The University of Melbourne

Doctoral Thesis

Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a Semileptonic Tag at the Belle Experiment

Giacomo Caria

https://orcid.org/0000-0002-3199-1754

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy
in the
School of Physics

June, 2019

Abstract

In recent years, several measurement of the ratios $\mathcal{R}(D) = \mathcal{B}(\bar{B} \to D\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B} \to D\ell^-\bar{\nu}_\ell)$ and $\mathcal{R}(D^*) = \mathcal{B}(\bar{B} \to D^*\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B} \to D^*\ell^-\bar{\nu}_\ell)$, where ℓ denotes an electron or a muon, have obtained results in disagreement with the theoretical predictions of the Standard Model (SM) of particle physics. When combining all experimental results together, the statistical significance of the discrepancy with the SM accounts to 4σ. If confirmed, this discrepancy would directly require an extension of the SM with new physics theories, to accommodate the experimental results.

This thesis reports a new independent measurement of the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$. The results are based on a data sample containing 772×10^6 $B\bar{B}$ pairs recorded at the $\Upsilon(4S)$ resonance with the Belle detector at the KEKB e^+e^- collider in Japan. The analysis uses a semileptonic reconstruction of the tag-side B-meson and leptonic τ decays.

The measured values are $\mathcal{R}(D) = 0.307 \pm 0.037$ (stat.) $\pm$ 0.016 (syst.) and $\mathcal{R}(D^*) = 0.283 \pm 0.018$ (stat.) $\pm$ 0.014 (syst.), which are in agreement with the SM predictions within 0.2σ and 1.1σ respectively. The $\mathcal{R}(D)$-$\mathcal{R}(D^*)$ combined result is in agreement with the SM predictions within 1.2σ.

The discrepancy of the experimental world average with the SM expectations decreases from 4 to 3σ when including these latest results, which represent to most precise measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ ever performed.

Declaration

This is to certify that:

- this thesis comprises only my original work towards the PhD except where indicated,
- due acknowledgement has been made in the text to all other material used,
- this thesis is less than 100,000 words in length, exclusive of tables, bibliographies and appendices.

Giacomo Caria

Preface

In recent years, accomplishments in experimental particle physics have been the result of increasingly large international scientific collaborative efforts as is the case for Belle experiment. As a result of these collaborative undertakings, it is incredibly rare that any substantial body of work is done in isolation. The standard procedure is to work in groups, each carrying out a specific purpose within the wider collaboration such as data acquisition, detector monitoring and development, particle reconstruction and event simulation. All these activities represent essential inputs to any scientific measurement in Belle. The results presented in this thesis are detailed here.

Chapters 1 is an original review of the theory of $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays, together with the measurement motivation and a review of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ previous measurements. Chapter 2 is a summary of the Belle experiment, with extracts from a number of public results about the Belle detector design and performance.
The author's original work is described in Chapters 3 to 8 which illustrate the work of the author conducted within the Belle collaboration, under the supervision of A/Prof. Phillip Urquijo. Unless specifically stated, all the information provided in these chapters and the scientific results represent original work of the author. The original work of the author represent unpublished material not yet submitted for publication.

Acknowledgements

Firstly, I am profoundly grateful to my supervisor A/Prof. Phillip Urquijo for the continuous support of my PhD, for his patience, motivation, and vast knowledge. His positive attitude has greatly helped me to overcome the difficult times of my PhD, and taught me that getting stuck on a problem is just an option and it can be easily avoided. Furthermore, I appreciate his constant encouragement to take responsibility for the research project I have undertaken.

I would like to thank all the friends I have made at CoEPP during these years: Brian Le, John Gargalionis, Laurence Spiller, Justin Tan, Eiasha Waheed, Scott Williams, including the italian crew: Tiziano Baroncelli, Tommaso Baroncelli, Marco Milesi, Francesco Nuti, Federico Scutti, Jafar Shojaii, Francesco Tenchini, Francesca Ungaro. Being able to talk and hang out with all of you has not only immensely enhanced my experience at CoePP, but also helped me keeping my sanity during this long journey that is a PhD.

Some of the important people that I would like to thank happen to live far away from this remote island in the middle of nowhere that I have chosen for this degree. In order for our friendships to be unaffected by such large differences of locations and time zones, they must be truly special people.

Last but not least, I express my gratitude to my family for always being present even though the long distance that separated us, and for constantly reminding me that I can always count on them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Standard Model of Particle Physics

1.1.1 Particles and Forces

The Standard Model (SM) of Particle Physics provides a description of nature through elementary blocks that make up the universe at its smallest scale. These blocks, or particles, are characterized by their spin, their mass, and the quantum numbers (charges), which determine their interactions. Generally speaking, fundamental particles can be split in two families, depending on their spin. Particles with a spin of 1/2 are called fermions.There are two types of fundamental fermions: leptons and quarks, which account for twelve different particles, arranged in three different generations.

Particles with a spin of 1 are called bosons. Fundamental bosons are responsible for mediating the interactions between fermions. The three type of forces described by the SM are, in order of strength, the weak, the electromagnetic and the strong interactions. Gravity is not incorporated in the SM, and its effects are assumed to be negligible at this scale, when compared to the other three interactions. The weak force is mediated by the massive positively charged $W^{\pm}$ and neutrally charged Z^0 bosons, and affects both leptons and quarks. The electromagnetic interaction is mediated by the massless photon and affects particles that carry electric charge. The strong force is mediated by the massless gluons, that couple to the color charge possessed only by quarks and gluons themselves. The properties of all fermions and bosons are summarized in Figure 1.1.

The SM is a gauge quantum field theory characterized by the gauge group $G_{SM} = SU(3) \times SU(2) \times U(1)$. The dynamics of the SM is conveniently described by a Lagrangian density $\mathcal{L}(x)$, normally simply called the Lagrangian. Gauge symmetry prevents a mass term for the gauge bosons to be added to the Lagrangian. While this does not cause any problem for the massless photon and gluons, a way of introducing a mass for the $W^{\pm}, Z^0$ bosons is needed. This is done through a process called spontaneous symmetry breaking [1, 2], which also gives mass also the leptons and introduces a new massive particle, the Higgs boson. Until the switch-on the LHC, the Higgs boson was

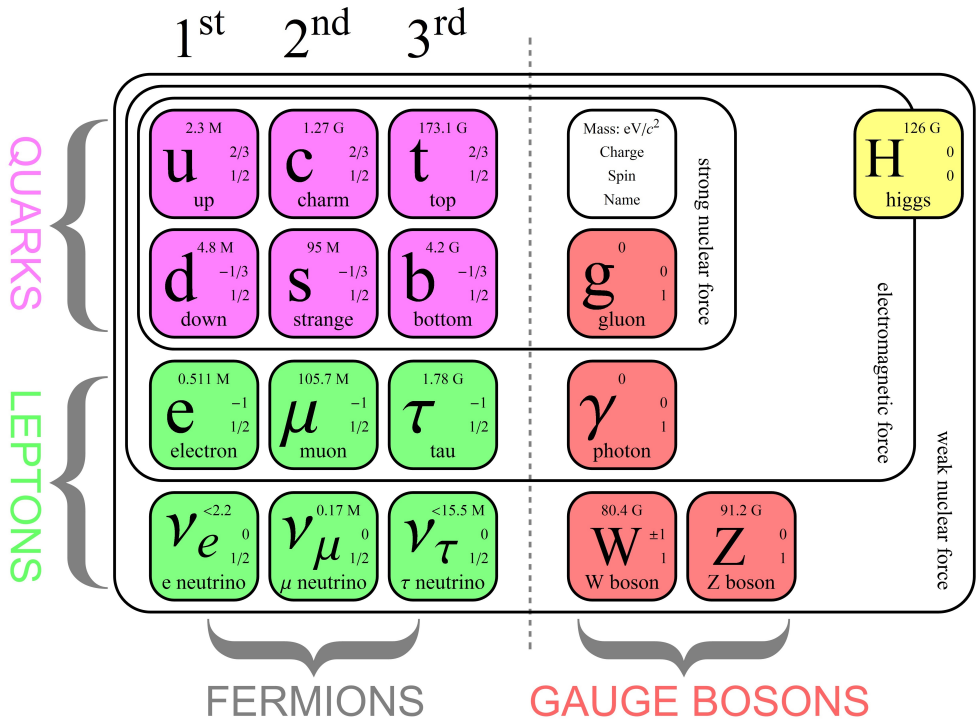


Figure 1.1: Summary scheme of the Standard Model: the mediator bosons of the fundamental interactions, the Higgs boson, and the three generations of fermions, divided into quarks and leptons.

considered the only missing block of the SM, and was discovered in 2012 by the ATLAS and CMS collaborations [3, 4].

1.1.2 Weak interactions and the CKM Matrix

The weak interaction is the only interaction of the SM that violates the parity symmetry. Formally, we can write the weak interaction amplitude as product of currents

$$\mathcal{M} = \frac{4G_F}{\sqrt{2}} J^\mu J^\dagger_\mu, \tag{1.1}$$

where G_F is the weak coupling constant, known as Fermi constant, and the current J^μ can be expressed as

$$J^\mu = \bar{\psi}\,\gamma^\mu \frac{1}{2}(1-\gamma^5)\psi. \tag{1.2}$$

This highlights its $V-A$ structure, which includes both a vector (V) term $\bar{\psi}\gamma^\mu\psi$ and an axial-vector (A) term $\bar{\psi}\gamma^\mu\gamma^5\psi$. Indeed, while the former is invariant under parity (P) transformations, the latter is not. Also the combination of parity and charge (C) symmetry is not respected by the weak interaction, a phenomenon known as CP violation.

Moreover, whenever the weak force is mediated by the $W^{\pm}$ boson, different flavors of quarks can mix. The complex, unitary 3×3 matrix that describes the quark flavor mixing is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [5], and describes the transition of down-type quarks from the mass eigenstates (d, s, b) to the weak eigenstates (d', s', b'). The transformation can be written as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.3}$$

where the elements V_{ij} with $i \in u, c, t$ and $j \in d, s, b$ determine the different strengths of the couplings between up and down type quarks. Four parameters define the CKM matrix: three mixing angles and one complex phase. In the SM, CP-violation is a direct consequence of the complex phase. CP-violating effects have been measured in the B-meson sector by the collaborations BaBar [6] and Belle [7], and led to the 2008 Physics Nobel Prize for Kobayashi and Maskawa. Even though CP violation has been found experimentally, its rate is too low to meet one of the Sakharov conditions for the observed matter-antimatter asymmetry in the universe [8].

1.1.3 Lepton Flavor Universality

Since the electron, muon and tau flavors of leptons share the same quantum numbers, there is nothing that sets them apart, if we neglect the differences in mass. For this reason it might seem that we have a simple replication of the same flavor and we expect the interactions of all leptons to to be of the same form and strength. This is expressed by the fact that the coupling constant G_F of the weak interaction is the same for all lepton flavors. The only difference in amplitudes of weak processes would then reside in the different volumes of phase space. This concept is called lepton flavor universality (LFU).

Since LFU is postulated by the SM, any observed deviations from it (violation), would be a clear sign of new physics. Many precision tests of LFU have been performed by several experiments, and they have all confirmed that LFU holds within experimental errors.

Muon decays and leptonic tau decays, whose Feynman diagram are shown in Figure 1.2, offer a platform for testing LFU. More specifically we are interested in testing that $G_F = \sqrt{2} g_w^2 / 8 M_W^2 = G_F^e = G_F^\mu = G_F^\tau$. Assuming non universal constants, we can write the total decay rate for the decay $\mu^- \to e^- \bar{\nu}_e \nu_\mu$ as [9]

$$\Gamma_{\mu \to e^- \bar{\nu}_e \nu_\mu} = \frac{1}{\tau_\mu} = \frac{G_F^e G_F^\mu m_\mu^5}{192 \pi^3}, \tag{1.4}$$

where τ_μ refers to the muon lifetime, and m_μ is the muon's mass. On the other hand, for the decay $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ we have to remember that the branching ratio is the fractional

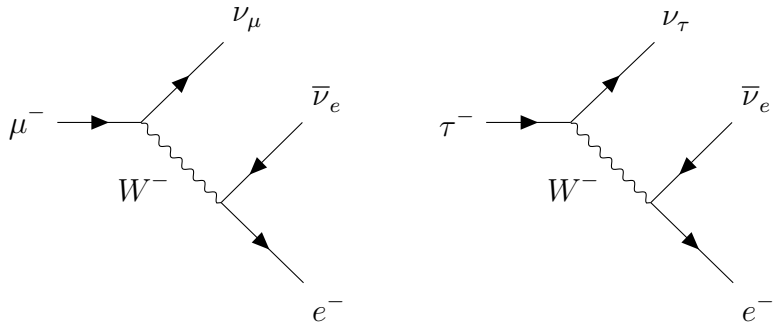


Figure 1.2: Feynman diagrams for muon decay (left) and leptonic tau decay (right).

number of decays via a specific mode, and is given by the ratio of the partial decay width to the total decay width. Hence, we can write

$$\frac{1}{\tau_\tau} = \frac{1}{\mathcal{B}(\tau \to e^- \bar{\nu}_e \nu_\tau)} \, \Gamma_{\tau \to e^- \bar{\nu}_e \nu_\tau} = \frac{1}{\mathcal{B}(\tau \to e^- \bar{\nu}_e \nu_\tau)} \frac{G_F^e G_F^\tau m_\tau^5}{192\pi^3}. \tag{1.5}$$

By dividing Equation 1.5 by Equation 1.4 we obtain

$$\frac{G_F^\tau}{G_F^\mu} = \frac{m_\mu^5 \tau_\mu}{m_\tau^5 \tau_\tau} \, \mathcal{B}(\tau \to e^- \bar{\nu}_e \nu_\tau). \tag{1.6}$$

All the quantities involved in Equation 1.6 are known with an excellent accuracy [10]:

$$\tau_\mu = 2.197034(21) \times 10^{-6}\,\mathrm{s}, \quad m_\mu = 105.658367(4)\,\mathrm{GeV},$$
$$\tau_\tau = (290.6 \pm 1.0) \times 10^{-15}\,\mathrm{s}, \quad m_\tau = 1776.82 \pm 0.16\,\mathrm{GeV}, \quad \mathcal{B}(\tau \to e^- \bar{\nu}_e \nu_\tau) = (17.82 \pm 0.04)\%,$$

and substituting for the numerical values we obtain

$$\frac{G_F^\tau}{G_F^\mu} = 1.0011 \pm 0.0015. \tag{1.7}$$

Hence, this experimental result (taken from [11] as all the other results in this section) confirms that the weak coupling is universal for all leptons. Similar results can be obtained for the other combinations of weak constants, as

$$\begin{aligned} \frac{G_F^\mu}{G_F^e} &= 1.0018 \pm 0.0014, \\ \frac{G_F^\tau}{G_F^e} &= 1.0030 \pm 0.0015. \end{aligned} \tag{1.8}$$

These measurements provide the most precise experimental results for LFU tests that involve the coupling of third lepton family to the first and second ones. Concerning

W and Z decays to leptons, several results can be listed, and generally the precision is about one order of magnitude higher in the Z decays. For the Z decays we have

$$\begin{aligned} \frac{\Gamma_{Z\to\mu^+\mu^-}}{\Gamma_{Z\to e^+e^-}} &= 1.0009 \pm 0.0028, \\ \frac{\Gamma_{Z\to\tau^+\tau^-}}{\Gamma_{Z\to e^+e^-}} &= 1.0019 \pm 0.0032, \end{aligned} \tag{1.9}$$

where LFU holds with a precision better than 0.3%, while for the W decays we have

$$\begin{aligned} \frac{\mathcal{B}(W\to e^-\bar{\nu}_e)}{\mathcal{B}(W\to \mu^-\bar{\nu}_\mu)} &= 1.004 \pm 0.008, \\ \frac{\Gamma_{W\to\tau^-\bar{\nu}_\tau}}{\Gamma_{W\to e^-\bar{\nu}_e}} &= 1.063 \pm 0.027, \\ \frac{\Gamma_{W\to\tau^-\bar{\nu}_\tau}}{\Gamma_{W\to\mu^-\bar{\nu}_\mu}} &= 1.070 \pm 0.026, \end{aligned} \tag{1.10}$$

where the first result is an average obtained assumed all uncertainties are fully uncorrelated and confirms LFU with a 0.8% precision, while the latter two that contain τ in the ratio are in tension with the SM prediction.

Another platform where LFU can be tested is the decays of B-mesons (composite particles made of a b quark and another lighter quark) to a charm meson (composite particles made of a c quark and another lighter quark) and a W boson, which in turn decays to a pair of leptons. This is the subject of the next section.

1.2 $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ Theory

1.2.1 $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ Decays in the Standard Model

The goal of this chapter is to detail the current understanding of the decays we aim at measuring, as presented in the theory of the Standard Model. It should be clear from this section that the SM leaves little room for uncertainty in the calculation of the decay rates of interest, which translates to a small uncertainty in the theoretical SM predictions for the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$.

The equations used in this Section follow from Reference [12]. In the SM semileptonic B-meson decays, whose Feynman diagram is shown in Figure 1.4, are described by the effective Lagrangian

$$\mathcal{L}_{\text{eff}} = -\frac{G_F}{\sqrt{2}} V_{cb} J^{\dagger\mu}_{\nu\ell} J_{cb\mu} + \text{h.c.}, \tag{1.11}$$

where V_{cb} represents the CKM matrix element for the transition from charm to bottom quarks, G_F is Fermi coupling constant, and the leptonic and quark currents are given

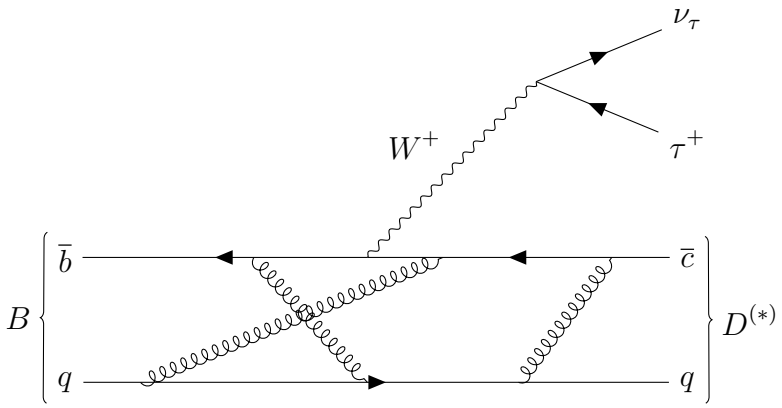


Figure 1.3: Feynman diagram for the semileptonic decay $B \to D^{(*)}\tau^+\nu_\tau$ in the SM. The unspecified spectator quark q can be either a u quark for the decay $B^+ \to \bar{D}^{(*)0}\tau^+\nu_\tau$ or a d quark for the decay $B^0 \to D^{(*)-}\tau^+\nu_\tau$.

as

$$J^\mu_{\nu\ell} \equiv \bar{\psi}_\nu \gamma^\mu (1-\gamma_5)\psi_\ell, \qquad J^\mu_{cb} \equiv \bar{\psi}_c \gamma^\mu (1-\gamma_5)\psi_b. \tag{1.12}$$

The amplitude for the process, or matrix element, is expressed as

$$\mathcal{M}^{\lambda_\ell}_{\lambda_M}(q^2,\theta_\ell) = -\frac{G_F V_{cb}}{\sqrt{2}} \langle \ell^-(p_\ell,\lambda_\ell)\bar{\nu}(p_\nu)| J^{\dagger\mu}_{\nu\ell} |0\rangle \langle M(p_M,\lambda_M)| J_{cb\mu} |\bar{B}(p_B)\rangle , \tag{1.13}$$

where M is either D or D^* and the particle helicities are $\lambda_M = 0$ for the D-meson, $\lambda_M = \pm 1, 0$ for the D^* meson and $\lambda_\ell = \pm 1/2$ for the lepton. We label the pseudoscalar state with $\lambda_M = s$ to distinguish it from the helicity-zero states of the D and the D^*. The matrix element depends on the squared four-momentum of the W-boson, q^2, where $q = p_B - p_M$, and the lepton decay angle in the virtual W frame, θ_ℓ.

By introducing the polarization vector of the virtual W, $\epsilon_W \equiv \epsilon(q,\lambda_W)$, we can rewrite the matrix element as a sum of the helicity states $\lambda_W = \pm 1, 0, s$ as

$$\mathcal{M}^{\lambda_\ell}_{\lambda_M}(q^2,\theta_\ell) = \frac{G_F V_{cb}}{\sqrt{2}} \frac{m_W^2}{m_W^2 - q^2} \sum_{\lambda_W} L^{\lambda_\ell}_{\lambda_W}(q^2,\theta_\ell) H^{\lambda_M}_{\lambda_W}(q^2), \tag{1.14}$$

where

$$H^{\lambda_M}_{\lambda_W}(q^2) = \epsilon^*_\mu(\lambda_W) \langle M(p_M,\lambda_M)| J^\mu_{cb} |\bar{B}(p_B)\rangle \tag{1.15}$$

$$L^{\lambda_\ell}_{\lambda_W}(q^2,\theta_\ell) = \epsilon_\mu(\lambda_W) \langle \ell(p_\ell,\lambda_\ell)\bar{\nu}_\ell(p_\nu)| J^{\dagger\mu}_{\nu\ell} |0\rangle \tag{1.16}$$

describe the decays $\bar{B} \to MW^*$ and $W^* \to \ell\bar{\nu}$ respectively.

The $V-A$ structure of the weak current dictates that for the decay of the pseudoscalar (spin 0) B-meson to a pseudoscalar D-meson final state, only the vector current

contributes, while for a D^*-meson vector (spin 1) final state, the vector current together with the axial-vector current contribute. Hence, the hadronic amplitudes are given as

$$\langle D|\, J^\mu_{cb}\,|\bar{B}\rangle = \langle D|\, V^\mu_{cb}\,|\bar{B}\rangle\,, \qquad \langle D^*|\, J^\mu_{cb}\,|\bar{B}\rangle = \langle D^*|\, V^\mu_{cb}\,|\bar{B}\rangle - \langle D^*|\, A^\mu_{cb}\,|\bar{B}\rangle\,. \tag{1.17}$$

In the hadronic currents $H^{\lambda_M}_{\lambda_W}$ the strong interaction between the quarks has to be taken into account, and as a consequence the decay rate calculations are noticeably more complicated. The strong coupling becomes large for low momentum transfer, which means that the decay kinematics, i.e. Feynman diagrams, of higher orders would have to be considered as well. For this reason, analytical solutions can not be found. The approach used to solve this problem is to form Lorentz invariant structures based on the four-momenta of the mesons, $p_B, p_D^{(*)}$, plus coefficient functions which depend solely on q^2. Those functions are called hadron transition *form factors* and are similar to those required to describe atomic scattering in nuclear physics.

