τ momentum in B_{sig} rest frame	45
missing momentum of event	41
number of tracks roe	20
invariant mass roe	10
transverse momentum roe	5
energy in ECL of roe	5
momentum of roe	3
number of ECL clusters in roe	2
energy of neutrals in ECL of roe	1
charge roe	1
number of π^0 in roe	0
number of ECL clusters in roe	0

Table 5.17: Importance of the signal-side and rest-of-the-event (roe) training variables in the B meson background suppression used in the final fit of the $B^0 \to \tau\mu$ channel.

Data - Monte Carlo Comparison of the Training Variables

Since this analysis is blinded, the agreement between real and Monte Carlo data is examined in the sideband region of the lepton momentum ($p_\ell < 2.2$ GeV).

Fig. 5.24 - 5.27 compare the distribution of the training variables between Monte Carlo and the sideband region of the real data. Most of them are in good agreement, apart from an overall scaling factor, which originates from the FEI calibration factors. These factors are only correct for events with the same composition of B_{tag} candidates as the signal events in the calibration study. Because the background yields are a free parameter in the final fit and applying the B meson background selection also alters the composition of the background events, no rescaling of the background Monte Carlo events is applied.

Some variables, like the number of neutral ECL clusters and the invariant mass of the rest of the event, exhibit visible discrepancies in addition to the overall scaling difference. In the following section, their impact on the shape of the lepton momentum distribution and on the signal efficiency is discussed.

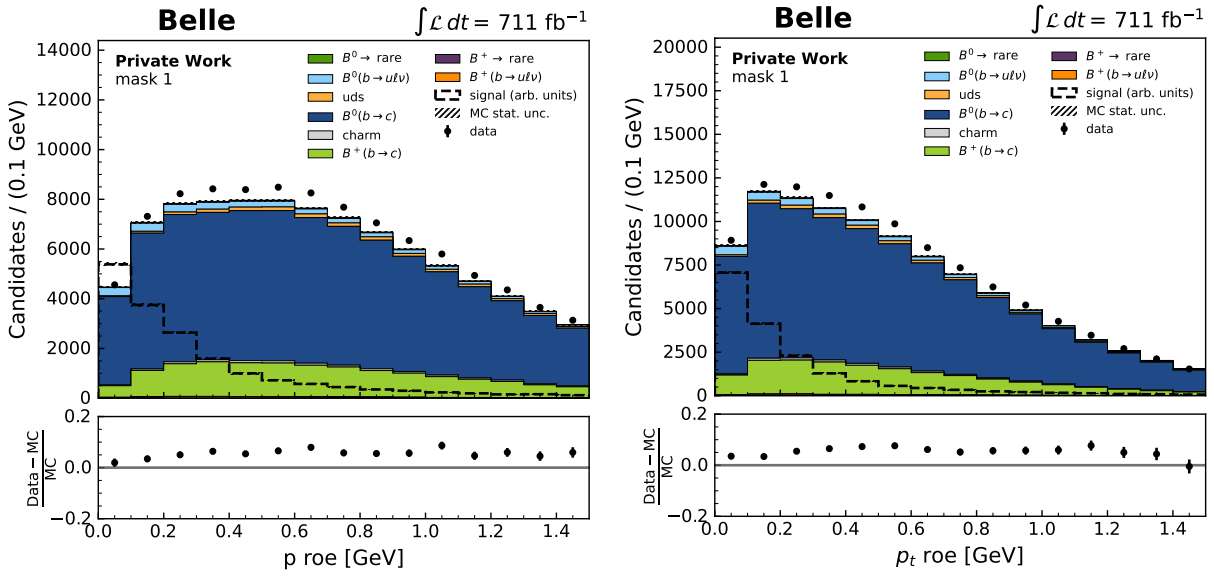

Figure 5.24: Rest-of-the-event (roe) training variables for mask 1 in the sideband region of the lepton momentum shown for signal (dashed black line) and background (colored stacked components) Monte Carlo, and real data (black points).

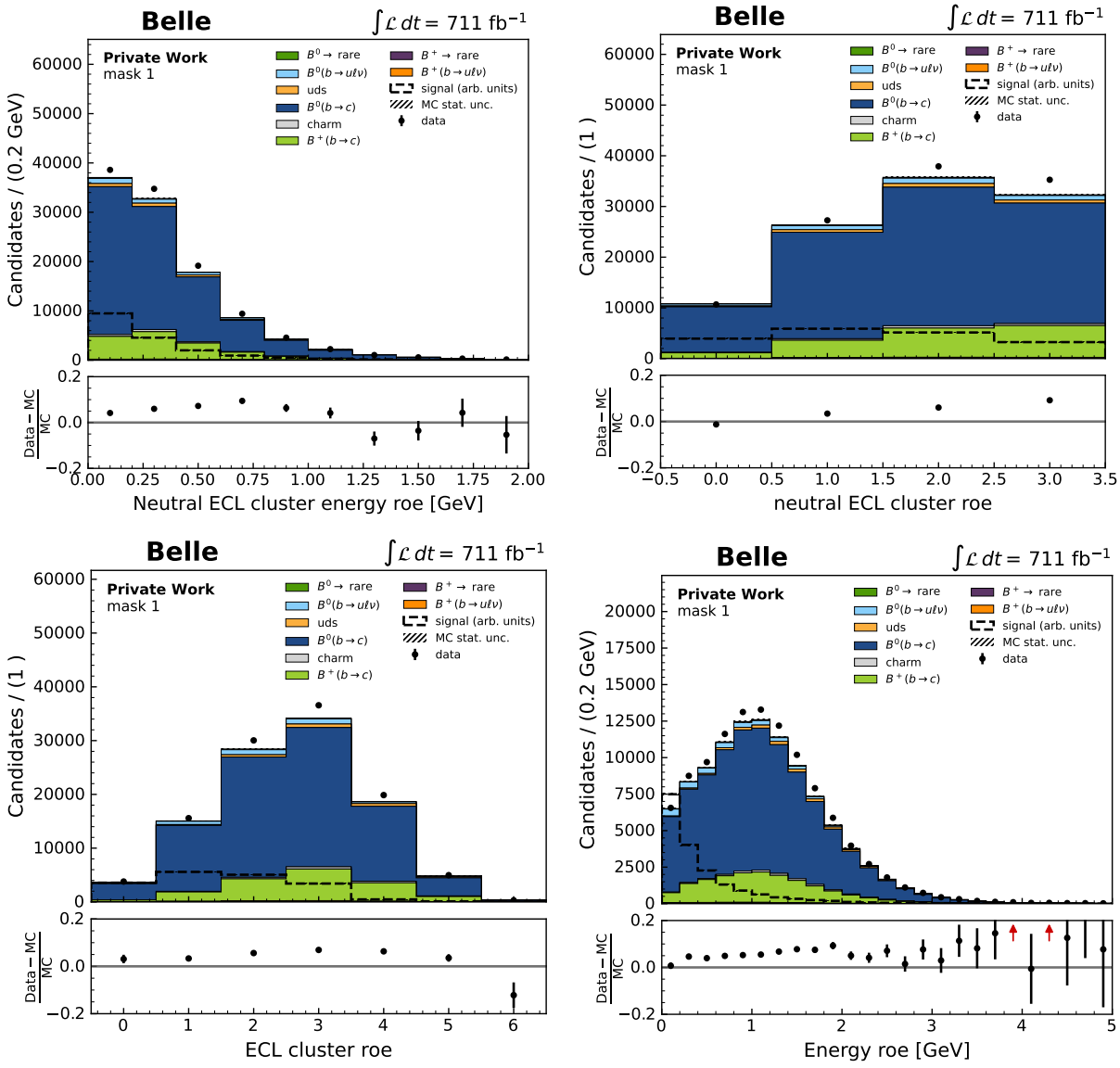

Figure 5.25: Rest-of-the-event (roe) training variables for mask 1 in the sideband region of the lepton momentum shown for signal (dashed black line) and background (colored stacked components) Monte Carlo, and real data (black points).

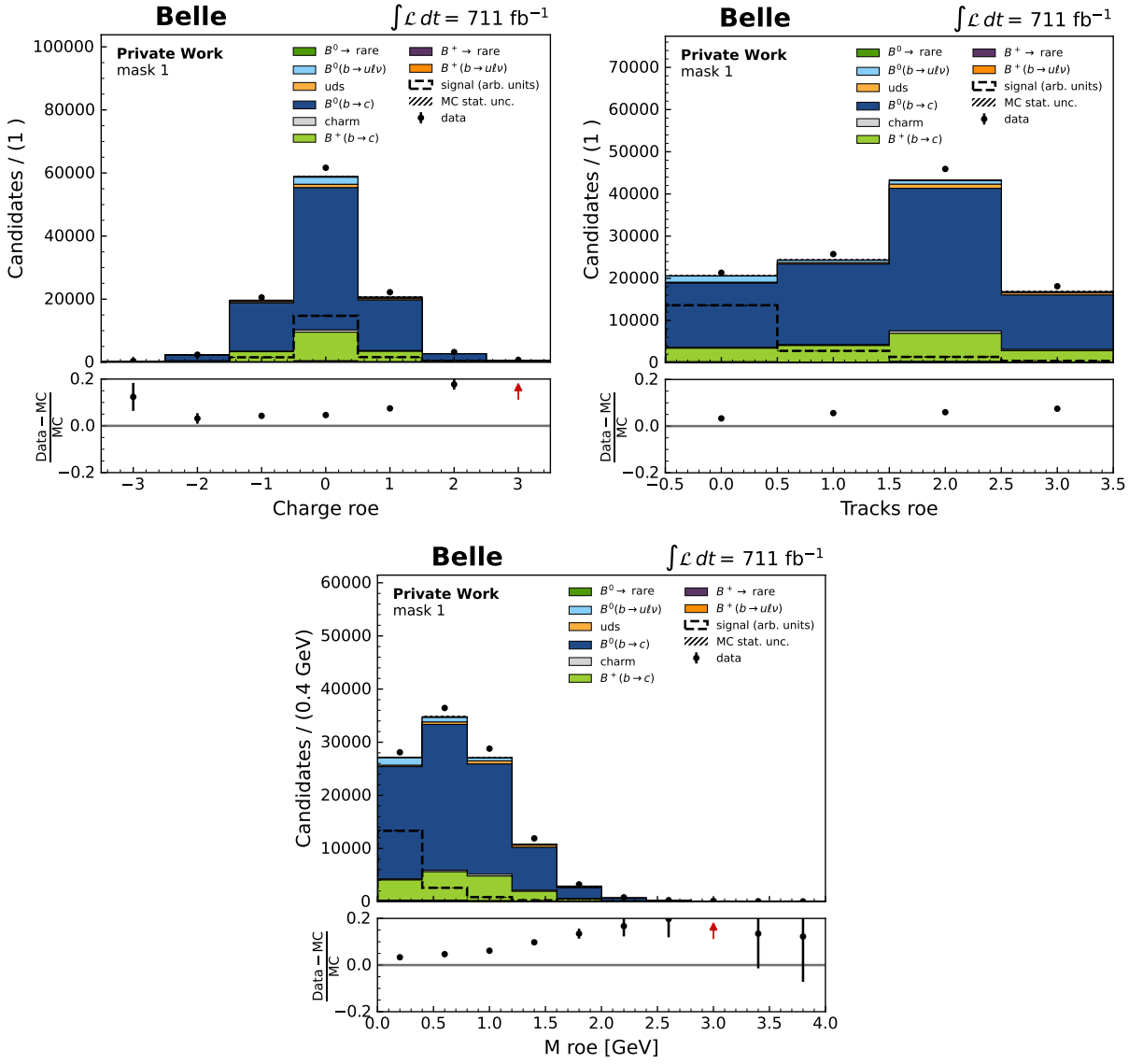

Figure 5.26: Rest-of-the-event (roe) training variables for mask 1 in the sideband region of the lepton momentum shown for signal (dashed black line) and background (colored stacked components) Monte Carlo, and real data (black points).

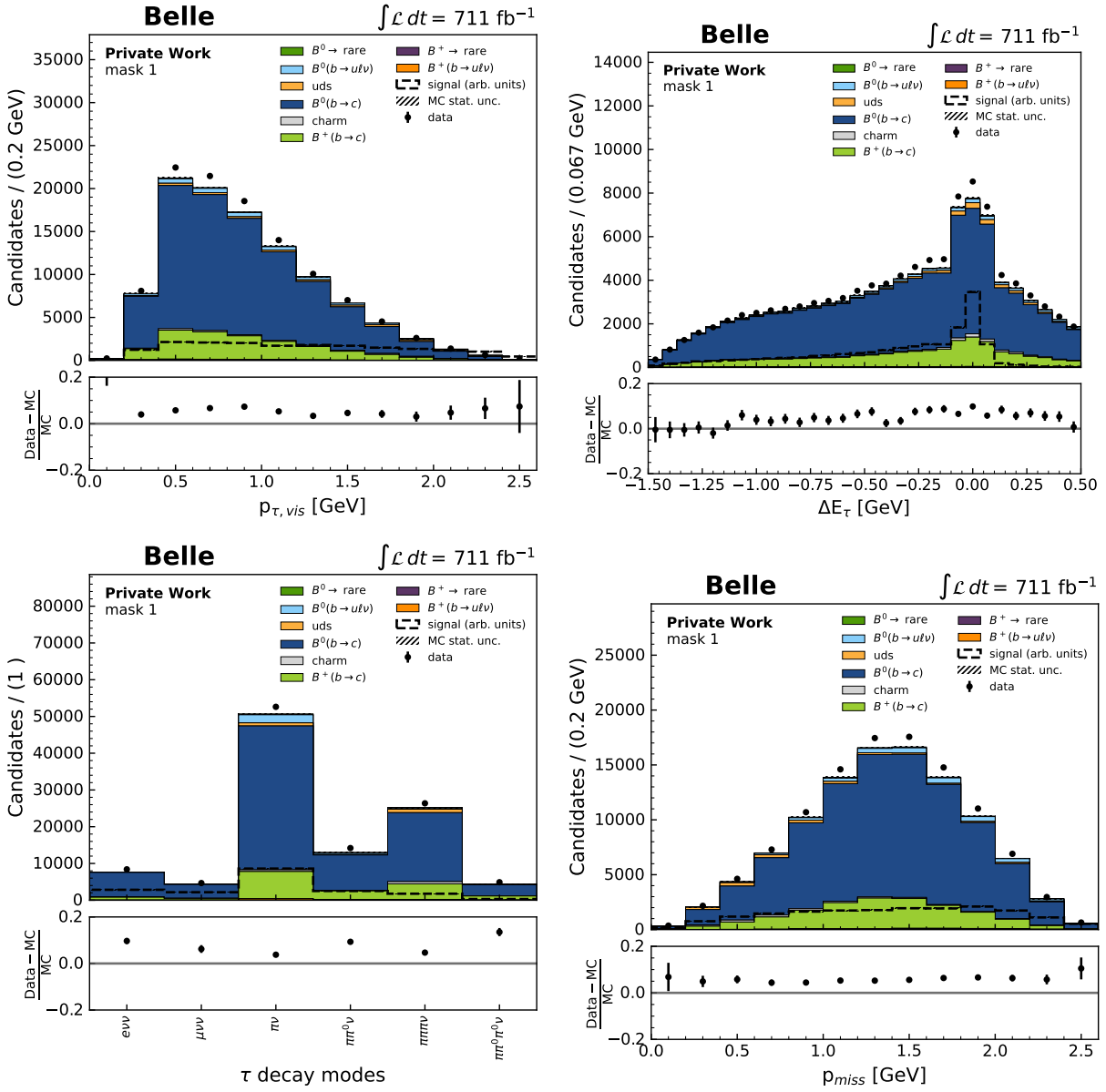

Figure 5.27: Signal-side training variables for mask 1 in the sideband region of the lepton momentum shown for signal (dashed black line) and background (colored stacked components) Monte Carlo, and real data (black points).

Data - Monte Carlo Reweighting

The observed discrepancies in the sideband region between real and background Monte Carlo data in the B meson background suppression variables can result in shape differences in the lepton momentum distribution. To determine the systematic uncertainty these discrepancies might introduce, an additional BDT for each B meson background suppression is trained with the same set of training variables (Fig. 5.19), classifying events as Monte Carlo (0) or real data events (1).

In the case of perfect agreement between real and background Monte Carlo data, the classifier output would peak at 0.5, indicating that no distinction is possible. Based on the classifier output c_{bdt}, the background Monte Carlo is reweighted to match the real data. The weight is defined as:

$$w = \frac{c_{bdt}}{1 - c_{bdt}}. \tag{5.15}$$

In the final fit, a systematic uncertainty on the signal and background shapes is assigned based on the difference in the lepton momentum between nominal and reweighted Monte Carlo events. Events in the sideband region of the lepton momentum between 1.7 GeV - 2.2 GeV in the B_{sig} rest frame are used in the training of the reweighting BDTs. Assuming the distribution of the training variables in the sideband region is identical to their distribution in the signal region, the impact of the shape differences obtained in the sideband region can be extrapolated to the signal region.

The full real data is split equally into training and test data, see Tab. 5.18.

Event Type	training data	test data
real data	55357	55736
background Monte Carlo	105771.0	104873.8
charm	1.41 %	1.18 %
uds	1.84 %	1.92 %
$B^0(b \to c)$	77.31 %	77.28 %
$B^+(b \to c)$	15.63 %	15.78 %
$B^0 \to$ rare	0.24 %	0.24 %
$B^+ \to$ rare	0.05 %	0.05 %
$B^0(b \to u\ell\nu)$	2.95 %	2.97 %
$B^+(b \to u\ell\nu)$	0.57 %	0.57 %

Table 5.18: Number of real data and weighted background Monte Carlo events scaled to the luminosity of the real data for the masks 1 and 2 used in the training of the reweighting BDTs.

The training and test data of the background Monte Carlo events consists of the same streams as the training and test data of the B meson background suppression. None of the real data events are excluded from the further analysis because the training result is only used to determine a systematic uncertainty for the fit.
Fig. 5.28 shows the distributions of the reweighting BDT classifier outputs of the respective B meson background suppressions in the final fit.

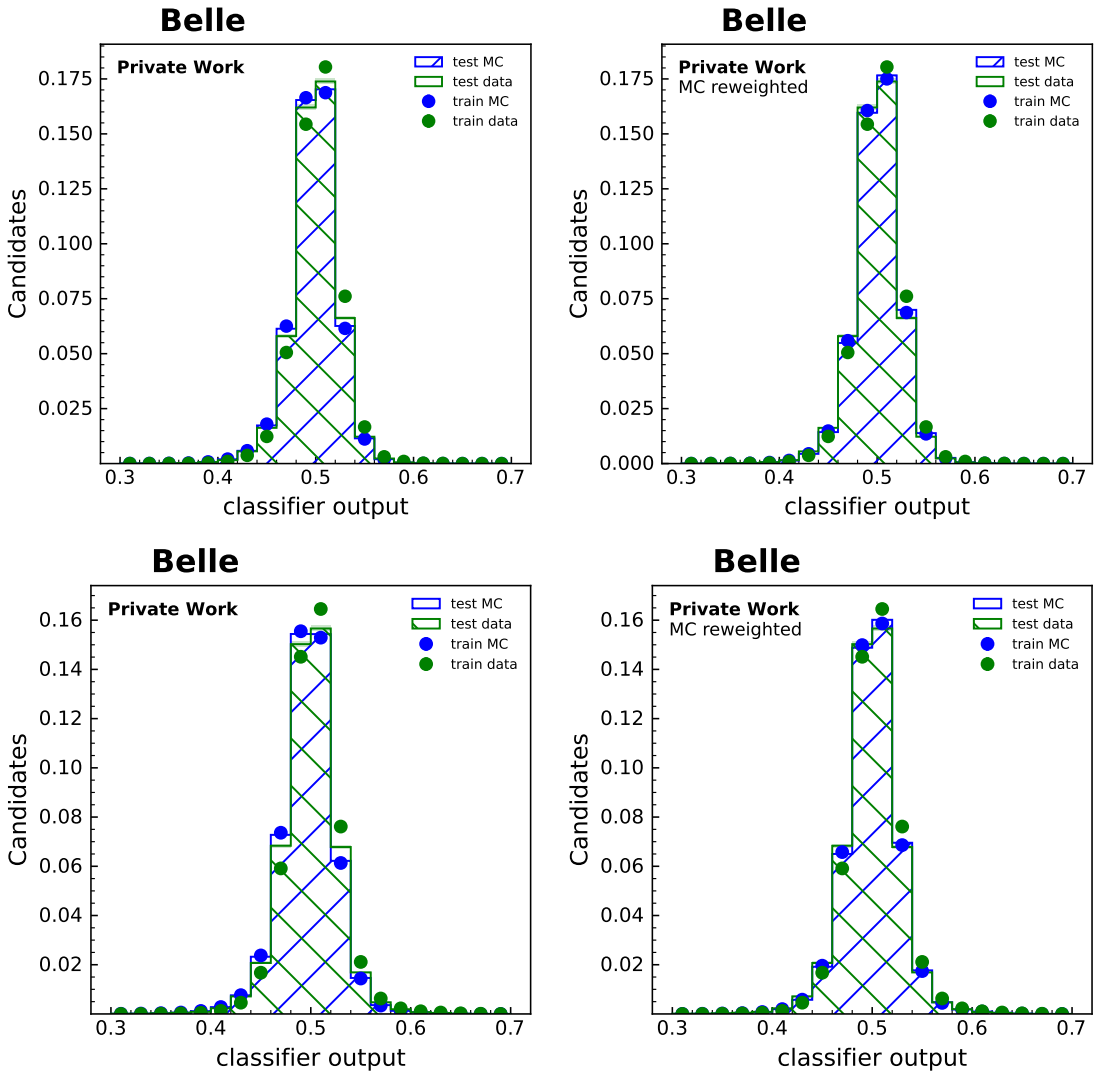

Figure 5.28: Classifier output of reweighting BDT before (left) and after (right) reweighting of the Monte Carlo events in the $B^0 \to \tau e$ channel (upper plots) and in the $B^0 \to \tau\mu$ channel (lower plots).

As expected, the classifier outputs are distributed around 0.5, which indicates an overall good agreement between real and background Monte Carlo data with a slight shift of the real data events toward higher classifier output values. The distributions on the left show the classifier output determined for the reweighting BDT, and the distributions on the right show the classifier output distribution after the test Monte Carlo data is reweighted.

The train and test data sets agree for Monte Carlo events but show some discrepancies in the real data. Since the real data is distributed evenly between the training and test data sets, reweighting the Monte Carlo data is the mean between the two distributions.

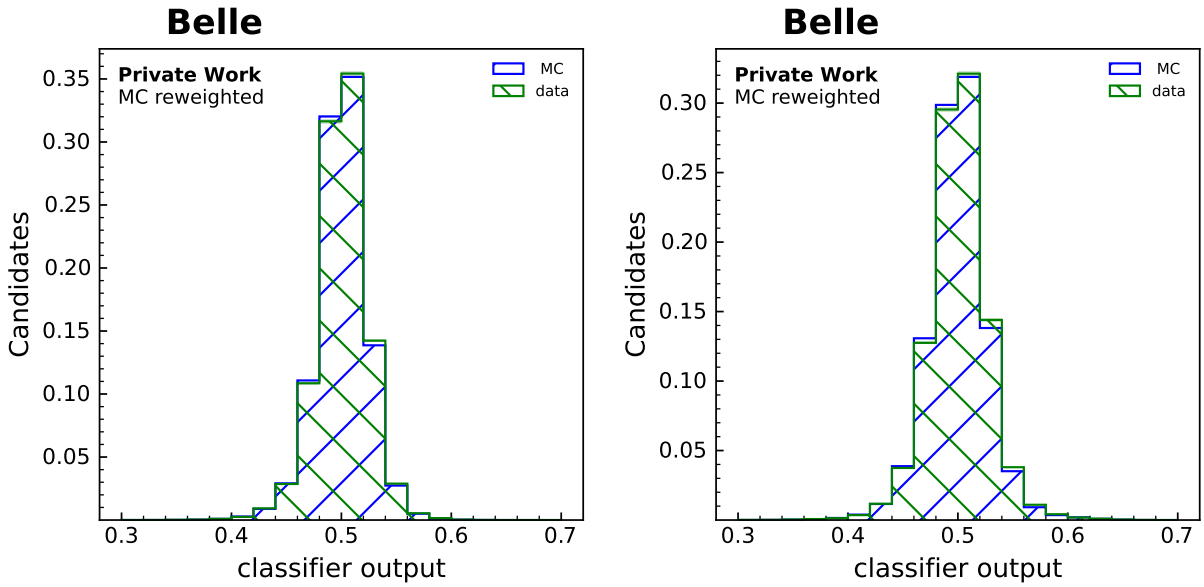

Figure 5.29: Classifier output of reweighting BDT before (left) and after (right) reweighting of the Monte Carlo events in the $B^0 \to \tau e$ channel (left) and in the $B^0 \to \tau \mu$ channel (right). The combined test and training Monte Carlo data are represented by the MC component, and the data component contains real data events in the training and test data sets.

Signal Efficiency Uncertainties due to the B Meson Background Suppression

The Monte Carlo signal efficiency for the selection on the B meson background suppression classifier output is defined as:

$$\epsilon_{\text{BDT}} = \frac{\text{N}_{\text{sig, after selection}}}{\text{N}_{\text{sig, before selection}}}. \tag{5.16}$$

Data - Monte Carlo discrepancies for signal events in the training variables can result in a different classifier output distribution, and thus, in a different signal efficiency in real data. Ideally, the uncertainty of the signal efficiency would be determined in the control channel with a similar topology. However, the various τ decays and the signal-side specific variables in the B meson background suppressions prevent finding one suitable control channel.

Therefore, the uncertainty on the signal efficiency, caused by differences in the detector simulation and the real detector, is determined with a reweighting technique relying on background Monte Carlo and real data in the signal sideband region, $p_\ell \in$ [1.7,2.2] GeV. First, every training variable is binned, including an under and overflow bin. In the next step, a weight for each bin is calculated, so that the data - Monte Carlo difference is the same for all bins while conserving the overall number of background Monte Carlo events. These weights are then used to reweight the signal Monte Carlo events.

Fig. 5.30 presents the reweighting procedure for one training variable, the energy in the rest of the event. The distribution on the left shows the nominal distribution, and the distribution on the right, the background and the signal Monte Carlo events are reweighted.

In the left histogram of Fig. 5.31, the distribution of the rest-of-the-event energy is shown for the nominal signal events in blue, and after the reweighting procedure in red. The change in the B meson background suppression classifier output distribution is shown in the histogram on the right.

The uncertainty on the signal efficiency is calculated by the differences of the efficiencies for the nominal and reweighted signal Monte Carlo events for each variable i:

$$\sigma_{\epsilon_i,\text{BDT}} = |\epsilon_{\text{BDT}} - \epsilon_{\text{BDT, reweighted i}}|. \tag{5.17}$$

Since some of the variables are correlated, a conservative approach for the total uncertainty on the B meson background suppression selection is determined via:

$$\sigma_{\epsilon,\text{BDT}} = \sum_i |\epsilon_{\text{BDT}} - \epsilon_{\text{BDT, reweighted i}}|. \tag{5.18}$$

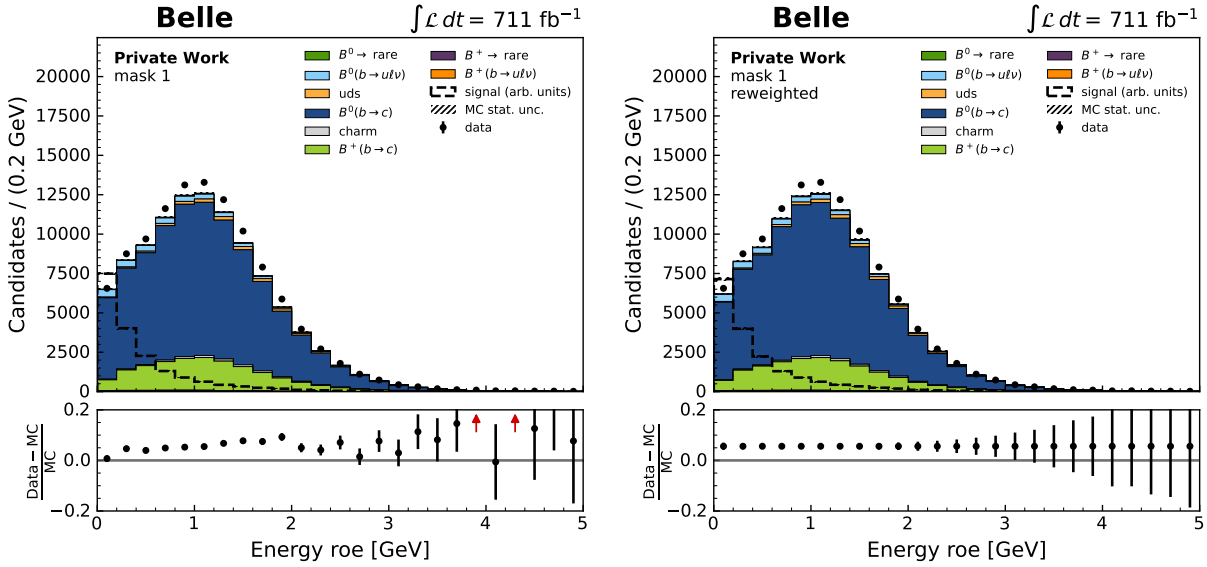

Figure 5.30: The energy of the rest of the event (roe) before (left) and after (right) reweighing the Monte Carlo events to the real data.

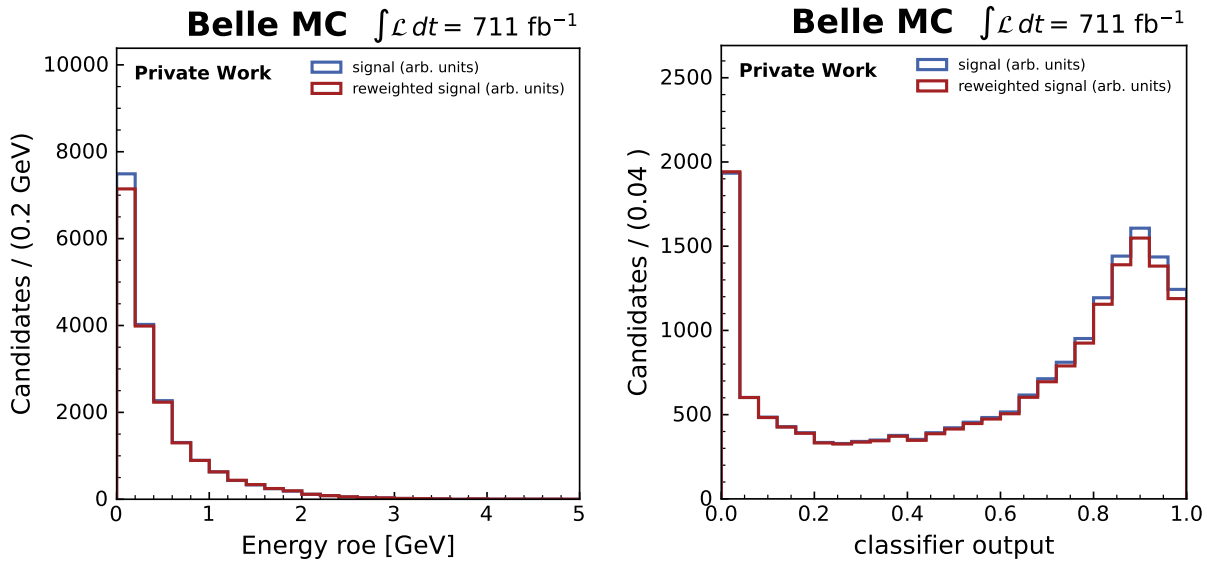

Figure 5.31: The left histogram shows the energy in the rest of the event for the nominal (blue) and reweighted (red) signal events. The right histogram shows the change in the classifier output distribution of the B meson background suppression for the nominal signal and the reweighted signal events based on the data - Monte Carlo discrepancy in the rest-of-the-event energy variable.

Chapter 6

Calibration of the Full Event Interpretation

This chapter describes the calibration of the Full Event Interpretation (FEI). Because the FEI is trained on Monte Carlo events, the difference in performance between Monte Carlo and real data needs to be determined. Therefore, a well-known and clean channel for the signal side is chosen, and the tag side is reconstructed using the hadronic FEI. Assuming that the signal and tag side are independent, all differences after applying the Monte Carlo correction to the signal side are due to the tag-side reconstruction. The total signal efficiency can be expressed as the product of the signal and tag side for both Monte Carlo and real data.

$$\epsilon = \epsilon_{\mathrm{B_{sig}}} \cdot \epsilon_{\mathrm{B_{tag}}} \tag{6.1}$$

The calibration is performed for events with a $\mathrm{B}^0 \to \mathrm{D}^{*-}(\to \mathrm{D}^0(\to K^+\pi^-)\pi^-)\ell^+\nu$ decay, where $\ell = \mathrm{e}, \mu$. This signal channel has the advantage of being nearly background-free with a branching ratio of 0.1367 ± 0.0032 %. Tab. 6.1 lists the individual branching ratios of the sub-decays. Only the continuum and B meson events with two $b \to c$ transitions are studied because the contributions from $\mathrm{B}(b \to u\ell\nu)$ and $\mathrm{B} \to$ rare events are negligible.

	$\mathcal{B}$
$\mathcal{B}$ $(\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu)$	5.1137 ± 0.1082 %
$\mathcal{B}$ $(\mathrm{D}^{*-} \to \mathrm{D}^0\pi^-)$	67.7 ± 0.5 %
$\mathcal{B}$ $(\mathrm{D}^0 \to K^+\pi^-)$	3.947 ± 0.030 %
product $\mathcal{B}$	**0.1367 ± 0.0032 %**

Table 6.1: Branching ratios of the calibration decay, with the isospin-averaged $\mathcal{B}(\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu)$ and the branching ratios of the subsequent decays according to the PDG [18].

Fig. 6.1 illustrates an example of an $\Upsilon(4\mathrm{S})$ resonance signal decay.

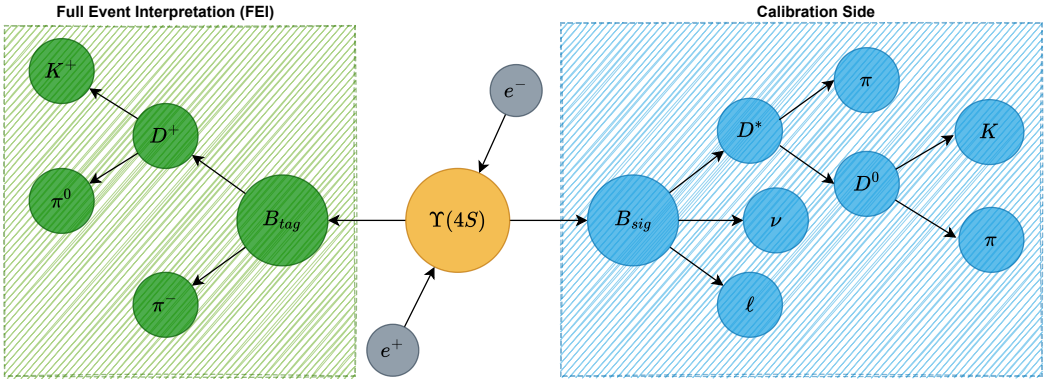

Figure 6.1: Sketch of an $\Upsilon(4S)$ decay used for the calibration of the FEI.

6.1 Reconstruction of $\Upsilon(4S)$

To avoid introducing differences in the B_{tag} composition between the calibration and signal channel, the same selections are applied when possible.

6.1.1 Event Selection

The event selection on the number of tracks, number of photons, and high energetic leptons in the event are the same as described in section 5.2 with a looser condition on the existence of lepton with momentum $p_{CMS} > 0.3$ GeV instead of $p_{CMS} > 0.5$ GeV. Tab. 6.2 lists the efficiencies for $B^0 \to D^{*-}\ell^+\nu$ events, where at least one of the two B mesons is generated in a $B^0 \to D^{*-}\ell^+\nu$ decay and for background events. The "other B^0 $(b \to c)$" component contains events where none of the two neutral B mesons decays in the $B^0 \to D^{*-}\ell^+\nu$ channel.

	$N_{tracks} < 18$	$N_\gamma < 18$	$p_{\ell,CMS} > 0.3$ GeV	all criteria
ϵ $B^0 \to D^{*-}\ell^+\nu$	99.99%	99.39%	78.18%	77.76%
ϵ other $B^0(b \to c)$	99.89%	96.91%	46.58%	45.88%
ϵ $B^+(b \to c)$	99.93%	96.22%	46.66%	45.86%
ϵ charm	100.00%	99.12%	24.80%	24.66%
ϵ uds	100.00%	99.59%	13.09%	13.04%

Table 6.2: Efficiency of the individual event selection on the number of tracks and photon, at least one lepton with leptonID > 0.9 and a CMS momentum > 0.3 GeV and the combination of all three selections for neutral B, charged B meson and continuum events. The $B^0 \to D^{*-}\ell^+\nu$ contains all events, where at least one of the B meson decays in the $B^0 \to D^{*-}\ell^+\nu$ channel.

6.1.2 B_{tag} Reconstruction

The B_{tag} is reconstructed with the hadronic FEI and must fulfill the same requirements described in section 5.2.2 of the $B^0 \to \tau\ell$ decays.

6.1.3 Signal Side Reconstruction

Tracks and clusters not used in the reconstruction of the B_{tag} are considered in the B_{sig} reconstruction.

Final State Particles

Each charged final state particle must pass the following track criteria $dr < 2$ cm and $|dz| < 4$ cm, to ensure their origin from the IP. Additionally, all charged particles apart from the charged pion from the D^* must satisfy a particle identification selection to increase the number of correctly identified particles and satisfy a selection on their momentum. These additional requirements are listed in Tab. 6.3.

	Selection Criteria
Electron	eID > 0.9 & $p_{CMS} > 0.3$ GeV
Muon	muID > 0.9 & $p_{CMS} > 0.3$ GeV
Charged pions from D^*	no additional selection
Charged pions from D^0	pion vs. kaon ID > 0.1 and $p_t > 0.1$ GeV
Kaon	pion vs. kaon ID < 0.9 and $p_t > 0.1$ GeV

Table 6.3: Selection criteria for the charged final state particles.

Because the charged pion originating from the D^* decay has a low momentum and does not necessarily reach all particle identification subdetectors, its particle identification is limited. Therefore, no particle identification selection is applied to those pions. The resulting purity for these pions is 12.4%.

Fig. 6.2 illustrates the composition of Monte Carlo particles for the different final state particle types. The leptons have a high purity of over 90%. The main contributors a fake muons are pions due to their similar mass. The pions and kaons show a fake rate between 30 - 40%, which decreases in the further reconstruction steps. Analogously to the electrons in the $B^0 \to \tau\ell$ analysis, the electrons are Bremsstrahlung-corrected.

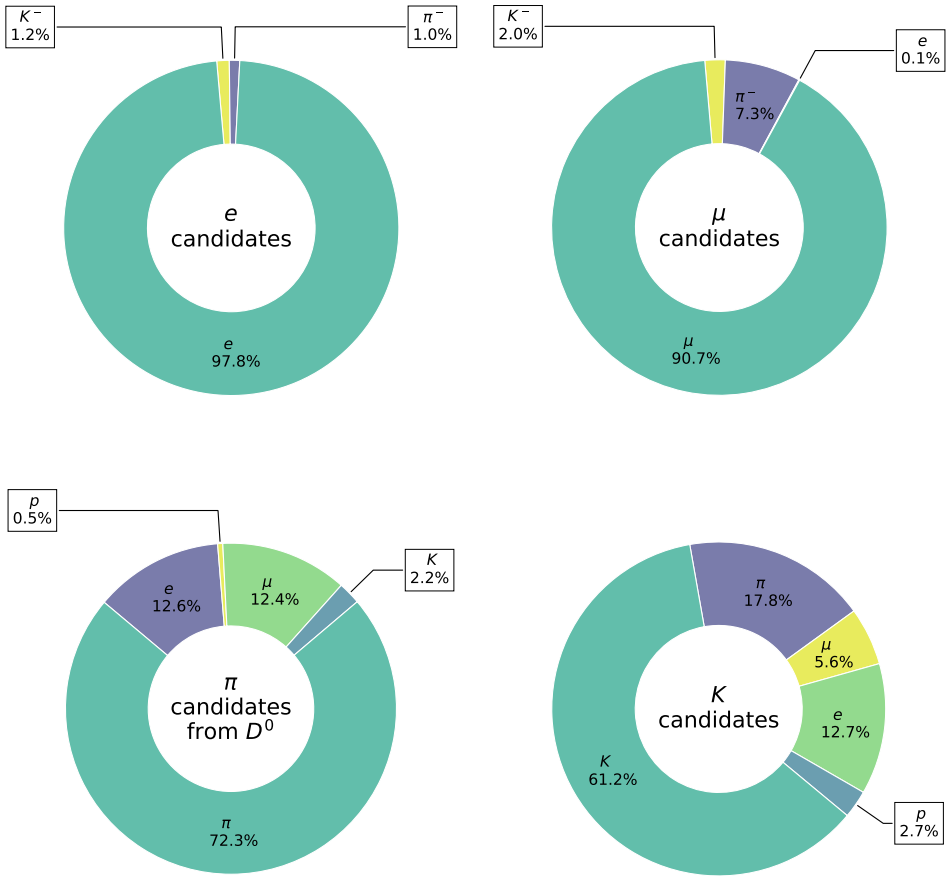

Figure 6.2: Composition of Monte Carlo particles in the different final state particles in events with at least one $B^0 \to D^{*-}\ell^+\nu$ decay.

D^0 Reconstruction

A D^0 candidate is formed from a pion and kaon candidate. The difference between the nominal and reconstructed D^0 mass serves as a good discriminator between correctly and wrongly reconstructed D^0 candidates, therefore a $|M_{rec} - M_{nom}| < 0.02$ GeV selection for the reconstructed D^0 is applied.

Fig. 6.3 compares the mass difference between correctly and wrongly reconstructed B_{sig} candidates in events with least one $B^0 \to D^{*-}\ell^+\nu$ decay. Wrongly reconstructed candidates can either be from events where the subsequent decay of the D^{*-} in the $B^0 \to D^{*-}\ell^+\nu$ decay is not generated in a complete signal decay chain or events where decay products are not assigned to the correct position in the decay chain.

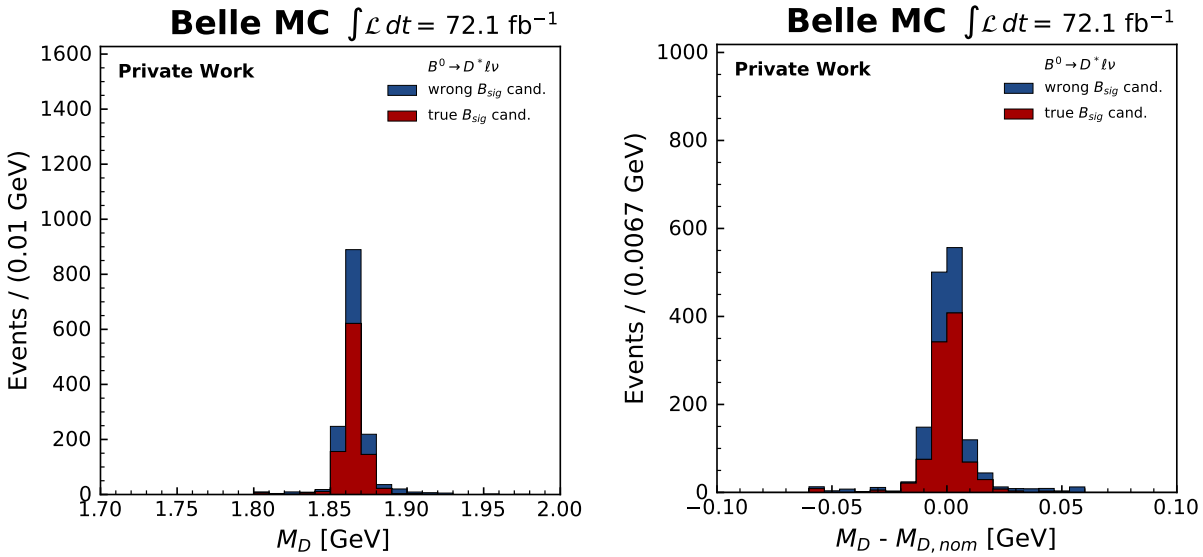

Figure 6.3: Invariant mass of the D^0 and its difference to the nominal mass in events with at least one $B^0 \to D^{*-}\ell^+\nu$ decay, shown for a subset of Monte Carlo data corresponding to the luminosity of 72 fb^{-1}. The red component represents the $\Upsilon(4S)$, where the B_{sig} was correctly reconstructed in the signal channel, and the blue component shows the B_{sig} candidates that were wrongly reconstructed.

To suppress ee$\to c\bar{c}$ background, all D candidates with a momentum larger than 2.5 GeV in the CMS frame are discarded because the D meson originating from B meson has a lower momentum than a D meson produced directly in the collision.

D^* Reconstruction

A D^* candidate is formed by combining a D^0 with a π^- candidate. The phase-space of the D^* decay is small because the D^0 mass is very close to the sum of the D^* and pion masses and their invariant mass difference is used to suppress background contributions. Fig. 6.4 presents the distributions of the invariant mass of the D^* and the invariant mass difference between the D^* and D^0. For correctly reconstructed D^0 and D^* candidates with the nominal masses of 1.864 GeV and 2.010 GeV [18], the invariant mass difference should be distributed around 0.146 GeV. Therefore, the mass difference is required to be between 0.142 and 0.148 GeV.

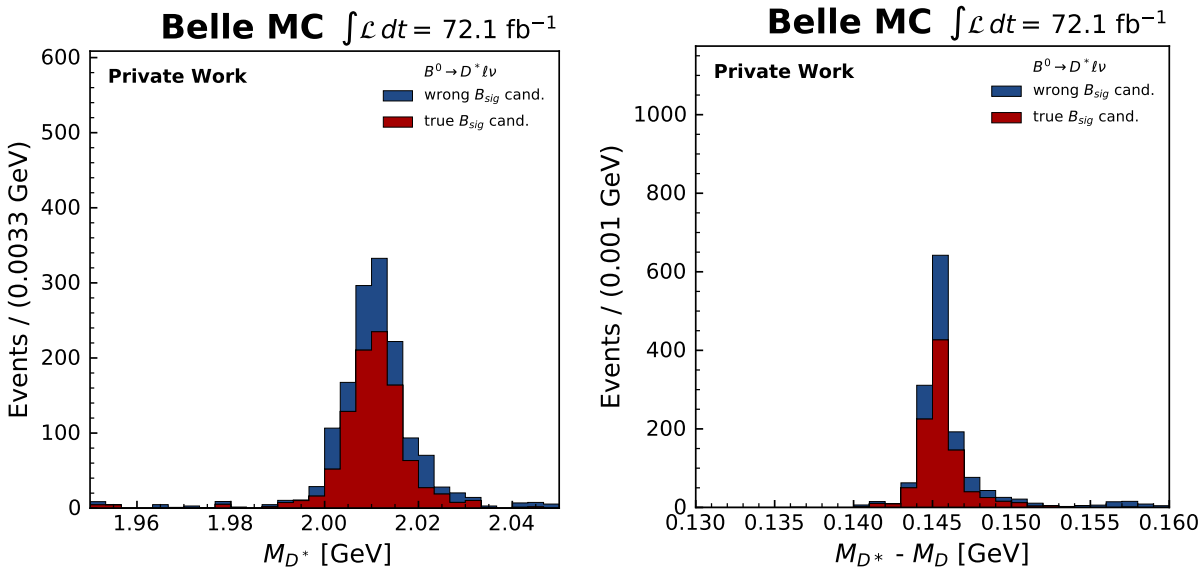

Figure 6.4: Invariant mass of the D^* and the difference between the invariant D^* and D^0 mass for correctly (red) and wrongly (blue) reconstructed B_{sig} candidates in events with at least one $B^0 \to D^{*-}\ell^+\nu$ decay, shown for a subset of Monte Carlo data corresponding to the luminosity of 72 fb^{-1}.

B_{sig} and $\Upsilon(4S)$ Reconstruction

The B_{sig} is reconstructed from a D^* and a light lepton and combined with the B_{tag} candidate to form an $\Upsilon(4S)$. The light lepton must have momentum $p > 0.6$ GeV in the lab frame.

6.2 Continuum Suppression and Best Candidate Selection

The same selection on the continuum suppression classifier output > 0.5 as for the events in the analysis $B^0 \to \tau\ell$ is applied. Afterward, the best candidate selection is implemented to select the $\Upsilon(4S)$ candidate in each event, which is most likely correctly reconstructed. The individual steps are:

I Select the B_{tag} with the highest signal probability (same as in signal $B^0 \to \tau\ell$ decay).

II Choose the D with the smallest invariant mass difference from its nominal value.

III Choose the D^* with the smallest invariant mass difference between D^* and D.

IV Select the B_{sig} lepton with the highest lepton identification.

V In case multiple $\Upsilon(4S)$ candidates remain, one is randomly selected.

In 95.42% of all cases, the best candidate selection chooses the correctly reconstructed $\Upsilon(4S)$ candidate. For most of the remaining cases, a wrongly reconstructed $\mathrm{B_{tag}}$ with a higher signal probability is selected over the correct $\mathrm{B_{tag}}$.

6.3 Fit Variable

The variable m^2_{miss} shows a good separation of signal and background events:

$$\begin{aligned} m^2_{miss} &= p^2_{miss} \\ \text{with } p_{miss} &= p_{\mathrm{B_{tag}}} - p_{\mathrm{B_{sig,vis}}} \text{ in CMS Frame} \\ p_{\mathrm{B_{tag}}} \equiv \left(E_{\mathrm{B_{tag}}}, \vec{p}_{\mathrm{B_{tag}}}\right) &= \left(\mathrm{E_{beam}}, \frac{\vec{p}_{\mathrm{B_{tag}}}}{|\vec{p}_{\mathrm{B_{tag}}}|}\sqrt{\mathrm{E^2_{beam}} - m^2_B}\right) \end{aligned} \tag{6.2}$$

Although the $\mathrm{B_{tag}}$ is not required to be correctly reconstructed, the determination of $p_{\mathrm{B_{tag}}}$ constrains its four-momentum with the beam energy and nominal B meson mass. Fig. 6.5 shows the comparison of the data - and Monte Carlo events in the m^2_{miss} distributions.

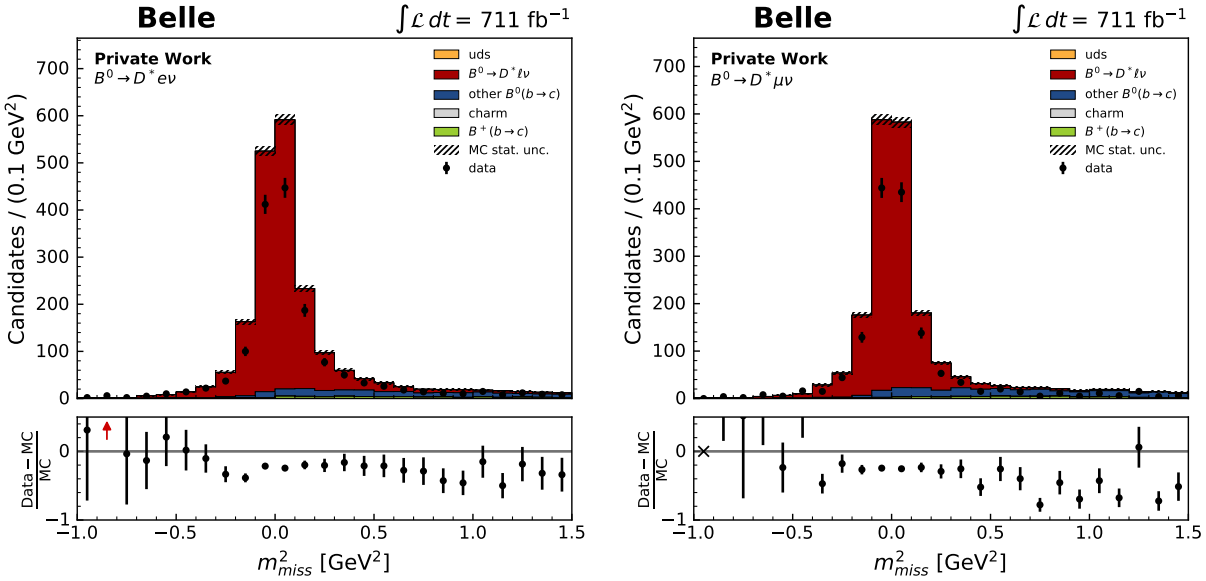

Figure 6.5: Distributions of the m^2_{miss} for Monte Carlo and real data. The signal Monte Carlo events are shown in red, and the signal is defined by a correctly reconstructed $\mathrm{B_{sig}}$. All other Monte Carlo events are represented by the other colored components. The left plot shows the m^2_{miss} distribution in the reconstructed $\mathrm{B^0} \to \mathrm{D^*}e\nu$ channel and on the right for the reconstructed $\mathrm{B^0} \to \mathrm{D^*}\mu\nu$ channel.

For correctly reconstructed signal events, the m^2_{miss} peaks at zero because the only undetected particle is the neutrino of the B_{sig} decay, and its invariant mass is zero. The finite resolution of the detector smears out the m^2_{miss} distribution.

The dominant Monte Carlo events are signal events, which contribute 88.65% of all Monte Carlo events. Overall, the Monte Carlo events overestimate the real data. Since all discrepancies between Monte Carlo and real data are corrected as described in section 4.5, the difference is solely attributed to a lower performance of the FEI on the real data than on Monte Carlo events.