We can parametrize the decay to the D pseudoscalar state in terms of the form factors $f_+(q^2)$ and $f_0(q^2)$ as

$$\begin{aligned}\langle D(p_D)|\, V^\mu_{cb}\,|B(p_B)\rangle = f_+(q^2)\left(p_B^\mu + p_D^\mu - \frac{m_B^2 - m_D^2}{q^2}q^\mu\right) \\ + f_0(q^2)\frac{m_B^2 - m_D^2}{q^2}q^\mu.\end{aligned} \tag{1.18}$$

Another common way to write the current is

$$\langle D(p_D)|\, V^\mu_{cb}\,|B(p_B)\rangle = f_+(q^2)(p_B^\mu + p_D^\mu) + f_-(q^2)(p_B^\mu - p_D^\mu), \tag{1.19}$$

where the relation with Equation 1.18 is given by

$$f_0(q^2) = f_+(q^2) + \frac{q^2}{m_B^2 - m_D^2}f_-(q^2). \tag{1.20}$$

The parametrization of the decay to the D^* vector final state is more complicated, and can be expressed in terms of the form factors $A_i(q^2)$ with $i = 0, 1, 2$ and V, as

$$\langle D^*(p_{D^*}, \epsilon_M)|\bar{c}\gamma_\mu b|B(p_B)\rangle = \frac{2iV(q^2)}{m_B + m_{D^*}}\,\epsilon_{\mu\nu\alpha\beta}\epsilon^{*\nu}p_B^\alpha p_{D^*}^\beta\,, \tag{1.21a}$$

$$\begin{aligned}\langle D^*(p_{D^*}, \epsilon)|\bar{c}\gamma_\mu\gamma_5 b|B(p_B)\rangle = 2m_{D^*}\,A_0(q^2)\frac{\epsilon^*\cdot q}{q^2}q_\mu + (m_B + m_{D^*})\,A_1(q^2)\left(\epsilon^*_\mu - \frac{\epsilon^*\cdot q}{q^2}q_\mu\right) \\ -A_2(q^2)\,\frac{\epsilon^*\cdot q}{m_B + m_{D^*}}\left((p_B + p_{D^*})_\mu - \frac{m_B^2 - m_{D^*}^2}{q^2}q_\mu\right)\,,\end{aligned} \tag{1.21b}$$

where $\epsilon_{\mu\nu\alpha\beta}$ is the Levi-Civita symbol and ϵ^* is the complex conjugated polarization

vector of the D^* vector meson. Thus, we can express the hadronic amplitudes as

$$
\begin{aligned}
H_{\pm}^{\pm}(q^2) &= (m_B + m_{D^*})A_1(q^2) \mp \frac{2m_B}{m_B + m_{D^*}}|\mathbf{p}\,|V(q^2)\,, \\
H_0^0(q^2) &= \frac{1}{2m_{D^*}\sqrt{q^2}}\left[(m_B^2 - m_{D^*}^2 - q^2)(m_B + m_{D^*})A_1(q^2) - \frac{4m_B^2|\mathbf{p}\,|^2}{m_B + m_{D^*}}A_2(q^2)\right]\,, \\
H_s^0(q^2) &= \frac{2m_B|\mathbf{p}\,|}{\sqrt{q^2}}A_0(q^2)\,.
\end{aligned}
\tag{1.22}
$$

We have assumed that since the parent B-meson has spin 0, the D^*-meson and the W-boson have the same helicity, i.e. $\lambda_W = \lambda_M$. These equations show that A_2 appears only in the H_0^0 definition, while V contributes only to $H_{\pm}^{\pm}$. On the other hand, A_1 plays a role in all three helicity amplitudes, and takes a dominant role for all of them at high q^2 values.

In the limit $m_\ell = 0$, i.e. when $\ell = e$ or μ, $\bar{B} \to D$ can be described by only one form factor f_+, since f_- is proportional to q^μ. Similarly, for the $\bar{B} \to D^*$ decay, H_s^0 can be ignored since the corresponding leptonic current $L_s^{\pm}$ vanishes, leaving only 3 relevant hadronic amplitudes for $\bar{B} \to D^*$.

The treatment of the form factors is a matter of ongoing theoretical activity. Several parametrizations of the form factors exist, such as the Boyd-Grinstein-Lebed [13] and the Caprini-Lellouch-Neubert [14] parametrization for the $B \to D^*$ decays, and the Bourrely-Caprini-Lellouch [15] parametrization for the $B \to D$ decays. The form factors may be written in form of an expansion in the complex variable z, which is defined as

$$
z = \frac{\sqrt{\omega + 1} - \sqrt{2}}{\sqrt{\omega + 1} + \sqrt{2}} \tag{1.23}
$$

where $w = m_B^2 + m_{D^*}^2 - q^2/2m_B m_{D^*}$. The expansion converges rapidly in the kinematical region of heavy hadron decays and the coefficients of the expansions are subject to unitarity bounds based on analyticity. All the parametrization listed above are constructed to satisfy the unitarity bounds, but the CLN approach differs mostly in its reliance on next-to-leading order heavy quark effective theory relations between the form factors.

Differential Decay Distribution and $\mathcal{R}(D^{(*)})$

After the calculation of the leptonic currents defined in Eq. 1.16 using the Dirac algebra of electroweak theory, we sum over the lepton's helicities and replacing them in Eq. 1.14, we can apply Fermi's golden rule and integrate over the full azimuthal range, obtaining

the differential decay rate

$$\frac{d\Gamma(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell\ )}{dq^2 d\cos\theta_\ell} = \frac{G_F^2|V_{cb}|^2|\mathbf{p}^*_{D^{(*)}}|q^2}{256\pi^3 m_B^2}\left(1-\frac{m_\ell^2}{q^2}\right)^2 \times \left[(1-\cos\theta_\ell)^2|H_+|^2 + (1+\cos\theta_\ell)^2|H_-|^2 + 2\sin\theta_\ell{}^2|H_0|^2 + \frac{m_\ell^2}{q^2}\Big(\sin\theta_\ell{}^2(|H_+|^2+|H_-|^2) + 2|H_s + H_0\cos\theta_\ell|^2\Big)\right], \tag{1.24}$$

where, as in the previous section, $H_\lambda \equiv H^{\lambda_M}_{\lambda_W}$ with $\lambda = \lambda_W = \lambda_M$ and the q^2 dependence of the hadronic decay amplitudes are omitted. We derive the q^2 spectrum by integrating over θ_ℓ as

$$\frac{d\Gamma(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell\ )}{dq^2} = \frac{G_F^2|V_{cb}|^2|\mathbf{p}^*_{D^{(*)}}|q^2}{96\pi^3 m_B^2}\left(1-\frac{m_\ell^2}{q^2}\right)^2 \times \left[(|H_+|^2+|H_-|^2+|H_0|^2)\left(1+\frac{m_\ell^2}{2q^2}\right) + \frac{3m_\ell^2}{2q^2}|H_s|^2\right]. \tag{1.25}$$

Finally, the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are extracted as follows:

$$\begin{aligned}\mathcal{R}(D^{(*)})(q^2) &\equiv \frac{d\Gamma(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau\ )/dq^2}{d\Gamma(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell\ )/dq^2} \\ &= \left(1-\frac{m_\tau^2}{q^2}\right)^2\left[\left(1+\frac{m_\tau^2}{2q^2}\right) + \frac{3}{2}\frac{m_\tau^2}{q^2}\frac{|H_s|^2}{|H_+|^2+|H_-|^2+|H_0|^2}\right].\end{aligned} \tag{1.26}$$

where $q^2_{\rm max} = (m_B - m_{D^{(*)}})^2$ and we are working under the assumption $q^2 \gg m_\ell$. From this expression it can be noted that the theoretical uncertainties on $\mathcal{R}(D^{(*)})$ will depend only on the tau and mesons masses, and on form factors that enter the helicity amplitude calculation.

$\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ Decays

An important source of background for the processes measured in this analysis are the decays of B mesons to orbitally excited P-wave charm mesons, labeled D^{**}. The D^{**} mesons typically decay to a D or D^* state by emission of one or more pions. Whenever the pion(s) is lost, this process can mimic the signal decay (i.e. $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$).

The model that describes the D-meson decay is the Heavy Quark Effective Theory (HQET) [16], a quantum field theory that deals with the strong interaction of a heavy quark with a light quark, such as in the meson system, i.e. a bound state of the heavy quark with any of the light u,d or s quarks.

If the light quark is labeled as q and the heavy one as Q, we can introduce the total angular momentum $\vec{J}$ and the q total angular momentum $\vec{j}_q$:

$$\begin{aligned} \vec{j}_q &= \vec{L} + \vec{S}_q \\ \vec{J} &= \vec{j}_q + \vec{S}_Q \\ S_q &= S_Q = 1/2 \end{aligned} \tag{1.27}$$

where $\vec{L}$ is the relative orbital momentum between Q and q, $\vec{S}_Q\,(\vec{S}_q)$ is the $Q(q)$-spin. Moreover, states are characterized by the label J^P, where the parity quantum number is defined as: $P = (-1)^J$.

The ground state is composed of a S-wave ($L = 0$) doublet, while the first excited state is a P-wave ($L = 1$) state with two values for the light quark angular momentum: $j_q = 1/2, 3/2$. When $j_q = 1/2$, we have a doublet with $J^P = 0^+, 1^+$, which corresponds to the D_1' and D_0^* states , while when $j_q = 3/2$ we have a doublet with $J^P = 1^+, 2^+$: the D_1 and D_2^* states. Conservation of parity and angular momentum conservation dictate the allowed decays allowed for each state: the D_1 and D_2^* states decay through a D-wave and have small decay widths ($\Gamma < 50$ MeV), while the D_1' and D_0^* states decay through an S-wave and are very broad ($\Gamma > 250$ MeV). The properties of the the D^{**} states are summarized in Table 1.1.

Semileptonic B decays to the narrow D_1 and D_2^* states have been studied by a number of experiments [17, 18, 19, 20, 21, 22, 23, 24]; while the broad states D_1' and D_0^* have been observed by the DELPHI [25], Belle [26] and Babar [27] experiments.

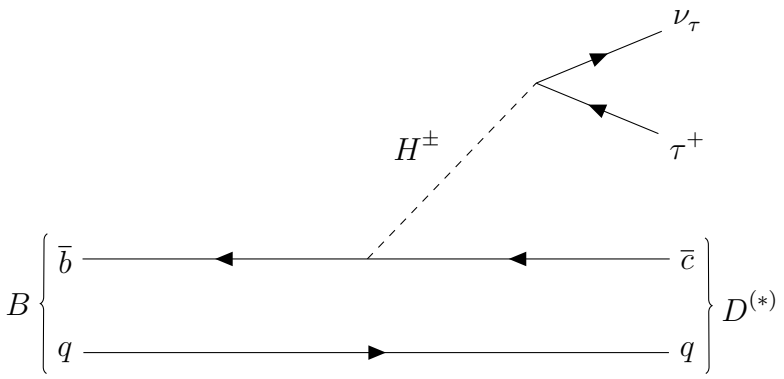
1.2.2 New Physics Scenarios

Effective Lagrangian

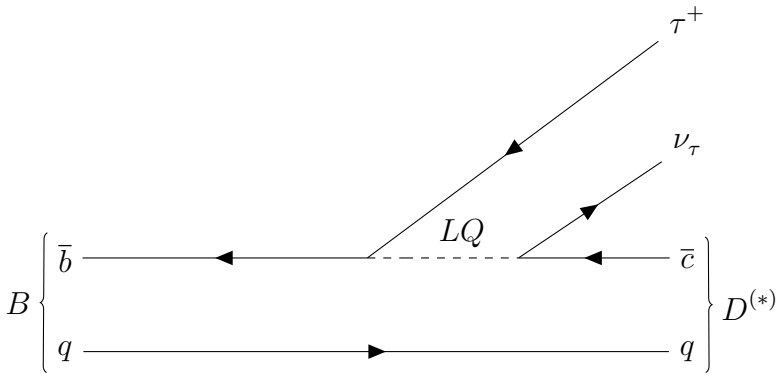
We outline a general way of parametrizing NP effects, based on the operator expansion in the effective Lagrangian. The term effective refers to the fact that this parametrization is meant to capture the low energy features of new physics and is only valid up to a certain energy scale. More specifically, the low energy behaviour of the new interactions is assumed to be a contact interaction between four fermions, that hides the high energy description of the interaction, in a fashion similar to the one used by Fermi to first describe the weak interaction. In this framework, an Operator Product Expansion (OPE) [28] is used to describe effects beyond the SM, in a model independent way.

The most general effective Lagrangian for the $b \to c\tau\bar{\nu}_l$ transition that contains all possible four-fermion operators of the lowest dimension, under the assumption that the neutrinos are left-handed, can be written as

$$\mathcal{L}_{\text{eff}} = -2\sqrt{2}G_F V_{cb} \sum_{l=e,\mu,\tau} \left[(\delta_{l\tau} + C_{V_1}^l)\mathcal{O}_{V_1}^l + C_{V_2}^l \mathcal{O}_{V_2}^l + C_{S_1}^l \mathcal{O}_{S_1}^l + C_{S_2}^l \mathcal{O}_{S_2}^l + C_T^l \mathcal{O}_T^l \right], \tag{1.28}$$



(a) Charged Higgs ($H^\pm$)



(b) Leptoquark (LQ)

Figure 1.4: Feynman diagrams for NP contributions to the the semileptonic decay $B \to D^{(*)}\tau^+\nu_\tau$.

Table 1.1: Summary of the current status of $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ branching ratio measurements, quoted as $\mathcal{B}^{\pm\text{stat}}_{\pm\text{sys}}(\%)$, as well as the combinations from the Belle and Babar experiments.

	Belle [20]	BaBar[22, 21]	Average[10]
$\mathcal{B}(B^+ \to \bar{D}_1'^0\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_1'^0 \to D^{*-}\pi^+)$	< 0.07	$0.27^{\pm0.04}_{\pm0.05}$	$0.27^{\pm0.04}_{\pm0.05}$
$\mathcal{B}(B^0 \to D_1'^-\ell^+\nu_\ell) \times \mathcal{B}(D_1'^- \to \bar{D}^{*0}\pi^-)$	< 0.5	$0.31^{\pm0.07}_{\pm0.05}$	$0.31^{\pm0.07}_{\pm0.05}$
$\mathcal{B}(B^+ \to \bar{D}_0^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_0^{*0} \to D^-\pi^+)$	$0.24^{\pm0.04}_{\pm0.06}$	$0.26^{\pm0.05}_{\pm0.04}$	0.25 ± 0.05
$\mathcal{B}(B^0 \to D_0^{*-}\ell^+\nu_\ell) \times \mathcal{B}(D_0^{*-} \to \bar{D}^0\pi^-)$	$0.20^{\pm0.07}_{\pm0.05}$	$0.44^{\pm0.08}_{\pm0.06}$	0.30 ± 0.12
$\mathcal{B}(B^+ \to \bar{D}_1^0\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_1^0 \to D^{*-}\pi^+)$	$0.42^{\pm0.07}_{\pm0.07}$	$0.297^{\pm0.017}_{\pm0.017}$ $0.29^{\pm0.03}_{\pm0.03}$	0.303 ± 0.020
$\mathcal{B}(B^0 \to D_1^-\ell^+\nu_\ell) \times \mathcal{B}(D_1^- \to \bar{D}^{*0}\pi^-)$	$0.54^{\pm0.19}_{\pm0.19}$	$0.27^{\pm0.04}_{\pm0.03}$	0.280 ± 0.028
$\mathcal{B}(B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_2^{*0} \to D^-\pi^+)$	$0.22^{\pm0.03}_{\pm0.04}$	$0.142^{\pm0.015}_{\pm0.015}$ $0.15^{\pm0.02}_{\pm0.02}$	0.153 ± 0.016
$\mathcal{B}(B^+ \to \bar{D}_2^{*0}\ell^+\nu_\ell) \times \mathcal{B}(\bar{D}_2^{*0} \to D^{*-}\pi^+)$	$0.18^{\pm0.06}_{\pm0.03}$	$0.087^{\pm0.011}_{\pm0.007}$ $0.15^{\pm0.02}_{\pm0.02}$	0.101 ± 0.024
$\mathcal{B}(B^0 \to D_2^{*-}\ell^+\nu_\ell) \times \mathcal{B}(D_2^{*-} \to \bar{D}^0\pi^-)$	$0.22^{\pm0.04}_{\pm0.04}$	$0.110^{\pm0.017}_{\pm0.008}$	0.121 ± 0.033
$\mathcal{B}(B^0 \to D_2^{*-}\ell^+\nu_\ell) \times \mathcal{B}(D_2^{*-} \to \bar{D}^{*0}\pi^-)$	< 3.0	$0.068^{\pm0.012}_{\pm0.05}$	0.068 ± 0.012

where the four-Fermi operators are defined as

$$\mathcal{O}^l_{V_1} = \bar{c}_L\gamma^\mu b_L\,\bar{\tau}_L\gamma_\mu\nu_{Ll}\,, \tag{1.29}$$
$$\mathcal{O}^l_{V_2} = \bar{c}_R\gamma^\mu b_R\,\bar{\tau}_L\gamma_\mu\nu_{Ll}\,, \tag{1.30}$$
$$\mathcal{O}^l_{S_1} = \bar{c}_L b_R\,\bar{\tau}_R\nu_{Ll}\,, \tag{1.31}$$
$$\mathcal{O}^l_{S_2} = \bar{c}_R b_L\,\bar{\tau}_R\nu_{Ll}\,, \tag{1.32}$$
$$\mathcal{O}^l_{T} = \bar{c}_R\sigma^{\mu\nu} b_L\,\bar{\tau}_R\sigma_{\mu\nu}\nu_{Ll}\,, \tag{1.33}$$

and $C^l_X\,(X = V_{1,2}, S_{1,2}, T)$ denote the Wilson coefficients of the operators $\mathcal{O}^l_X$, which in the SM are set to 0. The neutrino flavour is represented by l and all flavours are considered, $l = e$, μ or τ.

The general approach for the assessment of the compatibility of NP is to vary each operator and see how the new resulting decay rates predicted by the theory agree with the experimental values. Different families of NP models correspond to specific choices of the Wilson coefficients of Equation 1.28, and in following sections we outline two of most popular NP scenarios commonly used to explain the anomalies in $b \to c\tau\bar{\nu}_l$.

Table 1.2: Parameters $\xi_{d,u}$ in each type of 2HDMs.

	Type I	Type II	Type X	Type Y
ξ_d	$\cot^2\beta$	$\tan^2\beta$	-1	-1
ξ_u	$-\cot^2\beta$	1	1	$-\cot^2\beta$

2HDM

The SM relies on the minimal choice of a single scalar doublet that, acquiring a vacuum expectation value (VEV), gives masses to all the particles (except the neutrino) contained in the SM. Hence, one natural direction towards constructing behind-the-SM scenarios is to extend the SM scalar sector.

The simplest extension of this type is the two Higgs-doublet model (2HDM)[29], which has received a lot of attention mainly because the minimal supersymmetric Standard Model (MSSM) relies on it [30]. 2HDM scenarios have also been investigated to look for additional sources of CP violation for generating baryon asymmetry of the universe of sufficient size [31].

This model contains two complex doublets of scalar fields, H_1 and H_2. Three of the resulting eight degrees of freedom are used by the Goldstone bosons to provide the longitudinal modes of the bosons $W^\pm$ and Z, that become massive. As a consequence, five physical Higgs bosons remain: three neutral ones h_1, h_2, h_3 and two charged ones $H^\pm$.

In the 2HDM, the relevant Wilson coefficients introduced in Eq. (1.28) are given by

$$C^\tau_{S_1} = -\frac{m_b m_\tau}{m^2_{H^\pm}}\xi_d\,, \quad C^\tau_{S_2} = -\frac{m_c m_\tau}{m^2_{H^\pm}}\xi_u\,. \tag{1.34}$$

where $m_{H^\pm}$, m_b, m_c and m_τ represent the masses of the charged Higgs boson, b quark, c quark and tau, respectively. The parameters ξ_d and ξ_u are presented in Table 1.2. The ratio of two VEVs is defined as $\tan\beta = v_2/v_1$, where v_i denotes the vacuum expectation value (VEV) of H_i, and is an important parameter of the theory, as shown in the following sections.

Several types of 2HDM models exist: in type I, all masses of quarks and leptons are obtained by the VEV of H_2. In type II, H_2 provides mass to the up-type quarks, while H_1 provides mass to the down-type quarks and leptons. In type X, quark and lepton obtain their mass from H_2 and H_1 respectively. Finally, in type Y H_1 provides the masses of the down-type quarks, while other fermions obtain their masses from H_2.

2HDM - Type II

We will use the type II 2HDM (2HDM-II) to show how the decay rates for $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ are affected by the charged Higgs contribution. A Feynman diagram for the $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decay mediated by a charged Higgs is shown in Figure 1.4a. The matrix element created by the additional 2HDM-II Lagrangian terms that affect $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays are

$$\mathcal{M}^{\lambda_\tau}_{\lambda_M}(q^2,\theta_\tau) = \frac{G_F V_{cb}}{\sqrt{2}} L'^{\lambda_\tau}\left[\tan^2\beta \frac{m_b m_\tau}{m^2_{H^\pm}-q^2} H'^{\lambda_\tau}_R + \frac{m_c m_\tau}{m^2_{H^\pm}-q^2} H'^{\lambda_\tau}_L\right] \tag{1.35}$$

The square of $\sim \tan\beta/m^2_{H^\pm}$ enhances the 2HDM contributions to the matrix element of Equation 1.35, which is also proportional to the masses of the quarks and leptons that participate in the $H^\pm$ mediation. For this reasons, $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays are affected by NP contributions by a factor of of $(m_\tau/m_\mu)^2 \approx 280$ more than the $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays.

The leptonic and hadronic currents due to the charged Higgs mediation are defined as

$$\begin{aligned} L'^{\lambda_\tau}(q^2,\theta_\tau) &\equiv \langle \tau(p_\tau,\lambda_\tau)\bar{\nu}_\tau(p_\nu)|\,\bar{\tau}(1-\gamma_5)\nu_\tau\,|0\rangle\,, \\ H'^{\lambda_M}_{R,L}(q^2) &\equiv \langle M(p_M,\lambda_M)|\,\bar{c}(1\pm\gamma_5)b\,|\bar{B}(p_B)\rangle \end{aligned} \tag{1.36}$$

where the upper sign is for R and the lower sign for L. It is possible to rewrite these currents in terms of the SM scalar currents ($\lambda_W = s$) as

$$\begin{aligned} L'^{\lambda_\tau} &= \frac{\sqrt{q^2}}{m_\tau} L^{\lambda_\tau}_s \\ H'^{s}_{R,L} &= \frac{\sqrt{q^2}}{m_b - m_c} H'^{s}_s \\ H'^{0}_{R,L} &= \pm\frac{\sqrt{q^2}}{m_b + m_c} H'^{0}_s \\ H'^{\pm}_{R,L} &= 0 \end{aligned} \tag{1.37}$$

To differentiate the hadronic currents that contribute to $\bar{B} \to D\tau^-\bar{\nu}_\tau$ ($\lambda_D = s$) and to $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ ($\lambda_{D^*} = \pm, 0$), the helicity of the $D^{(*)}$ meson is specified. The currents with zero helicity are the only ones that contribute, given that the charged Higgs is a scalar (spin 0).

It is possible to express the total 2HDM scalar current as a function of the SM current as

$$H^{\text{2HDM}}_s \equiv H^{\text{SM}}_s \times \left(1 - \frac{\tan^2\beta \times q^2}{(m^2_{H^\pm}-q^2)(1\mp m_c/m_b)} + \frac{q^2}{(m^2_{H^\pm}-q^2)(m_c/m_b \mp 1)}\right) \tag{1.38}$$

where the upper sign is for $\bar{B} \to D\tau^-\bar{\nu}_\tau$ decays and the lower sign for $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ decays. For values of $m_{H^\pm}$ larger than 15 GeV/c^2 the average q^2 value is about 8 GeV/c^2 , which

means $m^2_{H^\pm} \gg q^2$. The region for $m^2_{H^\pm} \leq 15$ GeV/c^2 has already been excluded by $B \to X_s\gamma$ measurements [32, 33], so that the following approximate expression can be used

$$H_s^{\text{2HDM}} \approx H_s^{\text{SM}} \times \left(1 - \frac{\tan^2\beta}{m^2_{H^\pm}} \frac{q^2}{1 \mp m_c/m_b}\right) \tag{1.39}$$

To give an expression for $\mathcal{R}(D^{(*)})$ in the 2HDM-II as a function of $\tan\beta/m_{H^\pm}$, it is possible to substitute H_s^{SM} for H_s^{2HDM} in Equation 1.24. By discarding the NP dependence of the denominator in Equation 1.26 as negligible, the values of $\mathcal{R}(D^{(*)})$ in the 2HDM follow a parabola in the variable $\tan^2\beta/m^2_{H^\pm}$

$$R(D^{(*)})_{2HDM} = R(D^{(*)})_{SM} + A_{D^{(*)}} \frac{\tan^2\beta}{m^2_{H^\pm}} + B_{D^{(*)}} \frac{\tan^4\beta}{m^4_{H^\pm}} \tag{1.40}$$

The impact of a charged Higgs is larger on $\mathcal{R}(D)$ than on $\mathcal{R}(D^*)$, as a consequence of the fact that $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ decays receive contributions from the hadronic currents $H_\pm$, which weaken the NP contribution affecting only H_s^{2HDM}.