6.4 Fit

To determine the calibration factor C_{FEI}, which quantifies the performance difference between Monte Carlo and real data, the number of signal events in Monte Carlo and real data are extracted with a binned maximum likelihood fit of the m^2_{miss} distribution. Separate calibration factors are determined for the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channels.

$$C_{FEI} = \frac{N_{sig,data}}{N_{sig,MC}} \tag{6.3}$$

The signal and background shapes are described by two template probability density functions (PDFs) generated from the respective Monte Carlo expectation in the region of the m^2_{miss} between -1.0 GeV^2 and 1.5 GeV^2 in 25 equidistant bins.

Poisson uncertainties on the number of events in each bin of the PDF data are included as nuisance parameters. These nuisance parameters are uncorrelated between the fit bins and PDF components. In roofit [89], with which the fit was implemented, the method of including the Poisson uncertainty due to the finite size of PDF data is called the Barlow Beeston [90] method. In total, the model PDF contains two free yield parameters and 50 nuisance parameters.

Monte Carlo events corresponding to five times the Belle data luminosity (5 streams) are used to generate the signal and background PDF. The large PDF data sizes minimize uncertainty on the calibration factor caused by the finite size of PDF data. The Monte Carlo fit data consist of all other available Monte Carlo events, which are weighted to their expected rates for the luminosity of the Belle data. It contains one stream of continuum events, which contributes only a tiny fraction of background events, and four streams of $B(b \to c)$ events. Tab. 6.4 presents the number of streams in PDF data and Monte Carlo fit data.

	MC streams in PDF data	MC streams in fit data
continuum	0 - 4	5
B($b \to c$)	0 - 4	5,7,8,9

Table 6.4: The number of streams of the Monte Carlo data used for the creation of the template PDF and the fit data. Stream 6 was not included in the analysis since it was used in the FEI training and could bias the final result.

Since the B meson events are the dominant contributions, balanced template and fit data sets result in the smallest overall uncertainty on the calibration factors. Increasing the number of events in the fit data would result in larger uncertainties due to the finite size of the PDF data and vice versa.

6.4.1 Fit Validation

A toy study is performed to verify the accurate determination of the statistical uncertainty on the signal yield N_{sig} by the fit. In total, 10,000 fits are performed. For each fit, new fit and PDF data are generated based on the original signal and background PDFs. Subsequently, a new PDF is created from the PDF data, which is then fitted to the newly generated fit data. The signal and background data sets are generated separately from their respective PDFs. The process of the data set generation is presented in Fig. 6.6.

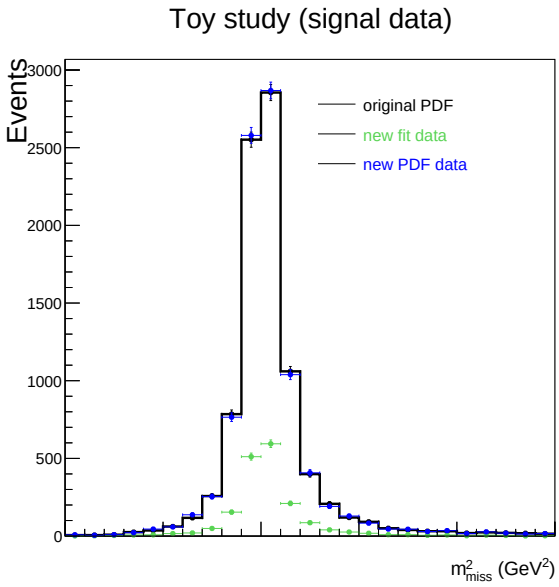

Figure 6.6: Illustration of the generation of a new template (blue) and fit data (green) from the original PDF data (black) for signal events in the $B^0 \to D^* e\nu$ channel.

The black distribution represents the original PDF data from which the signal PDF was created. A new data set is generated, where the number of events in each bin varies within their Poisson uncertainties according to the nominal distribution. The nominal distribution refers to the expected distribution for a specific number of events without any variations. For the new PDF data, shown in blue, the original PDF data serves as the nominal distribution. Similarly, the fit data set, shown in green, is generated with the total number of events being 1/5 of the original PDF data to match the size of the Belle data. The same procedure is applied for the background data sets.

Tab. 6.5 lists the nominal number of events used in the generation of the PDF and fit data.

	N_{sig} PDF	N_{sig} fit	N_{bkg} PDF	N_{bkg} fit
$B^0 \to D^* e\nu$	8836	1767	1225	245
$B^0 \to D^* \mu\nu$	8386	1677	1476	295

Table 6.5: Nominal number of signal (sig) and background (bkg) events used to generate the PDF and fit data in the toy study.

For each fit, a pull value is calculated:

$$\text{pull} = \frac{N_{\text{sig,gen}} - N_{\text{sig,fit}}}{\sigma_{N_{\text{sig,fit}}}} \tag{6.4}$$

where $N_{\text{sig, gen}}$ denotes the number of generated signal events in the fit data, $N_{\text{sig, fit}}$ and $\sigma_{N_{\text{sig, fit}}}$ are the signal yield and its uncertainty obtained in the fit, respectively. For an unbiased fit setup with correct uncertainty estimation, the pull distribution is described by a normal Gaussian distribution with a mean (μ) of zero and a standard deviation (σ) of one.

Fig. 6.7 shows the results of the toy study for the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channel. The pull distributions for both channels exhibit small negative biases, which are consistent with zero within their respective two σ uncertainties. Since the value of σ is less than one, indicating that the fit overestimates the uncertainty, the small bias is assumed to be negligible for the final result. The distributions of the fitted number of signal yields and the signal yield uncertainties are shown in Appendix B.

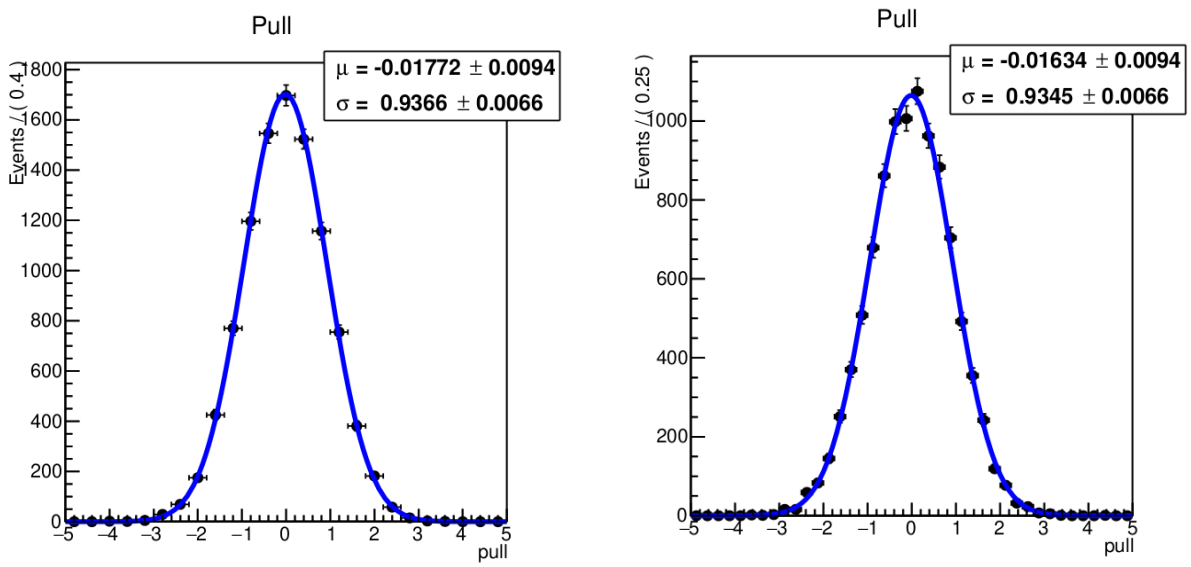

Figure 6.7: Pull distribution for the signal yields obtained in a toy study for the $B^0 \to D^* e\nu$ channel on the left and the $B^0 \to D^* \mu\nu$ channel on the right.

6.4.2 Fit Results

After the fit validation, the final fit is performed to extract the number of signal events in Monte Carlo and Belle data. The post-fit distributions are shown in Figs. 6.8 and 6.9 for the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channel. The fitted model PDF is in good agreement with the data. The yields only include the statistical uncertainty, and the systematic uncertainty arising from the finite size of the PDF data.

The fitted signal yields and the corresponding calibration factors for the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channels are listed in Tab. 6.6.

	$N_{\text{sig,PDF}}$	$N_{\text{sig,MC}}$	$N_{\text{sig,data}}$	C_{FEI}
$B^0 \to D^* e\nu$	1767	1745.44 ± 30.84	1360.78 ± 44.64	0.7796 ± 0.0291
$B^0 \to D^* \mu\nu$	1677	1653.62 ± 28.70	1301.57 ± 41.31	0.7871 ± 0.0285

Table 6.6: Number of signal events in the PDF data scaled to the luminosity to the real data, signal yields obtained from the fit of Monte Carlo and real data, and the corresponding calibration factors, where the systematic uncertainty due to the finite size of the PDF data is included.

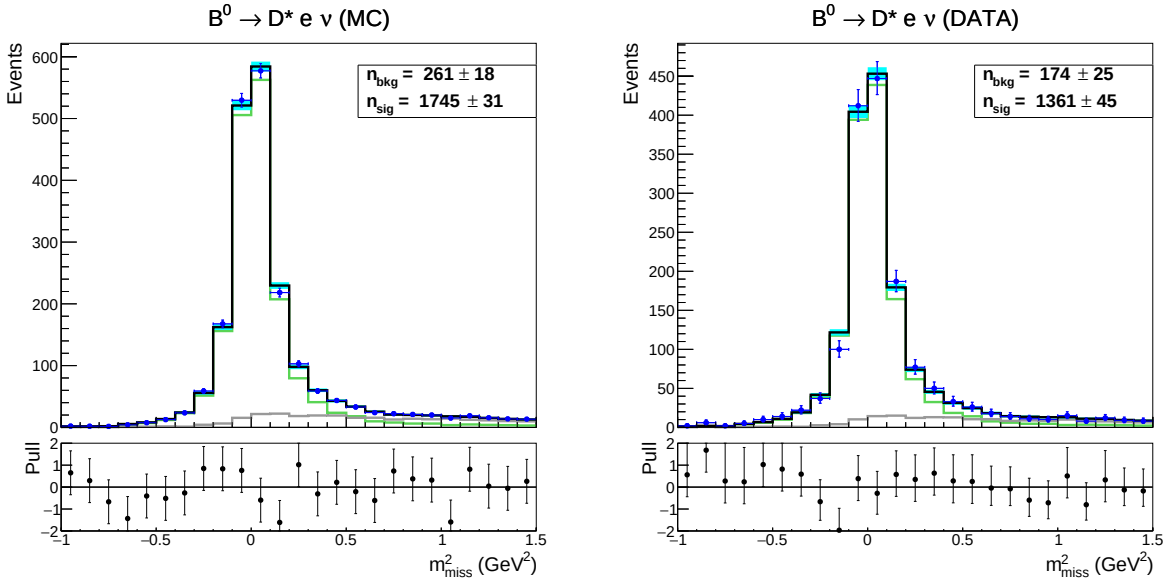

Figure 6.8: Post-fit distribution of the m^2_{miss} for $B^0 \to D^* e\nu$ channel in Monte Carlo (left) and real data (right), where the blue points describe the fit data, the black PDF the combined model PDF with the individual signal component in green and the background component in grey.

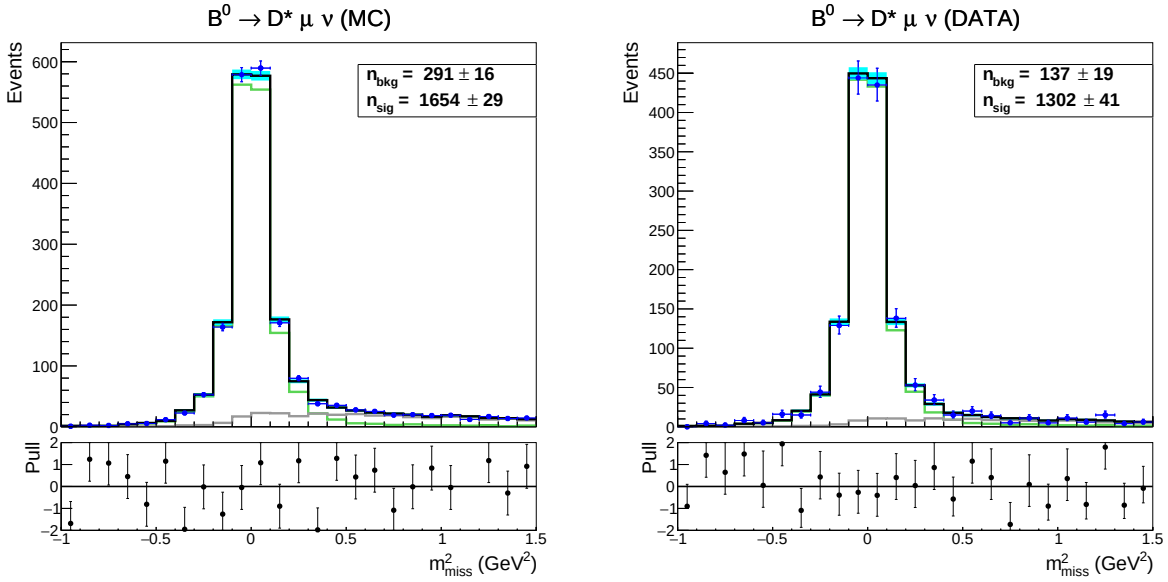

Figure 6.9: Post-fit distribution of the m^2_{miss} for $B^0 \to D^* \mu\nu$ channel in Monte Carlo (left) and real data (right), where the blue points describe the fit data , the black PDF the combined model PDF with the individual signal component in green and the background component in grey.

The signal yield obtained from the fit of Monte Carlo data agrees with the expected number of signal events in the PDF data scaled to the luminosity of the real data. No significant deviation was expected since all Monte Carlo streams are identical apart from statistical fluctuations. The majority of Monte Carlo events are $\mathrm{B}(b \to c)$ events, which are weighted by a factor of 1/4. Because the fit takes the event weights into account, the uncertainty of the Monte Carlo signal yield is lower than the signal yield uncertainty obtained in real data.

Assuming the signal yields are uncorrelated, the uncertainty of the calibration factors is determined via the Gaussian error propagation:

$$\sigma_{\mathrm{C_{FEI}}} = \sqrt{\left(\frac{\partial \mathrm{C_{FEI}}}{\partial \mathrm{N_{data}}}\sigma_{\mathrm{N_{data}}}\right)^2 + \left(\frac{\partial \mathrm{C_{FEI}}}{\partial \mathrm{N_{MC}}}\sigma_{\mathrm{N_{MC}}}\right)^2} \tag{6.5}$$

The Barlow Beeston fit introduces correlated uncertainties between Monte Carlo and real data since the systematic uncertainty due to the finite data size of the PDF data is the same for both fits. To test the impact of this systematic uncertainty, the fit is repeated without it. The results are shown in Tab. 6.7.

	$\mathrm{N_{sig,MC}}$	$\mathrm{N_{sig,data}}$	$\mathrm{C_{FEI}}$
$\mathrm{B^0} \to \mathrm{D^*}e\nu$	1745.48 ± 22.59	1362.00 ± 41.39	0.7803 ± 0.0258
$\mathrm{B^0} \to \mathrm{D^*}\mu\nu$	1653.62 ± 21.07	1300.56 ± 38.38	0.7865 ± 0.0253

Table 6.7: Signal yields and calibration factors obtained from the fit without the systematic uncertainty due to the limited data set size in the template creation.

There is an obvious decrease in the uncertainty on the Monte Carlo signal yield. However, the signal yield uncertainty from real data is the dominating source of the calibration factor uncertainty, which decreases only slightly.

6.4.3 Closure Test: Branching Ratio

To cross-check if the signal efficiency is correctly determined, the branching ratio of the signal decay is calculated from the obtained value of C_{FEI}, which only includes the statistical uncertainty, and the systematic uncertainty due to the limited size of the PDF data. The result is expected to be consistent with the signal branching ratio from the Monte Carlo generation, if no mistakes occur during the calculation.

The corresponding branching ratio for the signal decay is defined as:

$$\mathcal{B}(B^0 \to D^{*-}(\to D^0(\to K^+\pi^-)\pi^-)\ell^+\nu) = \frac{N_{sig,MC}}{2 \cdot N_{BB} \cdot \epsilon_{sig} \cdot f_{00}} \tag{6.6}$$

where $N_{sig,\,data}$ denotes the number of fitted signal events in real data, N_{BB} is the number of B meson pairs in data, f_{00} is the fraction of $\Gamma(\Upsilon(4S) \to B^0\bar{B}^0)/\Gamma(\Upsilon(4S) \to BB)$and ϵ_{sig} is the signal efficiency.

The signal efficiency is calculated from Monte Carlo events. It depends on the number of signal events after the final selection, $N_{sig,\,final}$, which are the signal events used for the creation of the signal template, the total number of generated signal events, $N_{sig,\,gen}$, and the calibration factor C_{FEI}.

$$\epsilon_{sig} = \frac{N_{sig,final} \cdot C_{FEI}}{N_{sig,\,gen}} \tag{6.7}$$

Since $N_{sig,\,gen}$ is scaled to its expectation in the real data, the number of generated signal events is defined as:

$$N_{sig,\,gen} = 2 \cdot \mathcal{B}_{calibration} \cdot N_{BB} \cdot f_{00} \tag{6.8}$$

with $\mathcal{B}_{calibration} = 0.1367\%$ (Tab. 6.1).

Tab. 6.8 lists the number of fitted signal events, the signal efficiencies and the corresponding branching ratios.

	$N_{sig,data}$	$N_{sig,final}$	ϵ_{sig}	$\mathcal{B}$
$B^0 \to D^* e\nu$	1360.78 ± 44.64	1767 ± 18	(0.134 ± 0.005)%	(0.1349 ± 0.0037) %
$B^0 \to D^* \mu\nu$	1301.57 ± 41.31	1677 ± 18	(0.129 ± 0.005)%	(0.1347 ± 0.0037)%

Table 6.8: Signal yields for the Barlow-Beeston fit on data, the signal events in Monte Carlo, the signal efficiency, and the branching ratio for the electron and $B^0 \to D^*\mu\nu$ channel.

The calculated branching ratio for the $B^0 \to D^* e\nu$ and $B^0 \to D^*\mu\nu$ channel agrees with the branching ratio in the Monte Carlo generation within its uncertainty.

6.4.4 Systematic Uncertainties

Additional uncertainties due to the uncertainties of the Monte Carlo corrections, see section 4.5, are divided into global uncertainties, which only alter the yields, and local uncertainties, which change the shape of the PDF.

Global Uncertainties

The uncertainty on the calibration factor due to a global uncertainty is calculated according to Eq. 6.5. The global uncertainty only affects the Monte Carlo signal yield, ($\sigma_{\mathrm{N_{data}}} = 0$) and its uncertainty on the Monte Carlo signal yield is defined as:

$$\sigma_{\mathrm{N_{MC}}} = \mathrm{N_{MC}} \cdot \sigma_{\mathrm{unc}} \tag{6.9}$$

where σ_{unc} is the relative uncertainty per event caused by a specific uncertainty source.

Track Finding Efficiency

For each of the four tracks on the signal side, an uncertainty on the track finding efficiency $\sigma_{\mathrm{track}} = 0.0035$ is assigned.

$$\sigma_{4\text{ tracks}} = 1 - (1 - \sigma_{\mathrm{track}})^4 \tag{6.10}$$

Branching Ratio of Signal Decay

The relative uncertainty due to the uncertainty of the signal decay branching ratio is calculated from the standard deviation uncertainty via:

$$\mathcal{B} = (0.1366 \pm 0.0032)\% \rightarrow \sigma_{\mathcal{B}} = 0.0237. \tag{6.11}$$

Local Uncertainties

Since local systematic uncertainties can alter the shape of the PDF, a new PDF is created from the reweighted Monte Carlo events, where the nominal weight of each event $= w_i^{\mathrm{nom}}$ is varied by $\delta w_i^{\mathrm{source}}$ for a specific uncertainty source. The event weight w_i is defined as:

$$w_i = w_i^{\mathrm{nom}} + \delta w_i^{\mathrm{source}}. \tag{6.12}$$

The new PDF is then fitted to the real data, whose distribution remains unchanged, and to the fit Monte Carlo data, where the events are varied analogously to the reweighted PDF data.

Low Momentum Efficiency

The corrections for finding a particle with a low momentum are divided into the six momentum bins, which are correlated. By diagonalizing the covariance matrix, six independent variations are calculated, each with an up and a down variation. For each variation j, the uncertainty on the calibration factor is determined by the larger uncertainty on the calibration factor of the down or up variation:

$$\sigma_{\text{var,j}} = \max_{\text{i = up, down}} |\text{C}_{\text{FEI}} - \text{C}_{\text{FEI,var j,i}}|. \tag{6.13}$$

The overall uncertainty of the six uncorrelated variations is determined with:

$$\sigma_{\text{source}} = \sqrt{\sum_j \sigma^2_{\text{var,j}}}. \tag{6.14}$$

Form Factors of $\text{B} \to D^{(*)}\ell\nu$ Decays

The same approach as for the low momentum finding efficiency is applied for the form factor uncertainties of $\text{B} \to \text{D}\ell\nu$ with five variations (five parameters of the BGL model) and $\text{B} \to \text{D}^*\ell\nu$ with six variations (six parameters of the BGL model).

Particle Identification

In the case of the particle identification uncertainties, the weights of the events vary within their uncorrelated statistical uncertainties, and 100% correlated systematic uncertainties of their respective pid bin uncertainty.

The nominal weight of each event i (w_i^{nom}) is determined by varying statistical (δw_i^{stat}) and systematic uncertainty (δw_i^{sys}).

$$w_i = w_i^{\text{nom}} + \delta w_i^{\text{stat}} + \delta w_i^{\text{sys}} \tag{6.15}$$

For the systematic uncertainty on the particle identification, the events are reweighted by:

$$\delta w_i^{\text{sys}} = G_{sys}^{\text{rand}}(0,1) \times \sigma_i^{\text{sys}} \tag{6.16}$$

Because the systematic uncertainties across pid bins are correlated 100%, the weight for each event is varied within the same random value of the normal Gaussian distribution multiplied by the systematic uncertainty of the respective pid bin.

The statistical uncertainty is uncorrelated across pid bins. Since it is notably smaller than the systematic uncertainty on the correction of the pid bins, an approximation is applied. This approach simplifies the implementations of statistical pid bin uncertainty by treating

the statistical uncertainty on each event weight as uncorrelated. For each event, the varied event weight is determined by a random value of a Gaussian, with the mean being the corresponding statistical uncertainty.

$$\delta w_i^{\text{stat}} = G_{\text{stat}}^{\text{rand}} \left(0, \sigma_i^{\text{stat}}\right) \tag{6.17}$$

This procedure is repeated for 400 fits for each correctly identified particle type, and a new calibration factor is determined from each fit. The distribution of the calibration factors is fitted with a Gaussian distribution, where one standard deviation (σ) is assigned as uncertainty on the calibration factor due to particle identification uncertainty. The distribution of the calibration factors for correctly identified leptons, pions, and kaons are shown in Fig. 6.10, 6.11, and in 6.12. The uncertainty on the calibration factor due to the lepton pid correction is the dominating uncertainty among all particle identification corrections.

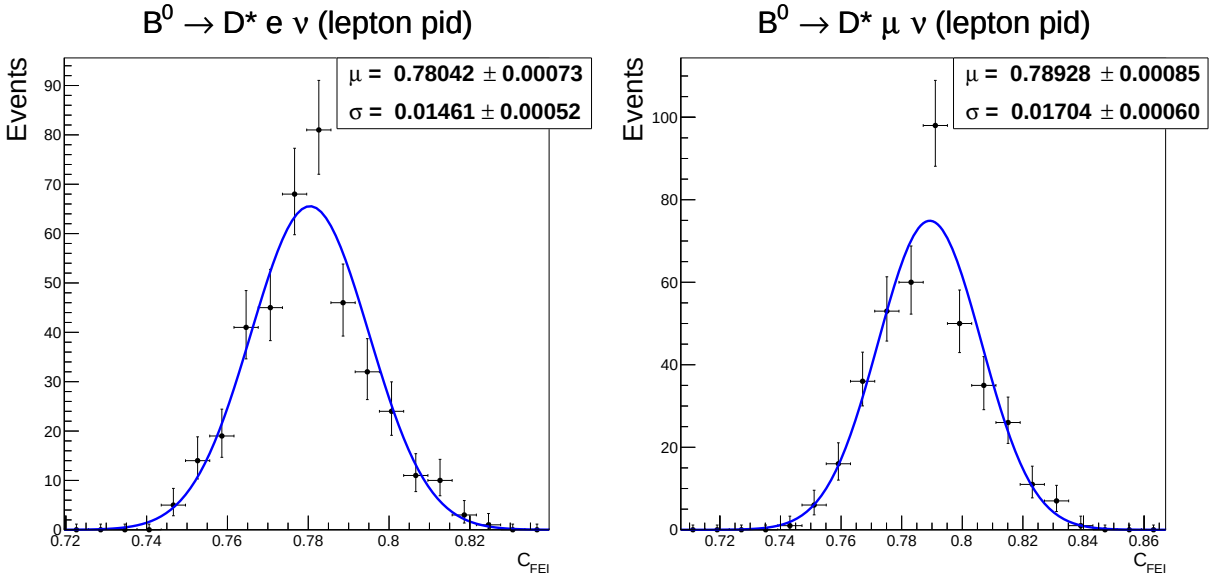

Figure 6.10: Distribution of calibration factors obtained from fits with varied lepton pid weights in the $B^0 \to D^* e\nu$ (left) and the $B^0 \to D^* \mu\nu$ (right) channel.

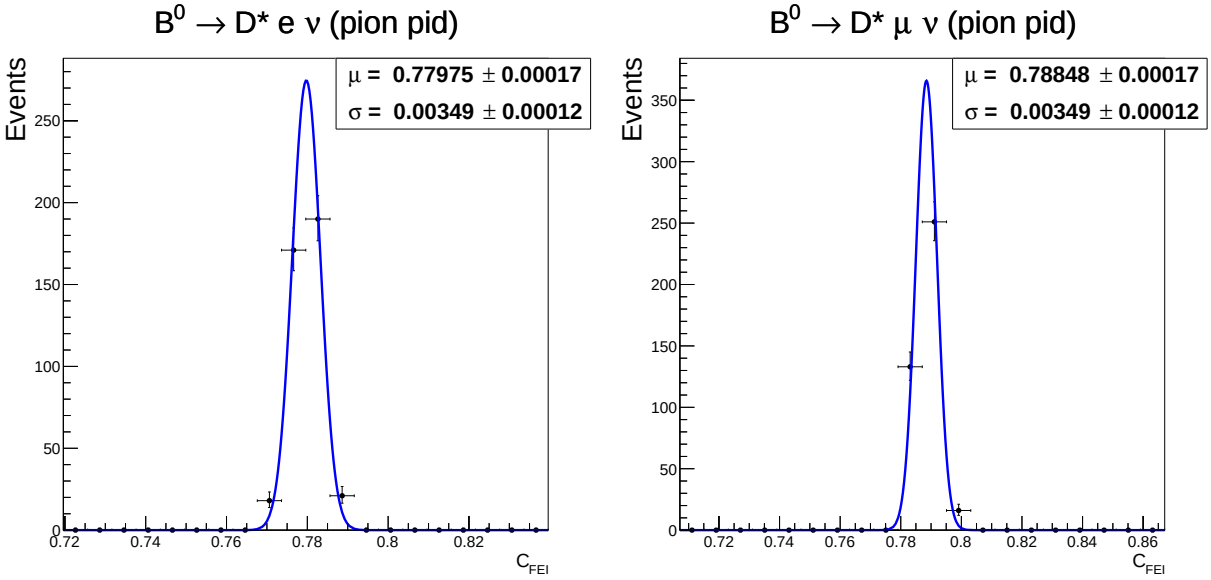

Figure 6.11: Distribution of calibration factors obtained from fits with varied pion pid weights in the $B^0 \to D^* e\nu$ (left) and the $B^0 \to D^* \mu\nu$ (right) channel.

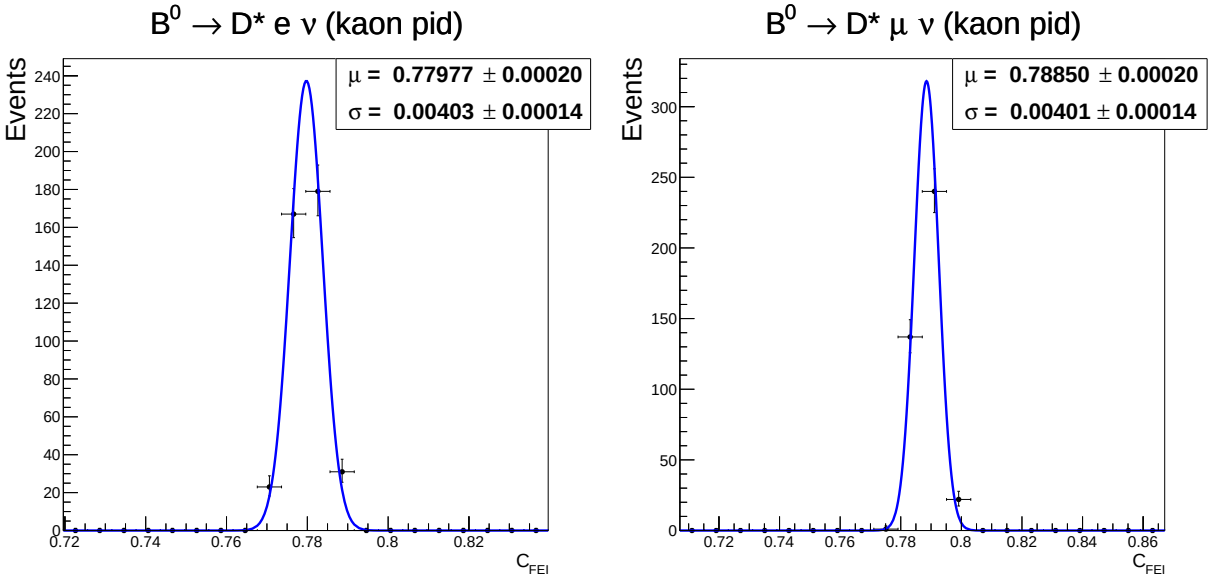

Figure 6.12: Distribution of calibration factors obtained from fits with varied kaon pid weights in the $B^0 \to D^* e\nu$ (left) and the $B^0 \to D^* \mu\nu$ (right) channel.

Wrongly identified particles, the so-called fakes, also have a corresponding uncertainty on the fake rate. Because the signal events are identified by a correctly reconstructed B_{sig}, all its decay products are also correctly reconstructed. Therefore, the shape of the signal PDF remains unchanged by varying the event weights according to the fake rate uncertainty. It only impacts the shape of the background PDF. Tab. 6.9 lists the composition of Monte Carlo particles for the individual final state particles with a pid selection in the background Monte Carlo data.

	MC particle	$B^0 \to D^* e\nu$	$B^0 \to D^* \mu\nu$
ℓ from B_{sig}	ℓ^-	97.88%	73.34%
	π^-	1.91%	23.65%
	K^-	0.11%	2.11%
K from D^0	K^-	98.83%	98.76%
	π^-	0.76%	0.94%
	μ^-	0.21%	0.24%
	other	0.20%	0.06%
π from D^0	π^-	98.81%	98.90%
	μ^-	0.89%	0.93%
	other	0.30%	0.17%

Table 6.9: Monte Carlo information on the final state particles on the signal side with a pid selection in background events.

Only for true kaons and pions identified as electrons and muons a fake rate is provided, with only a statistical uncertainty on the correction. For true pions identified as kaons and vice versa, the uncertainty on the fake rate has both statistical and systematic uncertainty. Analogously to the correctly identified particles, the events are reweighted, 400 fits are performed, and the uncertainty on the calibration factor is determined.

In principle, the uncertainties due to the number of B meson pairs N_{BB} and f_{00} uncertainty need to be included in the calibration factor uncertainties. However, as the uncertainties are identical in the calibration and signal study, they are only included in the signal $B^0 \to \tau\ell$ fit to avoid double-counting.

Overview of Uncertainties

The total uncertainty on the calibration factor is determined by combining all individual uncorrelated uncertainties via the Gaussian error propagation. Tab. 6.10 lists the calibration factor uncertainties for the individual sources in the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channels.

uncertainty source	$B^0 \to D^* e\nu$ $[10^{-2}]$	$B^0 \to D^* \mu\nu$ $[10^{-2}]$
fit uncertainty + finite size of PDF data	2.905	2.847
track finding efficiency	1.068	1.096
$\mathcal{B}(B^0 \to D^{*-}(\to D^0(\to K^+\pi^-)\pi^-)\ell^+\nu)$	1.845	1.863
low momentum efficiency	1.298	1.311
form factor $B \to D\ \ell\ \nu$	0.002	0.001
form factor $B \to D^*\ \ell\ \nu$	0.230	0.209
pid true lepton	1.461	1.704
pid fake lepton	0.046	0.073
pid true pion	0.349	0.349
pid fake pion	0.007	0.000
pid true kaon	0.403	0.401
pid fake kaon	0.002	0.006
total	4.145	4.211

Table 6.10: Overview of the absolute uncertainties on the calibration factors for the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channels.

The final result for the calibration factor is given in Tab. 6.11.

	C_{FEI}
$B^0 \to D^* e\nu$	0.7796 ± 0.0415
$B^0 \to D^* \mu\nu$	0.7871 ± 0.0421

Table 6.11: Result of C_{FEI} for the $B^0 \to D^* e\nu$ and $B^0 \to D^* \mu\nu$ channels.

6.5 Validation of FEI Calibration for the $B^0 \to \tau\ell$ Channel

Since the signal efficiency of $B^0 \to \tau\ell$ events on real data is unknown, its determination relies solely on signal Monte Carlo events, which are corrected to match the real data by implementing the Monte Carlo corrections. It is crucial to verify that the obtained correction factors are also valid for the reconstructed B_{tag} candidates in signal events. If the composition of the B_{tag} candidates agrees between the calibration and signal study, the calibration factors can be applied to the signal events. Therefore, B_{tag} variables between the calibration and signal channel in their respective signal regions are compared in the following section.
The signal region of the calibration and signal channel are defined as:

- calibration channel : -1.0 $GeV^2 < M_{miss} < 1.5$ GeV^2
- signal channel: 2.2 GeV $< p_\ell <$ 2.42 GeV

  \+ B meson background suppression selection of the final result

In the calibration channel, events with a correctly reconstructed B_{sig} are considered signal, while all events in the signal channel are identified as signal. Both calibration and signal channel do not require the B_{tag} to be correctly reconstructed.

The following Tab. 6.12 presents the fraction of events with a correctly reconstructed B_{tag} relative to all events for the electron and muon as the lepton on the signal side. Small deviations between the purities for the individual reconstructed τ channels in signal events are observed. Because the final fit in the signal channel combines all reconstructed τ decay channels, this purity is most important and is in good agreement with the purity of the calibration channel.

	Purity B_{tag}
true $B^0 \to D^* e\nu$	50.4%
$B^0 \to \tau e$	51.3%
true $B^0 \to \tau e$	53.8%
$B^0 \to \tau(\to e\nu\nu)e$	53.1%
$B^0 \to \tau(\to \mu\nu\nu)e$	52.3%
$B^0 \to \tau(\to \pi\nu)e$	50.2%
$B^0 \to \tau(\to \pi\pi^0\nu)e$	50.4%
$B^0 \to \tau(\to \pi\pi\pi\nu)e$	52.3%
$B^0 \to \tau(\to \pi\pi^0\pi^0\nu)e$	59.4%

	Purity B_{tag}
true $B^0 \to D^* \mu\nu$	51.3%
$B^0 \to \tau\mu$	51.0%
true $B^0 \to \tau\mu$	52.7%
$B^0 \to \tau(\to e\nu\nu)\mu$	52.5%
$B^0 \to \tau(\to \mu\nu\nu)\mu$	52.9%
$B^0 \to \tau(\to \pi\nu)\mu$	48.9%
$B^0 \to \tau(\to \pi\pi^0\nu)\mu$	52.7%
$B^0 \to \tau(\to \pi\pi\pi\nu)\mu$	51.9%
$B^0 \to \tau(\to \pi\pi^0\pi^0\nu)\mu$	54.9%

Table 6.12: Purity of B_{tag} in the calibration and signal channels with an electron(left) and a muon(right) as the lepton from the B_{sig}.

Fig. 6.13 - 6.14 compare the normalized distributions of the B_{tag} variables: M_{bc}, ΔE and the reconstructed B_{tag} decay modes, between the calibration (red) and signal channel (blue).

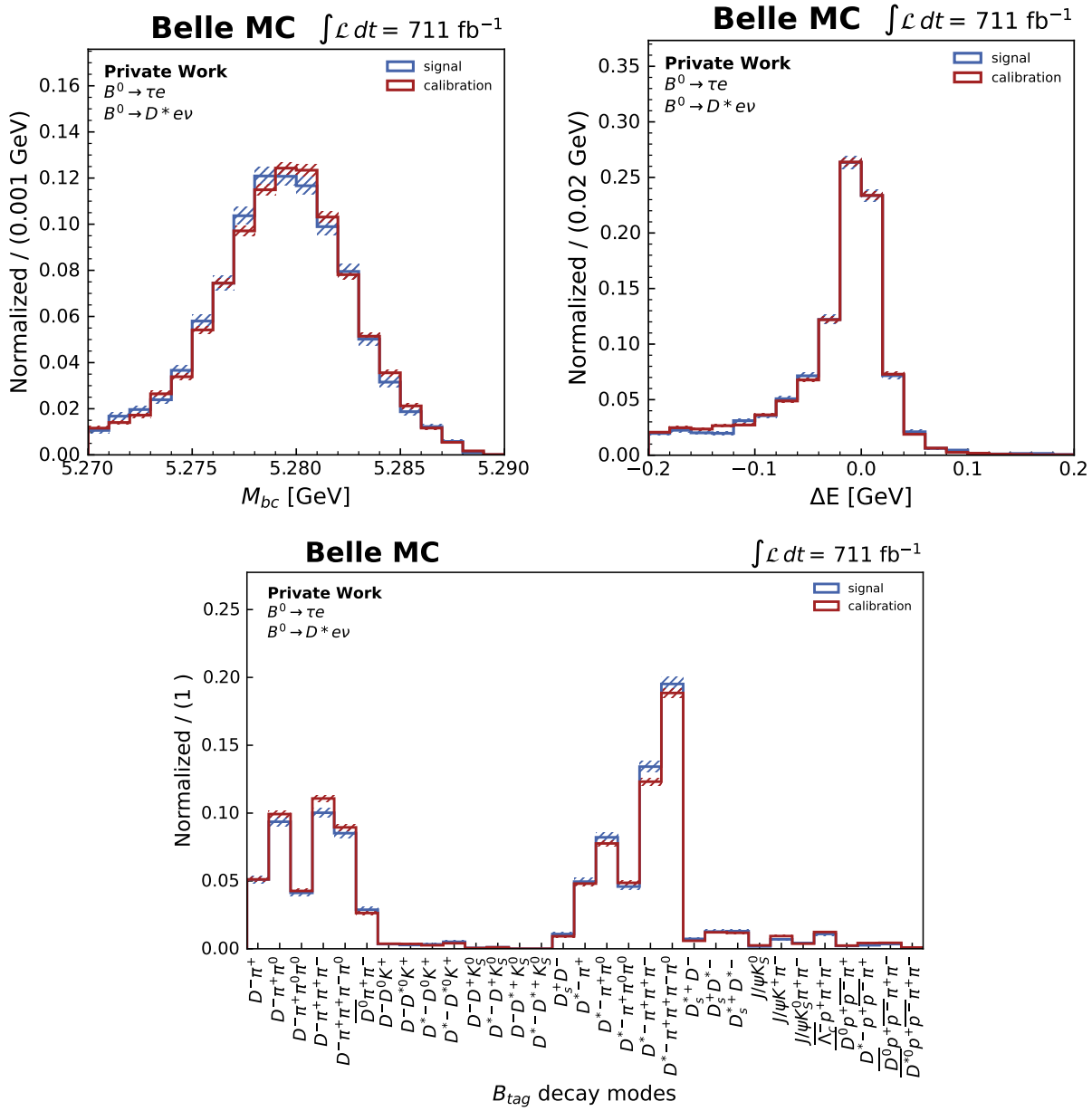

Figure 6.13: M_{bc} (left), ΔE (right) of B_{tag} (bottom), and the reconstructed B_{tag} decay modes (bottom) for calibration (red) and signal (blue) channel with an electron as the lepton on the signal side.

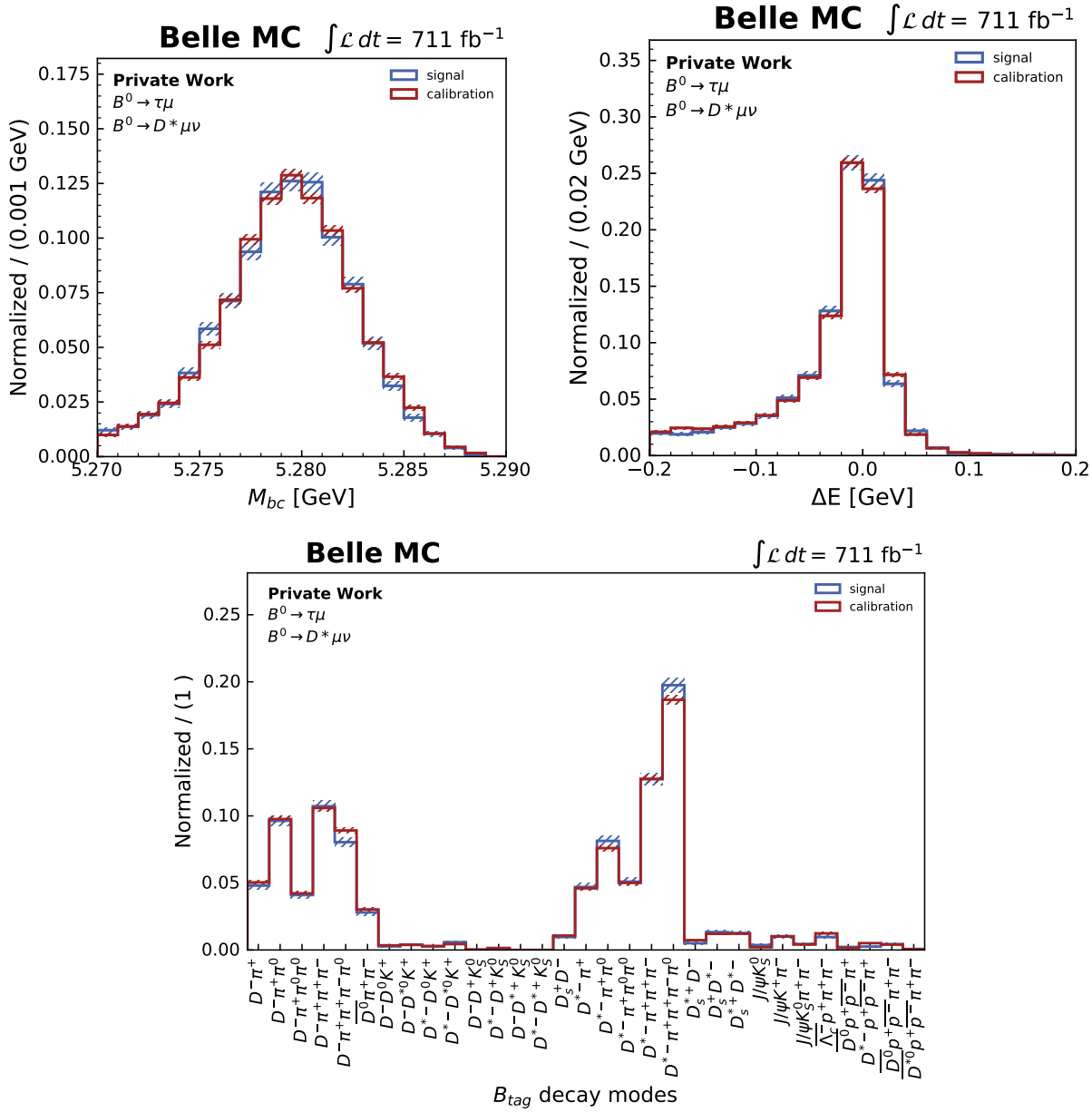

Figure 6.14: M_{bc} (left), ΔE (right) of B_{tag} (bottom), and the reconstructed B_{tag} decay modes (bottom) for calibration (red) and signal (blue) channel with a muon as the lepton on the signal side.

The distributions of M_{bc} and ΔE are in good agreement, while the reconstructed decay channels differ for some B_{tag} decay modes. To quantify this difference, the absolute differences between the normalized signal and calibration events in each reconstructed decay mode are added in a conservative approach via:

$$\sigma_{\text{conservative}} = \sum_i |\text{signal}_i - \text{calibration}_i| \,, \tag{6.18}$$

and assuming the B_{tag} decay modes are uncorrelated via:

$$\sigma_{\text{uncorrelated}} = \sqrt{\sum_i (\text{signal}_i - \text{calibration}_i)^2}, \tag{6.19}$$

where i denotes the reconstructed B_{tag} decay mode. The differences are listed in Tab. 6.13.

Channel	$\sigma_{\text{conservative}}$ (%)	$\sigma_{\text{uncorrelated}}$ (%)
$B^0 \to \tau e$	6.7	2.0
$B^0 \to \tau\mu$	5.3	1.6

Table 6.13: Values of $\sigma_{\text{conservative}}$ and $\sigma_{\text{uncorrelated}}$ for the electron and muon channel.

In conclusion, the FEI calibration factors are applicable for determining the signal efficiency, and no additional uncertainty due to discrepancies between the signal and calibration B_{tag} composition is added.

Chapter 7

Fit of the Lepton Momentum Distribution

This chapter presents the method for extracting signal events from real data, and determining the upper limits of the branching ratio using the statistical methods explained in chapter 2.1.

The branching ratio $\mathcal{B}$ is estimated as:

$$\mathcal{B} = \frac{\mathrm{N_{sig}}}{2 \cdot \mathrm{N_{BB}} \cdot \mathrm{f_{00}} \cdot \epsilon_{\mathrm{sig}}} \tag{7.1}$$

where $\mathrm{N_{sig}}$ denotes the number of signal events obtained in the fit, $\mathrm{N_{BB}}$ is the number of B meson pairs in the real data, $\mathrm{f_{00}} = \Gamma\left(\Upsilon(4\mathrm{S}) \to \mathrm{B^0\overline{B}^0}\right) / \Gamma\left(\Upsilon(4\mathrm{S}) \to \mathrm{B\overline{B}}\right)$, and ϵ_{sig} represents the signal efficiency. Since both B mesons in the event can decay into the signal mode, a factor of 2 is included in the denominator.

Because of the good separation between background and signal events, the lepton momentum in the $\mathrm{B_{sig}}$ rest frame is chosen as the fit variable. The fit region ranges from 1.7 GeV to 2.6 GeV. Probability density functions (PDFs) for signal and background shapes are created from the Monte Carlo expectation as template PDFs with 27 equidistant bins.

Since the $\mathrm{B}(b \to u\ell\nu)$ events exhibit a different shape than all other background components, two separate background PDFs are generated. The PDF named "ulnu bkg" describes the distribution of $\mathrm{B}(b \to u\ell\nu)$ events, while the "other bkg" PDF represents continuum, $\mathrm{B} \to$ rare and the $\mathrm{B}(b \to c)$ events. The combined background PDF is required to have no bin with zero entries to ensure fit stability. The expectation for signal events is represented by a single PDF, called the signal PDF. No distinction is made between events with correctly or wrongly reconstructed $\mathrm{B_{sig}}$ candidates because their distribution in lepton

momentum is similar, as shown in Fig. 7.1, and the same composition is expected in real data after applying all the Monte Carlo corrections.

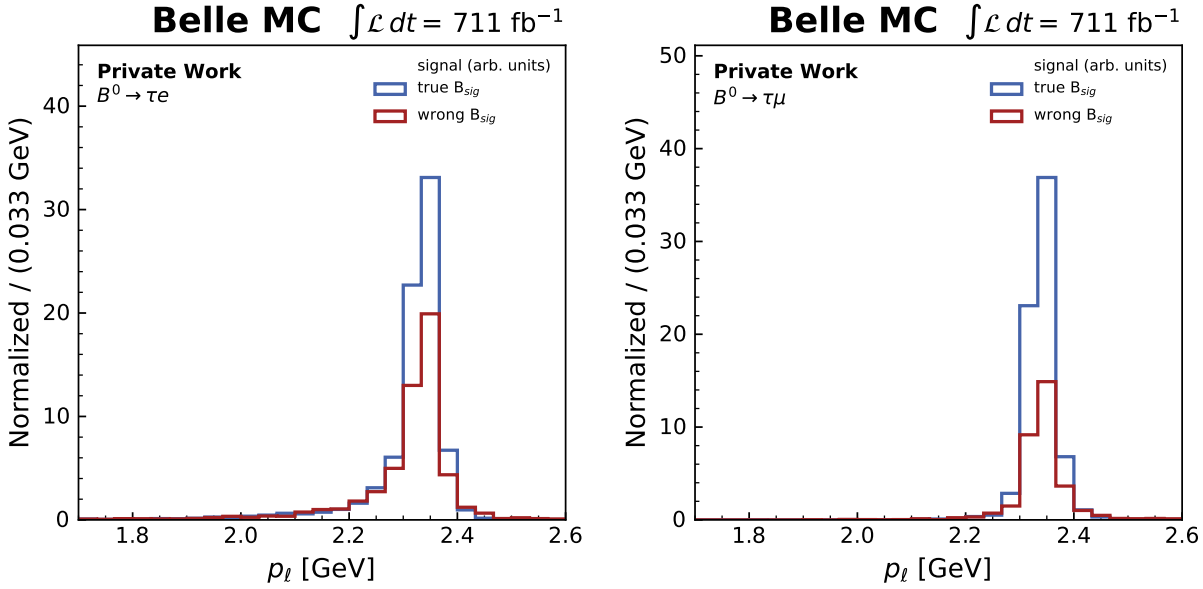

Figure 7.1: Normalized lepton momentum distribution in the B_{sig} rest frame for correctly (blue) and wrongly (red) reconstructed B_{sig} candidates in signal events for the $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) channel.

The fit is implemented as an extended binned maximum likelihood fit, with the pyhf fitting package [91; 92] and cabinetry [93]. In pyhf, all the systematic uncertainties on the PDFs can be included as nuisance parameters, which is necessary for calculating the upper limit of the signal branching ratio. Since no signal events are expected in the real data, the signal extraction and expected upper limit are tested on Asimov data, with zero signal events.

For each B meson background suppression, the expected upper limit is determined separately for the $B^0 \to \tau e$ and $B^0 \to \tau\mu$ channel. Additionally, to ensure that the fit also converges on real data, the fit and upper limit determination are performed on the real data without unblinding the signal region or the fit result. In the end, the B meson background suppression with the best upper limit on the Asimov data and a converging fit on real data is chosen. This chapter only includes the results and distributions of the chosen B meson background suppression, but the same procedure was implemented for all other B meson suppressions.

Fig. 7.2 and 7.3 show the Monte Carlo events in the fit region and the pre-fit distributions with the three PDFs components and Asimov data with zero signal events.

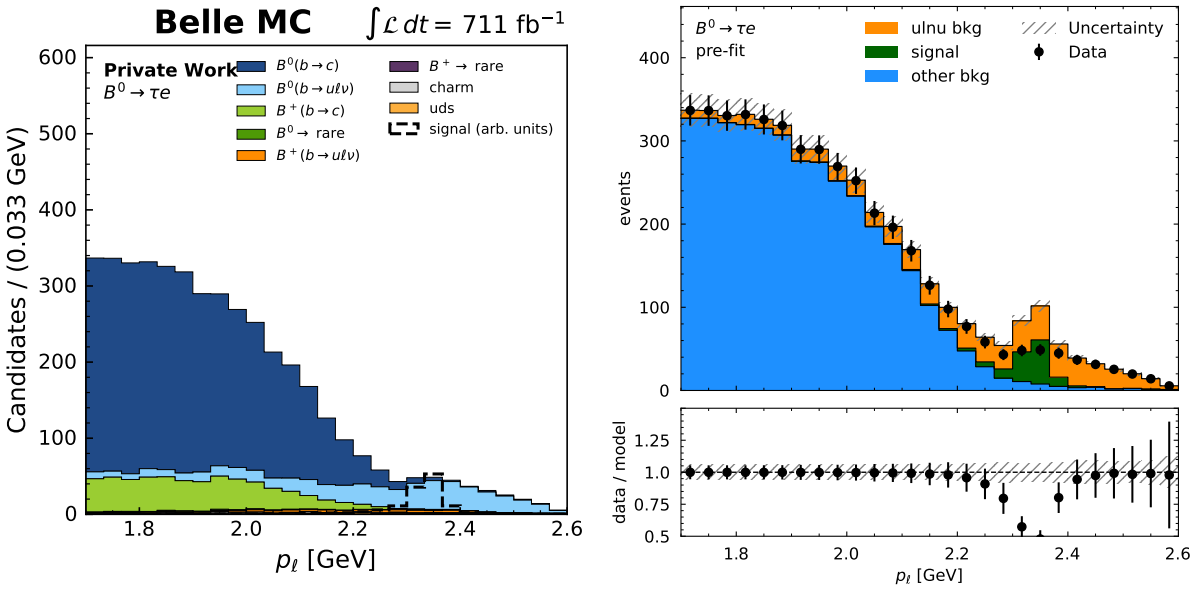

Figure 7.2: Distribution of the lepton momentum in the B_{sig} rest frame for the individual Monte Carlo contributions (left) and the pre-fit distribution with the three fit components (right) for the $B^0 \to \tau e$ channel.