As shown in Figure 1.5, the values of $\mathcal{R}(D^{(*)})$decrease until the charged Higgs contribution to H_s^{2HDM} equals the SM contribution, which happens at $\tan\beta/m_{H^\pm} \sim 0.1\,\text{c}^2/\text{GeV}$ for $\bar{B} \to D\tau^-\bar{\nu}_\tau$ decays and at $\tan\beta/m_{H^\pm} \sim 0.2\,\text{c}^2/\text{GeV}$ for $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ decays. For large values of $\tan\beta/m_{H^\pm}$, the charged Higgs contribution completely dominates and $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ increase rapidly. Based on several theoretical works and their

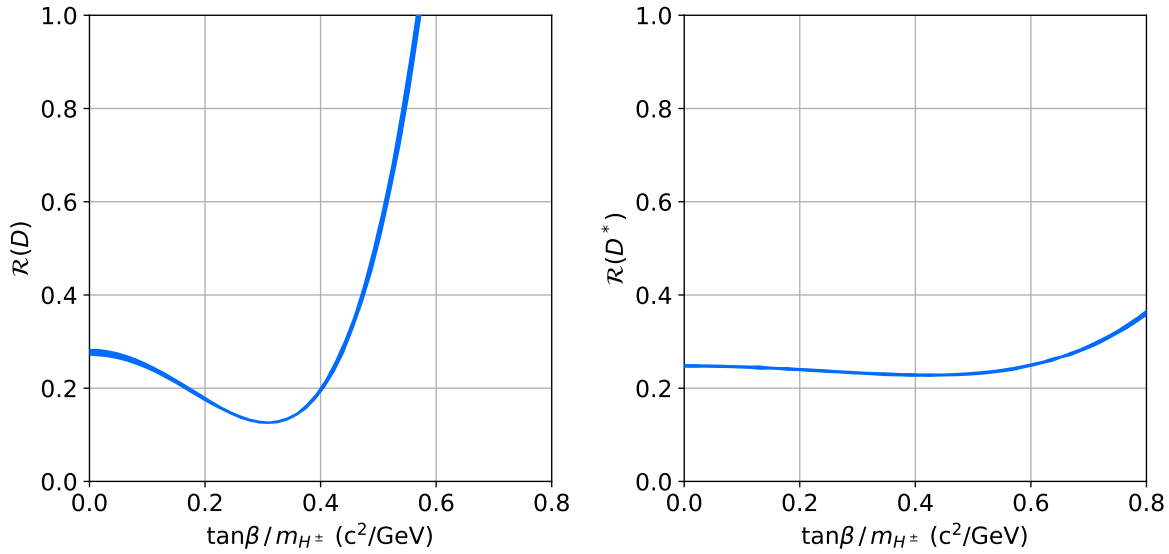


Figure 1.5: Values of $\mathcal{R}(D)$ (left) and $\mathcal{R}(D^*)$ (right) in the 2HDM-II as a function of the $\tan\beta/m_{H^\pm}$ parameter ratio, obtained using the `flavio` package [34].

experimental data, the BaBar collaboration shows that the 2HDM-II is excluded at 99.8% confidence level (CL) [35]. Interestingly, Belle's results are compatible with the 2HDM-II in the $\tan\beta/m^2_{H^\pm}$ region around 0.45 c^2/GeV and zero [36].

Leptoquarks

Leptoquarks (LQs) are new scalar and vector bosons that provide a unified description of quarks and leptons as an extension of the SM, since they decay into a quark and a lepton. Their properties are non-zero baryon and lepton numbers, color, and have a fractional electric charge.

It is possible to reduce the number of possible LQ states, and arrive at the conclusion that only ten LQ models are invariant under the SM gauge group. Among them, six LQ bosons are relevant for the process $b \to c\tau\bar{\nu}$ [37]. Since LQs would couple to the two quarks and two leptons in the process, they could be good candidates to explain the $\mathcal{R}(D^{(*)})$ anomalies. A Feynman diagram for the $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decay mediated by LQs is shown in Figure 1.4b.

One finds the general Wilson coefficients for all possible types of leptoquarks contributing to the $b \to c\tau\bar{\nu}$ process at the energy scale $\mu = M_X$ (where X represents a leptoquark) are

$$C^l_{V_1} = \frac{1}{2\sqrt{2}G_F V_{cb}} \sum_{k=1}^{3} V_{k3} \left[\frac{g^{kl}_{1L} g^{23*}_{1L}}{2M^2_{S_1^{1/3}}} - \frac{g^{kl}_{3L} g^{23*}_{3L}}{2M^2_{S_3^{1/3}}} + \frac{h^{2l}_{1L} h^{k3*}_{1L}}{M^2_{U_1^{2/3}}} - \frac{h^{2l}_{3L} h^{k3*}_{3L}}{M^2_{U_3^{2/3}}} \right], \tag{1.41a}$$

$$C^l_{V_2} = 0\,, \tag{1.41b}$$

$$C^l_{S_1} = \frac{1}{2\sqrt{2}G_F V_{cb}} \sum_{k=1}^{3} V_{k3} \left[-\frac{2g^{kl}_{2L} g^{23*}_{2R}}{M^2_{V_2^{1/3}}} - \frac{2h^{2l}_{1L} h^{k3*}_{1R}}{M^2_{U_1^{2/3}}} \right], \tag{1.41c}$$

$$C^l_{S_2} = \frac{1}{2\sqrt{2}G_F V_{cb}} \sum_{k=1}^{3} V_{k3} \left[-\frac{g^{kl}_{1L} g^{23*}_{1R}}{2M^2_{S_1^{1/3}}} - \frac{h^{2l}_{2L} h^{k3*}_{2R}}{2M^2_{R_2^{2/3}}} \right], \tag{1.41d}$$

$$C^l_{T} = \frac{1}{2\sqrt{2}G_F V_{cb}} \sum_{k=1}^{3} V_{k3} \left[\frac{g^{kl}_{1L} g^{23*}_{1R}}{8M^2_{S_1^{1/3}}} - \frac{h^{2l}_{2L} h^{k3*}_{2R}}{8M^2_{R_2^{2/3}}} \right], \tag{1.41e}$$

where h^{ij} and g^{ij} are the dimensionless couplings; S_1, S_3, and R_2 are scalar leptoquark bosons; U_1, U_3, and V_2 are vector leptoquark bosons; the index i (j) indicates the generation of quarks (leptons); and V_{k3} denotes the CKM matrix elements. The LQ electric charge is denoted by its upper index. These six leptoquark bosons (S_1, S_3, R_2, U_1, U_3, and V_2) can contribute to $\bar{B} \to D^{(*)}\tau\bar{\nu}$.

As can be seen in Equations (1.41a)-(1.41e), several leptoquark bosons with several combinations of the couplings can contribute to $b \to c\tau\bar{\nu}_l$. Those contributions can be classified as

- $C^l_{S_2} = -4C^l_T$ mediated by S_1 boson with nonzero value of $(g_{1L}g^*_{1R})$,
- $C^l_{S_2} = 4C^l_T$ by R_2 boson with $(h_{2L}h^*_{2R})$,
- $C^l_{V_1}$ by S_1, S_3, U_1, or U_3 bosons with $(g_{1L}g^*_{1L})$, $(g_{3L}g^*_{3L})$, $(h_{1L}h^*_{1L})$, or $(h_{3L}h^*_{3L})$,

- $C_{S_1}^l$ by U_1 or V_2 bosons with $(h_{1L}h_{1R}^*)$ or $(g_{2L}g_{2R}^*)$.

To evaluate the effects on the observables $R(D)$ and $R(D^*)$, the running effect of $C_Y^l(\mu)$ (Y denotes the types of the effective operators) from $\mu = M_X$ to $\mu = \mu_b$, where μ_b is the mass scale of the bottom quark, must be taken into account. A scale dependence in the scalar $S_{1,2}$ and tensor T currents exist and is approximately evaluated as

$$C_{S_{1,2}}(\mu_b) = \left[\frac{\alpha_s(m_t)}{\alpha_s(\mu_b)}\right]^{-\frac{12}{23}} \left[\frac{\alpha_s(m_{\rm LQ})}{\alpha_s(m_t)}\right]^{-\frac{4}{7}} C_{S_{1,2}}(m_{\rm LQ})\,, \tag{1.42}$$

$$C_T(\mu_b) = \left[\frac{\alpha_s(m_t)}{\alpha_s(\mu_b)}\right]^{\frac{4}{23}} \left[\frac{\alpha_s(m_{\rm LQ})}{\alpha_s(m_t)}\right]^{\frac{4}{21}} C_T(m_{\rm LQ})\,, \tag{1.43}$$

where $\alpha_s(\mu)$ is the running QCD coupling at a scale μ.

The branching ratios of $\bar{B} \to D^{(*)}\tau\bar{\nu}$ can be calculated, given hadronic form factors that are precisely estimated with use of the heavy quark effective theory. The formulae in terms of the helicity amplitudes are found, e.g., in References [38, 39].

1.3 Measurement Motivation

Semileptonic B-mesons decays are well understood in the SM and, as a consequence, they offer good opportunities for testing the SM by using experimental data. As discussed in the previous section, several NP models, such as charged Higgs and leptoquarks, could mediate the $b \to c\tau\bar{\nu}$ process. Therefore, $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ is an optimal candidate to search for NP effects. A Feynman diagram for this decay is shown in Figure 1.3.

The experiments Belle, BaBar and LHCb have found that measured values for the branching fraction ratios

$$\mathcal{R}(D) = \frac{\mathcal{B}\left(\bar{B} \to D\tau^-\bar{\nu}_\tau\right)}{\mathcal{B}\left(\bar{B} \to D\ell^-\bar{\nu}_\ell\right)} \tag{1.44}$$

and

$$\mathcal{R}(D^*) = \frac{\mathcal{B}\left(\bar{B} \to D^*\tau^-\bar{\nu}_\tau\right)}{\mathcal{B}\left(\bar{B} \to D^*\ell^-\bar{\nu}_\ell\right)}, \tag{1.45}$$

exceed the SM expectation values in a consistent way, where ℓ refers to both electrons and muons, and the notation $D^{(*)}$ is used to refer to both the D and D^* states.

As shown in the previous Sections, the calculation of the SM values of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are straightforward, and the most recent predictions for these ratios are [40]

$$\mathcal{R}(D)_{\rm SM} = 0.299 \pm 0.003 \tag{1.46}$$

$$\mathcal{R}(D^*)_{\rm SM} = 0.258 \pm 0.005, \tag{1.47}$$

while the world averages for the experimental results prior to this analysis are [40]

$$\mathcal{R}(D)_{\rm exp} = 0.407 \pm 0.039 \pm 0.024 \tag{1.48}$$

$$\mathcal{R}(D^*)_{\rm exp} = 0.306 \pm 0.013 \pm 0.007. \tag{1.49}$$

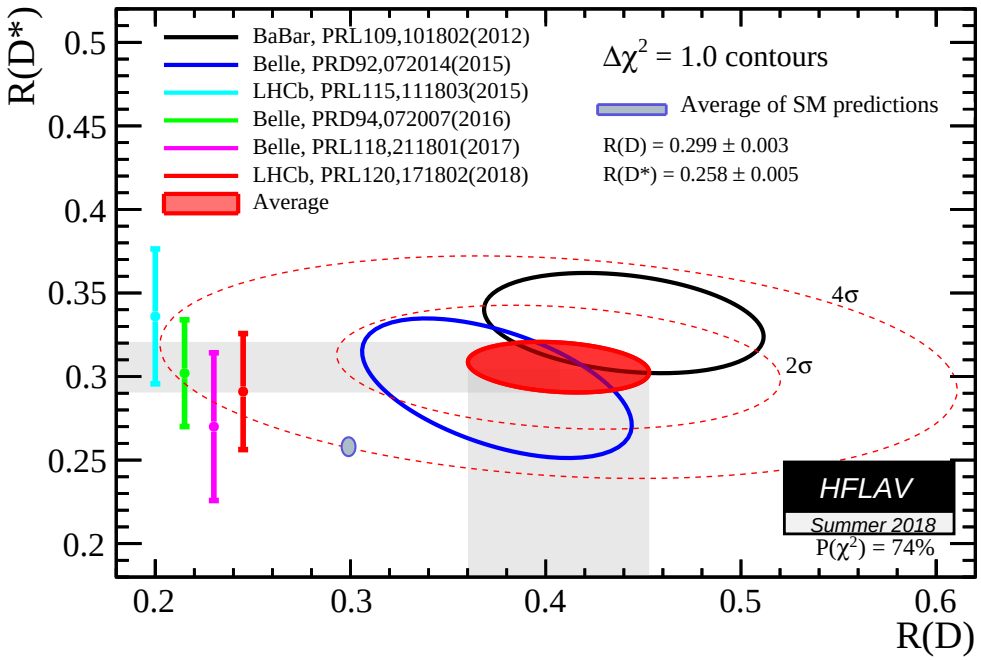


Figure 1.6: Theoretical SM expectation for the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ ratios are compared with all experimental results and their average, as of June 2018 [40].

The statistical significance of the world average discrepancy with the SM predictions is $\sim 2.3\sigma$ and $\sim 3.0\sigma$ for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ respectively, while the discrepancy accounts to 3.9σ when combining the results of the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ analysis. Both experimental and theoretical values for the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ ratios are displayed in Figure 1.6, and their numerical values are listed in Table 1.3.

1.3.1 Procedure outline

Our goal is to measure the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ ratios defined in Equations 1.44 and 1.45. We define the $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decay as the *signal mode* and the $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decay as the *normalization mode*.

At Belle B-mesons are produced in pairs through the decay of the $\Upsilon(4S)$ resonance, produced in e^+e^- collisions at a known center-of-mass (CM) energy. The reconstruction of one B-meson is the role of the *tagging* and, as a consequence, we refer to this B-meson as the B_{tag} . Even though the B_{tag} will not be explicitly used for the final measurement extraction, it ensures one half of the full $\Upsilon(4S)$ is well reconstructed, therefore limiting the degrees of freedom of the reconstruction of the remaining B-meson and increasing the chances it will be reconstructed correctly. Indeed, we reconstruct the other B-meson in the event in the $\bar{B} \to D\ell^-\bar{\nu}_\ell$ and $\bar{B} \to D^*\ell^-\bar{\nu}_\ell$ channels where $\ell^- = \mu^-$ or e^-, and we call this B-meson the B_{sig} . We apply a series of selection criteria to make sure we are

left only with B_{sig} candidates that are likely to be either the signal or the normalization mode.

Owing to its short lifetime ($2.906(10) \times 10^{-13}$ s), the τ lepton has a flight mean length of only 87 μm, which means that it decays inside the beam pipe, before reaching any subdetector. Its large mass, 1.78 GeV/c^2 allows for hadronic decays, which make up for around 65 % of the total branching ratio, however we only consider the cleaner leptonic τ decays such as $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$. In these decays the final state particles of signal and normalization modes are the same, and, as a consequence, the same reconstruction and selection procedure is applied to both modes.

By measuring the ratio of these two modes we eliminate many sources of systematic uncertainties for the measurement, and therefore increase the power of the compatibility test with SM predictions. On the experimental side, this results from similarities between the kinematics of the decay products of signal and normalization modes, up to the differences introduced by the fact that the τ lepton is heavier than the light leptons, this corresponds to a cancellation of experimental efficiency uncertainties between the two modes. On the theoretical side, this is due to cancellation of factors such as $|V_{cb}|$ in the ratio, since it is present in both the signal and normalization mode decay rates. The ratio measurements also lead to a partial cancellation of theoretical uncertainties related to hadronic effects.

Using Monte Carlo simulation samples we build Probability Distribution Functions (PDFs) for all relevant categories of events that end up in our B_{sig} sample, that is signal, normalization and various backgrounds events. We extract the number of signal and normalization events reconstructed in the data sample by finding the best fit of the PDFs yields that represents the measured data sample, and by scaling it to the full branching ratios using efficiency factors, we can finally measure $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$.

1.4 Previous Measurements

Between 2007 and 2010 several measurements from the Belle and BaBar experiments found evidence for the $\bar{B} \to D\tau^-\bar{\nu}_\tau$ and $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ decays and provided the first measurements for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$İn 2012 BaBar produced a measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ using an hadronic tag, which showed a large discrepancy with the SM predictions, of 2.0σ and 2.7σ respectively [41]. In 2015 Belle conducted the same measurement, which, despite having a smaller combined significance of 1.8σ, also showed a similar excess to the one seen by BaBar [36]. In 2016 Belle performed a semileptonic tag study that was limited to $\mathcal{R}(D^{*+})$ and once again saw an excess for $\mathcal{R}(D^*)$, of 1.6σ.

Due to the difficulty in reconstructing of electrons at LHCb, they only used the the muonic channel $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$ for a first measurement in 2015, which showed a 2σ discrepancy with the SM [42]. On the other hand, a subsequent measurement that used the hadronic $\tau^- \to \pi^+\pi^-\pi^-(\pi^0)\nu_\tau$ was performed in 2018 and found to be compatible with the SM within 1σ [43].

All these measurements are limited by the statistical sizes of the experimental

datasets and therefore are not particularly significant if taken singularly. However when combined in the world average they produce a discrepancy of 3.9σ with the SM predictions [40].

Results for the previous measurements of the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are summarized in Table 1.3. Notably, more studies have focused on $\mathcal{R}(D^*)$ than on $\mathcal{R}(D)$ and in particular, no measurement that uses a semileptonic tag has ever been performed for $\mathcal{R}(D)$. This thesis will focus on a semileptonic tag and will perform simultaneous measurements of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ using the Belle dataset, which allows to extract their correlations and provide information about their potential disagreement with the SM predictions in the 2D plane $\mathcal{R}(D)$ vs $\mathcal{R}(D^*)$.

Table 1.3: Summary of the previous measurements for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ produced by the B-factories Belle and BaBar, and by LHCb, together with the world average and the SM expectations provided by the HFLAV group. The first error is statistical and the second one is systematic.

Experiment	Integrated luminosity	Tag	τ mode	$\mathcal{R}(D)$	$\mathcal{R}(D^*)$
BaBar 12 [41]	433 fb^{-1}	Hadr	$\ell\nu\nu$	$0.440 \pm 0.058 \pm 0.042$	$0.332 \pm 0.024 \pm 0.018$
Belle 15 [36]	711 fb^{-1}	Hadr	$\ell\nu\nu$	$0.375 \pm 0.064 \pm 0.026$	$0.293 \pm 0.038 \pm 0.015$
LHCb 15 [42]	3 fb^{-1}	-	$\ell\nu\nu$	-	$0.336 \pm 0.027 \pm 0.030$
Belle 16 [44]	711 fb^{-1}	Semilep	$\ell\nu\nu$	-	$0.302 \pm 0.030 \pm 0.011$
Belle 17 [45]	711 fb^{-1}	Hadr	$\pi\nu, \rho\nu$	-	$0.270 \pm 0.035 \pm 0.027$
LHCb 18 [43]	3 fb^{-1}	-	3π	-	$0.291 \pm 0.019 \pm 0.029$
Average [40]				$0.407 \pm 0.039 \pm 0.024$	$0.306 \pm 0.013 \pm 0.007$
SM [40]				0.299 ± 0.003	0.258 ± 0.005

1.4.1 Previous Belle Semileptonic Tag Analysis

The previous Belle analysis that used a semileptonic tag to measure $\mathcal{R}(D^*)$ was limited to the $\bar{B}^0 \to D^{*+}\ell^-\nu$ channel for both the signal side and the tag side [44]. In addition, signal side and tag side reconstruction were identical and only the variable $\cos\theta_{B,D^{(*)}\ell}$ (presented in Section 3.2.1) was used for selection of correctly reconstructed B_{tag} candidates. This analysis extends the study not only to $B^- \to D^{*0}\ell^-\nu$ decays for $\mathcal{R}(D^*)$, which are more difficult to measure due to the lower purity of the D^{*0} sample, but also measures $\mathcal{R}(D)$. Moreover, the tag channels are extended to include both $\bar{B} \to D\ell^-\bar{\nu}_\ell$ and $\bar{B} \to D^*\ell^-\bar{\nu}_\ell$ decays which ensures a larger signal reconstruction efficiency and stronger statistical significance. The tag reconstruction is performed

by a dedicated algorithm that relies on multivariate analysis methods. As a consequence, we have access to a much more efficient method to select well reconstructed B_{tag} candidates, namely the classifier output, which can be combined with the variable $\cos\theta_{B,D^{(*)}\ell}$.

Chapter 2

Experimental Apparatus

2.1 KEKB Accelerator

The KEKB accelerator [46, 47] was located in Tsukuba, Japan, and provided beams of electrons and positrons with asymmetric energies of 8 and 3.5 GeV respectively. The electrons are evaporated off a filament, accelerated to 8 GeV in the Linear Accelerator (LINAC) and then injected into the electron ring. Part of the electron beam is diverted into a target where electron-positron pairs are produced. Upon their separation from the companion electrons, the positrons are accelerated to 3.5 GeV and injected into the positron ring. Electrons and positrons are then grouped into bunches spaced approximately 60 cm apart, a distance that corresponds to about 2 ns in time.

The resulting center-of-mass (CM) energy was $\sqrt{s} = 10.58$ GeV, which corresponds to the $\Upsilon(4S)$ resonance. This resonance is a bound state of a b and a $\bar{b}$ quark and it decays almost exclusively to a pair of B-mesons (with a branching fraction greater than 96%). Thus, these e^+e^- experiments are known as B-factories. Almost all the energy apart from a 20 MeV excess, is used for the mass of the B-mesons, so that in the CM frame they are produced almost at rest. Given that the main goal of the Belle experiment was to perform a time-dependent analysis of CP violation in the B-meson sector, which involves the measurement of the position of B decay vertices and the determination of their flight times, it was necessary to have B-meson vertices displaced from the interaction point (IP). In order to achieve this requirement, the B-mesons particles were boosted. Indeed, the choice of asymmetric energies provides a Lorentz boost in the CM system of $\beta\gamma = 0.425$, defined as

$$\beta\gamma = \frac{E_- - E_+}{2\sqrt{E_- E_+}}, \tag{2.1}$$

where E_-, E_+ are the energies of the electron and positron beams, respectively.

The resulting mean distance between the B decays vertexes, Δz, along the beam direction was equal to 200 μm, with a corresponding time distance $\Delta t = \Delta z/\beta\gamma c$. The

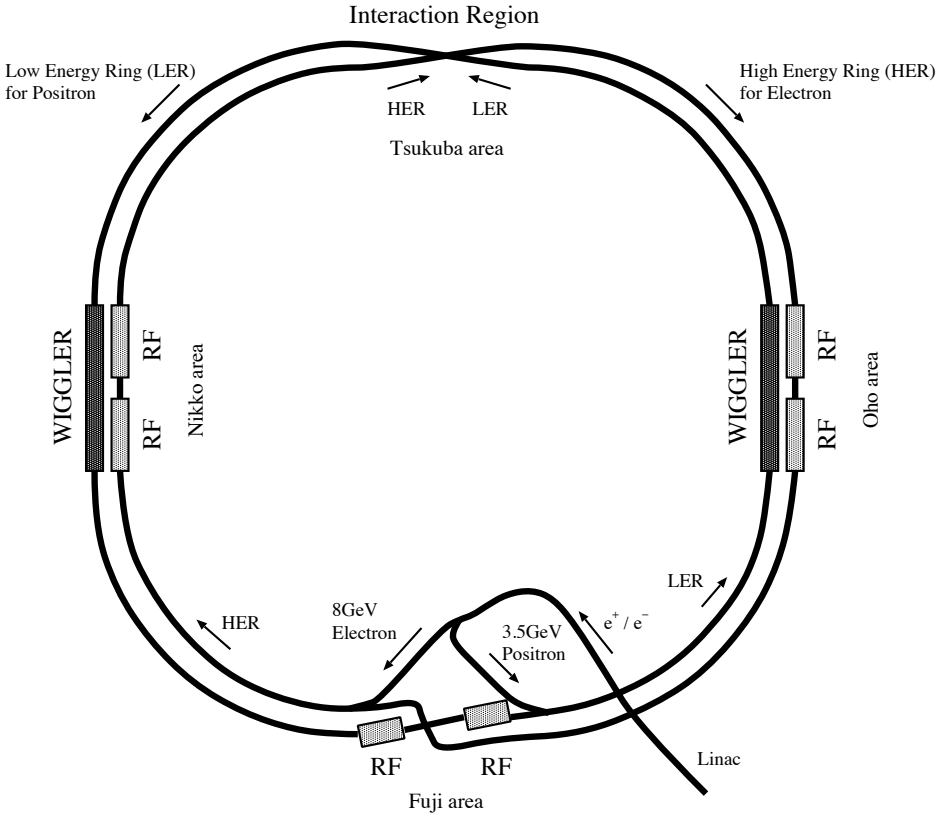


Figure 2.1: Schematic diagram of the KEKB electron (high energy) and positron (low energy) storage rings and the Linac that feeds them. The Belle detector is located at the Tsukuba area and surrounds the Interaction Region.

instantaneous luminosity of an accelerator is defined as

$$L = \frac{N_b n_{e^-} n_{e^+} f}{A_{\text{eff}}}, \tag{2.2}$$

where n_{e^-} and n_{e^+} represent the numbers of electrons and positrons in each particles bunch respectively, the number of bunches is given by N_b, the circulation frequency is given by f, and A_{eff} is the cross-section of the overlapping transverse area of the beams measured at the IP. The design goal for the instantaneous luminosity of the KEKB accelerator was $L = 1.0 \times 10^{34}\text{cm}^{-2}\,\text{s}^{-1}$, which translates into a production rate of 10 $B\bar{B}$ pairs per second. This expectation was largely exceeded, setting the world record for the instantaneous luminosity of an e^+e^- collider at $L = 2.1 \times 10^{34}\text{cm}^{-2}\,\text{s}^{-1}$.