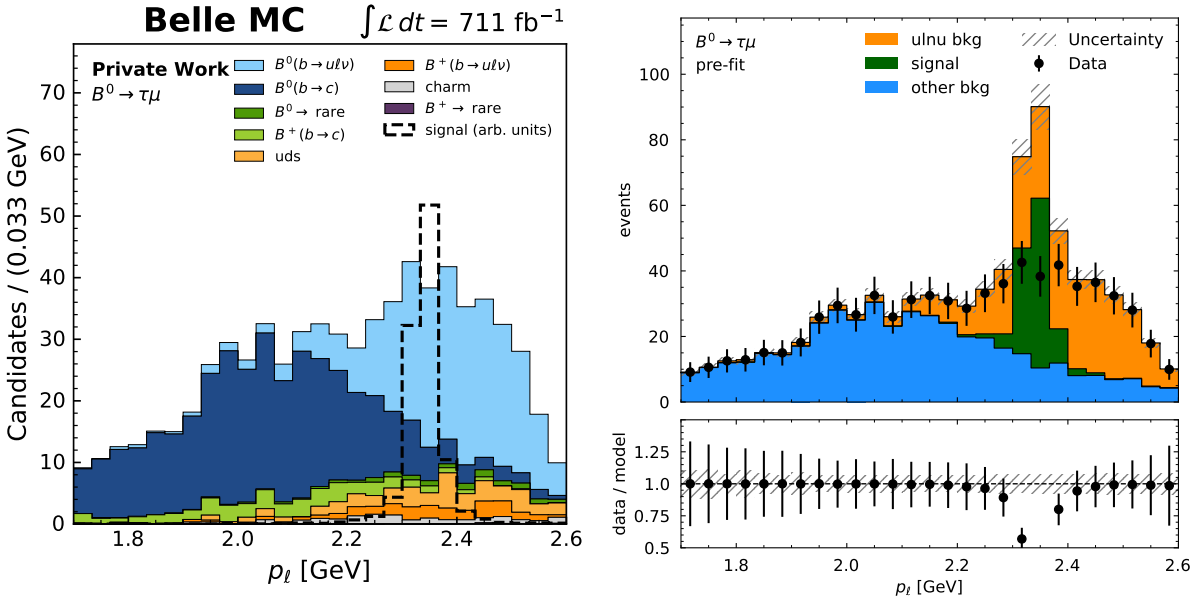

Figure 7.3: Distribution of the lepton momentum in the B_{sig} rest frame for the individual Monte Carlo contributions (left) and the pre-fit distribution with the three fit components (right) for the $B^0 \to \tau\mu$ channel.

The background Monte Carlo events used in the PDF creation are weighted to match their expectations in the real data, and signal events are scaled down by a factor of 50, corresponding to a branching ratio of $6.67 \cdot 10^{-5}$. The weighted number of events for the three PDFs are provided in Tab. 7.1 for the $B^0 \to \tau e$ and $B^0 \to \tau \mu$ channel.

	$B_{sig} \to \tau e$	$B_{sig} \to \tau \mu$
signal	133.96	105.26
ulnu background	563.08	278.30
other background	3771.58	430.82

Table 7.1: Monte Carlo signal and background events used for the PDF creation in the $B^0 \to \tau e$ and the $B^0 \to \tau \mu$ channel.

7.1 Model Parameters

The model PDF $f\left(x_i; \vec{\theta}\right)$, combined from the three PDFs, contains the parameters $\vec{\theta}$, which can be divided into free parameters $\vec{\mu}$ and constrained parameters $\vec{\chi}$.

$$f\left(x_i; \vec{\theta}\right) = f\left(x_i; \vec{\mu}, \vec{\chi}\right) \tag{7.2}$$

The free parameters are the signal and two background strengths. The strength of a component i is defined as:

$$\mu_i = \frac{N_{\text{fit}_i}}{N_{\text{PDF}_i}}$$

where N_{fit_i} represents number of fitted events for component i and N_{PDF_i} denotes the number of events of component i used in the PDF creation. The constrained parameters include additional uncertainties on the PDF shape and are discussed in the following section. They rely on the Monte Carlo corrections described in section 4.5.

7.2 Global Uncertainties

Global uncertainties preserve the shape of the PDF and only modify the normalization. They are Gaussian-constrained parameters centered around the mean of one and a width of σ, which denotes the relative uncertainty on the normalization due to a specific uncertainty source.

Track and π^0 Finding Efficiency

Each charged particle on the signal side is assigned an uncertainty $\sigma_{\text{track}} = 0.0035$, resulting in a global uncertainty, which is defined as:

$$\sigma_{\text{all,track}} = \frac{\mathrm{N}_2 \cdot (1-\sigma_{\text{track}})^2 - \mathrm{N}_2 + \mathrm{N}_4 \cdot (1-\sigma_{\text{track}})^4 - \mathrm{N}_4}{\mathrm{N}_2 + \mathrm{N}_4} \tag{7.3}$$

where N_2 denotes the number of events with two charged particles, which includes all events with reconstructed τ decays apart from $\tau \to \pi\pi\pi$ decays with three charged pions. These events with four charged tracks are denoted by N_4.

The finding efficiency for π^0 in Monte Carlo is corrected by the ratio $\mathrm{R} = 0.957 \pm 0.008 \pm 0.023$ per π^0 on the signal side, where the second term corresponds to the statistical and the third term on the systematic uncertainty. The combined uncertainty on the ratio is 0.024 with a relative uncertainty on the ratio of $\sigma_{\pi^0} = 0.025$. Based on the number of events with one π^0 ($\mathrm{N}_{1\pi^0}$) or two π^0 ($\mathrm{N}_{2\pi^0}$) in the τ decays, an uncertainty on the normalization is calculated:

$$\sigma_{\pi^0,\text{total}} = \frac{\mathrm{N}_{1\pi^0} \cdot (1-\sigma_{\pi^0}) - \mathrm{N}_{1\pi^0} + \mathrm{N}_{2\pi^0} \cdot (1-\sigma_{\pi^0})^2 - \mathrm{N}_{2\pi^0}}{\mathrm{N}_{\text{total}}} \tag{7.4}$$

where $\mathrm{N}_{\text{total}}$ denotes all events.

N_{BB} and f_{00}

For the signal PDF, additional constraints are included, which depend on uncertainties on the number of $\mathrm{B\overline{B}}$ pair events in the real data:

$$\mathrm{N}_{\mathrm{BB}} = 771,581,000 \pm 10,566,000 \to \sigma_{\mathrm{N}_{\mathrm{BB}}} = 0.0137 \tag{7.5}$$

and the uncertainty on f_{00}

$$\mathrm{f}_{00} = 0.486 \pm 0.006 \to \sigma_{\mathrm{f}_{00}} = 0.0123 \tag{7.6}$$

B Meson Background Suppression and FEI

Because the signal efficiency is calculated on signal Monte Carlo events, it depends on its correct representation of the signal events in real data. The selection of a specific value of the B meson background suppression classifier output introduces an additional global uncertainty on the signal efficiency, which was determined with the approach explained in chapter 5.3.6. Tab. 7.2 lists the relative uncertainties on the signal efficiency for the $\mathrm{B}^0 \to \tau e$ and $\mathrm{B}^0 \to \tau\mu$ channel. The uncertainty due to the FEI calibration is included for all three PDF components and given in Tab. 7.3.

	$\sigma_{\text{BDT eff.}}$
$B^0 \to \tau e$	0.0206
$B^0 \to \tau \mu$	0.0690

Table 7.2: Uncertainty on the signal efficiency introduced by the B meson background suppression selection.

	FEI correction	σ_{FEI}
$B^0 \to \tau e$	0.7796±0.0415	0.0532
$B^0 \to \tau \mu$	0.7871±0.0421	0.0535

Table 7.3: FEI calibration factor and relative uncertainty σ_{FEI}.

7.3 Local Uncertainties

In addition to global uncertainty, local uncertainties altering the shape of the PDFs are implemented. Therefore, the difference in the lepton momentum distribution in the Monte Carlo data used for the PDF generation with nominal and varied corrections is determined. Each Monte Carlo event is assigned a nominal weight, which is the product of all individual correction factors. A variation of the weight is calculated based on the uncertainty introduced by a specific source.

$$w_i = w_{\text{nom},i} + \delta w_i \tag{7.7}$$

where $w_{\text{nom},i}$ is the nominal weight for an event i and δw_i is the uncertainty on the event weight due to a specific uncertainty source.

The shape uncertainties are divided into correlated and uncorrelated shape uncertainties. Correlated shape uncertainties add one extra parameter to the PDF because the variation of the shape is correlated across all lepton momentum bins. In contrast, uncorrelated parameters introduce 27 new parameters because each lepton momentum bin is treated independently. If the correction factor depends on multiple correlated parameters, uncorrelated variations are determined by diagonalizing the covariance matrix, as explained in section 4.5.2.

Finite Size of the Monte Carlo Data

Because the PDFs are created from finite Monte Carlo data, a Poisson uncertainty is assigned to the number of events in each bin. These uncertainties are uncorrelated shape uncertainty, resulting in 27 Poisson-constrained parameters per fit component. If a bin has zero entries for a fit component, the Poisson uncertainty is zero.

Lepton Identification & Lepton Fake Rate

The corrections on the electron and muon identification for correctly identified leptons have an associated statistical and systematic uncertainty for each particle identification (pid) bin.

Since the systematic uncertainty on the lepton identification is assumed to be 100% correlated across pid bins for each identified lepton type, the uncertainty on the entries across the lepton momentum bins and the individual PDF components is 100% correlated. The parameters for the electron and muon identification are Gaussian-constrained.

The statistical uncertainty of lepton identification correction is uncorrelated for different pid bins. Therefore, a random value from a normal Gaussian distribution is drawn for each pid bin i and multiplied by its statistical uncertainty $\sigma^{\text{stat}}_{\text{pid bin } i}$. The uncertainty on the event weight is determined as follows:

$$\delta w^{\text{stat}}_{\text{pid bin } i} = \sigma^{\text{stat}}_{\text{pid bin } i} \cdot G^{\text{rand}}_{\text{stat}}\ (0,1) \tag{7.8}$$

Afterward, a new lepton momentum distribution is created with the varied weights. This procedure is repeated 50 times. For each lepton momentum bin i, an uncertainty is assigned, calculated as the standard deviation of the number of entries in bin i for the 50 variations. Fig. 7.4 illustrates the variations in the lepton momentum distribution, where the differences caused by the statistical uncertainty of the lepton identification are enhanced by a factor of 50 to make them visible.

This approach assumes independence between lepton momentum bins, which is not entirely correct because leptons from the same pid bin can contribute to different lepton momentum bins. However, this approximation is chosen since the statistical uncertainty is smaller than the systematic uncertainty, and the overall uncertainty is smaller than 1%. In the fit, the statistical uncertainty on the electron and muon identification are combined using Gaussian error propagation to reduce the total number of parameters.

The provided uncertainty on the lepton fake rate correction is solely statistical and included analogously to the previously described statistical uncertainty on lepton identification. It only includes fake rate corrections for true pions and kaons, which fall into the lepton pid bins. All other particles are assigned a lepton fake rate correction of one with an uncertainty of zero. This only occurs in background events, and the maximal fraction of non-true pions or kaons being identified as leptons is less than 0.4%. Tab. 7.4 and Tab. 7.5 list the composition of Monte Carlo particles for the different light leptons types on the signal side, separately for background and signal events.

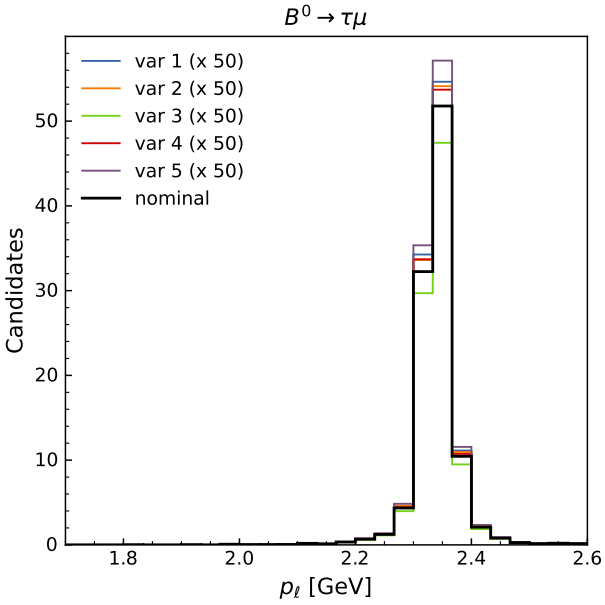

Figure 7.4: Illustration of the procedure to determine the uncertainty due to the statistical uncertainty on the lepton identification, shown for the signal Monte Carlo data in the $B^0 \to \tau\mu$ channel. The lepton momentum distribution is shown with nominal weights (black) and the five varied weights (colors), where the variations are enhanced by a factor of 50 to make them visible.

Lepton origin	MC particle	reconstructed $B^0 \to \tau e$	reconstructed $B^0 \to \tau\mu$
$B^0 \to \tau\ell$	ℓ^-	99.29%	88.09%
	π^-	0.67%	8.66%
	K^-	0.00%	3.20%
	other	0.04%	0.05%
$\tau \to e\nu\nu$	e^-	98.89%	99.47%
	π^-	0.85%	0.53%
	K^-	0.21%	0.00%
	other	0.05%	0.00%
$\tau \to \mu\nu\nu$	μ^-	88.15%	86.24%
	π^-	10.61%	12.31%
	K^-	0.97%	1.07%
	other	0.27%	0.38%

Table 7.4: Monte Carlo information on signal-side light lepton candidates in background events.

Lepton origin	MC particle	reconstructed $B^0 \to \tau e$	reconstructed $B^0 \to \tau \mu$
$B^0 \to \tau \ell$	ℓ^-	99.98%	99.87%
	π^-	0.02%	0.13%
$\tau \to e\nu\nu$	e^-	99.69%	99.93%
	π^-	0.31%	0.07%
$\tau \to \mu\nu\nu$	μ^-	95.14%	96.86%
	π^-	4.44%	3.00%
	K^-	0.42%	0.14%

Table 7.5: Monte Carlo information on signal-side light lepton candidates in signal events.

Low Momentum Particle Finding Efficiency

For particles with a $p_t < 0.2$ GeV, a low momentum finding efficiency correction in six distinct low momentum bins is applied. Since the uncertainties on the correction factors are partially correlated across the low momentum bins, six uncorrelated variations are determined by diagonalizing the covariance matrix, resulting in six Gaussian-constrained parameters in the model PDF, which are shared across all PDF components and lepton momentum bins.

Form Factors

Furthermore, uncertainties arising from the uncertainties of the form factor parameters of $B \to D^{(*)}\ell\nu$ decays are included. The uncorrelated variations of the form factor parameters are calculated by diagonalizing the covariance matrix, resulting in five variations for $B \to D\ell\nu$ and six variations for the $B \to D^*\ell\nu$ form factor parameters. In total, eleven Gaussian-constrained parameters are added to the model PDF, where the parameter of the same variation represents the correlation between the lepton momentum bins. The Gaussian-constrained parameters are shared across PDF components.

Branching Ratio of τ Decays

Six Gaussian-constrained parameters are added to the signal PDF to account for the uncertainties of τ branching ratios. These uncertainties are calculated by varying the weight of signal events based on Monte Carlo information of the generated τ decays. For each reconstructed τ decay mode i, the events generated in the decay mode i are reweighted according to their branching ratio uncertainty σ_i:

$$\mathcal{B}_i \to \mathcal{B}_i + \sigma_i \tag{7.9}$$

To conserve the total τ branching ratio, all events generated in not reconstructed (nr) τ decay modes are reweighted with the opposite sign of σ_i:

$$\mathcal{B}_{nr} \rightarrow \mathcal{B}_{nr} - \sigma_i \tag{7.10}$$

The constrained parameters due to the uncertainties of the τ branching ratios are assumed to be uncorrelated.

Branching Ratio of Resonant $B \rightarrow X_u \ell\nu$ Decays

The same procedure as applied for the τ branching ratio is implemented for the uncertainties on the exclusive branching ratios of B decays via the $(b \rightarrow u\ell\nu)$ transitions. Here, the non-resonant decay modes replace the not-reconstructed τ decay modes. Tab. 7.6 lists the branching ratio and the respective uncertainties for the resonant $B \rightarrow X_u \ell\nu$ decays.

B meson	X_u particle	$\mathcal{B}\,(\times 10^{-3})$
Neutral	π^+	0.150 ± 0.006
	ρ^+	0.294 ± 0.021
Charged	π^0	0.078 ± 0.003
	ρ^0	0.158 ± 0.011
	η	0.039 ± 0.005
	η'	0.023 ± 0.008
	ω	0.119 ± 0.009

Table 7.6: Branching ratios of the resonant $B \rightarrow X_u \ell\nu$from Tab. 4.5.

Since most of the resonant $B(b \rightarrow u\ell\nu)$ events arise from neutral B meson events, which decay in the resonant $B \rightarrow \pi\ell\nu$ and $B \rightarrow \rho\ell\nu$ channels, only three constrained parameters are included in the "ulnu bkg" PDF. The third parameter contains the uncertainty on the shape for all additional resonant decays of the charged B mesons $B \rightarrow X_{u,\ other\ res}\ell\nu$ with $X_{u,\ other\ res} = \eta, \omega, \eta'$.

Fig. 7.5 - 7.6 show the composition of the different $B \rightarrow X_u \ell\nu$ events in the $B^0 \rightarrow \tau e$ and $B^0 \rightarrow \tau\mu$ channel.

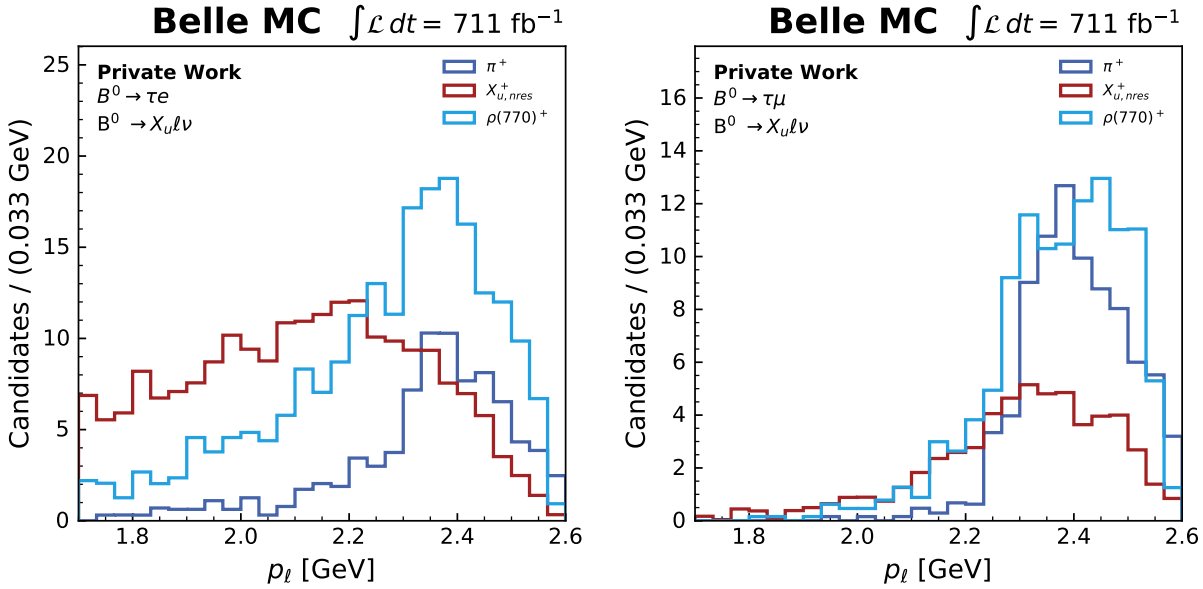

Figure 7.5: Distribution of the lepton momentum for $B^0(b \to u\ell\nu)$ events in the decays for the reconstructed $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) channel.

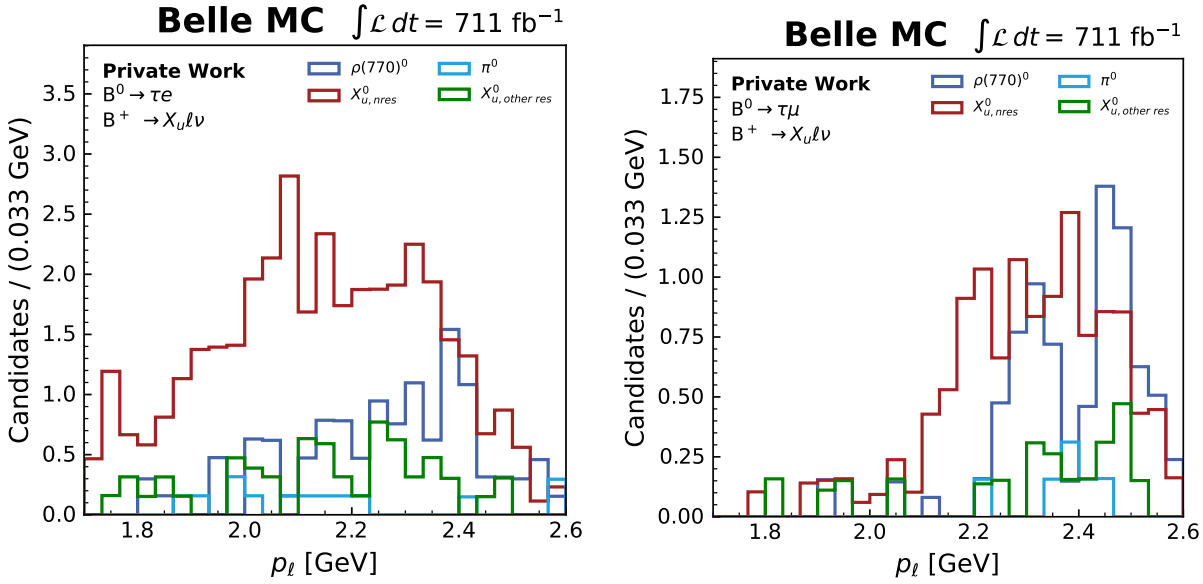

Figure 7.6: Distribution of the lepton momentum for $B^+(b \to u\ell\nu)$ events in the decays for the reconstructed $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) channel.

Branching Ratio of Inclusive $B \to X_u \ell \nu$ Decays

The non-resonant $B \to X_u \ell \nu$ decays are generated according to the inclusive branching ratio. Because the branching ratio also has an associated uncertainty, new hybrid weights for the non-resonant $B \to X_u \ell \nu$ events are calculated by varying the inclusive branching ratio by its uncertainty. The correlated shape difference in the lepton momentum distribution is included as a Gaussian parameter in the "ulnu bkg" PDF.

Model for Non-resonant $B \to X_u \ell \nu$ Decays

The uncertainties on the DFN model parameters introduce two Gaussian-constrained parameters for the correlated shape uncertainties on the "ulnu bkg" PDF for non-resonant B ($b \to u\ell\nu$) events. Additionally, a third shape uncertainty is included, representing how the choice of the theory model in the generation of the non-resonant $b \to u\ell u$ decays influences the lepton momentum distribution. Therefore, the non-resonant B($b \to u\ell\nu$) events are reweighted according to the BLNP model prediction, and the difference in the lepton momentum distribution with respect to the DFN model prediction is included as a correlated shape uncertainty.

Data - Monte Carlo Differences in the Background Suppression Variables

A correlated shape uncertainty is included, which takes into account the data - Monte Carlo discrepancies in the training variables of the B meson background suppression. By calculating the difference between the lepton momentum distributions with and without reweighting the Monte Carlo events to match the data as described in section 5.3.6, a Gaussian-constrained parameter is determined. The constrained parameters are shared across the PDF components.

7.4 Fit Validation

A toy study is implemented to validate that the fit correctly estimates the statistical uncertainties of the signal yields. Since no signal events are expected in real data, this validation is performed on Asimov data with zero signal events. A fit data set corresponding to a luminosity of 711 fb^{-1} is generated from the same background Monte Carlo events used in the background PDF creation.

For each bin i of the lepton momentum, the nominal number of entries N_i is varied by the random value drawn from a Poisson distribution with $\mu = N_i$. Fig. 7.7 illustrates the nominal distribution from which the three variations of the fit data sets are generated. If a bin has negative entries, it is replaced by zero entries.

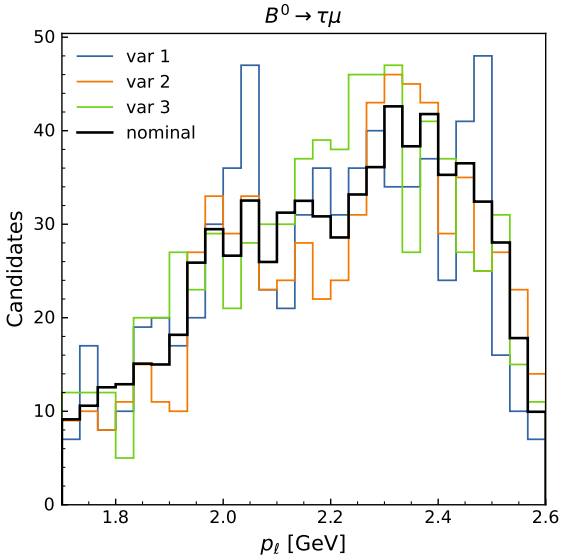

Figure 7.7: The nominal distribution of the lepton momentum for the muon channel is shown in black, and the colored distributions represent three toy data sets containing zero signal events.

The original model PDF is then fitted to the toy data, where the only parameters included in the PDF are the three yield parameters. The complete procedure, starting from the fit data generation, is repeated 30,000 times. The pull for each of these fits is determined as:

$$\text{pull} = \frac{\text{N}_{\text{sig,fit}} - \text{N}_{\text{sig,gen}}}{\sigma_{\text{N}_{\text{sig,fit}}}}. \tag{7.11}$$

The pull distributions of the signal events for the $\text{B}^0 \to \tau e$ and the $\text{B}^0 \to \tau\mu$ channel are shown in Fig. 7.8.

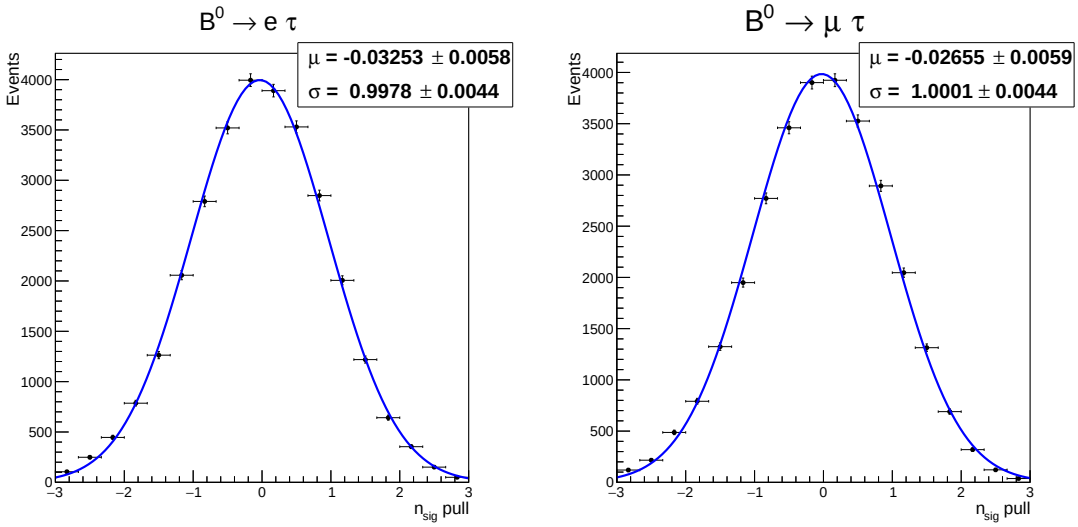

Figure 7.8: Pull distribution of signal events, where the toy fit data consists only of background events, shown for the $B^0 \to \tau e$ (left) and $B^0 \to \tau \mu$ (right) channel. The fitted mean μ and standard deviation σ are shown in the upper right of each plot.

Fig. 7.9 shows the distribution of the signal strength uncertainties obtained in the toy study.

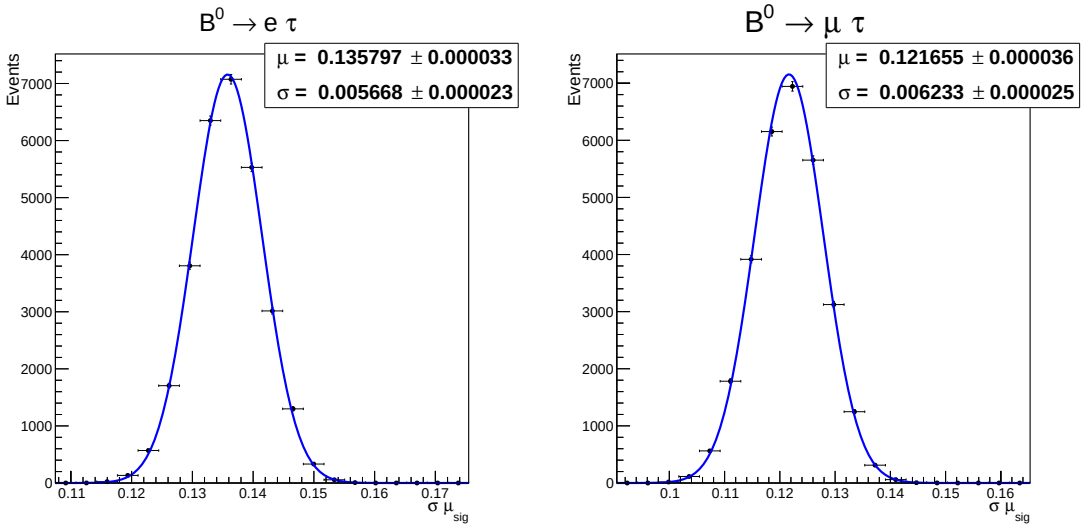

Figure 7.9: Distribution of the signal strength uncertainty obtained in the toy study. The toy fit data consists only of background events, shown for the $B^0 \to \tau e$ (left) and $B^0 \to \tau \mu$ (right) channel. The fitted mean μ and standard deviation σ are shown in the upper right of each plot.

In the case of a correct uncertainty estimation, the pull distribution is described by a normal Gaussian distribution. The unbinned maximum likelihood fit of the pull distributions shows that the width is consistent with one, but the mean is slightly biased toward negative values. Since including an overall global uncertainty on the signal strength $\sigma_{\text{bias}} = |\mu_{\text{pull}}| \cdot \sigma_{\mu}$ has a negligible impact compared to the other uncertainties, no additional uncertainty due to the observed bias is added to the signal PDF.

In another study, the stability of the fit is tested by generating Asimov data sets with zero to twenty signal events, while the background events correspond to their expectation from Monte Carlo events. A fit of the model PDF, including all parameters, is performed for each Asimov data set. Fig. 7.10 compares the generated signal events to the fitted signal yields. The fit is assumed to be stable since it returns the expected numbers of fitted signal events, and all fits for the 20 different Asimov data sets converged.

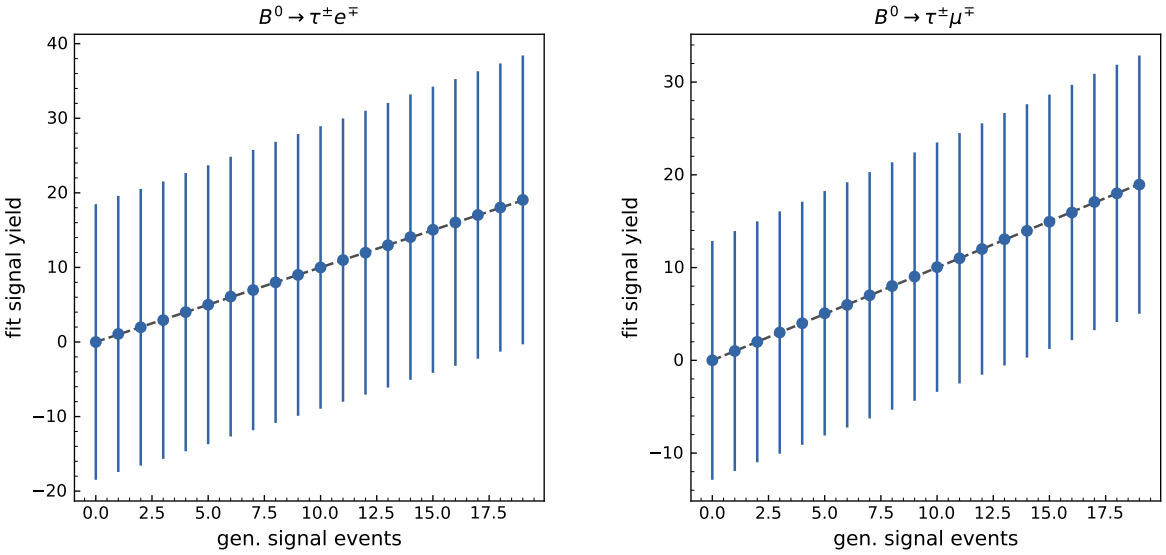

Figure 7.10: Stability test of fit performed on the Asimov fit data with 0 to 20 signal events. The distribution on the left shows the result for the $B^0 \to \tau e$ channel and on the right for the $B^0 \to \tau \mu$ channel. The uncertainties on the signal yield include all systematic uncertainties.

7.5 Fit on Asimov Data

With Asimov data, the expected behavior of the fit can be examined. The results of the fits on Asimov data with zero signal events are presented in Fig. 7.11. As expected the model and data are in perfect agreement.

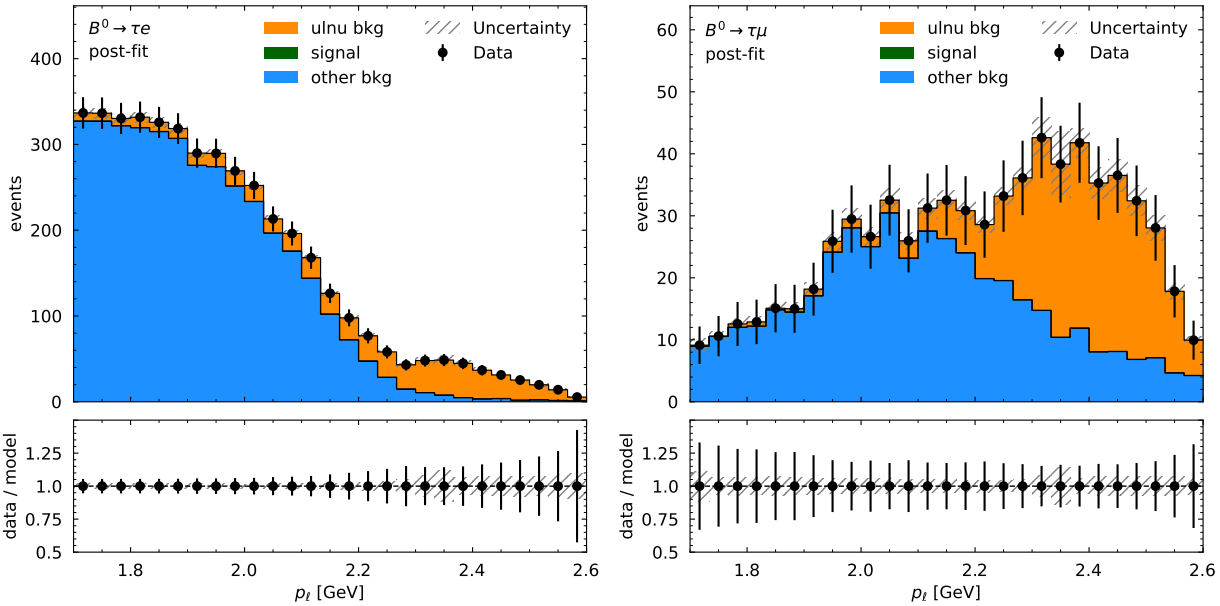

Figure 7.11: Post-fit distribution on Asimov data with zero signal events for the $B^0 \to \tau e$ channel on the left and for the $B^0 \to \tau \mu$ channel on the right.

Tab. 7.7 lists the signal and background yields with their average expected uncertainties. The background yields are equal to the background events used in the background PDF generation and the signal events are zero.

	$B^0 \to \tau e$	$B^0 \to \tau \mu$
signal	0.00 ± 18.47	0.00 ± 12.81
ulnu background	563.08 ± 48.48	278.30 ± 27.96
other background	3771.58± 69.29	430.82 ± 27.04

Table 7.7: Fitted yields for signal and background components on the Asimov data set for the $B^0 \to \tau e$ and $B^0 \to \tau \mu$ channel.

7.5.1 Expected Upper Limit on the Signal Branching Ratio

The expected upper limit on the branching ratio is determined based on the CL_s method implemented in pyhf. The upper limit determination is tested on the Asimov data containing zero signal events.

In 16 hypothesis tests with signal strengths varying in equidistant steps between 0.00 and 0.30, an upper limit on the branching ratio with a 90% confidence level is determined using the $\tilde{q}$ test statistic. This test statistic only considers positive signal yields. The upper limit scan is presented in Fig. 7.12, and the obtained upper limits are listed in Tab. 7.8.

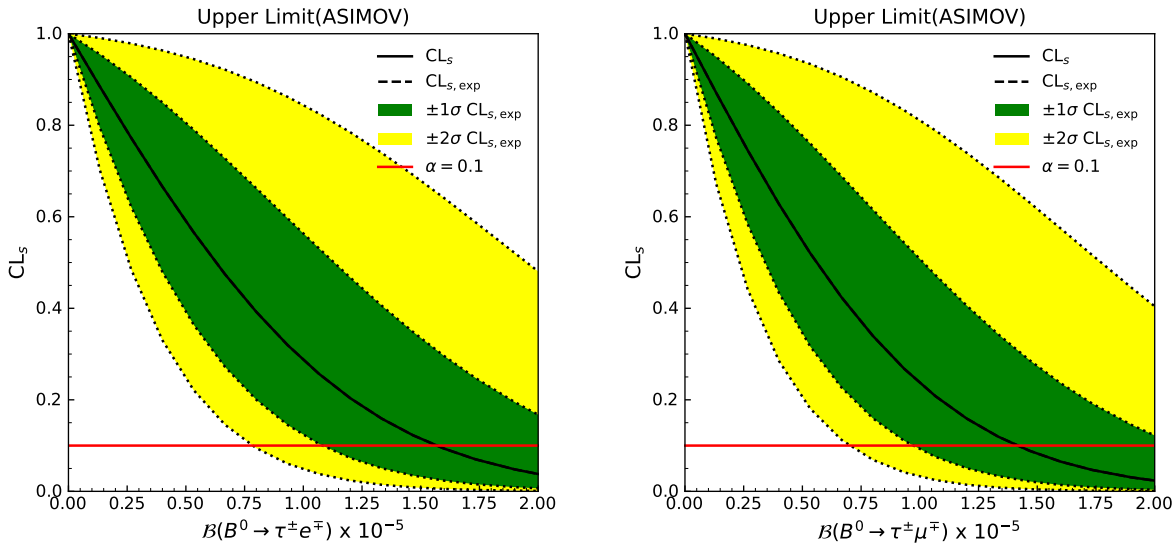

Figure 7.12: Upper limit scan on branching ratio for $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) for a fit on Asimov data with zero signal events. In the upper limit plot, the observed CL_s values are compared with the expected CL_s values for various signal branching ratio hypotheses. The observed CL_s values are obtained from fits to the fit data, while the expected CL_s values are determined on pseudo experiments. Additionally, the expected CL_s values are determined with variations corresponding to $\pm$ one σ (green band) and $\pm$ two σ (yellow band) to represent the uncertainty due to statistical fluctuations. The red line denotes the significance α of the test. The branching ratio value, where the curve of the observed CL_s values and α intersect, denotes the upper limit on signal branching. The true branching ratio value is expected to be below this upper limit with a probability of 90%.

	$B^0 \to \tau e$	$B^0 \to \tau \mu$
$\mathcal{B}$ limit 90% CL	$1.57 \cdot 10^{-5}$	$1.42 \cdot 10^{-5}$

Table 7.8: Upper limit on branching ratios at 90% CL for the electron $B^0 \to \tau e$ and $B^0 \to \tau \mu$ channel obtained on Asimov data with zero signal events.

The impact of the individual systematic uncertainties sources on the upper limit is shown in Fig. 7.13. To estimate the impact, the fit is repeated on the Asimov data with zero signal events, and additional systematic uncertainties are included one after the other, starting with only the yields as parameters of the model PDFs.

For the $B^0 \to \tau e$ and the $B^0 \to \tau \mu$ channel, the upper limit on the branching ratio is limited by the data size of the fit data. Including all the systematic uncertainties only alters the upper limit by $0.02 \cdot 10^{-5}$ – $0.03 \cdot 10^{-5}$.

The systematic uncertainties with the highest impact are the FEI calibration uncertainty and the B meson background suppression uncertainties. In the $B^0 \to \tau e$ channel, the discrepancies for background Monte Carlo and real data in the training variables increase the upper limit. Meanwhile, the uncertainty on the signal efficiency is the dominating source for the increased upper limit on the $\mathcal{B}(B^0 \to \tau \mu)$.

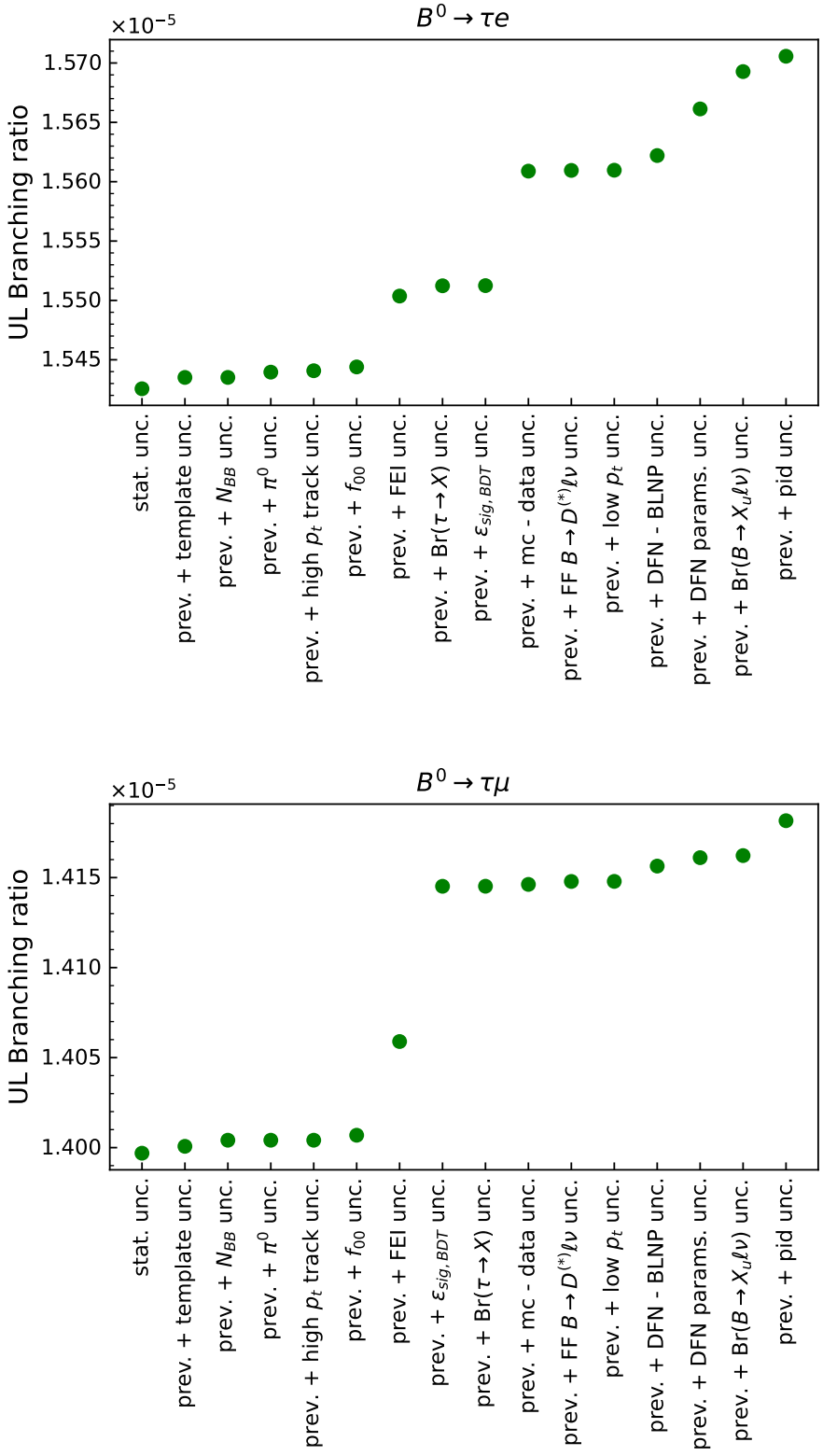

Figure 7.13: Impact of the systematic uncertainties on the upper limit on the branching ratio for $B^0 \to \tau e$ (upper) and $B^0 \to \tau\mu$ (lower) for a fit on Asimov data with zero signal events.

Chapter 8

Results

This chapter presents the unblinding process and the fit to the Belle data. First, a subset of the data is unblinded to evaluate its agreement with Monte Carlo data. For the partial unblinding, the data set of experiment 55 is chosen, which comprises approximately 10% of the Belle data. Afterward, the Belle data is unblinded, and the upper limit on the branching ratio of $B^0 \to \tau e$ and $B^0 \to \tau \mu$ decays is determined.

8.1 Partial Unblinding of Experiment 55

In the process of partial unblinding, the training variables of the B meson background suppression and the lepton momentum in the B_{sig} rest frame are studied, where no significant deviations from the Monte Carlo expectation were observed.

Therefore, the lepton momentum distribution is fitted, and the upper limit on the branching ratios is determined for the Monte Carlo and real data of experiment 55 to check the reliability of the fit. Since the fit data sets only contain a low number of events, especially for the $B^0 \to \tau \mu$ channel, and the fit did not converge if all nuisance parameters were included, the model PDF only contained systematic uncertainty due to limited data size of the Monte Carlo data used in the PDF generation. The obtained results still give a good estimate for the upper limit because the size of the fit data set is the limiting factor.

Fit of $B^0 \to \tau e$

Tab. 8.1 lists the yields for the signal and background contributions obtained in the fit to Monte Carlo and real data in the $B^0 \to \tau e$ channel. The signal yields are consistent with zero within their uncertainties.

	Monte Carlo data	real data
n_{sig}	-3.71 ± 5.91	-2.95 ± 4.77
$n_{u\ell\nu}$	76.06 ± 16.10	46.35 ± 13.48
n_{other}	378.43 ± 21.80	411.73 ± 22.10

Table 8.1: Yields obtained from fits to Monte Carlo and real data of experiment 55 in the $B^0 \to \tau e$ channel.

Fig. 8.1 presents the post-fit distributions of the lepton momentum distribution for the Monte Carlo data and the real data set. The model and data are in good agreement.

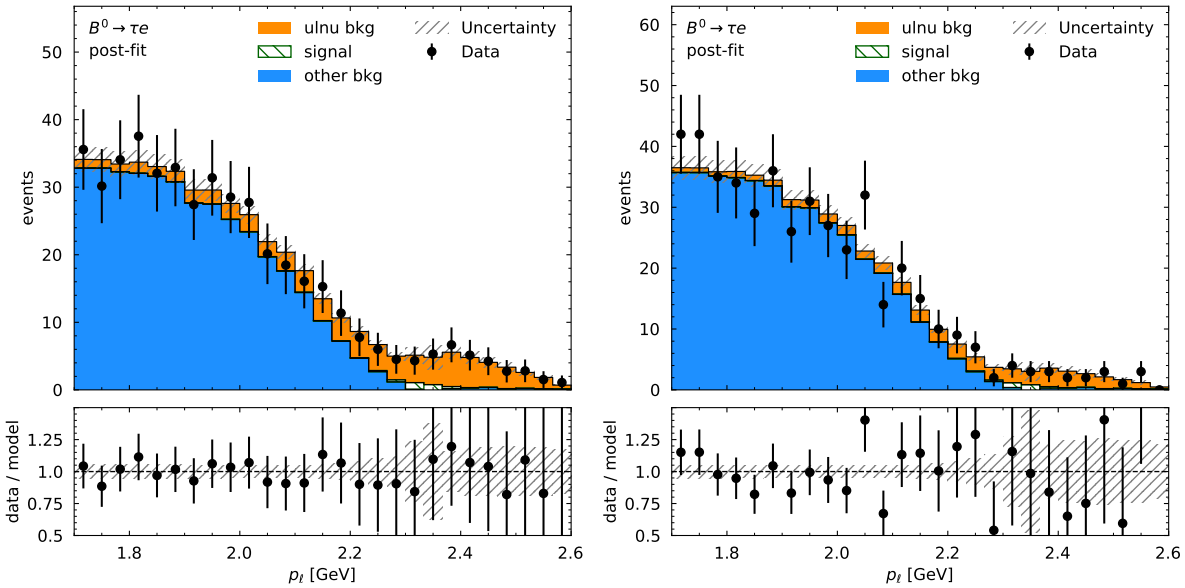

Figure 8.1: Post-fit distribution of the lepton momentum for Monte Carlo data (left) and real data (right) for experiment 55 in the $B^0 \to \tau e$ channel. Signal events are shown in green for positive signal yields and in white-green for negative yields.

Additionally, the upper limits on the $\mathcal{B}(B^0 \to \tau e)$ in Monte Carlo and real data are determined. The expected and observed upper limits with a confidence level of 90% for the Monte Carlo and real data are listed in Tab. 8.2.

$\mathcal{B}$ Limit 90% CL	Monte Carlo data	real data
observed	$4.36 \cdot 10^{-5}$	$3.62 \cdot 10^{-5}$
expected	$5.49 \cdot 10^{-5}$	$4.56 \cdot 10^{-5}$

Table 8.2: Upper limits on the $\mathcal{B}(B^0 \to \tau e)$ obtained from fits to Monte Carlo and real data of experiment 55.

The upper limits obtained from Monte Carlo and real data show no unexpected differences. The scans of the upper limits are presented in Fig. 8.2. Since the extracted signal yields are negative, the observed upper limits are also below the expected upper limits, but they lie in the one-σ band of the expectation.

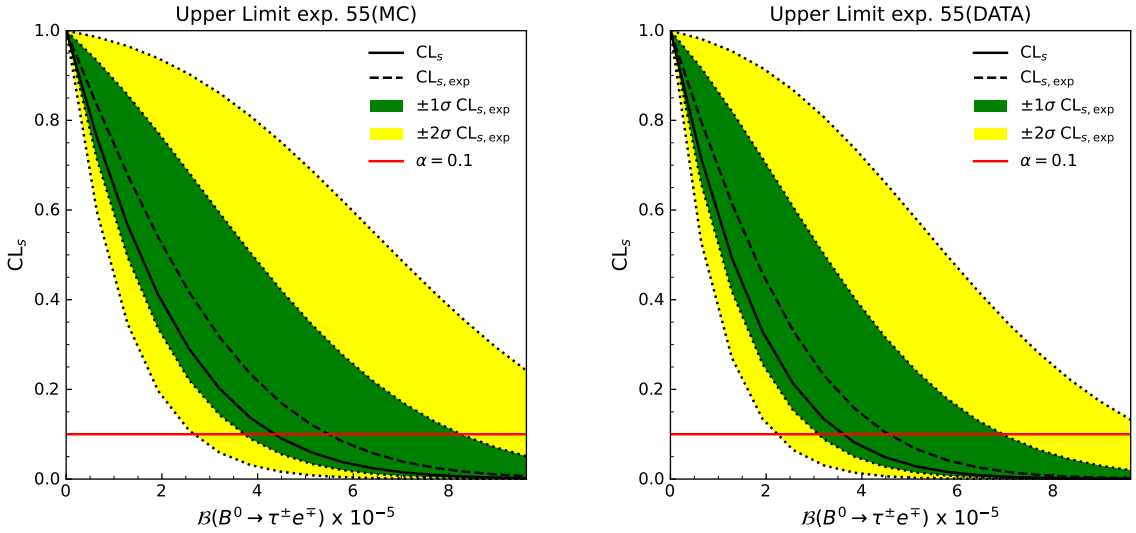

Figure 8.2: Upper limit scan on the branching ratio for Monte Carlo (left) and real data (right) in experiment 55 for the $B^0 \to \tau e$ channel.

Fit of $B^0 \to \tau\mu$

Tab. 8.3 provides the fitted signal and background yields determined from fits to Monte Carlo and real data in the $B^0 \to \tau\mu$ channel. For Monte Carlo data, the signal yield is consistent with zero within its uncertainty. However, the signal yield for the real data is above zero and only consistent with zero within its 1.2 σ uncertainty.

	Monte Carlo data	real data
n_{sig}	-1.83 ± 3.82	5.34 ± 4.52
$n_{u\ell\nu}$	32.33 ± 9.11	22.49 ± 8.04
n_{other}	43.77 ± 8.51	41.20 ± 8.05

Table 8.3: Yields obtained from fits to Monte Carlo and real data of experiment 55 in the $B^0 \to \tau\mu$ channel.

The post-fit distributions of the lepton momentum for Monte Carlo and real data are shown in Fig. 8.3. The Monte Carlo data is in good agreement with the model, while the fit of the real data is affected by the low number of events and statistical fluctuations.