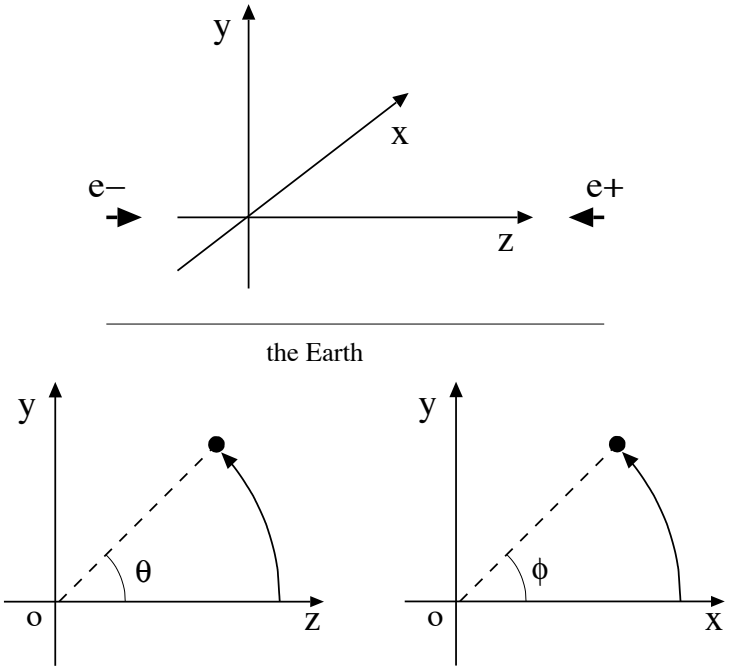


Figure 2.2: The coordinate system used when describing the Belle detector and directional parameters.

2.2 Belle Detector

The Belle detector [48] was a multi-purpose detector with a solid angle acceptance of 4π, made of several components, each of them with a specific task for the reconstruction and identification of the particles that are produced in the collisions. The schematic longitudinal cross-section of the Belle detector is shown in Figure 2.3. Each component's location has been chosen to best serve its purpose. The detector is divided into the barrel part and the forward and backward endcaps. The location of the forward (backward) part is defined by direction of the incoming e^- (e^+) beam.

The coordinate system for the Belle detector is defined in Figure 2.2. In this coordinate system, the z-axis runs along the beam line and is in the opposite direction of the e^+ beam, the y-axis is vertical and points upward, while the x-axis horizontal, with its positive direction pointing away from the centre of the accelerator ring. The polar angle θ is measured from the positive z-axis, while the azimuthal angle ϕ is measured from the positive x-axis and lies in the xy plane. The radial dimension r lies in the xy plane and is measured as $r = \sqrt{x^2 + y^2}$.

The Belle experiment has been designed to seach for CP violation in the B-meson decays. In addition, its physics program also included the precise measurements of decays of bottom mesons, charm mesons and τ leptons. It also searched for forbidden or rare processes in the SM.

Belle started the data taking in June 1999. Thereafter, for 6-9 months of every year data taking was continuous until June 2010, which marks the final shutdown of the experiment. The dataset is divided in *Experiments*, where each Experiment started after a major shutdown. The first version of the Silicon Vertex Detector was used for recording Experiments 7 - 27, while the rest is recorded with the second version [49]. The Belle detector is shown in the photographs of Figure 2.4.

2.2.1 Beam Pipe

The beam pipe contains the IP of the electron and positron beams. Its volume determines the distance of the closest SVD layer, and therefore it has to be carefully designed. The beam pipe consists of a double-wall beryllium cylinder, with an inner radius of 20 mm and an outer radius of 23.5 mm, where each wall is only 0.5 mm thick, and helium gas for cooling circulates in the inner gap of 2.5 mm, to counterbalance the heat produced by beam effects. The outer cylinder is covered with a 20 μm thick gold foil to reduce background from synchrotron radiation. In 2003 the beam was upgraded to have an outer radius of 15 mm and a gold foil plating of 10 μm. A detailed description of the beam pipe is given in Ref. [48]. A cross section and side view of the beam pipe is shown in Figure 2.5.

2.2.2 Silicon Vertex Detector

The innermost component, the Silicon Vertex Detector (SVD), has the role of recording accurate information about the first length of charged tracks, with the goal of providing the fundamental building blocks for the reconstruction of the decay vertices. This component is thus fundamental for all time-dependent analyses, which rely on knowing the decay coordinates and hence the decay time of the B-mesons. With the Lorentz boost provided by the asymmetric beam energies of the KEKB accelerator, the decay length difference between two B-mesons is about 200 μm, which is measurable by the SVD, since it provides a resolution of 100 μm for the z-vertex position.

The SVD is composed of several layers of double-sided silicon-strip detectors (DSSDs), which are depleted pn-junctions ionized by the passage of charged particles. The p^+ and n^+ strips are located in directions parallel and perpendicular to the beam line, respectively, therefore providing a measurement of the track in the $r\phi$ and z directions.

The first version of the SVD, labeled SVD1, was used from 1999 to 2003, and consisted of three layers of DSSDs. Subsequently, the diameter of the beam pipe was reduced, allowing for hit measurements closer to the IP and one layer of the DSSDs was added to the SVD, which was then called SVD2, see Figure 2.6. Coincidentally the angular coverage of the SVD was improved to reach a range of $17^\circ < \theta < 150^\circ$. The radii of the 4 layers SVD2 are 20, 43.5, 70 and 80 mm. This upgrade led to an improvement in the z-vertex resolution of low-momentum tracks of about 20% with respect to SVD1. Since the luminosity of the KEKB accelerator increased after the

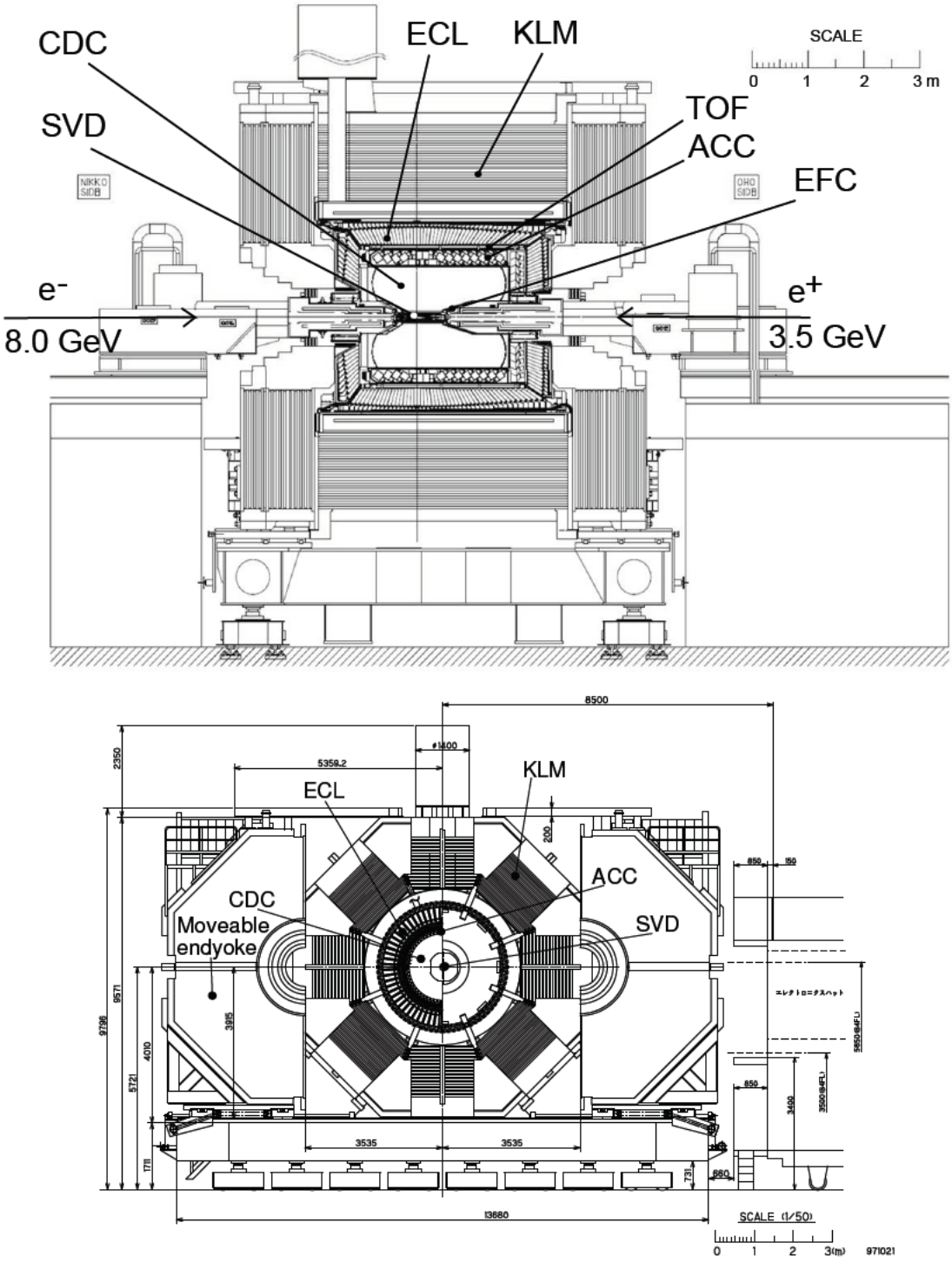


Figure 2.3: At the top (bottom) : longitudinal (transverse) cross sections of the Belle detector [49].

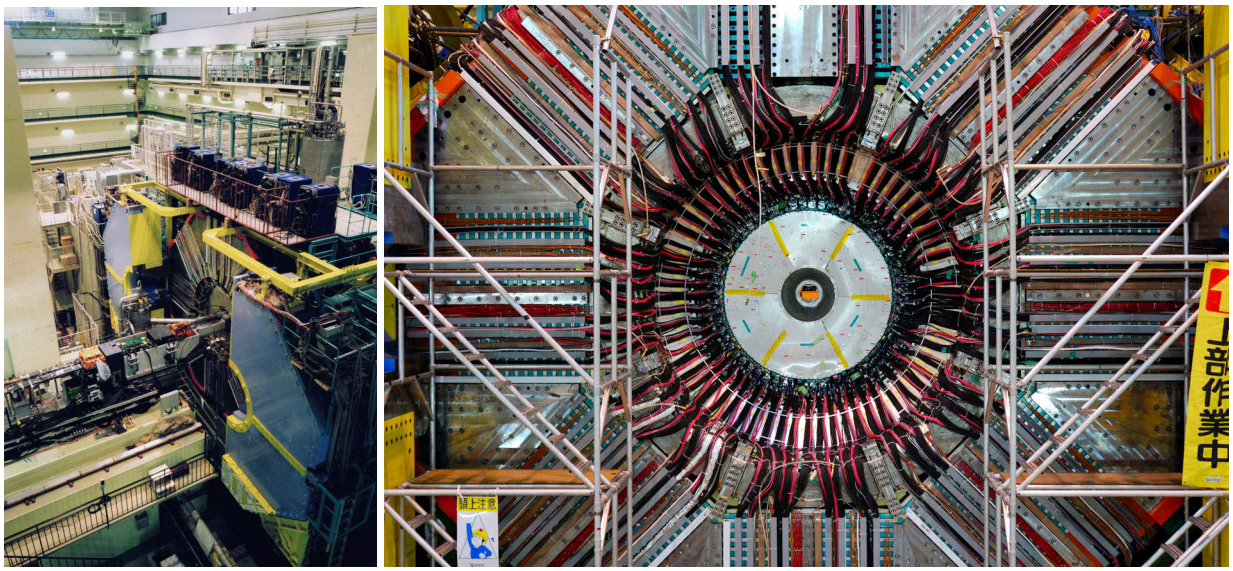


Figure 2.4: The Belle detector is shown in the Tsukuba detector hall on the left. The detector is shown from the beamline view on the right. From the inner to the outer part, the CDC end flange, ACC PMT's, ECL modules and the KLM modules are shown. [49].

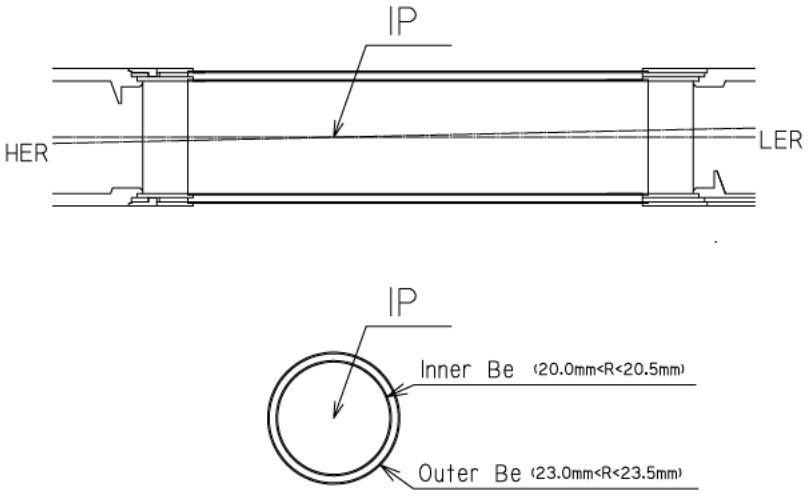


Figure 2.5: A cross section and side view of the beam pipe.

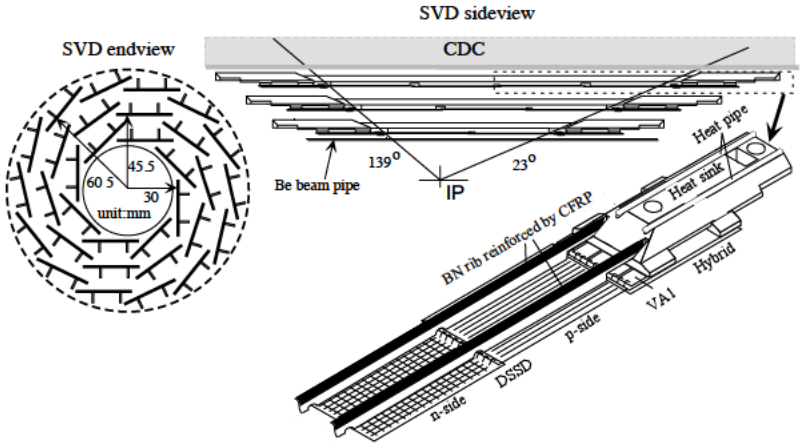


Figure 2.6: Detector configuration of the SVD. [50]

SVD2 installation, the vast majority of the Belle data (85%) was recorded with SVD2.

2.2.3 Central Drift Chamber

The Central Drift Chamber (CDC) serves the purpose of measuring the charged particle tracks and provide energy loss information to help distinguishing the type of charged particle that crossed its volume. Low-momentum tracks, which do not reach the particle identification (PID) system, can be identified using only the CDC. The momentum is measured by exploiting the curvature of the particle's trajectory which is caused by a 1.5 T solenoid. The CDC surrounds the SVD, with an angular coverage of $17^\circ < \theta < 150^\circ$, a length of 2.4 m, and inner and outer radii of, respectively, 83 mm and 888 mm. Its geometry shows an asymmetry in the z-direction to account for the Lorentz boost of the beams. The layout of the CDC can be seen in Figure 2.7.

The working principle of the CDC is that any charged particle will ionize the gas mixture that fills it (50% He and 50% C_2H_6). Electrons and ions in the gas will then drift toward charged wires which are present in large number. The strong electric field in proximity of the wires causes an avalanche effect that increases the signal recorded by the wires. The CDC has a total of 8400 drift cells. The wires run in two directions to provide hit information with $r\phi$ and z coordinates, which, as is the case for the SVD hits, will be processed and assembled by dedicated track finding algorithms. The cell structure and the arrangement are shown in Figure 2.8.

The information from the electric signal recorded by the wires can be used to the measure the energy loss of the charged particle, and is combined with the theoretical knowledge of energy loss for velocity-dependent particles (Bethe-Bloch formula) to provide good separation between kaons and pions up to particle momenta of approximately 1.5 GeV/c^2 . Figure 2.9, shows a plot of dE/dx versus momentum for experimental data. A particle's energy loss is dependent on its boost, for a given momentum. Hence,

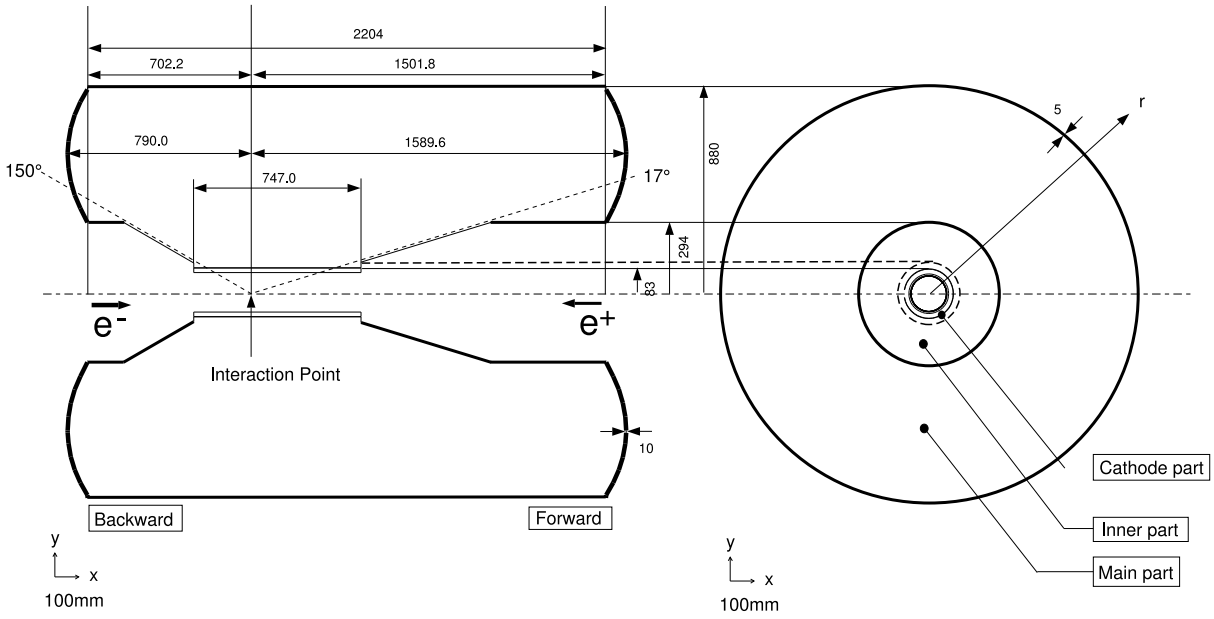


Figure 2.7: Geometrical layout of the CDC [48].

dE/dx will differ between particle types and can be used for particle identification. The separation between kaons, pions, protons and electrons is shown in Figure 2.9.

2.2.4 Aerogel Cherenkov Counter

The Aerogel Cherenkov Counter (ACC) provides useful information to distinguish kaons from pions by using fine mesh-type photomultiplier tubes (PMTs) to measure the light of the Cherenkov cone, which is produced when a charged particle travels in a medium with a velocity larger than that of light in the same medium. The effective momentum range for the kaon/pion separation is 1.2 GeV/$c < p <$ 3.5 GeV/c , thus complementing the energy loss information from the CDC. Mathematically, we can set the relationship

$$n > \frac{1}{\beta} = \sqrt{1 + \left(\frac{m}{p}\right)^2}, \tag{2.3}$$

which defines the condition for the emission of Cherenkov light for a particle with mass m, momentum p, and velocity β, when n is the refractive index of the medium. In particular, the Cherenkov threshold of electron is a few MeV/c order while that of pion is between 0.5 and 1.0 GeV/c , depending on the refractive index. The ACC threshold counters cover a polar angle of $33.3^\circ < \theta < 127.9^\circ$ in the barrel, and $13.6^\circ < \theta < 33.4^\circ$ in the forward endcap 2.10. The barrel (endcap) part comprises 60 (12) identical sectors in the ϕ direction, and 16 (19) modules can be found in each sector. One module has a typical size of approximately $120 \times 120 \times 120$ mm^3, and is occupied by a silica aerogel radiator. One or two fine-mesh PMT(s) read out each counter to detect Cherenkov

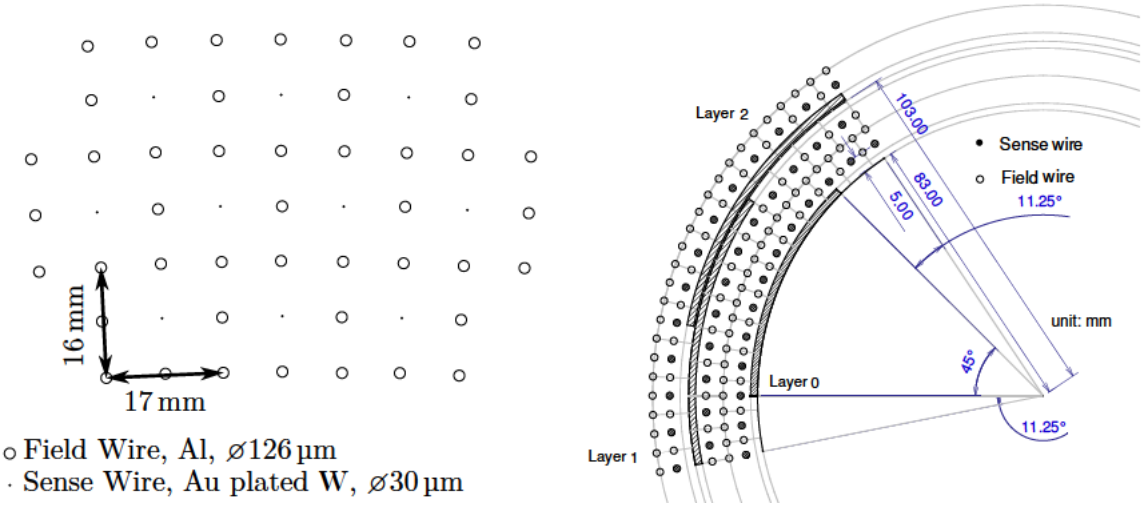


Figure 2.8: Cell structure and arrangement of wires in the CDC [48].