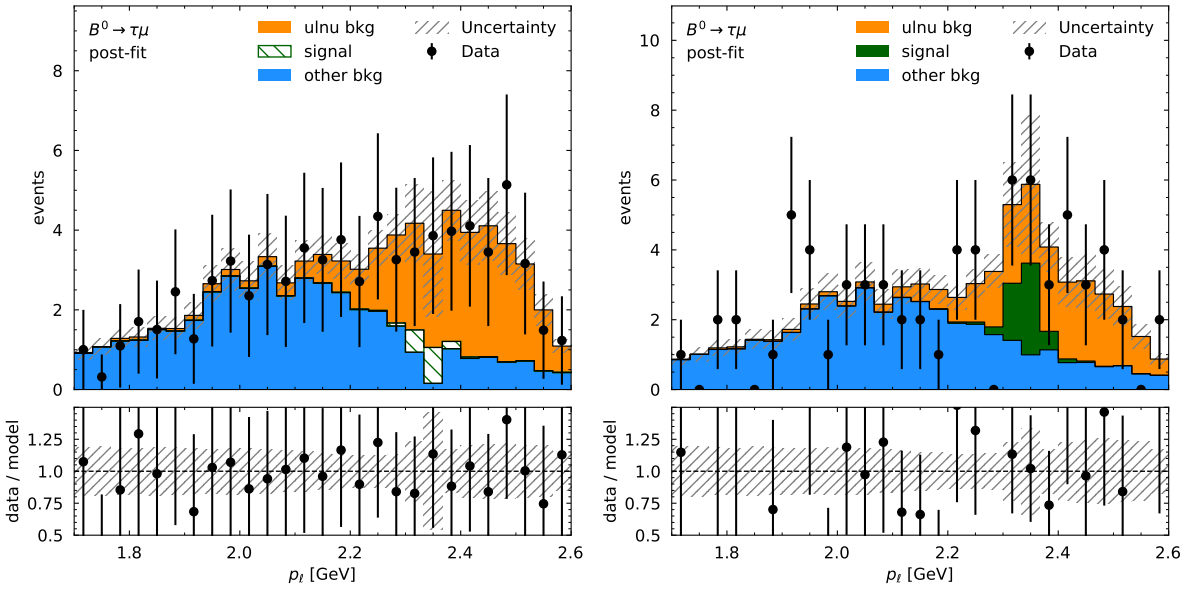

Figure 8.3: Post-fit distribution of the lepton momentum for Monte Carlo data (left) and real data (right) for experiment 55 in the $B^0 \to \tau\mu$ channel. Signal events are shown in green for positive signal yields and in white-green for negative yields.

The expected and observed upper limits on the $\mathcal{B}(B^0 \to \tau\mu)$ with a confidence level of 90% for the Monte Carlo and real data are listed in Tab. 8.4.

$\mathcal{B}$ Limit 90% CL	Monte Carlo data	real data
observed	$3.96 \cdot 10^{-5}$	$7.43 \cdot 10^{-5}$
expected	$4.83 \cdot 10^{-5}$	$4.67 \cdot 10^{-5}$

Table 8.4: Upper limits on the $\mathcal{B}(B^0 \to \tau\mu)$ obtained in experiment 55 for Monte Carlo and real data.

Similar to the $B^0 \to \tau e$ fit, the observed Monte Carlo upper limit is below the expected upper limit but lies within its one-σ uncertainty band. The upper limit observed in real data is above the expected upper limit because the fitted signal yield was above zero. Fig. 8.4 depicts the upper limits scans for Monte Carlo and real data.

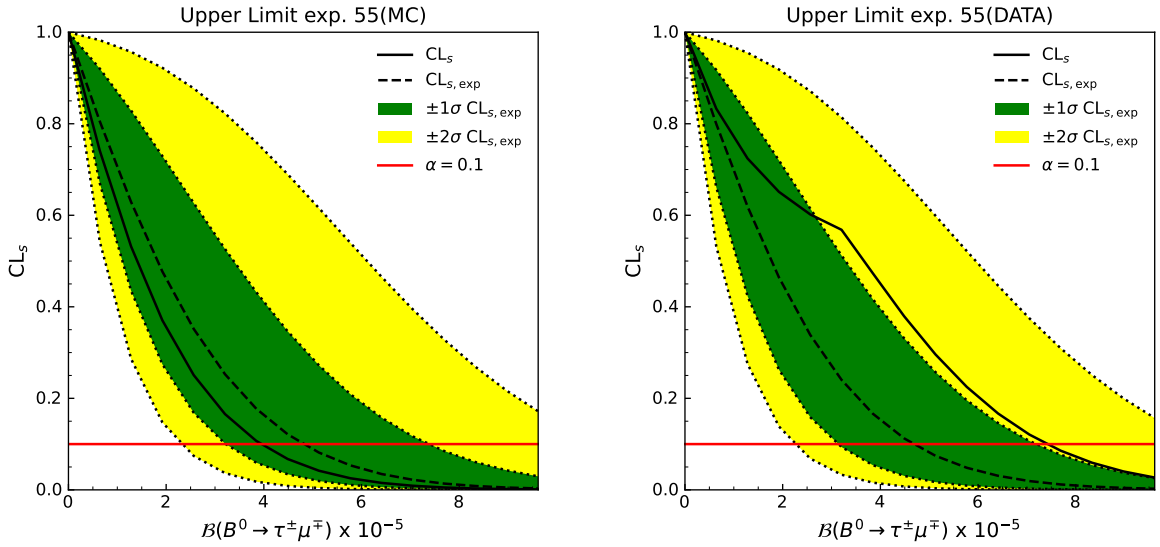

Figure 8.4: Upper limit scan on the branching ratio for Monte Carlo (left) and real data (right) in experiment 55 for the $B^0 \to \tau\mu$ channel.

8.2 Fit of the Belle Data

Because no unexpected behavior was observed during the partial unblinding, the Belle data is unblinded. No significant deviations in the training variables are observed. Similar to the partial unblinding, the distributions of the lepton momentum and the B meson background suppression classifier output are compared for the full fit range of the lepton momentum in the B_{sig} rest frame. The classifier output and the lepton momentum distributions are shown in Fig. 8.5 and in Fig. 8.6.

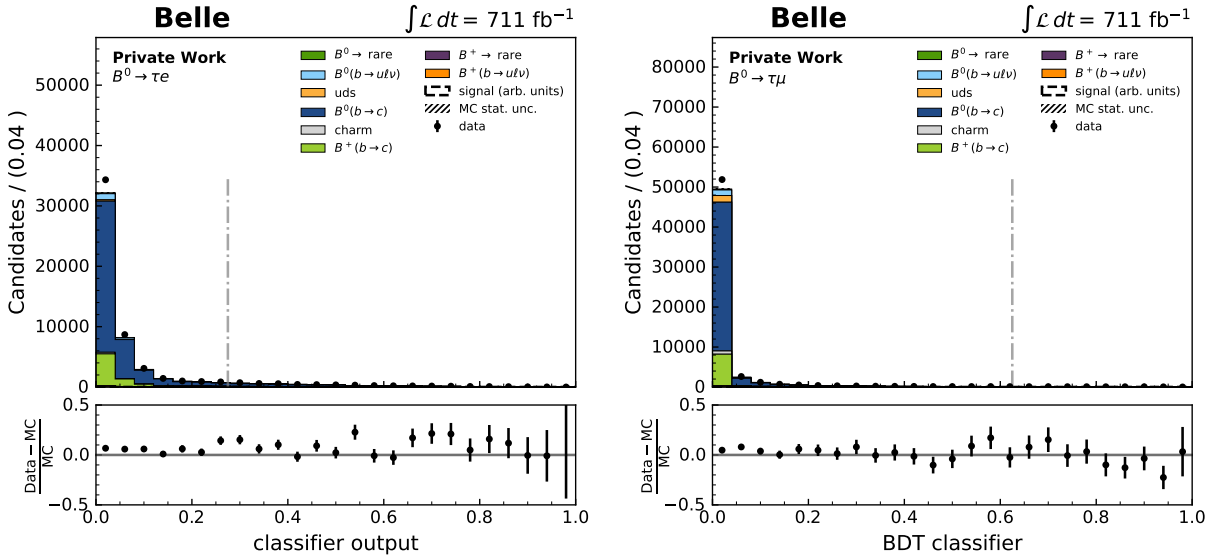

Figure 8.5: Distribution of the B meson background suppression classifiers after unblinding for the $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ channel (right) on the Belle data. The dashed grey line marks the selection on the classifier.

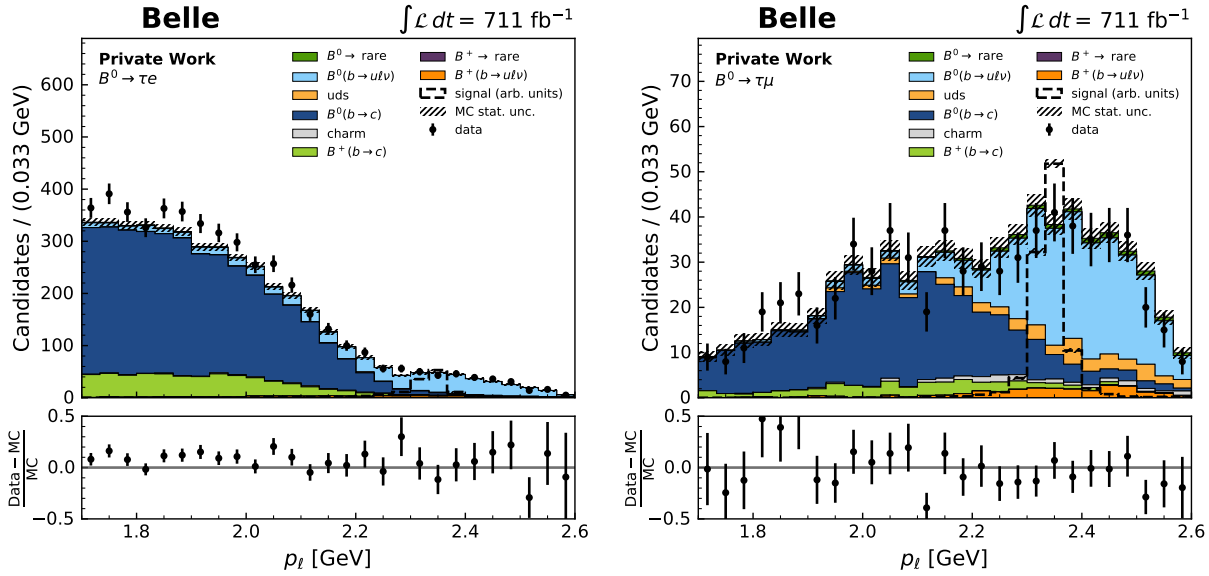

Figure 8.6: Distribution of the lepton Momentum after unblinding $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ channel (right) on the Belle data.

Afterward, the fit is performed on the Belle data. The post-fit distributions are presented in Fig. 8.7, which exhibit a good agreement between model and data.

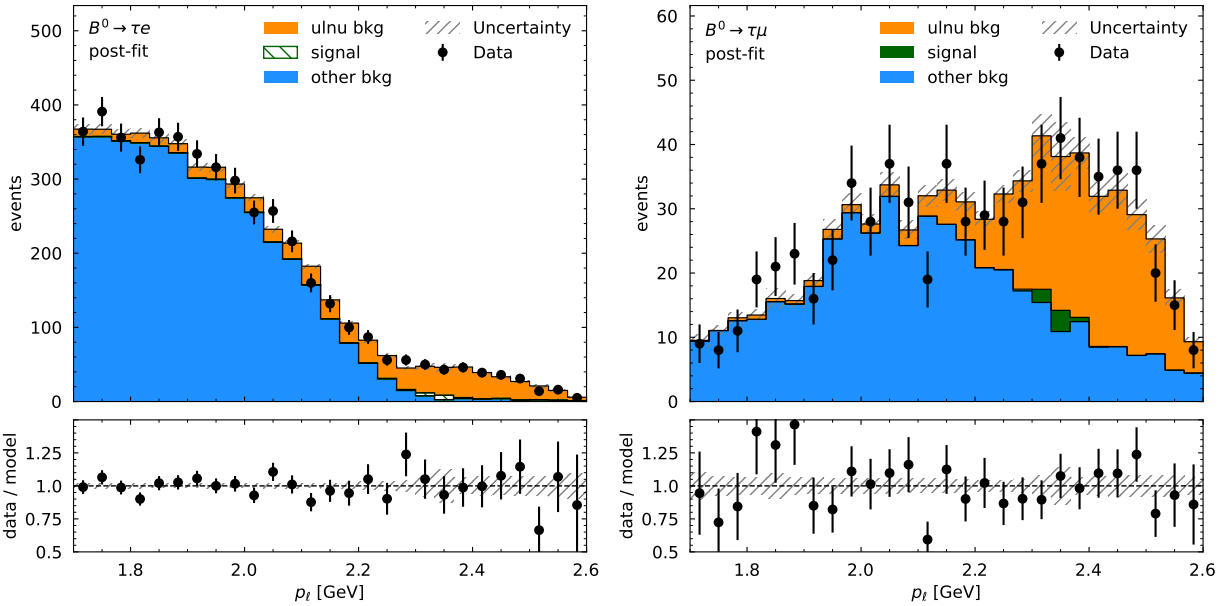

Figure 8.7: Post-fit distribution of the lepton momentum for the $B^0 \to \tau e$ (left) and the $B^0 \to \tau\mu$ channel for the unblinded Belle data. Signal events are shown in green for positive signal yields and in white-green for negative yields.

Tab. 8.5 lists the obtained yields on the signal and background components and the p-value for the goodness of fit test for the Belle data and, as a comparison, the expectation obtained from a fit to Asimov data with zero signal events.

For the $B^0 \to \tau e$ channel, the signal yield is negative but consistent with zero within its uncertainty with a p-value of 0.61 for the goodness of fit test. In the fit of the $B^0 \to \tau\mu$ channel, a positive signal yield is extracted, which is consistent with zero within its uncertainty. The fit has a p-value of 0.53 for the goodness of fit test.

	$B^0 \to \tau e$		$B^0 \to \tau\mu$	
	Belle data	Asimov data	Belle data	Asimov data
n_{sig}	-15.06 ± 18.60	0.00 ± 18.47	6.71 ± 12.48	0.00 ± 12.81
$n_{u\ell\nu}$	597.98 ± 49.09	563.08 ± 48.48	239.07 ± 27.65	278.30 ± 27.96
n_{other}	4121.06 ± 71.53	3771.58 ± 69.29	451.20 ± 27.91	430.82 ± 27.04

Table 8.5: Fitted yields for the fit of the unblinded Belle data and on Asimov data with zero signal events in the $B^0 \to \tau e$ and $B^0 \to \tau\mu$ channel.

Fig. 8.8 illustrates the correlations between the three yield parameters. In particular, a high correlation between the "ulnu bkg" and the "signal" yield parameters is present because the "ulnu bkg" is the dominating background contribution in the signal region.

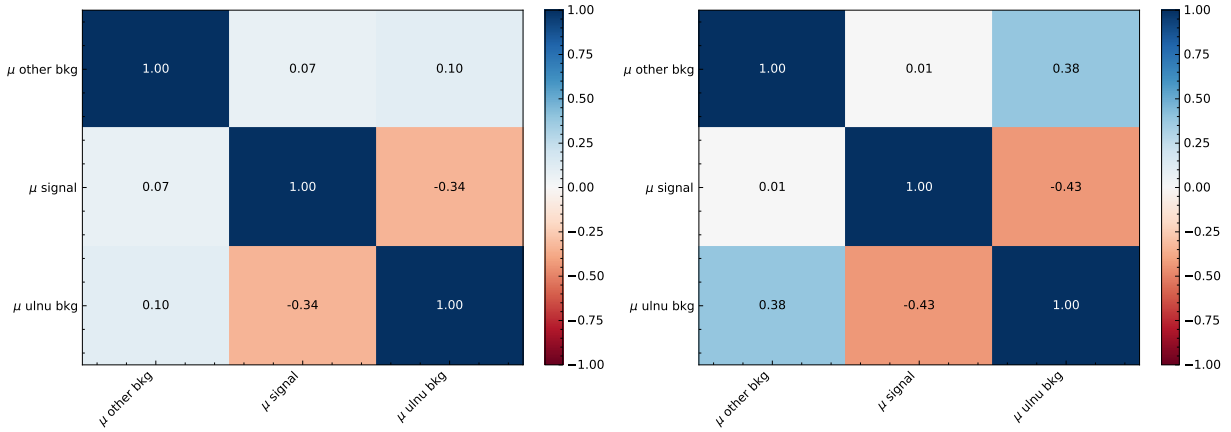

Figure 8.8: Correlation matrix of the yield parameters of the fit of the Belle data in the $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) channel.

In the upper limit scan of the $\mathcal{B}(B^0 \to \tau e)$ and $\mathcal{B}(B^0 \to \tau\mu)$, the $\tilde{q}$ test statistic is used, which allows only positive signal yields. Tab. 8.6 summarizes the final results of the upper limit on the branching ratio of the $B^0 \to \tau e$ and $B^0 \to \tau\mu$ decays.

$\mathcal{B}$ Limit 90% CL	$B^0 \to \tau e$	$B^0 \to \tau\mu$
observed	$1.20 \cdot 10^{-5}$	$1.67 \cdot 10^{-5}$
expected	$1.60 \cdot 10^{-5}$	$1.37 \cdot 10^{-5}$

Table 8.6: Upper limits on the branching ratio for the $B^0 \to \tau e$ and $B^0 \to \tau\mu$ decays obtained from the fit to the Belle data.

Fig.8.9 presents the upper limits scan on the branching ratios of the $B^0 \to \tau e$ and $B^0 \to \tau\mu$ decays obtained from Belle data. The solid black line, defined by the observed CL_s values for the hypothesized values of the branching faction, lies within the one-σ band, indicating no statistically significant departure of the data from the expectation.

As expected from the previous result of the signal yields, the upper limit determined in the $B^0 \to \tau e$ channel is below the expected limit and above the expected limit for the $B^0 \to \tau\mu$ channel.

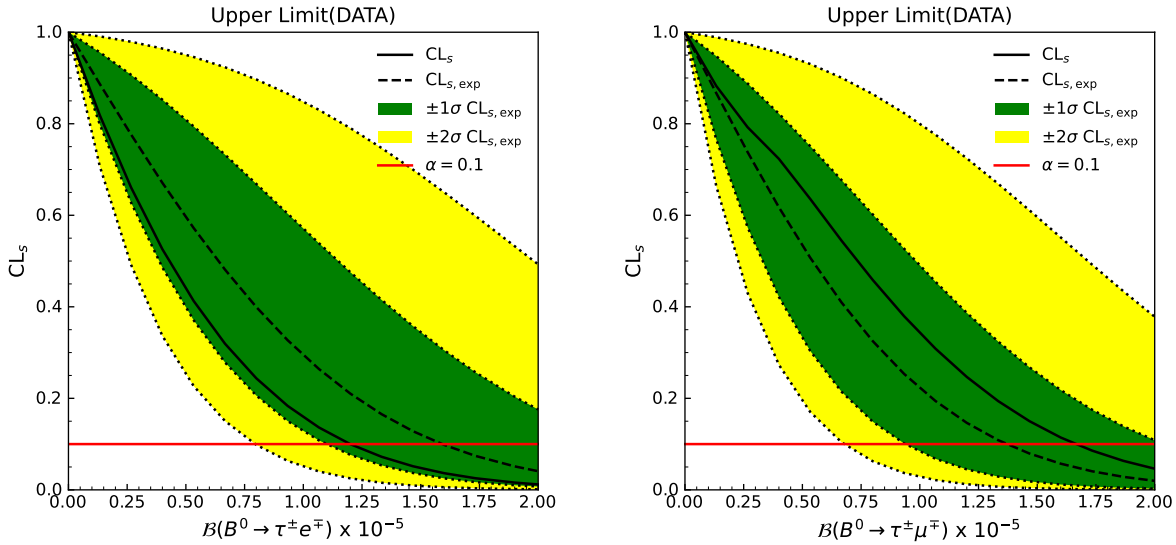

Figure 8.9: Upper limit scan on the branching ratio for the $B^0 \to \tau e$ (left) and $B^0 \to \tau \mu$ channel (right) obtained from the Belle data.

Uncertainty due to Continuum Suppression Selection

Because the continuum suppression might perform differently on real data compared to Monte Carlo events and thus alter the signal efficiency or the shapes of the PDFs, the whole analysis is repeated with cs classifier output > 0.4 and cs classifier output > 0.6 selections. The determined upper limits on the branching ratios, see Tab. 8.7, lie within the one-σ band of the nominal result. Therefore, no additional uncertainty is included in the model PDF. The fit results are shown in the Appendix C.

	$\mathcal{B}$ Limit 90% CL	$B^0 \to \tau e$	$B^0 \to \tau \mu$
cs classifier	observed	1.20×10^{-5}	1.52×10^{-5}
output > 0.4	expected	1.58×10^{-5}	1.37×10^{-5}
cs classifier	observed	1.24×10^{-5}	1.62×10^{-5}
output > 0.6	expected	1.58×10^{-5}	1.35×10^{-5}

Table 8.7: Comparison of the upper limit on the branching ratio for the $B^0 \to \tau e$ and $B^0 \to \tau \mu$ channel obtained from a fit on the Belle data set with the two varied cs classifier selections.

8.3 Discussion of the Results

The obtained upper limit on the branching ratio of $B^0 \to \tau e$ decays of $1.2 \cdot 10^{-5}$ improves the best upper limit of $1.6 \cdot 10^{-5}$ (Belle). For the branching ratio of $B^0 \to \tau\mu$, the upper limit of $1.2 \cdot 10^{-5}$ determined at LHCb remains the best upper limit.

The most limiting factor on the upper limit is the size of the fit data, followed by the systematic uncertainty on the B meson background suppression and the FEI calibration. To determine more stringent upper limits, the size of the fit data needs to increase.

An estimation of the sensitivity for this analysis approach is performed on a data set corresponding to a luminosity of 5 ab^{-1}, which corresponds to the expected Belle II data in the next few years. The fit data only contains the predicted background events. They are determined from the obtained background yields in the Belle data with a luminosity and scaled to their expected rates in a data set for luminosity of 5 ab^{-1}. All the systematic uncertainties in the model PDF are unchanged and the upper limit scan is carried out. The expected upper limits are reduced by a factor of two compared to the Belle data and are given in Tab. 8.8.

	$B^0 \to \tau e$	$B^0 \to \tau\mu$
n_{sig}	0.00 ± 52.59	0.00 ± 33.31
$n_{u\ell\nu}$	4205.64 ± 159.53	1681.31 ± 76.51s
n_{other}	28980.35 ± 210.80	3172.97 ± 78.39
$\mathcal{B}$ Limit 90% CL	$6.28 \cdot 10^{-6}$	$5.02 \cdot 10^{-6}$

Table 8.8: Expected signal and background yields and the expected upper limits obtained from a fit of Asimov data with zero signal events and the expected background events for a luminosity of 5 ab^{-1}.

Additionally, new analysis strategies can be invented to increase the low efficiency of the full hadronic FEI ($\mathcal{O}(10^{-3})$), while still maintaining a good separation power between signal and background events. Since the goal of Belle II is collecting a data set corresponding to the luminosity of 50 ab^{-1}, more stringent upper limits on the branching ratio are expected in the future.

Bibliography

[1] M. Peskin, *Concepts of Elementary Particle Physics*. Oxford Master Series in Condensed Matter Physics Series. Oxford University Press, 2019.

[2] **BaBar** Collaboration, D. Boutigny et al., *The BABAR physics book: Physics at an asymmetric B factory*. 1998.

[3] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.

[4] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587.

[6] Wikipedia contributors, "Standard model — Wikipedia, the free encyclopedia." `https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`, 2024. [Online; accessed 15-April-2024].

[7] A. Salam and J. C. Ward, "Weak and electromagnetic interactions," *Il Nuovo Cimento (1955-1965)* **11** no. 4, (Feb, 1959) 568–577.

[8] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics* **22** no. 4, (1961) 579–588.

[9] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[10] Y. Nambu, "Quasi-Particles and Gauge Invariance in the Theory of Superconductivity," *Phys. Rev.* **117** (Feb, 1960) 648–663.

[11] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Phys. Rev.* **127** (Aug, 1962) 965–970.

[12] A. Pich, "The Standard Model of Electroweak Interactions; rev. version,". https://cds.cern.ch/record/819632.

[13] F. m. c. Englert, "Nobel Lecture: The BEH mechanism and its scalar boson," *Rev. Mod. Phys.* **86** (Jul, 2014) 843–850.

[14] G. Aad and et. al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B* **716** no. 1, (2012) 1–29.

[15] S. Chatrchyan and et. al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B* **716** no. 1, (2012) 30–61.

[16] N. Cabibbo, "Unitary symmetry and leptonic decays," *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533.

[17] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics* **49** no. 2, (02, 1973) 652–657.

[18] **Particle Data Group** Collaboration, R. L. Workman et al., "Review of Particle Physics," *PTEP* **2022** (2022) 083C01.

[19] M. Neubert, "Heavy-Quark Effective Theory," arXiv:hep-ph/9610266 [hep-ph].

[20] L. Canetti, M. Drewes, and M. Shaposhnikov, "Matter and antimatter in the universe," *New Journal of Physics* **14** no. 9, (Sep, 2012) 095012.

[21] **Super-Kamiokande** Collaboration, Y. Fukuda and et. al., "Evidence for Oscillation of Atmospheric Neutrinos," *Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567.

[22] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge," *Zh. Eksp. Teor. Fiz.* **34** (1957) 247.

[23] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π Decay of the K_2^0 Meson," *Phys. Rev. Lett.* **13** (Jul, 1964) 138–140.

[24] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the Unified Model of Elementary Particles," *Progress of Theoretical Physics* **28** no. 5, (Nov, 1962) 870–880.

[25] T. Kajita, "Nobel lecture: Discovery of atmospheric neutrino oscillations," *Rev. Mod. Phys.* **88** (Jul, 2016) 030501.

[26] **Super-Kamiokande** Collaboration, K. Abe et al., "Evidence for the Appearance of Atmospheric Tau Neutrinos in Super-Kamiokande," *Phys. Rev. Lett.* **110** (May, 2013) 181802.

[27] P. F. de Salas, D. V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena, C. A. Ternes, M. Tórtola, and J. W. F. Valle, "2020 global reassessment of the neutrino oscillation picture," *Journal of High Energy Physics* **2021** no. 2, (Feb, 2021) 71.

[28] M. Aker et al., "Direct neutrino-mass measurement with sub-electronvolt sensitivity," *Nature Physics* **18** no. 2, (Feb, 2022) 160–166.

[29] M. Ardu and G. Pezzullo, "Introduction to Charged Lepton Flavor Violation," *Universe* **8** no. 6, (2022) .

[30] A. D. Smirnov, "Vector leptoquark mass limits and branching ratios of K_L^0,B^0,$B_s \to l_i^+ l_j$ decays with account of fermion mixing in leptoquark currents," *Modern Physics Letters A* **33** no. 03, (2018) 1850019.

[31] J. C. Pati and A. Salam, "Lepton number as the fourth "color"," *Phys. Rev. D* **10** (Jul, 1974) 275–289.