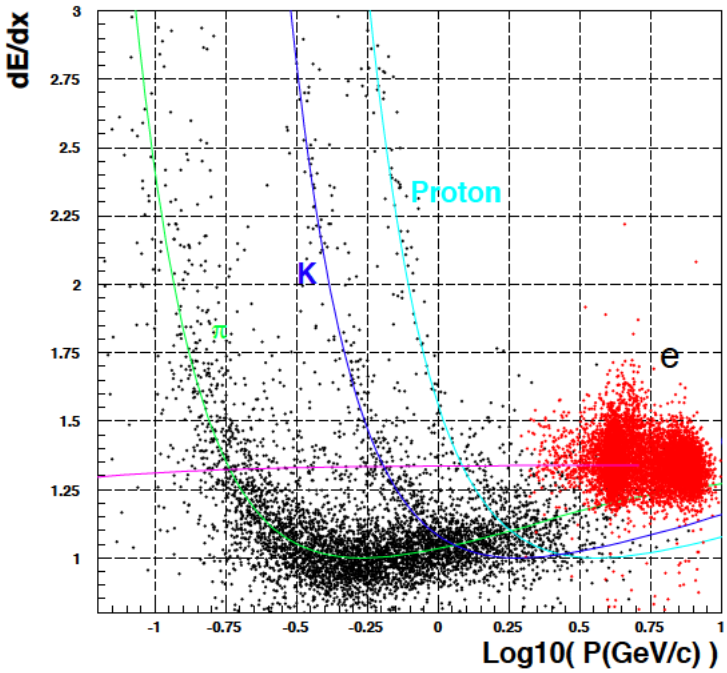


Figure 2.9: A plot of dE/dx versus momentum for experimental data. The mean dE/dx is shown as a function of momentum for several particle species.

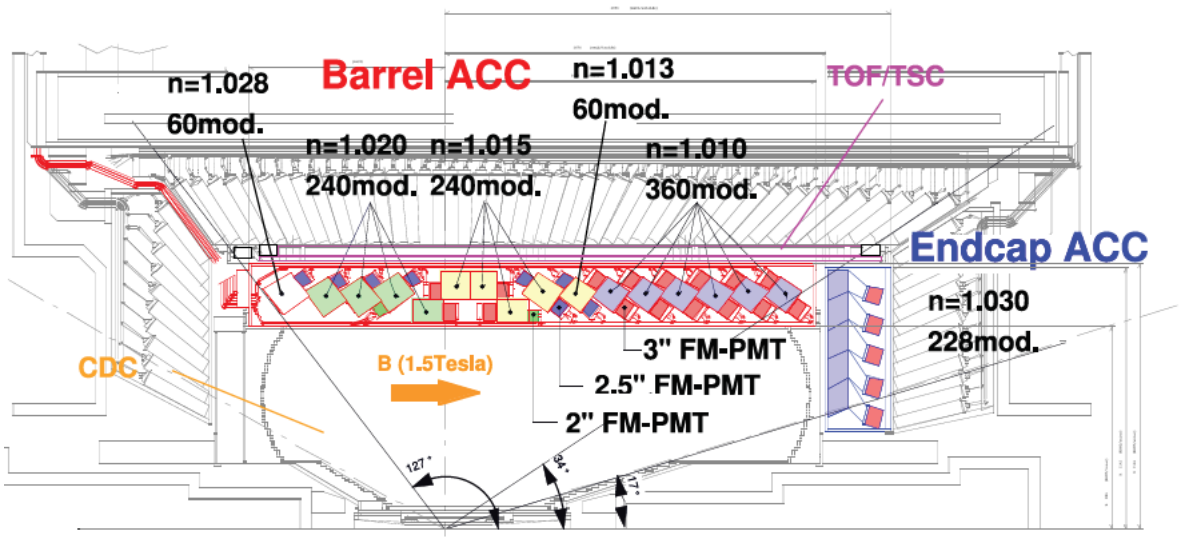


Figure 2.10: Arrangement of the ACC modules [48].

light in an axial magnetic field of 1.5 T.

2.2.5 Time-of-Flight Counter

The Time-of-Flight (TOF) counter measures the time difference between the collision of the beams and the instant where a charged particle crosses its volume. Combined with the determination of the momentum performed by the CDC, this measurement of the time-of-flight, T, allows us to determine the mass of the particle, m, by using the equation:

$$m = \frac{p}{c}\sqrt{\left(\frac{cT}{L}\right)^2 - 1}, \tag{2.4}$$

where p is the momentum of the particle and L is the distance traveled by the particle on its helical path from the interaction point to the TOF.

The TOF system consists of a barrel of 128 plastic scintillator counters and provides particle identification for particles below 1.2 GeV/c^2 . Since the TOF has a very short time resolution, around 100 ps, it is also used as a signal for the trigger system. Two TOF counters and one thin trigger scintillation counter (TSC) can be found in each module. The TOF counter is read out at both ends by fine-mesh PMTs. The angular acceptance is $33^\circ < \theta < 121^\circ$. 0.28 GeV/$c$ is the minimum transverse momentum necessary to reach a TOF counter.

2.2.6 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECL) collects and measures the energy deposited by charged particles and photons. The ECL consists of 8736 Thallium-doped Caesiu-

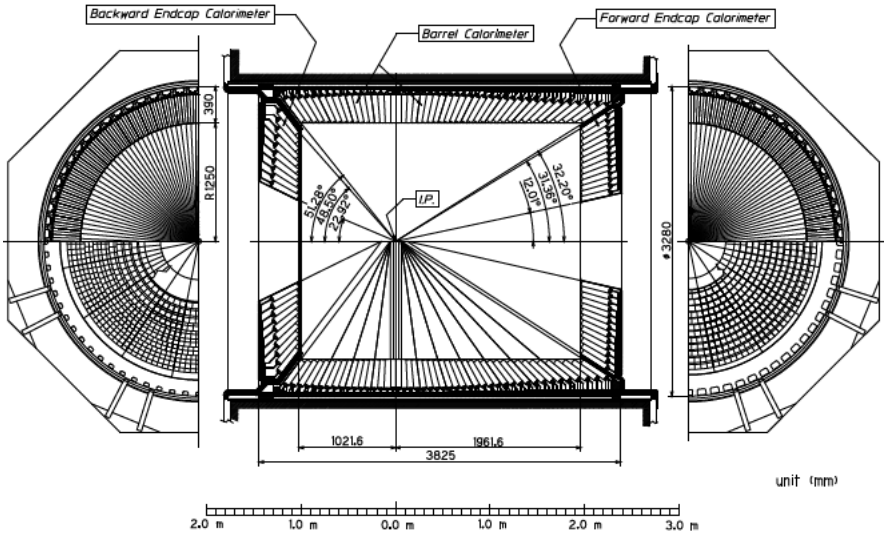


Figure 2.11: The Belle Electromagnetic Calorimeter [48].

miodide (CsI(Tl)) crystal counters. The cylindrical CsI(Tl) crystals are 30 cm long, which is equivalent to 16.2 radiation lengths, and point towards the IP. The ECL is composed of barrel and end cap parts. The barrel consists of 6624 crystals and the end caps contain 1152 (960) crystals in the forward (backward) directions. The scintillation light produced by particles in the crystals is detected with silicon photodiodes. The angular coverage of the ECL is $17^\circ < \theta < 150^\circ$ (total solid-angle coverage of 91% of 4π) and its geometry is shown in Figure 2.11.

The ECL energy resolution varies from 4% at 100 MeV to about 1.6% at 8 GeV. Such an energy and angular resolution provides a π^0 mass resolution equal to approximately 4.5 MeV/c^2 . The ratio of the ECL shower energy to the track momentum, E/p, is a fundamental parameter used for electron/hadron separation.

Electrons and photons deposit their energy in the crystals producing electromagnetic showers via bremsstrahlung and pair production. Also heaver particles like pions, muons and kaons deposit energy in the ECL, however they will do so less effectively. As a consequence, the ratio E/p will be close to unity for electrons but small for other particles, therefore introducing an additional electron identification method.

2.2.7 Superconducting Solenoid Magnet

The superconducting solenoid magnet is 4.4 m in length and 3.4 m in diameter. It has a 1.5 T magnetic field points which in the z-direction and permeates all subdetectors that within its volume, namely all components except for the KLM. The coil is made of a single layer Niobium-Titanium-Copper (NbTi/Cu) alloy that is located in a high purity aluminum stabilizer, cooled down to superconducting temperatures by circulating liquid helium. The purpose of the magnetic field is to bend the trajectories of charged particles, to allow for accurate measurements of their momentum as explained in the

previous sections.

2.2.8 K_L^0 and Muon Detector

The K_L^0 meson and muon detector is embedded in the iron structure that supports the entire detector. It is a sampling calorimeter that uses the layers of the iron structure as interaction material for particles to cause the production of electromagnetic showers. The iron layers are interspersed with layers of resistive plate counters (RPCs) that are used to detect and measure these showers. The KLM is therefore able to provide directional information for particles such as K_L^0 mesons and muons but poor, unusable, energy resolution.

Each RPC layer has two modules that comprise two glass-electrodes with a high bulk resistivity ($10^{10}\Omega$ cm) and a gas filled gap. A signal on the external read out strips in the ϕ and θ directions is caused by a charged particle passing through the gap. These signals can then be used to determine location and time of the ionization.

The K_L^0 mesons deposit energy mainly in the KLM. Indeed, the 14 iron layers of the KLM and provide 3.9 interaction lengths for K_L^0 mesons, which add to the 0.8 interaction lengths provided by the ECL. As a consequence, a KLM cluster that is not associated to a CDC track is identified as a K_L^0 meson and mainly used as a veto. Muons with large enough momentum (above $\approx$ 600 MeV/c^2) are able to reach the KLM and, having passed through the CDC, the combination of KLM cluster plus CDC track provides a very strong identification criterion for muons.

2.2.9 Trigger and Data Acquisition

Since many signals recorded by the subdetectors do not originate from events that are interesting for the scope of the experiment, and since the data acquisition system has a finite input rate, it is important to discriminate against events that are not relevant. For this reason only events that pass several criteria, or triggers, are retained, and correspond to events such as $e^-e^+ \rightarrow \Upsilon(4S) \rightarrow b\bar{b}$, $e^-e^+ \rightarrow q\bar{q}$, $e^-e^+ \rightarrow \tau^-\tau^+$, $e^-e^+ \rightarrow \gamma\gamma$, Bhabha scattering and others. The main sources of background events are undesirable interactions between the electron-positron beams, collisions of a beam particle with residual gas molecules or with the beam-pipe, and synchrotron radiation from the beams. The cross-sections and corresponding rates for physical processes of interest are listed in Table 2.1. Bhabha and two-photons events are recorded to measure the instantaneous luminosity and to study the detector responses. However, the rates of these processes are very large, which requires to prescale these trigger rates by a factor $\sim$100.

At first, a Level-1 hardware trigger scans all the subdetectors, excluding the SVD, and looks for fundamental signals for the relevant physics events, deciding whether to keep the event within 2.2 μs of the beam crossing. A Level-3 software trigger operates online with fast track reconstruction algorithms, to reject events where tracks originate far from the IP or events with an energy higher than 3 GeV/c^2 deposited in the ECL.

Physics process	Cross section (nb)	Rate (Hz)
$\Upsilon(4S) \rightarrow b\bar{b}$	1.1	12
$u\bar{u}(\gamma)$	1.6	16
$d\bar{d}(\gamma)$	0.4	4
$d\bar{d}(\gamma)$	0.4	4
$d\bar{d}(\gamma)$	1.3	13
$\mu^+\mu^- + \tau^+\tau^-$	1.6	16
Bhabha ($\theta_{\rm lab} \geq 17^\circ$)	44	4.4^a
$\gamma\gamma(\theta_{\rm lab} \geq 17^\circ)$	2.4	0.24^a
2γ processes ($\theta_{\rm lab} \geq 17^\circ$)	~ 15	$\sim 35^b$
Total	~ 67	~ 96

Table 2.1: Total cross section and trigger rates for a luminosity $L = 10^{34}$ cm^{-2}s^{-1} are given for several Υ(4S) physics processes [48]. We use the superscript a to indicate when values are pre-scaled by a factor 1/100 and superscript b to indicate the applied threshold of $p_t \geq 0.3$ GeV/c .

The Level-3 trigger reduces the event rate by $50 \sim 60\%$. and all events that pass this trigger are written to storage. The efficiency for hadronic events after the Level-3 trigger is more than 99%, while the total data rate is reduced by a factor two. Finally a further selection of the events is performed offline by a Level-4 trigger that processes the raw data, applying light selection criteria and providing basic reconstruction algorithms, like the calculation of track properties, as the first step for data analysis. The data size is further reduced for specific physics processes, and the skimmed data are stored as Mini-DST (MDST) files.

2.3 Particle Identification Algorithms

The particle identification at Belle exploits likelihood ratios [49]. Given a candidate particle α, likelihoods for hadron identification are calculated based on dE/dx information from the CDC ($L_\alpha^{\rm CDC}$), time of flight from the TOF ($L_\alpha^{\rm TOF}$) and the number of photons from the ACC ($L_\alpha^{\rm ACC}$), respectively. Then, the likelihood ratios

$$L(\alpha : \beta) = \frac{L_\alpha^{\rm CDC} L_\alpha^{\rm TOF} L_\alpha^{\rm ACC}}{L_\alpha^{\rm CDC} L_\alpha^{\rm TOF} L_\alpha^{\rm ACC} + L_\beta^{\rm CDC} L_\beta^{\rm TOF} L_\beta^{\rm ACC}} \tag{2.5}$$

are calculated and used for identification.

Practically, this means that pions (kaons) can be selected by requiring a low (high) value of $L(K : \pi)$. Normally, the threshold used for the likelihood ratios is optimized

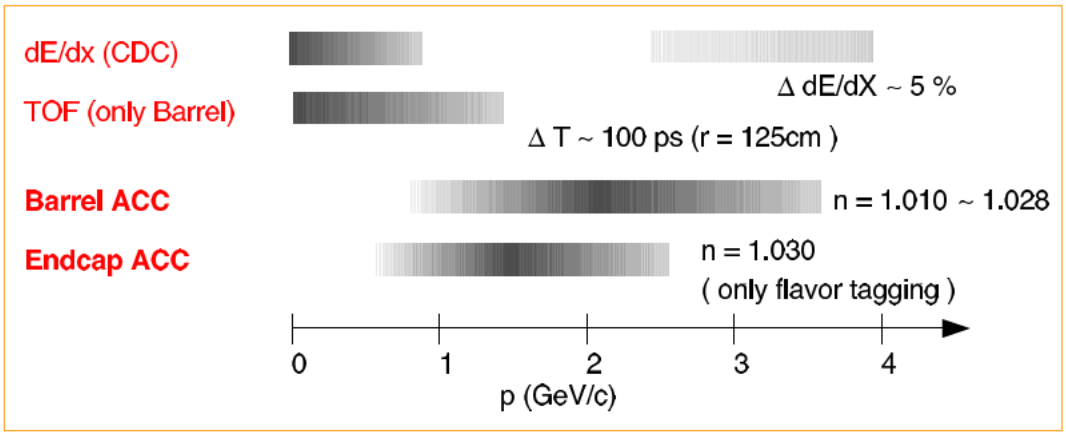


Figure 2.12: The momentum coverage of each of the subdetectors used in charged hadrons PID.

depending on the analysis requirements, as it is the case for this thesis. Figure 2.12 shows the momentum coverage of the different subdetectors used in hadrons PID.

For electron identification, $L_\alpha^{\rm CDC}$ and $L_\alpha^{\rm ACC}$ are used together with information from the ECL (E/p, and transverse shower shape) in order to provide likelihood ratios. The momentum dependence of averaged dE/dx and E/p for electrons and pions, and the shower shape $E9/E25$ distributions for electrons and pions are shown in Figure 2.13.

For muon identification, reconstructed hits in the KLM are matched to the extrapolation of the track built using the CDC information. The difference ΔR between the measured and expected range of the track, together with the χ^2_r constructed from the transverse deviations of all hits associated to the track are used for this comparison. ΔR and χ^2_r are the two variables used to build PDFs for the likelihoods for the muon, pion, and kaon hypotheses.

Finally, the likelihood ratio $L_\mu/(L_\mu + L_\pi + L_K)$ is used as a discriminating variable. The muon identification efficiency and fake rate as a function of momentum are shown in Figure 2.14. The effect of momentum threshold, ~ 600 MeV/c^2 , necessary for muons to reach the KLM is clearly visible in the plot. In particular, muons are also identified with 90% efficiency (2% fake rate) for charged tracks with momenta larger than 0.8 GeV/c^2 .

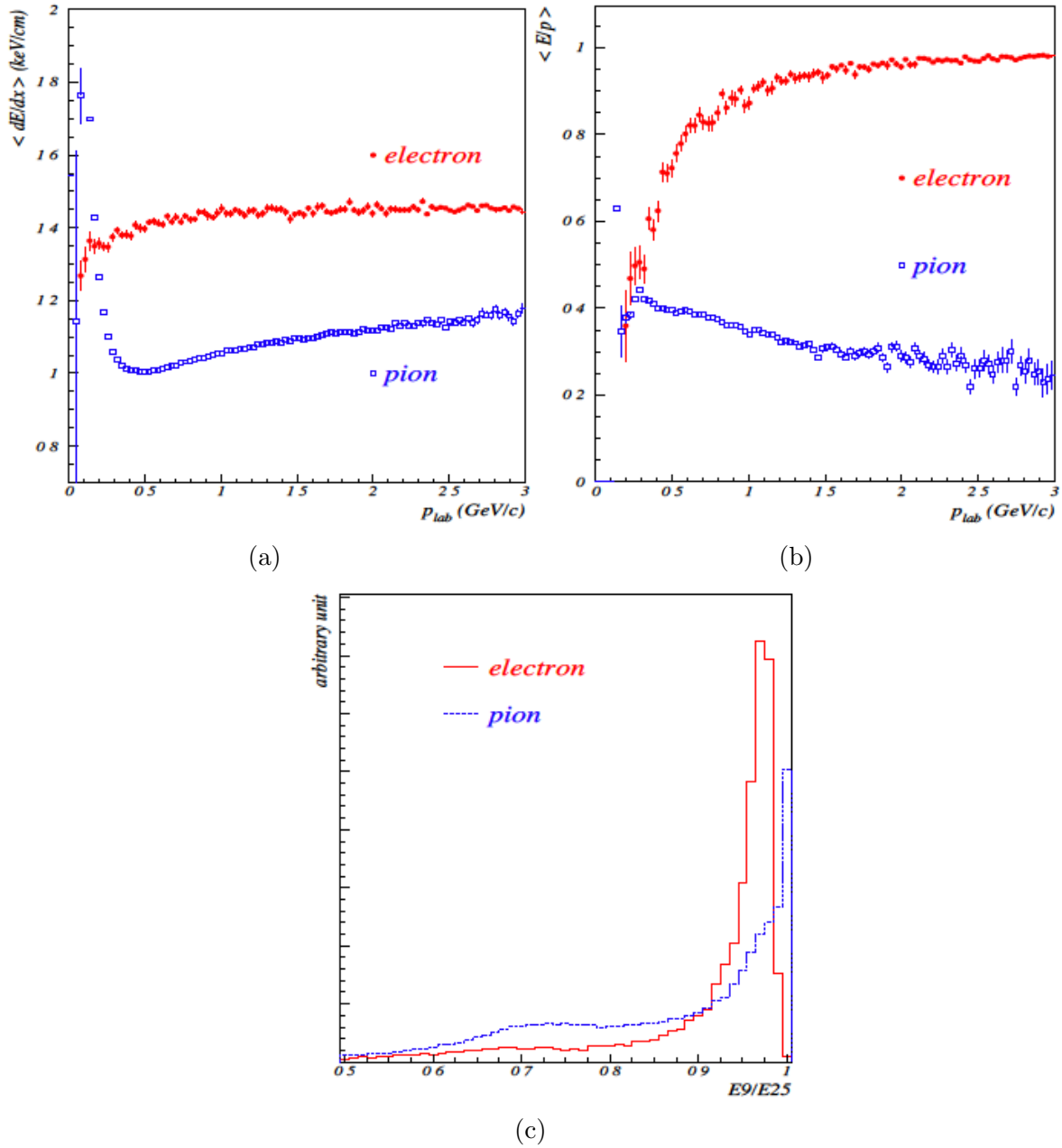


Figure 2.13: Mean dE/dx as a function of momentum (a), and E/p for electrons and pions (b), and shower shape $E9/E25$ for electrons and pions (c). The shower shape is defined as the ratio of energy deposit in 3×3 counters ($E9$) and 5×5 counters ($E25$) surrounding a peak energy counter. This variable is used to parametrize the lateral distribution of the ECL shower [51].

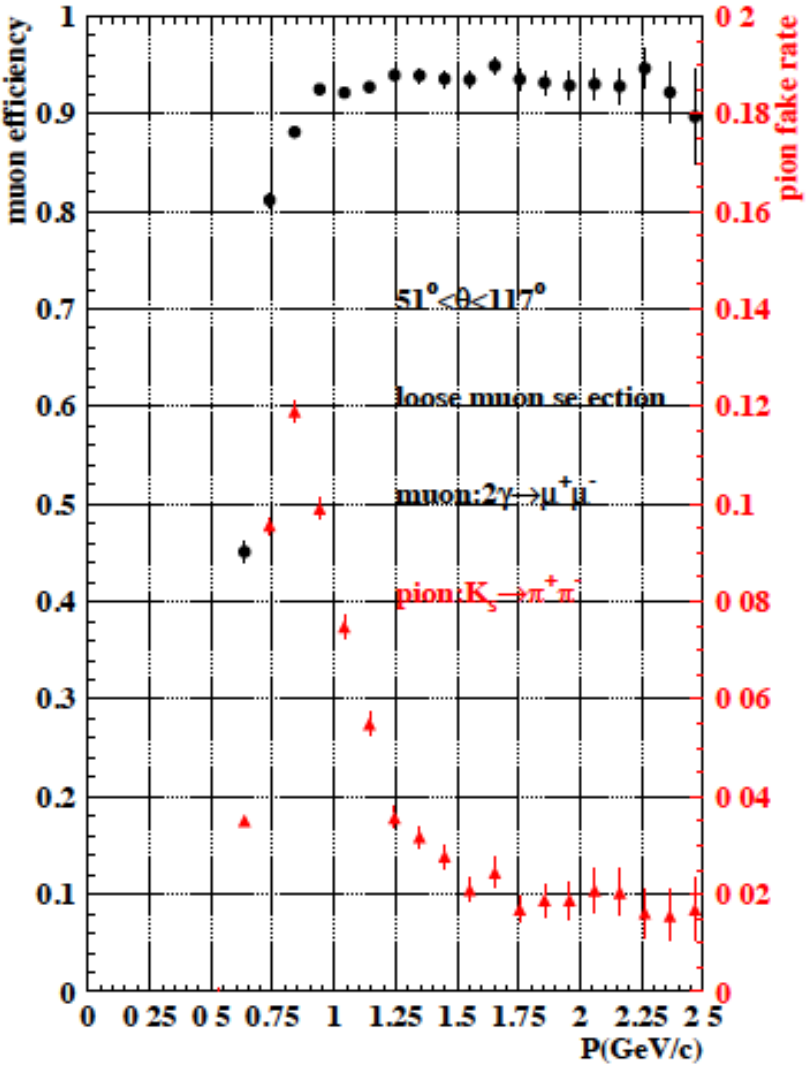


Figure 2.14: Muon identification efficiency and fake rate as a function of momentum [52].

Chapter 3

Event Reconstruction

3.1 Data Samples

3.1.1 Data

We analyze the full $\Upsilon(4S)$ on-resonance Belle data sample with an integrated luminosity of 711 fb^{-1}, containing $772 \times 10^6 B\bar{B}$ pairs. We estimate the non-$B\bar{B}$ background contribution by analyzing the off-resonance data set with an integrated luminosity of 79 fb^{-1}, recorded at $\sqrt{s} = 10.52\,\mathrm{GeV}/c^2$, i.e. 60 MeV/c^2 below the on-resonance energy. In order to reduce the computational time required by the analysis of the data sample, events that are not relevant for the analysis are discarded by the `HadronBJ` skim, which is one of the skims mainly used in Belle for analyses of B and charm mesons.

The `HadronBJ` skim requirements act mainly on track multiplicity and visible energy [49]. The event must have at least 3 charged particle tracks with a transverse momentum greater than 0.1 GeV/c that originate from the region around the interaction point, a requirement defined by the criteria $|\Delta r| < 2$ cm and $|\Delta z| < 4$ cm. The track parameters Δz (Δr) describe the distance of the track's origin in a direction along (perpendicular to) the beam direction of the interaction point. In addition, the sum of the energy of charged tracks and reconstructed photons (E^*_{vis}) must be greater than 20% of the CM energy, $\sqrt{s}$. We use an asterisk to label all observables measured in the CM frame. These two requirements allow for the complete rejection of beam gas background and two-photon events. The requirements $|\Delta r| \leq 3.5$ cm and $|\Delta z| \leq 1.5$ cm for the primary vertex position of the event greatly help in reducing even more the beam gas background. Radiative Bhabha and higher multiplicity QED events are rejected by requiring that two or more ECL clusters are detected at large angle ($-0.7 < \cos\theta^* < 0.9$), the average ECL cluster energy to be smaller than 1 GeV, and the total ECL cluster energy (E^*_{sum}) to be below 80% of $\sqrt{s}$. Additionally, given that some of the τ-pair, beam gas and two photon events have a low energy sum, E^*_{sum} is required to be greater than 18% of $\sqrt{s}$.

3.1.2 Simulated Data

We use Monte-Carlo (MC) simulation to produce data to tune all aspects of the analysis such as selection criteria and determining the best approach for the signal extraction. More specifically, we use MC samples produced by the Belle collaboration, which corresponds to 10 times the integrated luminosity of Belle data available The software packages `EvtGen` [53] and `Pythia` [54] are used to simulate the physics of the decay processes,the detector simulation is performed by the package `GEANT3` [55], while the package `PHOTOS` [56] is used to model final state radiation and add it to the simulation.

3.2 Tag Side Reconstruction

3.2.1 *B*-meson Reconstruction

B-meson decays can be categorized in two main groups: hadronic and semileptonic decays. In hadronic decays all final state particles can be reconstructed by the detector, allowing for the full reconstruction of the B-meson and its properties. This gives access to the full power of the beam-related discriminating variables, which allow for the reconstruction of a sample of B-mesons with high purity. In a e^+e^- collider such as KEKB, the initial state of the $\Upsilon(4S)$ is very well known and as a consequence its two-body decay $\Upsilon(4S) \to B\bar{B}$ is kinematically constrained: in the center of mass (CM) frame each B-meson carries the beam energy $E^*_{beam} \equiv \sqrt{s}/2$ where $\sqrt{s}$ is the total energy of the e^+e^- system in the CM frame.

In order to identify B-meson decay candidates, two variables are defined, the beam-energy-constrained mass, $M_{\rm bc}$, and the energy difference, ΔE. The former is defined as

$$M_{\rm bc} \equiv M_B = \sqrt{E^{*2}_{\rm beam} - \mathbf{p}^{*2}_B}, \tag{3.1}$$

where $\mathbf{p}^{*2}_B$ is the momentum of the B-meson in the CM frame, calculated using the momenta of their decay products, and we have substituted B-meson energy with $E^*_{\rm beam}$. For correctly reconstructed B-mesons M_{bc} should peak at the nominal B-meson mass, i.e. $5.279\,\mathrm{GeV}/c^2$. We use an asterisk to denote all observables measured in the CM frame. The difference between the measured energy of the reconstructed B-meson and the beam energy is

$$\Delta E \equiv E^*_B - E^*_{\rm beam}, \tag{3.2}$$

where E^*_B is the reconstructed energy of the B-meson and this variable's distribution should instead peak at zero for correctly reconstructed candidates.

Generally, with hadronic tags the reconstructed sample of B-mesons is very pure, as previously explained. However, hadronic tagging suffers from low tagging efficiency since hadronic decays can have a high track multiplicity and a single hadronic decay has a branching ratio of $\mathcal{O}(10^{-3})$).

Semileptonic decays, on the other hand, are characterized by high branching ratios, since a typical semileptonic B-meson decay has a branching fraction of $\mathcal{O}(10^{-2})$, and a

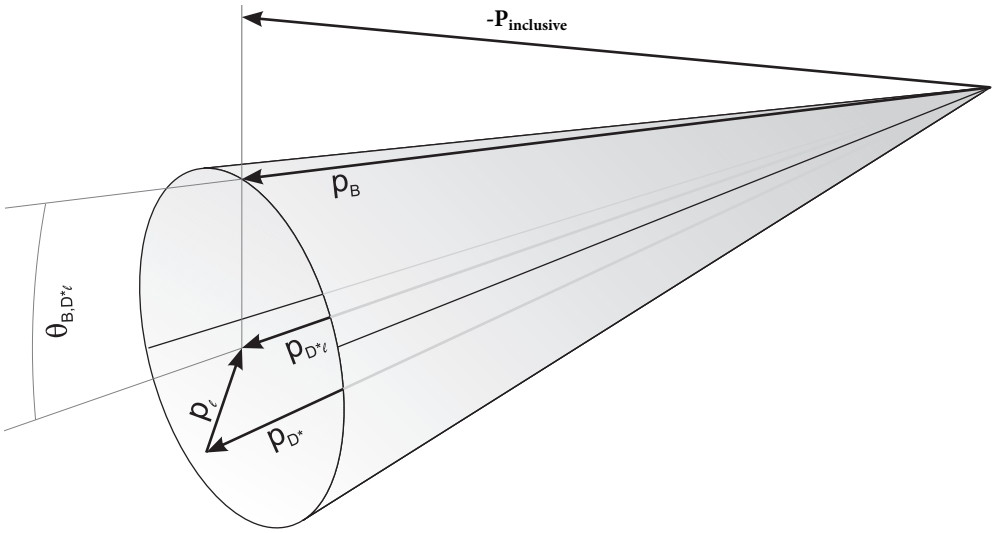


Figure 3.1: Schematic representation of the quantities involved in the definition of the variable $\cos\theta_{B,D^{(*)}\ell}$.

relatively simple decay chain, which leads to high efficiency. However, the downside is that at least one neutrino is included in the decay chain. Neutrinos can not be detected by the Belle detector and imply missing four-momentum in the decay, which forbids the use of strong background discriminators, eventually leading to low purity.

Nevertheless, there are kinematic constraints that can be used to discriminate against background in the case of semileptonic decays. Under the assumption that the neutrino is the only missing particle, the cosine of the angle between the inferred direction of the reconstructed B-meson and that of the $D^{(*)}\ell$ system is

$$\cos\theta_{B,D^{(*)}\ell} = \frac{2E_B^* E^*_{D^{(*)}\ell} - m_B^2 - m^2_{D^{(*)}\ell}}{2|\mathbf{p}_B^*||\mathbf{p}^*_{D^{(*)}\ell}|}, \tag{3.3}$$

where $|\mathbf{p}_B^*| = \sqrt{E_B^* - m_B^2}$. The quantities $E^*_{D^{(*)}\ell}$, $\mathbf{p}^*_{D^{(*)}\ell}$ are calculated from the reconstructed $D^{(*)}\ell$ system. This is a powerful discriminator since events where only one neutrino is missing will lie in the region $(-1, 1)$, while the background does not have this restriction and populates a much wider region. The quantities involved in the definition of this variable are shown in Figure 3.1, and the distribution of normalization, signal and background events is shown in Figure 3.2. Another variable that can be used in the case of semileptonic decay is the squared missing mass

$$m^2_{\text{miss}} = p_\nu^2 = (p_B - p_{D^{(*)}} - p_\ell)^2 \tag{3.4}$$

which in the case of a single particle that has not been reconstructed, corresponds to the squared mass of this missing particle. To solve the above equation it is necessary to measure the B-meson four-momentum, p_B. However, since we reconstruct the B_{tag} in

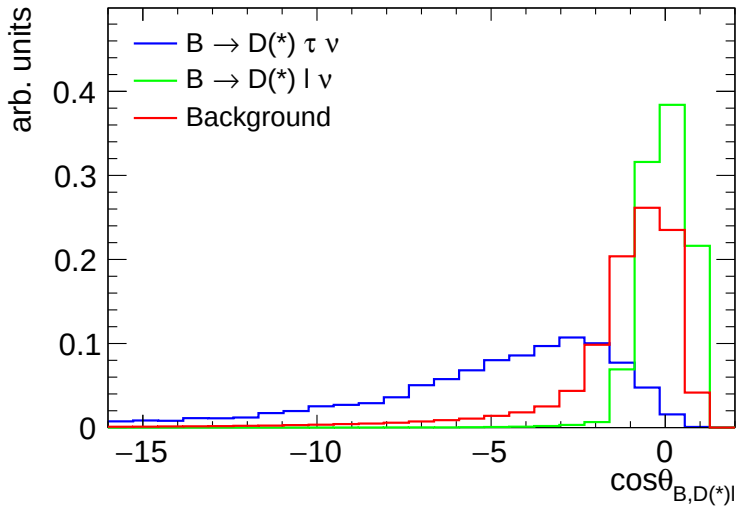


Figure 3.2: $\cos\theta_{B,D^{(*)}\ell}$ distributions for signal, normalization and background events, for B decays to both D and D^* mesons. For the normalization mode only one neutrino is missing and the distribution lies in the region $(-1, 1)$, with smearing effects due to the reconstruction resolution.

a semileptonic decay too, the direction of the motion of the B_{sig} is not known. On the other hand, we know that in $\Upsilon(4S)$ decays the B-meson momentum is small (typically around $0.34\,\text{GeV}/c^2$) compared to the average values of lepton and charm meson momenta. As a consequence, the B-meson momentum can be considered negligible and the squared missing mass can written as

$$m^2_{\text{miss}} = \left(\frac{\sqrt{s}}{2} - E_{D^{(*)}} - E_\ell\right)^2 - (\boldsymbol{p}_{D^{(*)}} - \boldsymbol{p}_\ell)^2. \tag{3.5}$$

In the case of events with only one missing neutrino this variable peaks at zero, while signal events that include a $\tau \to \ell\nu\nu$ decay have a broad distribution that peaks above zero, and can used for background suppression. The distribution of normalization, signal and background events is shown in Figure 3.3. Lastly, a third type of tagging method, which yields a higher signal efficiency, is the inclusive B_{tag} reconstruction. In this method, the signal side is reconstructed first and then the tag side is reconstructed using all remaining particles, and the event is kept when certain selection criteria are satisfied, without however verifying the consistency of the B_{tag} with any specific B-meson decays. The B-meson candidate reconstructed through this procedure is then checked for consistency with a B-meson decay using variables like M_{bc} and ΔE, together with a check for the consistency of the full event.

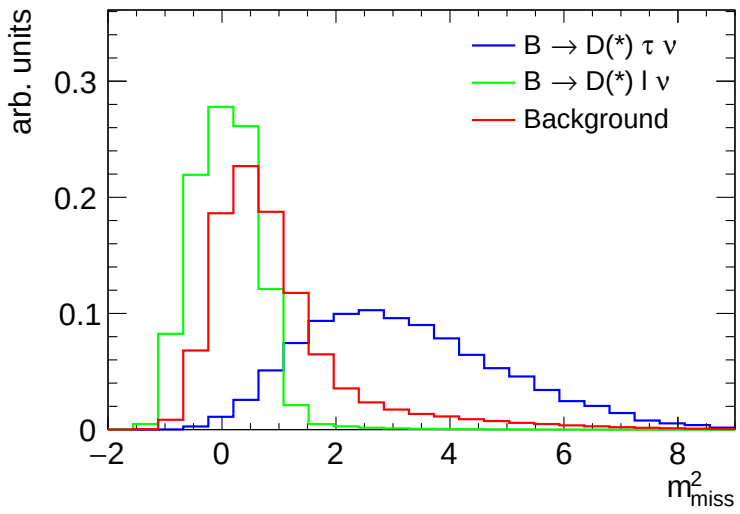


Figure 3.3: m^2_{miss} distributions for signal, normalization and background events, for B decays to both D and D^* mesons. For the normalization mode only one neutrino is missing and the distribution peaks at 0.

3.2.2 *B*-tag Reconstruction

The B_{tag} that accompanies the B_{sig} is reconstructed through a semileptonic tagging algorithm [57], which has been developed in the software framework of the Belle II experiment. Using a BDT classifier, the algorithm is trained on simulated data to recognize the properties of correctly reconstructed particles, and reconstructs B-meson candidates in a hierarchical fashion.

This means that each composite particle is assigned a certain likelihood of being a particle (Signal Probability) using both a set of discriminating variables specific to this variable and the particle's daughters classification that has performed at a previous stage of the algorithm. This procedure is iterated over all daughters in the B-meson tree, until the B_{tag} is reconstructed and is assigned a *signal probability*. The variables used for the classification of the B_{tag} are listed in Table 3.1. The algorithm's approach is depicted in Figure 3.4. This work is the very first Belle $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ analysis that a semileptonic tagging algorithm for B_{tag} reconstruction.

3.2.3 *B*-tag Selection

The main peculiarity of this analysis is that we only select B_{tag} candidates reconstructed through semileptonic decays. The B-meson and charm meson decay channels on the tag side are summarized respectively in tables 3.2 and 3.4, respectively. While the D^*-meson decay channels are the same as on the signal side, and for the the B-meson reconstruction we use the same channels as for the normalization decay of the signal side, the D-meson decay channels for the B_{tag} differ from the ones used for the B_{sig}

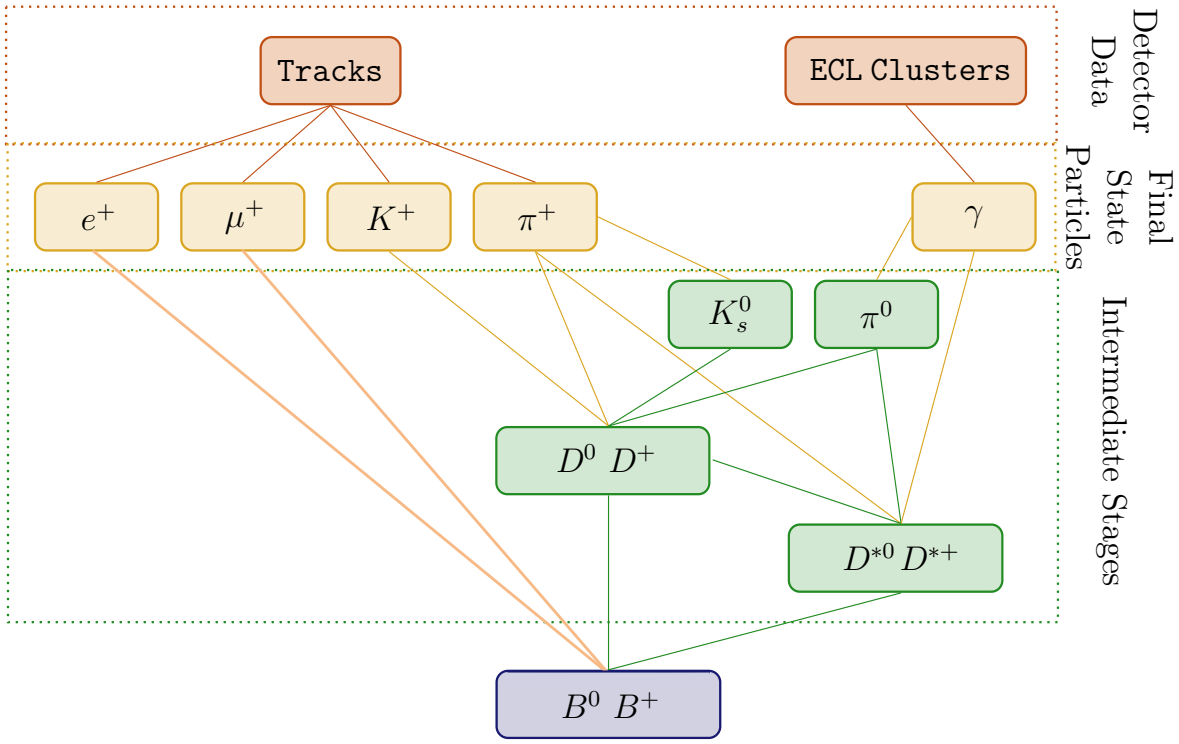


Figure 3.4: Hierarchical approach of the B_{tag} reconstruction algorithm, which in this figure flows from top to bottom. Adapted from [57].

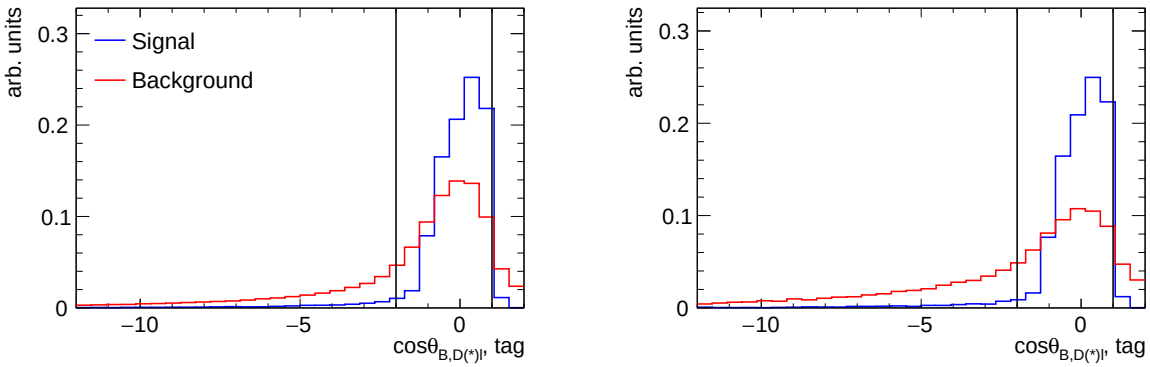
reconstruction. We veto $B \to D^{(*)}\tau(\to \ell\nu\nu)\nu$ events on the tag side by selecting $-2 < \cos\theta_{B,D^{(*)}\ell} < 1.0$. The distribution of $\cos\theta_{B,D^{(*)}\ell}$ on the tag side is shown in Figure 3.5. We select well reconstructed B_{tag} candidates by requiring their tagging classifier output to be larger than $10^{-1.5}$. The latter selection criterion allows for an excellent rejection of mis-reconstructed B_{tag} candidates with mis-assigned charge, which would introduce a large systematic uncertainty to thge final measurement. The distribution of tagging classifier output is shown in Figure 3.6. By preferring B_{tag} candidates with good quality we limit the chance that particles missed by the B_{tag} reconstruction will interfere with the B_{sig} reconstruction, therefore increasing the purity ofB_{sig} .

Table 3.1: Variables used for the training of the semileptonic tagging algorithm. Throughout the table $\mathbf{p}_B$ refer to the B_{tag} three-momentum and $\mathbf{p}_d$ to a generic daughter's three-momentum. Whenever a variable is not dimensionless, its dimension is given within parentheses.

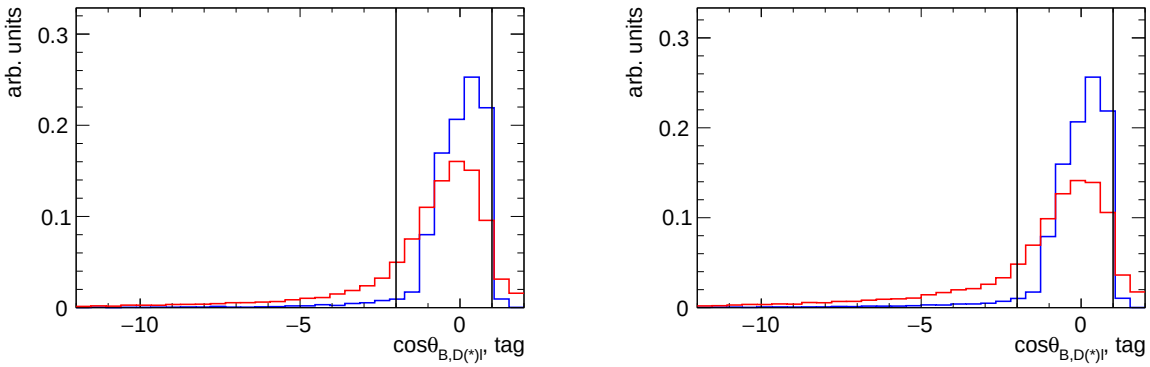
B-meson variables	χ^2 of vertex fit
	3D track distance relative to IP (cm)
	Significance of 3D track distance
	Transverse distance relative to IP (cm)
	x-axis track distance relative to IP (cm)
	y-axis track distance relative to IP (cm)
	z-axis track distance relative to IP (cm)
	ΔE (GeV)
B-meson daughter variables	χ^2 of vertex fit
	3D track distance relative to IP (cm)
	Signal probability
	$\mathbf{p}_d$ in CM frame (GeV/c)
	Cosine of the angle between $\mathbf{p}_B$ and $\mathbf{p}_d$ in the B rest frame
	Cosine of the angle between $\mathbf{p}_d$ and the vector connecting the IP and fitted daughter vertex
	Cosine of the angle between $\mathbf{p}_d$'s of each pair of daughters
	Decay mode ID

3.3 Signal Side Reconstruction

On the signal side we aim to reconstruct the signal mode $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$, providing a relative measurement of its branching ratio with respect to the normalization mode

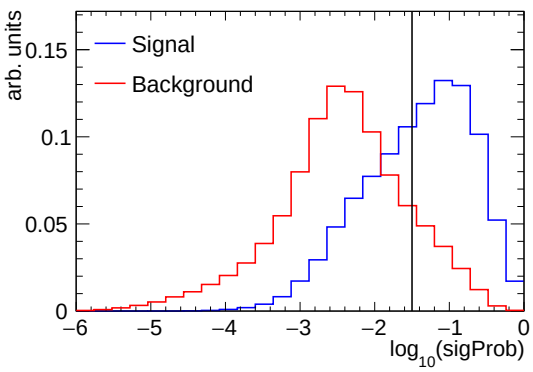


(a) $B^- \to D^0 \ell^- \nu$ (b) $\bar{B}^0 \to D^+ \ell^- \nu$

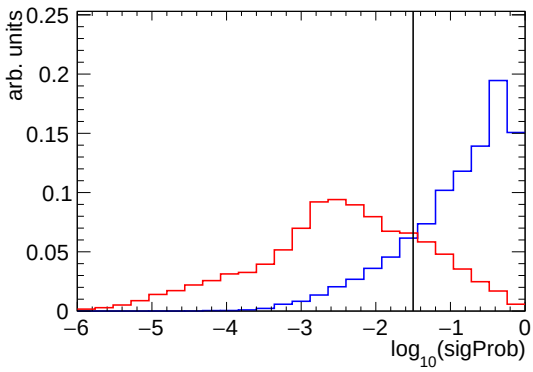


(c) $B^- \to D^{*0} \ell^- \nu$ (d) $\bar{B}^0 \to D^{*+} \ell^- \nu$

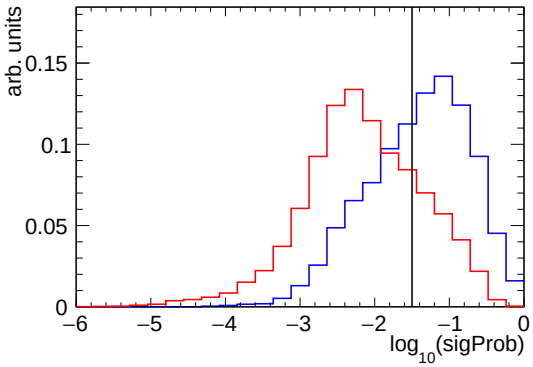
Figure 3.5: $\cos\theta_{B,D^{(*)}\ell}$ distributions on the tag side. Correctly reconstructed B_{tag} candidates are shown in blue, mis-reconstructed candidates in red. The upper and lower threshold values used for the candidates selection are shown with vertical black lines.