[32] I. Doršner, S. Fajfer, A. Greljo, J. Kamenik, and N. Košnik, "Physics of leptoquarks in precision experiments and at particle colliders," *Physics Reports* **641** (2016) 1–68. Physics of leptoquarks in precision experiments and at particle colliders.

[33] **Belle** Collaboration, H. Atmacan et al., "Search for $B^0 \to \tau^\pm \ell^\mp$ ($\ell = e, \mu$) with a hadronic tagging method at Belle," *Phys. Rev. D* **104** no. 9, (2021) L091105.

[34] **LHCb** Collaboration, R. Aaij and et. al., "Search for the Lepton-Flavor-Violating Decays $B_s^0 \to \tau^\pm \mu^\mp$ and $B^0 \to \tau^\pm \mu^\mp$," *Phys. Rev. Lett.* **123** (Nov, 2019) 211801.

[35] G. Cowan, *Statistical data analysis*. 1998.

[36] L. Lista, "Practical Statistics for Particle Physicists," *CERN Yellow Reports: School Proceedings* (2017) Vol 5 (2017): Proceedings of the 2016 European School of High–Energy Physics.

[37] S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The annals of mathematical statistics* **9** no. 1, (1938) 60–62.

[38] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical society* **54** no. 3, (1943) 426–482.

[39] T. Keck, "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification," `arXiv:1609.06119 [cs.LG]`.

[40] H. Deng, Y. Zhou, L. Wang, and C. Zhang, "Ensemble learning for the early prediction of neonatal jaundice with genetic features," *BMC Medical Informatics and Decision Making* **21** (Dez, 2021) .

[41] T. Keck et al., "The Full Event Interpretation," *Computing and Software for Big Science* **3** no. 1, (Feb, 2019) 6.

[42] T. Keck, *Machine learning algorithms for the Belle II experiment and their validation on Belle data*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2017.

[43] A. Abashian et al., "The Belle detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **479** no. 1, (2002) 117–232. Detectors for Asymmetric B-factories.

[44] A. J. Bevan et al., "The Physics of the B Factories," *The European Physical Journal C* **74** no. 11, (Nov, 2014) .

[45] J. Brodzicka et al., "Physics achievements from the Belle experiment," *Progress of Theoretical and Experimental Physics* **2012** no. 1, (Dez, 2012) .

[46] **Belle Collaboration** Collaboration, K. Abe et al., "Observation of mixing-induced CP violation in the neutral B meson system," *Phys. Rev. D* **66** (Aug, 2002) 032007.

[47] R. Abe et al., "The new beampipe for the Belle experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **535** no. 1, (2004) 558–561. Proceedings of the 10th International Vienna Conference on Instrumentation.

[48] Z. Natkaniec et al., "Status of the Belle silicon vertex detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **560** no. 1, (2006) 1–4. Proceedings of the 13th International Workshop on Vertex Detectors.

[49] **Belle II Collaboration** Collaboration, T. Abe et al., "Belle II Technical Design Report," `arXiv:1011.0352 [physics.ins-det]`.

[50] B. Casey, "HadronB." Belle Internal Document.

[51] D. J. Lange, "The EvtGen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **462** no. 1, (2001) 152–155. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[52] T. Sjöstrand, P. Edén, C. Friberg, L. Lönnblad, G. Miu, S. Mrenna, and E. Norrbin, "High-energy-physics event generation with Pythia 6.1," *Computer Physics Communications* **135** no. 2, (2001) 238–259.

[53] E. Barberio and Z. Was, "PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0," *Computer Physics Communications* **79** no. 2, (1994) 291–308.

[54] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, *GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version*. CERN, Geneva, 1987. https://cds.cern.ch/record/1119728.

[55] R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, *GEANT: Detector Description and Simulation Tool*. CERN Program Library. CERN, Geneva, Oct, 1994. http://cds.cern.ch/record/1082634.

[56] F. D. Fazio and M. Neubert, "$B \to X_u \ell \nu_{\bar{\ell}}$ decay distributions to order α_s," *Journal of High Energy Physics* **1999** no. 06, (1999) 017.

[57] Y. Amhis et al., "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018," *The European Physical Journal C* **81** no. 3, (Mar, 2021) .

[58] M. Tanabashi et al., "Review of Particle Physics," *Phys. Rev. D* **98** (Aug, 2018) .

[59] K. Nishimura, "Number of B events in Hadron B(J).". https://belle.kek.jp/secured/nbb/nbb.html [Accessed: 2024-03-10].

[60] "Remarks on Generated Generic MC samples.". https://belle.kek.jp/group/software/MC/genericmc.html [Accessed: 2024-03-10].

[61] I. Caprini, L. Lellouch, and M. Neubert, "Dispersive bounds on the shape of $B \to D^{(*)} \ell \nu$ form-factors," *Nucl.Phys.* **B530** (1998) 153–181.

[62] C. G. Boyd, B. Grinstein, and R. F. Lebed, "Constraints on Form Factors for Exclusive Semileptonic Heavy to Light Meson Decays," *Physical Review Letters* **74** no. 23, (1995) 4603–4606.

[63] C. G. Boyd, B. Grinstein, and R. F. Lebed, "Precision corrections to dispersive bounds on form factor," *Physical Review D* **56** no. 11, (1997) 6895–6911.

[64] M. Prim, "b2-hive/effort v0.1.0," 2020. https://doi.org/10.5281/zenodo.3965699.

[65] B. Aubert et al., "Determination of the form factors for the decay $B^0 \to D^{*-}\ell^+\nu_\ell$ and of the CKM matrix element $|V_{cb}|$," *Phys. Rev. D* **77** (2008) .

[66] R. Glattauer et al., "Measurement of the decay $B \to D\ell\nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$," *Physical Review D* **93** no. 3, (Feb, 2016) .

[67] A. Abdesselam et al., "Precise determination of the CKM matrix element $|V_{cb}|$ with $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle," `arXiv:1702.01521 [hep-ex]`.

[68] **Particle Data Group** Collaboration, "Review of Particle Physics," *Progress of Theoretical and Experimental Physics* **2020** no. 8, (Aug, 2020) .

[69] S. Duell, "Correction of BF(B $\to X_c\ell\nu$) in generic MC." Belle II Internal Website. [Accessed: 2024-04-11].

[70] L. Cao et al., "Measurements of partial branching fractions of inclusive $B \to X_u\ell^+\nu_\ell$ decays with hadronic tagging," *Phys. Rev. D* **104** (Jul, 2021) 012008.

[71] A. L. Kagan and M. Neubert, "QCD anatomy of B to X_s γ decays," *The European Physical Journal C - Particles and Fields* **7** no. 1, (1999) 5–27.

[72] B. O. Lange, M. Neubert, and G. Paz, "Theory of charmless inclusive B decays and the extraction of V_{ub}," *Physical Review D* **72** no. 7, (Oct, 2005) .

[73] B. Bhuyan, "High p_T Tracking Efficiency Using Partially Reconstructed D Decays." Belle Internal Document, 2010.

[74] W. Dungel, "Systematic investigation of the reconstruction efficiency of low momentum $\pi^\pm$ and π^0." Belle Internal Document, 2011.

[75] L.Hinz, C.Jacoby, E.Nakano, and J.Wicht, "Lepton efficiency and systematic error for experiments 21 to 27." Belle Internal Document, 2004.

[76] L. Hinz, "Lepton ID efficiency correction and systematic error." Belle Internal Document, 2006.

[77] P. Urquijo, "Semi-inclusive semi-leptonic B decays." Belle Internal Document, 2012.

[78] S. Nishida, "Study of kaon and pion identification using inclusive D^* sample." Belle Internal Document, 2005.

[79] S. Ryu, "Study of π^0 efficiency using $\tau^- \to \pi^-\pi^0\nu$ events." Belle Internal Document, 2012.

[80] **BABAR Collaboration** Collaboration, B. Aubert and et.al., "Searches for the decays $B^0 \to l^{\pm}\tau^{\mp}$ and $B^+ \to l^+\nu$ ($l = e$, μ) using hadronic tag reconstruction," *Phys. Rev. D* **77** (May, 2008) 091104.

[81] T. B. I. Collaboration, "Belle II Analysis Software Framework (basf2)," 2022.

[82] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, "The Belle II Core Software: Belle II Framework Software Group," *Computing and Software for Big Science* **3** no. 1, (Nov., 2018) . http://dx.doi.org/10.1007/s41781-018-0017-9.

[83] M. Gelb et al., "B2BII: Data Conversion from Belle to Belle II," *Computing and Software for Big Science* **2** no. 1, (Nov, 2018) 9.

[84] M. Röhrken, *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2012.

[85] F. Metzner, *Preparation of a Measurement of $\mathcal{R}(D^{(*)})$ with Leptonic and Hadronic FEI Tag at the Belle Experiment*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2022.

[86] D. M. A. et al., "Search for exclusive charmless hadronic B decays," *Phys. Rev. D* **53** (Feb, 1996) 1039–1050.

[87] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes," *Phys. Rev. Lett.* **41** (Dec, 1978) 1581–1585.

[88] X. Zhou, S. Du, G. Li, and C. Shen, "TopoAna: A generic tool for the event type analysis of inclusive Monte-Carlo samples in high energy physics experiments," *Computer Physics Communications* **258** (2021) 107540.

[89] W. Verkerke and D. P. Kirkby, "The roofit toolkit for data modeling,". https://arxiv.org/abs/physics/0306116.

[90] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples," *Computer Physics Communications* **77** no. 2, (1993) 219–228.

[91] L. Heinrich, M. Feickert, and G. Stark, "scikit-hep/pyhf: v0.7.3," 2023. https://doi.org/10.5281/zenodo.8256635.

[92] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer, "pyhf: pure-Python implementation of HistFactory statistical models," *Journal of Open Source Software* **6** no. 58, (2021) 2823.

[93] A. Held, M. Feickert, H. Schreiner, lhenkelm, A. Hollands, N. Simpson, and muellerr, "scikit-hep/cabinetry: v0.5.2," 2023. https://doi.org/10.5281/zenodo.7791483.

[94] D. Ferlewicz, P. Urquijo, and E. Waheed, "Revisiting fits to $B^0 \to D^{*-}\ell^+\nu_\ell$ to measure $|V_{cb}|$ with novel methods and preliminary LQCD data at nonzero recoil," *Phys. Rev. D* **103** (Apr, 2021) 073005.

A BGL Correlation matrix

The correlation matrices for the parameters were communicated privately with the authors of the paper [94].

	(Nom. $\pm\sigma_{\text{stat}}$) $\times 10^3$	Stat. Correlation Matrix					
$\tilde{a}_0^g$	1.00 ± 0.02	1.000	-0.937	-0.128	0.069	-0.081	0.161
$\tilde{a}_1^g$	-2.35 ± 0.61	-0.937	1.000	0.127	-0.222	0.110	-0.192
$\tilde{a}_0^f$	0.511 ± 0.004	-0.128	0.127	1.000	-0.800	-0.751	0.624
$\tilde{a}_1^f$	0.67 ± 0.17	0.069	-0.222	-0.800	1.000	0.443	-0.354
$\tilde{a}_0^{F_1}$	0.30 ± 0.06	-0.081	0.110	-0.751	0.443	1.000	-0.978
$\tilde{a}_1^{F_1}$	-3.68 ± 1.26	0.161	-0.192	0.624	-0.354	-0.978	1.000

Table A.1: Statistical correlation matrix for the fit with BGL parametrization in the configuration $(1, 1, 2)$.

	(Nom. $\pm\sigma_{\text{syst}}$) $\times 10^3$	Syst. Correlation Matrix					
$\tilde{a}_0^g$	1.00 ± 0.02	1.000	-0.940	-0.132	0.085	-0.077	0.158
$\tilde{a}_1^g$	-2.35 ± 0.66	-0.940	1.000	0.129	-0.228	0.107	-0.189
$\tilde{a}_0^f$	0.511 ± 0.013	-0.132	0.129	1.000	-0.806	-0.755	0.629
$\tilde{a}_1^f$	0.67 ± 0.30	0.085	-0.228	-0.806	1.000	0.452	-0.362
$\tilde{a}_0^{F_1}$	0.30 ± 0.08	-0.077	0.107	-0.755	0.452	1.000	-0.977
$\tilde{a}_1^{F_1}$	-3.68 ± 1.20	0.158	-0.189	0.629	-0.362	-0.977	1.000

Table A.2: Systematic correlation matrix for the fit with BGL parametrization in the configuration $(1, 1, 2)$.

B FEI Calibration: Toy Study

Additional information obtained in the toy study of the FEI calibration:

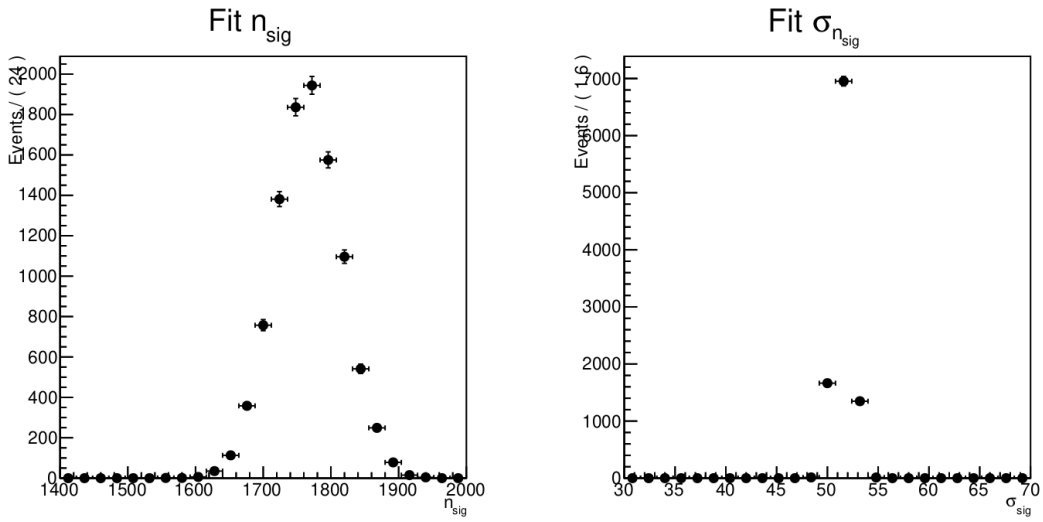

Figure B.1: Distributions of the signal events and the signal uncertainties obtained in the toy study of the FEI calibration in the $B^0 \to D^* e\nu$ channel.

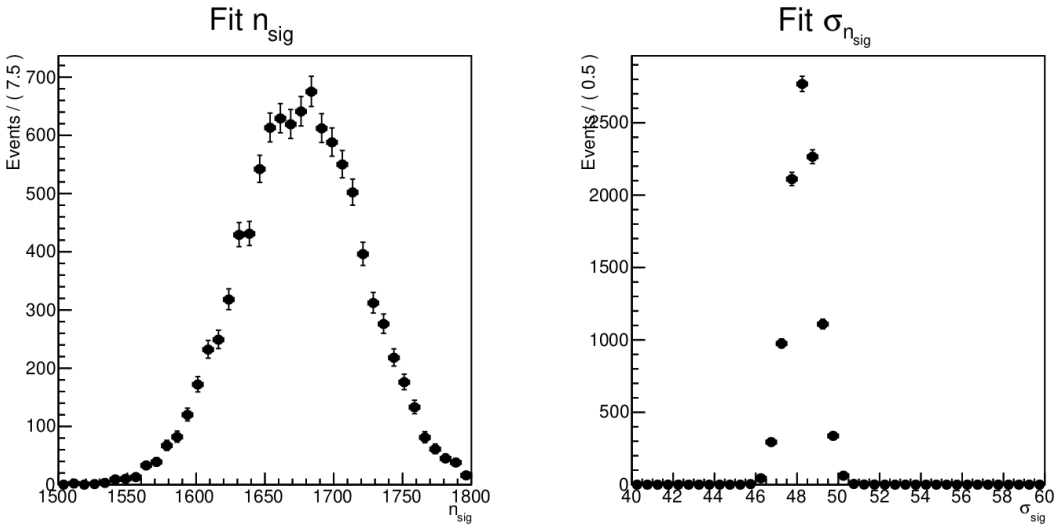

Figure B.2: Distributions of the signal events and the signal uncertainties obtained in the toy study of the FEI calibration in the $B^0 \to D^* \mu\nu$ channel.

C Continuum Suppression Selection

C.1 classifier output > 0.4

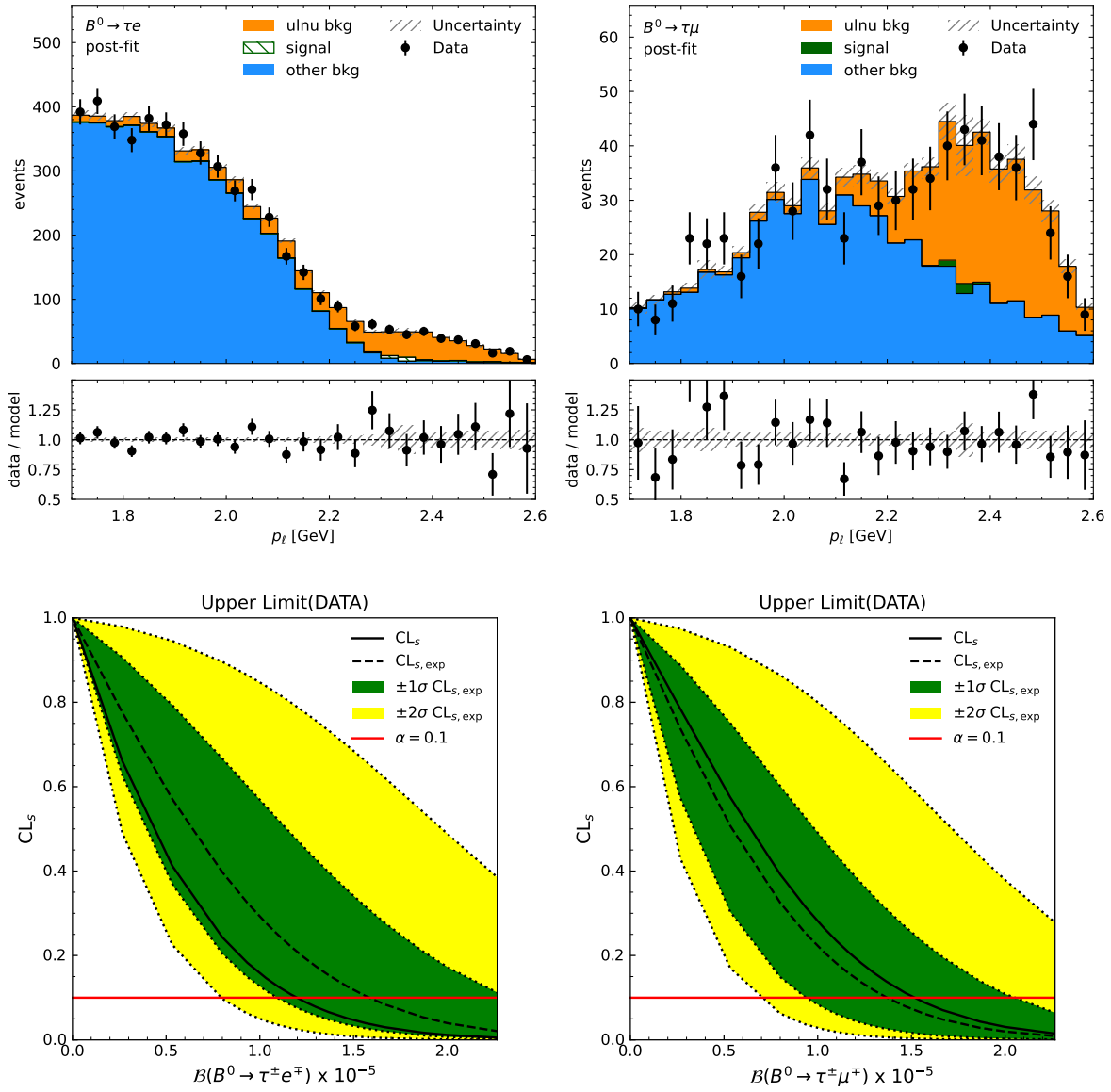

Figure C.1: Post-fit distribution for the fit on full Belle data (upper) and the upper limit scans on branching ratio (lower) for the $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) channels with the cs classifier output > 0.4 selection.

	$B^0 \to \tau e$	$B^0 \to \tau \mu$
n_{sig}	-15.34 $\pm$ 19.07	3.70 $\pm$ 12.97
$n_{u\ell\nu}$	633.57 $\pm$ 50.99	255.84 $\pm$ 28.82
n_{other}	4328.60 $\pm$ 73.67	489.46 $\pm$ 29.08

Table C.1: Fit results on full Belle data for the cs classifier output > 0.4 selection.

C.2 classifier output > 0.6

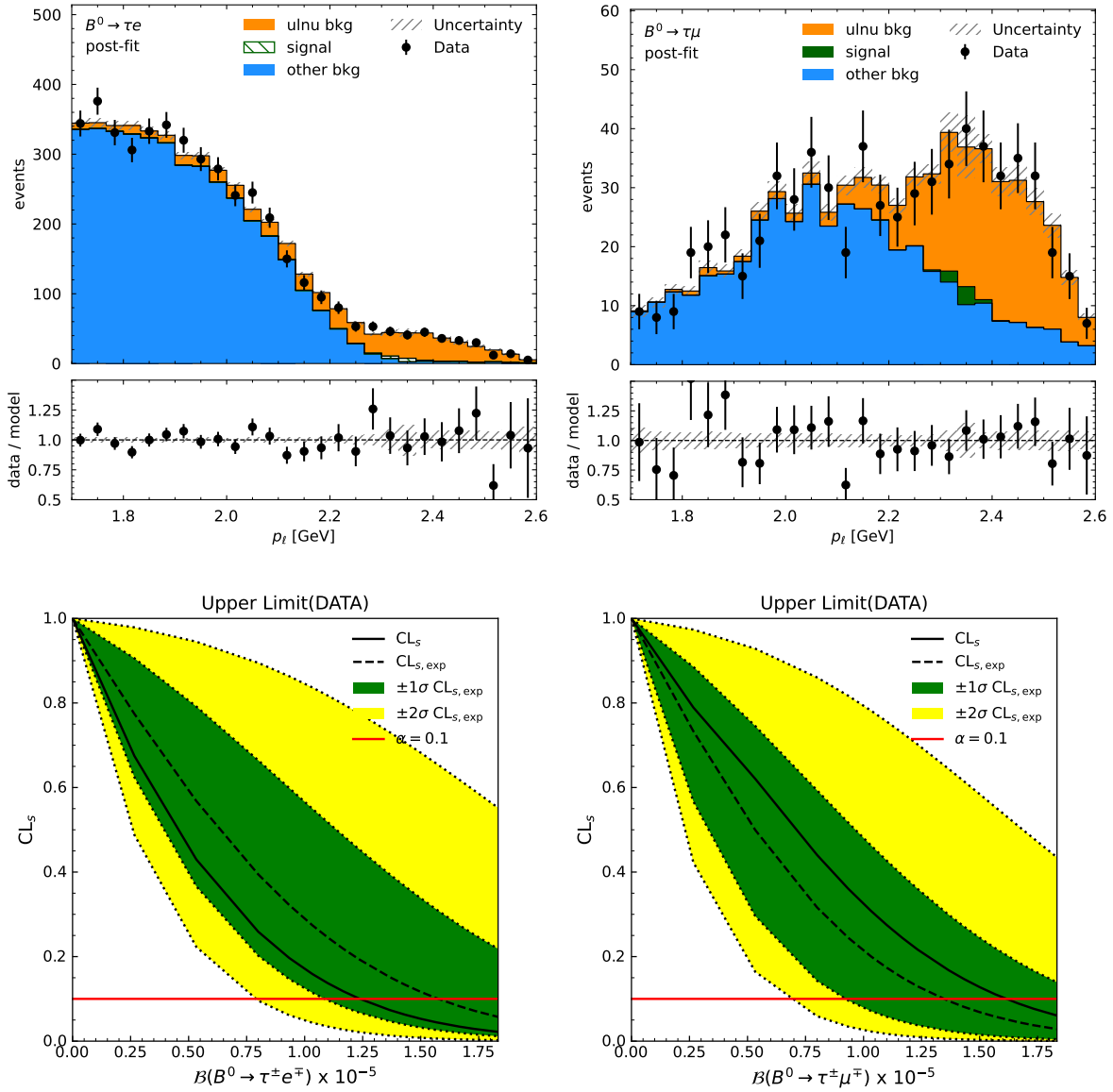

Figure C.2: Post-fit distribution for the fit on full Belle data (upper) and the upper limit scans on branching ratio (lower) for the $B^0 \to \tau e$ (left) and $B^0 \to \tau\mu$ (right) channels with the cs classifier output > 0.6 selection.

	$B^0 \to \tau e$	$B^0 \to \tau \mu$
n_{sig}	-12.50 $\pm$ 18.00	6.15 $\pm$ 12.19
$n_{u\ell\nu}$	554.11 $\pm$ 46.91	237.29 $\pm$ 26.59
n_{other}	3886.37 $\pm$ 69.07	424.58 $\pm$ 26.69

Table C.2: Fit results on full Belle data for the cs classifier output > 0.6 selection.

Danksagung

Ich möchte mich bedanken bei

- Prof. Thomas Kuhr für all deine Hilfe während meiner vierundeinhalb Jahre langen Promotion. Ich schätze sehr, dass du dir immer Zeit genommen hast und ein sehr nettes und unterstützendes Arbeitsumfeld geschaffen hast. Außerdem möchte ich mich bei dir bedanken, dass ich zu der B2GM nach Japan fliegen konnte und dabei auch den Belle II Detektor sehen konnte.

- Prof. Otmar Biebel für das Erstellen des Zweitgutachtens und allen Mitgliedern meiner Prüfungskommission für die Bereitschaft an der Prüfung mitzuwirken.

- Dr. Thomas Lück für die Betreuung während meiner Promotion. Danke, dass du dir so viel Zeit genommen hast um meine ganzen Fragen zu beantworten und Ideen für meine Analyse zu diskutieren. Dass du mir während der Zeit im Home Office 2020 angeboten hast mein Beteuer zu werden, hat mir sehr geholfen die Zeit gut und produktiv zu überstehen. Außerdem möchte ich mich natürlich auch für das Lesen meiner Arbeit und ausführlichen Anmerkungen bedanken.

- Dr. Felix Metzner für das Teilen deines Codes für die Formfaktoren Umgewichtung und das beantworten aller Fragen dazu.

- AG Kuhr for being an incredibly nice and supportive group. I really enjoyed our time at conferences and in the office, despite the freezing winters and hot summers, which was yet another reason to eat more ice cream. The times we spent together will always be fond memories of mine.

- Meinen Freunden aus dem Studium. Ich habe die letzten 10 Jahre mit euch sehr genossen und durch euch wurde das ganze Physikstudium viel leichter. Eure Unterstützung, eure Freundschaft und die gemeinsamen Erlebnisse haben meine Studienzeit unvergesslich gemacht.

- Meiner Familie. Danke, dass ich mich immer auf euch verlassen kann und für eure ganze uneingeschränkte Unterstützung.