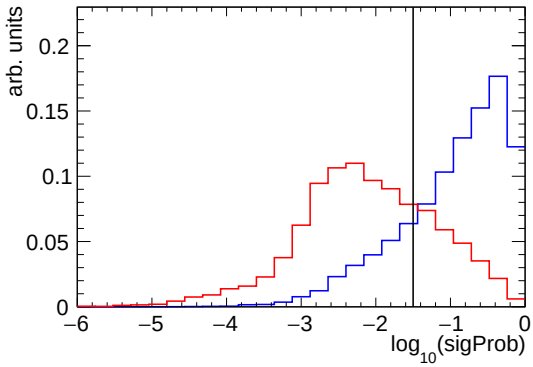
(a) $B^- \to D^0 \ell^- \nu$



(b) $\bar{B}^0 \to D^+ \ell^- \nu$



(c) $B^- \to D^{*0} \ell^- \nu$



(d) $\bar{B}^0 \to D^{*+} \ell^- \nu$

Figure 3.6: Tagging classifier output distributions for the B_{tag} . Correctly reconstructed B_{tag} candidates are shown in blue, mis-reconstructed candidates in red. The upper and lower threshold values used for the candidates selection are shown with vertical black lines.

Table 3.2: The B_{tag} decay channels are listed together with their relative branching ratios [40]. The first error is statistical and the second one is systematic.

B^0 channel	Branching ratio (%)	B^+ channel	Branching ratio (%)
$D^- \ell^+$	$2.20 \pm 0.04 \pm 0.09$	$\bar{D}^0 \ell^+$	$2.33 \pm 0.04 \pm 0.09$
$D^{*-} \ell^+$	$4.88 \pm 0.01 \pm 0.10$	$\bar{D}^{*0} \ell^+$	$5.59 \pm 0.02 \pm 0.19$

Table 3.3: The D^*_{tag} decay channels are listed together with their relative branching ratios taken from Ref. [10].

D^{*0} channel	Branching ratio (%)	D^{*+} channel	Branching ratio (%)
$D^0 \pi^0$	64.7 ± 0.9	$D^0 \pi^+$	67.7 ± 0.5
$D^0 \gamma$	64.7 ± 0.9	$D^+ \pi^0$	30.7 ± 0.5

$\bar{B} \to D^{(*)} \ell^- \bar{\nu}_\ell$, which is therefore also reconstructed in this analysis. We choose to reconstruct only τ leptonic decays, $\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau$, which account for approximately 35.2 % of the total τ branching ratio, so that the final state particles reconstructed for the signal and normalization sides are the same, given that neutrinos escape detection. Am event display for a reconstructed signal event is shown in Figure 3.7.

Only long lived particles such as $e^\pm, \mu^\pm, \pi^\pm, K^\pm, \gamma$, n, p, $\bar{p}$ and K^0_L deposit energy in the detector, which means that the reconstruction of B-meson and charm mesons, which decay inside the beam pipe, has to be done indirectly by grouping together final state particles and summing their momenta, assessing their consistency with the hypothesis of the mesons that is been reconstructed, using discriminating variables. In the next section we describe the selection applied to final state particles, charm mesons and B-mesons.

3.3.1 Final State Particles Selection

Charged final state particles used for the reconstruction of charm mesons and B-mesons, i.e. $e^\pm, \mu^\pm, \pi^\pm, K^\pm$ are selected using their particle identification (particle ID) probabilities, which are provided by a combination of subdetector measurements (as explained in Section 2.3), and the values for the track parameters Δz and Δr. For electrons, we also constrain the particles' momentum in the lab frame to be larger than 200 MeV/c^2 , since the electron candidate sample contains a large number of π and K at small momenta mimicking electrons, given that electrons in this momentum range can not reach the subdetectors (particularly the EC) responsible for their identification and as a consequence can not be well identified by the particle ID variables.

Neutral pions are reconstructed in the channel $\pi^0 \to \gamma\gamma$ and the π^0 energy resolution is improved by performing a mass-constrained vertex fit of the two photon candidates

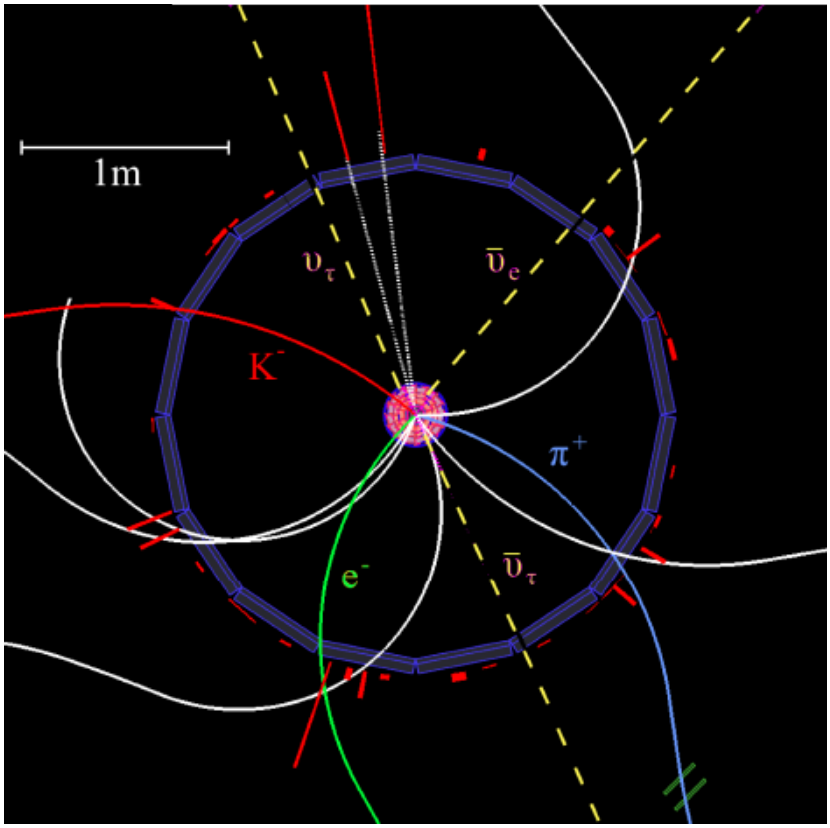


Figure 3.7: Event display illustrating the reconstruction of a signal decay: trajectories of charged particles are shown as colored solid lines, energy deposits in the calorimeters are depicted by red bars. The display is an end view perpendicular to the beam axis with the SVD in the center (small orange circle) and the TOF (dark purple polygon). This is an event with $B^- \to D^0\tau^-\bar{\nu}_\tau$, $D^0 \to K^-\pi^+$ and $\tau^- \to e^-\bar{\nu}_e\nu_\tau$, and the B^+ decaying to five charged particles (white solid lines) and two photons. The trajectories of undetected neutrinos are marked as dashed yellow lines [58].

Table 3.4: The D_{tag} decay channels are listed together with their relative branching ratios taken from Ref. [10].

D^0 channel	Branching ratio (%)
$K^-\,\pi^+\,\pi^0$	13.9 ± 0.5
$K^-\,\pi^+\,\pi^+\,\pi^-$	8.07 ± 0.23
$K^0_S\,\pi^+\,\pi^-\,\pi^0$	5.2 ± 0.6
$K^-\,\pi^+\,\pi^+\,\pi^-\,\pi^0$	4.2 ± 0.4
$K^-\,\pi^+$	3.88 ± 0.05
$K^0_S\,\pi^+\,\pi^-$	2.83 ± 0.20
$\pi^-\,\pi^+\,\pi^0$	1.47 ± 0.09
$K^0_S\,\pi^0$	1.19 ± 0.04
$\pi^-\,\pi^+\,\pi^0\,\pi^0$	1.01 ± 0.09
$\pi^-\,\pi^+\,\pi^+\,\pi^-$	0.745 ± 0.022
$K^-\,K^+\,K^0_S$	0.451 ± 0.034
$K^-\,K^+$	0.396 ± 0.008
$K^-\,K^+\,\pi^0$	0.338 ± 0.021

D^+ channel	Branching ratio (%)
$K^-\,\pi^+\,\pi^+$	9.46 ± 0.24
$K^0_S\,\pi^+\,\pi^0$	7.24 ± 0.17
$K^-\,\pi^+\,\pi^+\,\pi^0$	6.14 ± 0.16
$K^0_S\,\pi^+\,\pi^+\,\pi^-$	3.05 ± 0.09
$K^0_S\,\pi^+$	1.53 ± 0.06
$K^-\,K^+\,\pi^+$	0.996 ± 0.026
$K^+\,K^0_S\,K^0_S$	0.46 ± 0.21
$\pi^+\,\pi^0$	0.124 ± 0.006
$\pi^+\,\pi^+\,\pi^-\,\pi^0$	0.117 ± 0.008

to the nominal π^0 mass. Usually, the χ^2 of a vertex fit is used to identify the correct reconstruction of composite particles. In this specific case, photons are reconstructed as calorimeter clusters and do not provide useful position information for the vertex determination. However, when using mass constraints, the photons can affect the χ^2 minimization, since they contribute to the momentum of the composite particle.

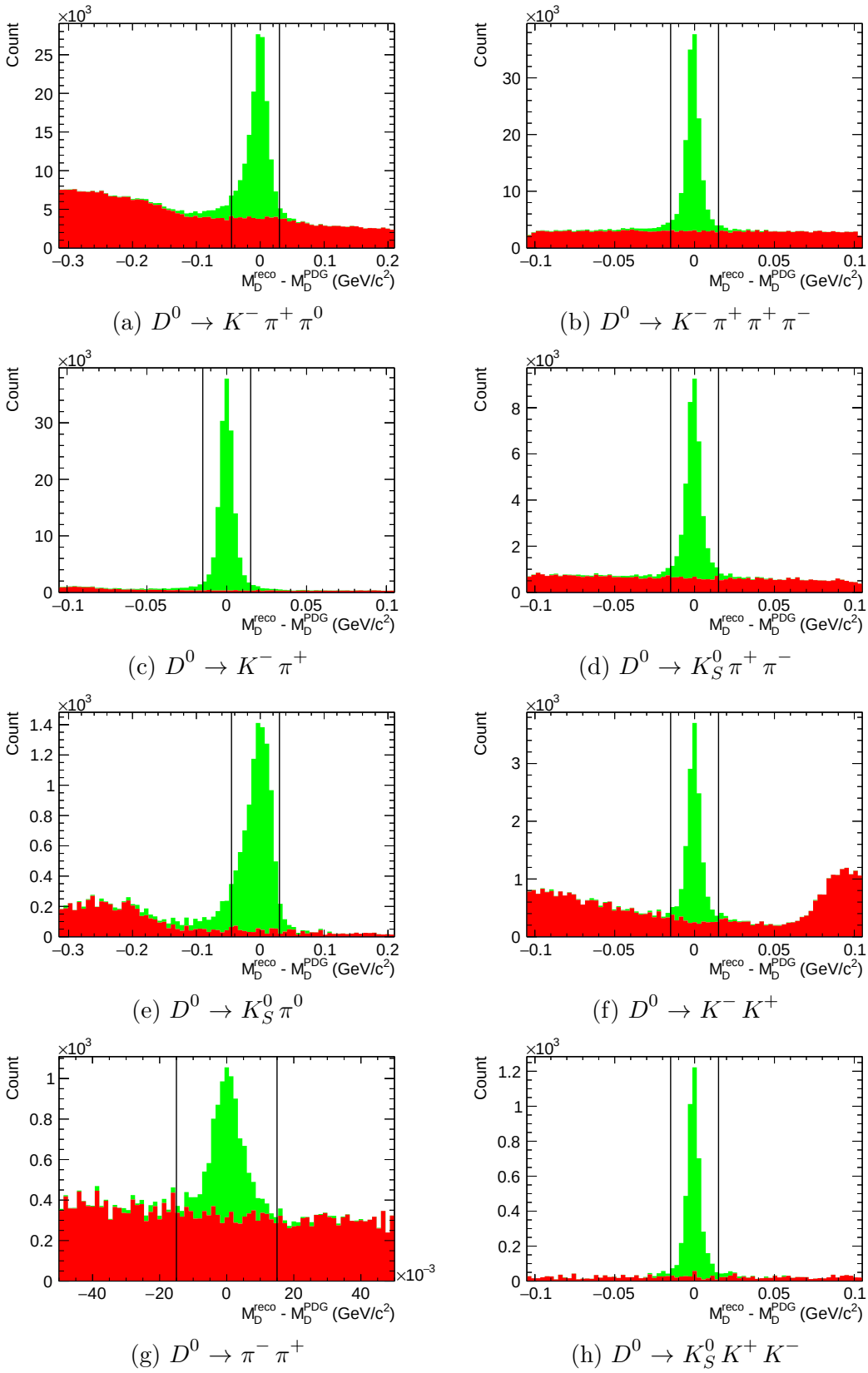
For the π^0 from D-meson decays, we require the photon daughter energies to be greater than 50 MeV. The minimum energies for ECL clusters used for photon reconstruction are 50, 100, and 150 MeV from the barrel, forward, and backward ECL regions respectively, in order to veto photons produced by the beam background. Low energy π^0 candidates from D^* are reconstructed using looser energy requirements: one photon must have an energy of at least 50 MeV, while the companion photon must have a minimum energy of 20 MeV. Since the combinatorics for pairs of gammas is very high, we remove redundant π^0 candidates that share photon daughters with other π^0 candidates of higher energy.

K^0_S-mesons are reconstructed in the channel $K^0_S \to \pi^+\,\pi^-$ and their selection is performed using an algorithm based on a NeuroBayes neural network [59], which exploits its output variables with the standard selection criteria $nb_{\text{vlike}} > 0.5$ (likelihood of candidate being a particle that decays into a pair of daughters) and $nb_{\text{nolam}} > -0.4$ (likelihood of candidate not being a Lambda particle). All the final state particles selection criteria are summarized in table 3.5.

3.3.2 *D*-meson Selection

The channels used for the reconstruction of the D^+ and D^0 mesons are summarized in Table 3.6. The combined reconstructed branching fractions are $\sim 30\%$ and $\sim 22\%$, for D^0 and D^+ , respectively.

The main source of background for D-mesons is labelled *combinatorial*, because it arises from reconstruction of a D-meson candidate using final state particles that actually originated from different mothers. The same type of background is present whenever a composite particle is reconstructed. All D channels are vertex-fit using an algorithm based on a Kalman filter [60] and the χ^2 of the fit is used to reject combinatorial background. Moreover, a selection criterion centred on 0 on the difference between the reconstructed invariant D mass and the nominal value, is applied to reduce the background contribution. These selection criterion values depend on the decay channel, where a looser selection criterion is applied whenever a π^0 is present in the D-meson decay channel, since the π^0 causes a decrease in the resolution of the dM distribution. The differential mass distributions of all D-meson channels are shown in Figures 3.8 and 3.9 for D^0-mesons and D^+-mesons, respectively. The D-meson selection is summarized in Table 3.7.



(a) $D^0 \to K^- \pi^+ \pi^0$

(b) $D^0 \to K^- \pi^+ \pi^+ \pi^-$

(c) $D^0 \to K^- \pi^+$

(d) $D^0 \to K^0_S \pi^+ \pi^-$

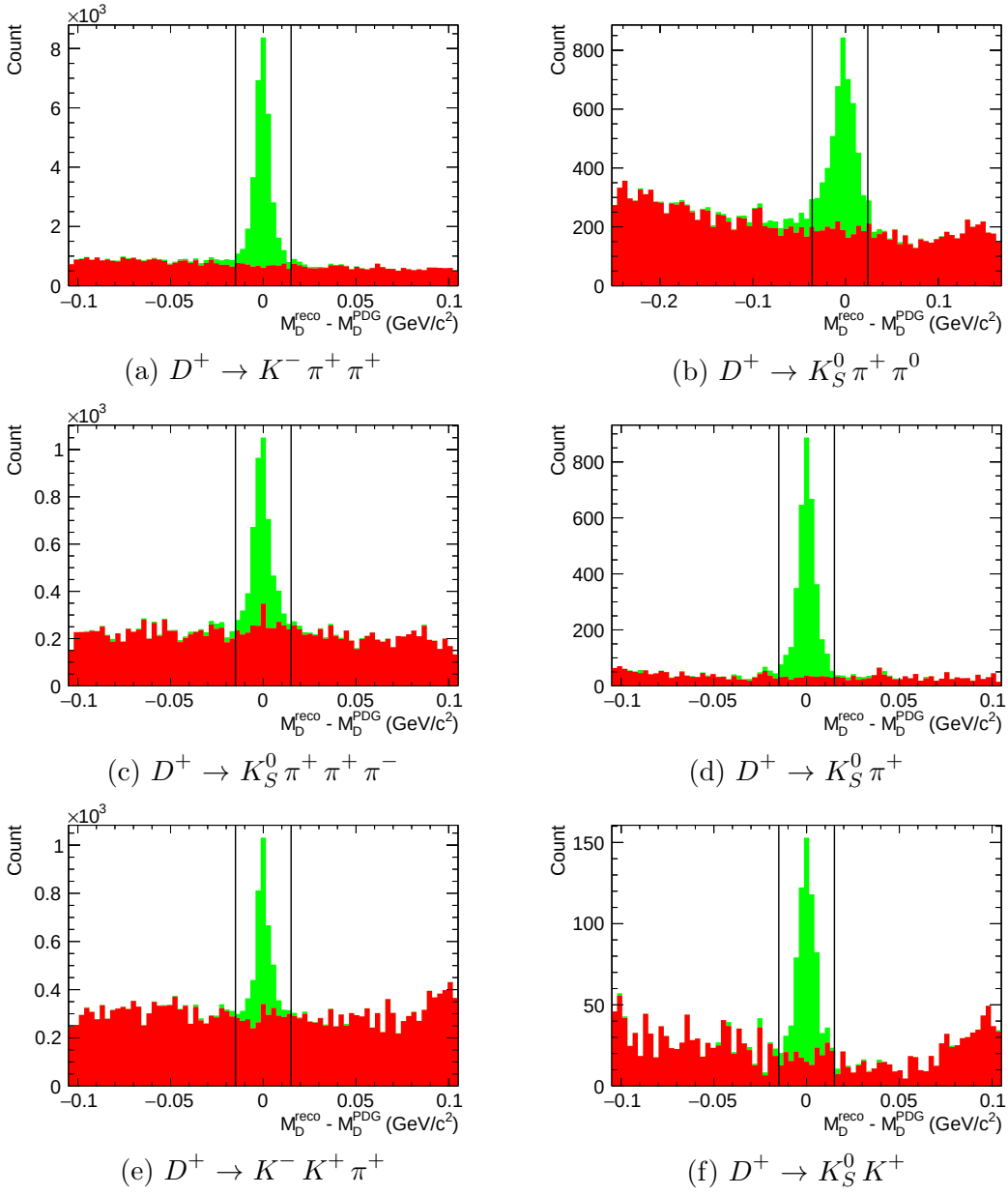
(e) $D^0 \to K^0_S \pi^0$

(f) $D^0 \to K^- K^+$

(g) $D^0 \to \pi^- \pi^+$

(h) $D^0 \to K^0_S K^+ K^-$

Figure 3.8: Difference between the reconstructed D^0-meson invariant mass and its PDG value. Correctly reconstructed candidates are shown in green, and mis-reconstructed candidates in red. Candidate selection thresholds are shown with black lines.

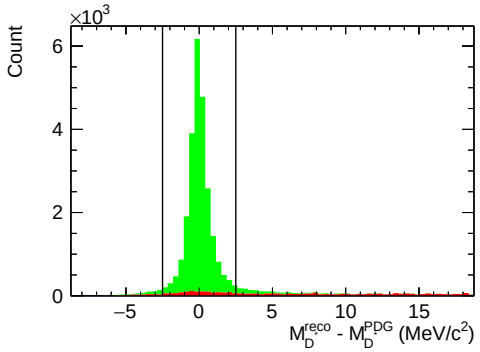


(a) $D^+ \to K^- \pi^+ \pi^+$ (b) $D^+ \to K^0_S \pi^+ \pi^0$

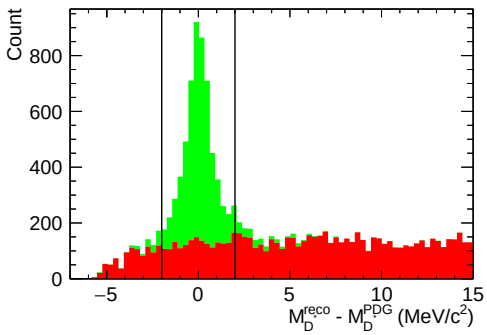
(c) $D^+ \to K^0_S \pi^+ \pi^+ \pi^-$ (d) $D^+ \to K^0_S \pi^+$

(e) $D^+ \to K^- K^+ \pi^+$ (f) $D^+ \to K^0_S K^+$

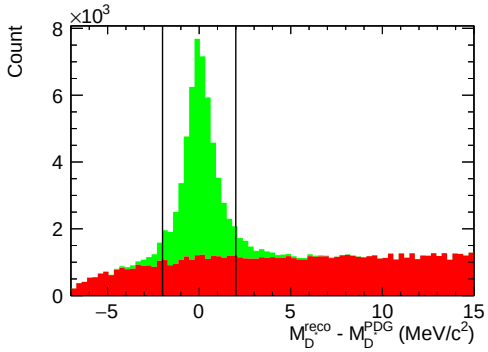
Figure 3.9: Distributions of the difference between the reconstructed D^+-meson invariant mass and its PDG value, for all D^+ channels. In the plots, correctly reconstructed D^+-meson candidates are shown in green, and mis-reconstructed candidates in red. The lower and upper threshold used for the candidate selection are shown with vertical black lines.



(a) $D^{*+} \to D^0 \pi^+$



(b) $D^{*+} \to D^+ \pi^0$



(c) $D^{*0} \to D^0 \pi^0$

Figure 3.10: Distributions of the difference between the reconstructed D^*-meson invariant mass and its PDG value, for all D^* channels. In the plots, correctly reconstructed D^*-meson candidates are shown in green, and mis-reconstructed candidates in red. The lower and upper threshold used for the candidate selection are shown with vertical black lines.

Table 3.5: Selection criteria for final state particles.

Particle	Variable type	Selection criterion
$K^{\pm}$	Particle ID	KID > 0.1
$e^{\pm}$	Particle ID	eID > 0.3
	Kinematics	$p_{\rm lab} > 0.200$ GeV/c
$\mu^{\pm}$	Particle ID	muID > 0.5
all charged	Track parameter	$\Delta r < 2.0$ cm
	Track parameter	$\lvert\Delta z\rvert < 5.0$ cm
π^0	Invariant mass	$0.120 < M(\,\mathrm{GeV}/c^2\,) < 0.150$
	E_γ	$> 50/100/150$ MeV (barrel/fwd/back)
	$cos\theta_{\gamma\gamma}$	> 0
	Kinematics	$p_{\rm lab} > 0.200\,\mathrm{GeV}/c$
π^0_{slow}	Invariant mass	$\lvert M_{\gamma\gamma} - M_{\pi^0}^{\rm PDG}\rvert < 0.010\,\mathrm{GeV}/c^2$
	E_γ, high	> 50 MeV
	E_γ, low	> 20 MeV
K^0_S	Invariant mass	$0.483 < M(\,\mathrm{GeV}/c^2\,) < 0.513$
	K^0_S Optimizer	$nb_{\rm vlike} > 0.5$ and $nb_{\rm nolam} > -0.4$

Table 3.6: D-meson decay channels for the $B_{\rm sig}$, with branching ratios taken from Ref. [10].

D^0 channel	Branching ratio (%)
$K^-\,\pi^+\,\pi^0$	13.9 ± 0.5
$K^-\,\pi^+\,\pi^+\,\pi^-$	8.07 ± 0.23
$K^-\,\pi^+$	3.88 ± 0.05
$K^0_S\,\pi^+\,\pi^-$	2.83 ± 0.20
$K^0_S\,\pi^0$	1.19 ± 0.04
$K^0_S\,K^+\,K^-$	0.451 ± 0.034
$K^-\,K^+$	0.396 ± 0.008
$\pi^-\,\pi^+$	0.1402 ± 0.0026

D^+ channel	Branching ratio (%)
$K^-\,\pi^+\,\pi^+$	9.46 ± 0.24
$K^0_S\,\pi^+\,\pi^0$	7.24 ± 0.17
$K^0_S\,\pi^+\,\pi^+\,\pi^-$	3.05 ± 0.09
$K^0_S\,\pi^+$	1.53 ± 0.06
$K^-\,K^+\,\pi^+$	0.996 ± 0.026
$K^0_S\,K^+$	0.295 ± 0.015

Table 3.7: Selection criteria for D-mesons.

Particle	Decay channel	Variable type	Selection criterion
D^+	with π^0	Invariant Mass	$-36\,\mathrm{MeV}/c^2 < M_D^{\mathrm{reco}} - M_D^{\mathrm{PDG}} < 24\,\mathrm{MeV}/c^2$
D^+	without π^0	Invariant Mass	$-15\,\mathrm{MeV}/c^2 < M_D^{\mathrm{reco}} - M_D^{\mathrm{PDG}} < 15\,\mathrm{MeV}/c^2$
		Vertex fit	pValue > 0 (successful fit)
D^0	with π^0	Invariant Mass	$-45\,\mathrm{MeV}/c^2 < M_D^{\mathrm{reco}} - M_D^{\mathrm{PDG}} < 30\,\mathrm{MeV}/c^2$
D^0	without π^0	Invariant Mass	$-15\,\mathrm{MeV}/c^2 < M_D^{\mathrm{reco}} - M_D^{\mathrm{PDG}} < 15\,\mathrm{MeV}/c^2$
		Vertex fit	pValue > 0 (successful fit)

3.3.3 D^*-meson Selection

We reconstruct D^{*+} meson decays in the channels $D^0\,\pi^+$ and $D^+\,\pi^0$. For the first channel, the D^{*+} vertex is fit using the previously fit D^0 vertex adding a $\pi^+_{\rm slow}$ track to the fit. This improves the purity for this channel. On the other hand, the vertex-fit for the $D^+\,\pi^0$ channel does not perform well, given the lack of directional information provided by a π^0 and therefore, we do not perform a vertex-fit for this channel. Moreover, for the $D^+\,\pi^0$ channel we loosen the photon selection for the slow π^0 with respect to the π^0 used for D-meson reconstruction, as mentioned in section 3.5. However, the asymmetry between the energies of the photons emitted by the slow π^0, defined as:

$$A_\gamma = \frac{|E_1 - E_2|}{(E_1 + E_2)}, \tag{3.6}$$

where E_1, E_2 are the two photon energies, is required to be smaller than 0.6, a value chosen by maximising the figure of merit for this distribution, as shown in Figure 3.11. We reconstruct D^{*0}-meson decay in the channel $D^0\,\pi^0$. The π^0 candidates are selected through the procedure explained in the previous paragraph. We do not apply a vertex fit to the D^{*0}, since no information is gained with this procedure, due to the presence of a π^0.

Given that in D^* decays there is only a small amount of phase space available for the momentum of the emitted π, the signal D^* decay can be effectively selected using a selection criterion on the $\Delta M \equiv M(D\pi) - M(\pi)$, where $M(D^{*+}) - M(D^+) = 140.66 \pm 0.10\,\text{MeV}/c^2$ and $M(D^{*0}) - M(D^0) = 145.42 \pm 0.01\,\text{MeV}/c^2$ [10]. The differential mass distributions of all D^*-meson channels are shown in Figure 3.10. The selection criteria for the different channels for this variable are summarized in Table 3.8.

3.3.4 B-signal Selection

We require $B_{\rm sig}$ candidates to have $\cos\theta_{B,D^{(*)}\ell} < 1.0$, since for both normalization and signal events the distributions of this variable do not extend any further than this limit on the positive axis range. On the other hand, we do not impose any lower bound on this variable since small values of $\cos\theta_{B,D^{(*)}\ell}$ are a signature of the signal events.

Table 3.8: Selection criteria for D^* mesons.

Particle	Decay channel	Selection criterion
D^{*0}	$D^0\pi^0$	$\lvert M^{\rm reco}_{D^*} - M^{\rm PDG}_{D^*}\rvert < 2.0\,\text{MeV}/c^2$
D^{*+}	$D^0\pi^+$	$\lvert M^{\rm reco}_{D^*} - M^{\rm PDG}_{D^*}\rvert < 2.5\,\text{MeV}/c^2$
	$D^+\pi^0$	$\lvert M^{\rm reco}_{D^*} - M^{\rm PDG}_{D^*}\rvert < 2.0\,\text{MeV}/c^2$

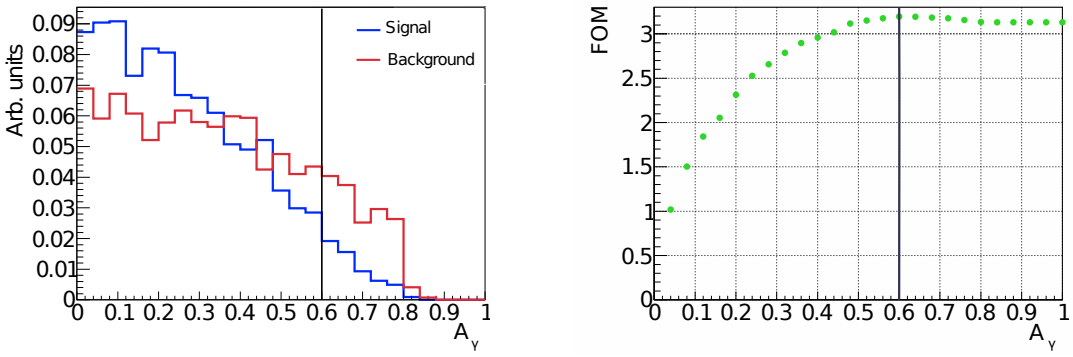


Figure 3.11: On the left, distribution of the gamma asymmetry variable A_γ for correctly reconstructed D^* meson candidates, labeled as signal, and mis-reconstructed candidates, labeled as background. On the right, the figure of merit (FOM) is shown as a function of the threshold for this variable.

3.4 Full Event Selection

We limit the background from *continuum events* generated by $e^+e^- \to q\bar{q}$ events (where the q refers to a c, s, u, d quark) by choosing B_{tag} and B_{sig} candidates that have a value of the variable R_2 (defined as the ratio between the 2nd and 0th Fox-Wolfram moments [61]) smaller than 0.4. This is justified by the fact that on-resonance $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ events have a spherical symmetry, as a consequence of the small momentum of the B-mesons in the CM frame, and tend to have R_2 values close to zero. In contrast, for $q\bar{q}$ events, the quarks are produced with a very large initial momentum, yielding a jet-like topology for the event. The distribution of the R_2 variable is shown in Figure 3.12. After B_{sig} and B_{tag} have been paired up to reconstruct a full $\Upsilon(4S)$ decay, we apply a final and stronger event selection. An event where one (or both of) the B-mesons is misreconstructed can be likely identified by looking at information from the Rest Of the Event (ROE). The presence of charged tracks or neutral ECL clusters that have not been used by the reconstruction of any of the B-mesons, i.e. they belong to the ROE, is very likely the signature for the misreconstruction of the B-mesons. For this reason almost all the full event selection deals with ROE selection criteria.

It is important to adequately remove false extra tracks and extra clusters from the ROE, by which we mean: clusters from beam background (E_{ECL} selection criterion) and tracks from photon conversion that are generated at material surfaces (tracks impact parameters selection criterion). The neutral ECL clusters which are allowed in the ROE have energies:

- $E > 50$ MeV , ECL barrel;
- $E > 150$ MeV, ECL backward end-cap;
- $E > 100$ MeV, ECL forward end-cap.

We only select events that satisfy the following requisites:

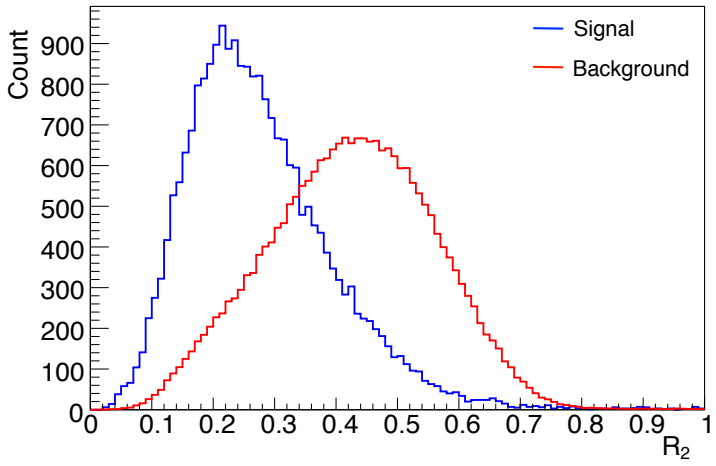


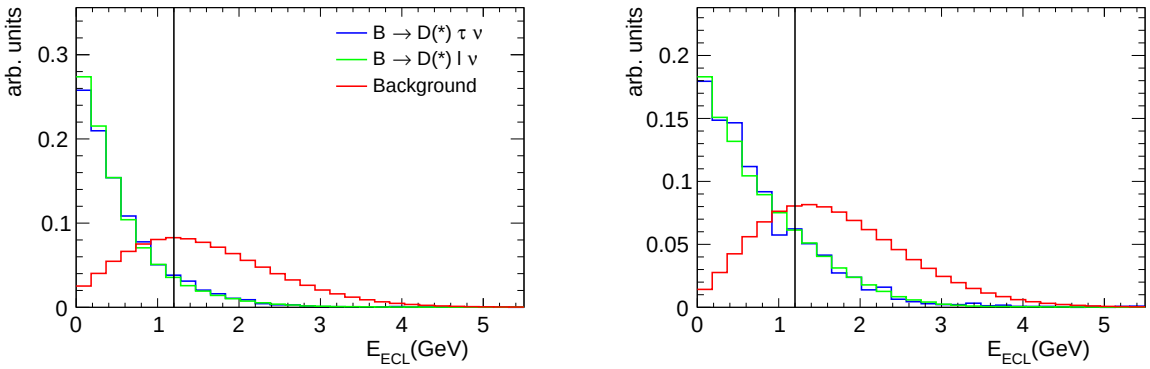
Figure 3.12: Distribution of the R_2 variable used for continuum suppression. In the plot the signal represents the $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ events, while the background represents the $e^+e^- \to q\bar{q}$ events (where the q refers to a c, s, u, d quark).

- $E_{\text{ECL}} < 1.2$ GeV (Figure 3.13);
- no extra tracks in the ROE (Figure 3.14);
- no extra K^0_S in the ROE (Figure 3.15);
- no extra π^0 in the ROE (Figure 3.16);

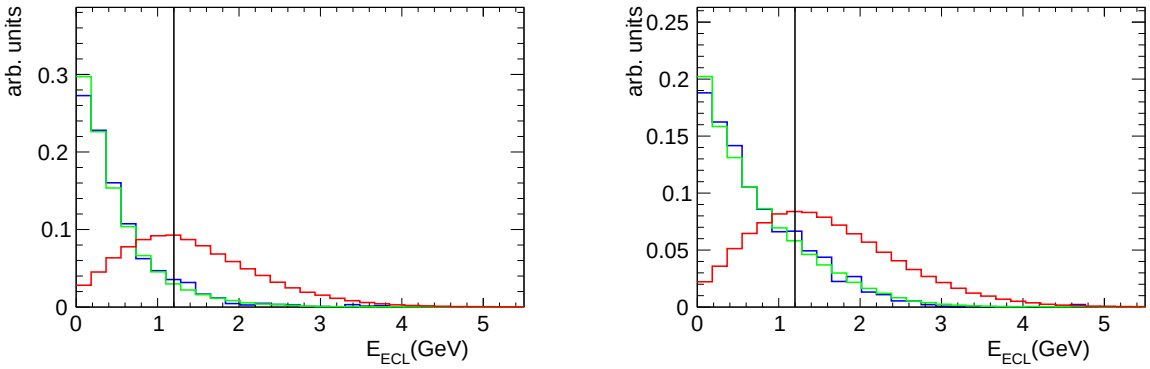
The K^0_S candidates in the ROE can be formed with tracks emerging from outside of the interaction region and the π^0 selections for the ROE is the same as the one used for the π^0 from D-mesons. We further discriminate against background by selecting the three-momentum of the $D^{(*)}$ signal meson to be smaller than $2.0\,\text{GeV}/c^2$ in the CM frame (Figure 3.17). This selection criterion also removes a significant fraction of normalization events, which form a dominant background. Lastly, we veto events where both B_{sig} and B_{tag} contain a slow π^0.

3.4.1 Best $\Upsilon(4S)$ Candidate Selection

Given that for a given event, more than one pair of B_{sig} and B_{tag} candidates will combine, thus producing multiple $\Upsilon(4S) \to B_{\text{sig}}\ B_{\text{tag}}$ candidates, we apply a best $\Upsilon(4S)$ candidate selection. We select the $\Upsilon(4S)$ candidate(s) that contains the B_{tag} candidate with the highest tagging classifier output, and whenever we have multiple $\Upsilon(4S)$ candidates that survive this selection, we choose the $\Upsilon(4S)$ candidate that contains the B_{sig} with the highest p-value for the D meson vertex fit quality. The expected yields for the signal, normalization and background events, together with their efficiencies are given in Table 3.9. The complete "cut flow" for each signal channel is shown in tables 3.10 to 3.13. The numbers in the tables for the "Signal", "Normalization" and

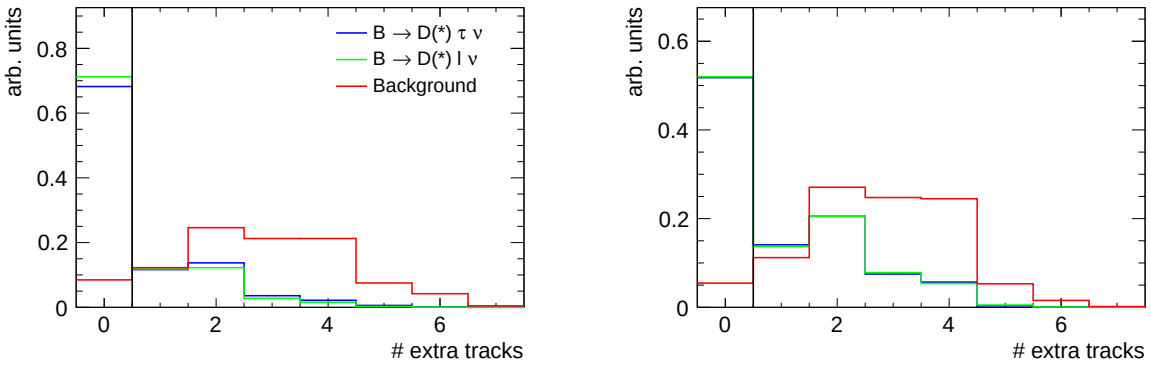


(a) $D^0\ell^-$ (b) $D^+\ell^-$



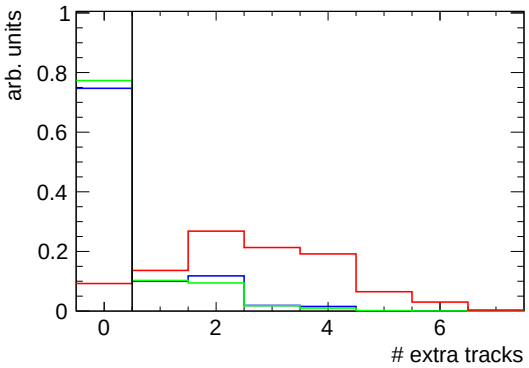
(c) $D^{*0}\ell^-$ (d) $D^{*+}\ell^-$

Figure 3.13: Distributions of the neutral energy left in calorimeter after $\Upsilon(4S)$ reconstruction, denoted by $E_{\rm ECL}$. The upper threshold value used for the candidates selection is shown with a vertical black line.

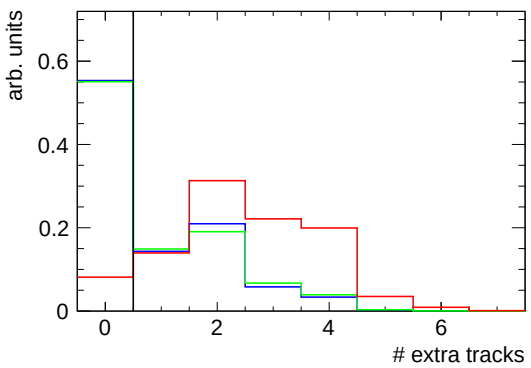


(a) $D^0\ell^-$

(b) $D^+\ell^-$

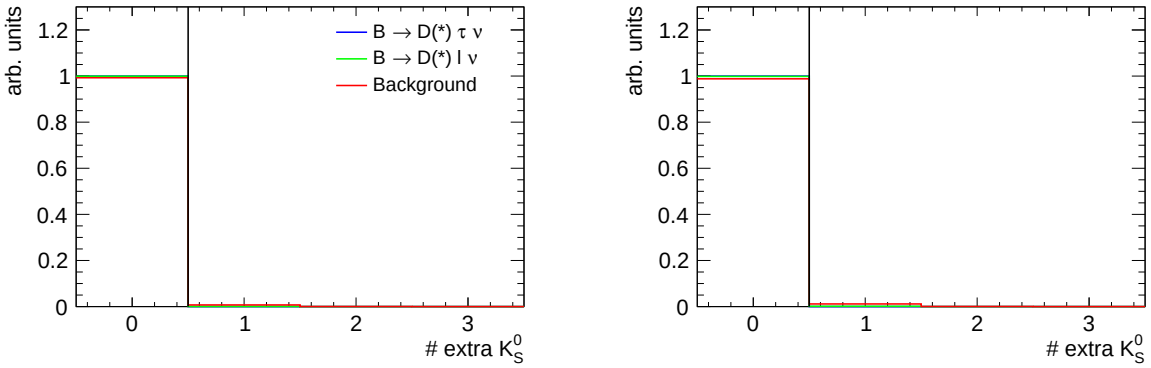


(c) $D^{*0}\ell^-$



(d) $D^{*+}\ell^-$

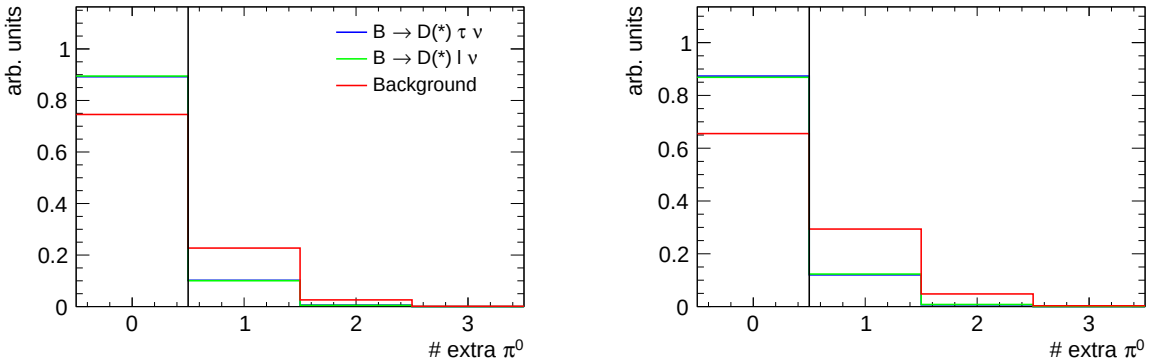
Figure 3.14: Distributions of the reconstructed number of tracks in the Rest of the Event. The upper threshold value used for the candidates selection is shown with a vertical black line.



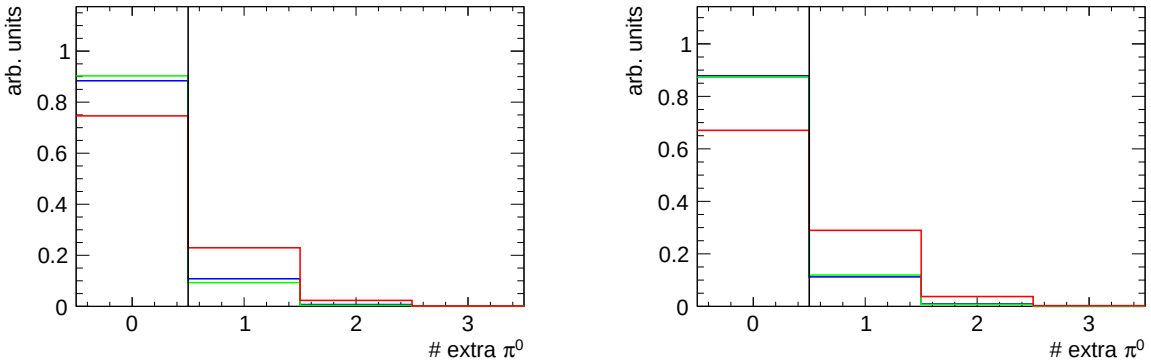
(a) $D^0\ell^-$ (b) $D^+\ell^-$

(c) $D^{*0}\ell^-$ (d) $D^{*+}\ell^-$

Figure 3.15: Distributions of the reconstructed number of extra K^0_S in the Rest of the Event. The upper threshold value used for the candidates selection is shown with a vertical black line.

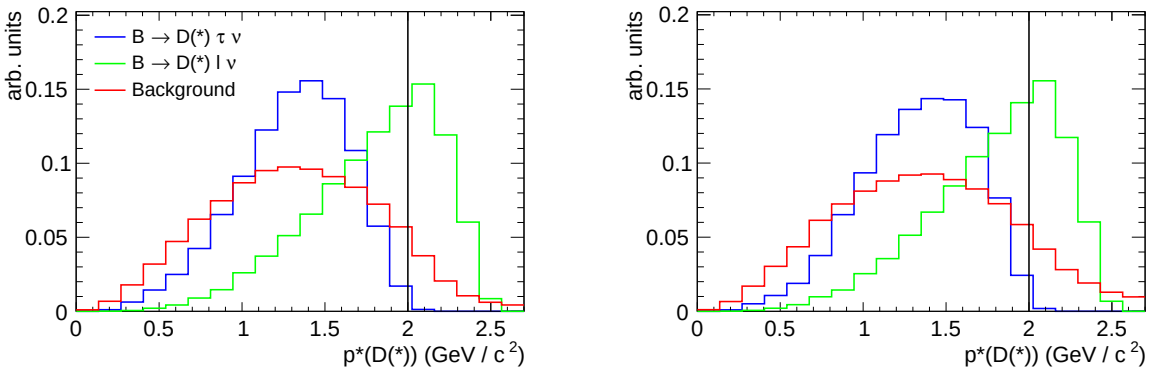


(a) $D^0\ell^-$ (b) $D^+\ell^-$

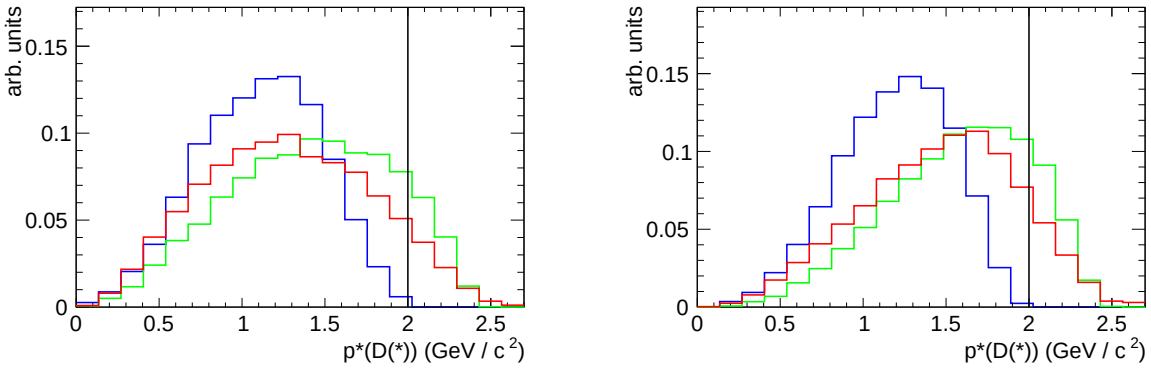


(c) $D^{*0}\ell^-$ (d) $D^{*+}\ell^-$

Figure 3.16: Distributions of the reconstructed number of extra π^0 in the Rest of the Event. The upper threshold value used for the candidates selection is shown with a vertical black line.



(a) $D^0\ell^-$ (b) $D^+\ell^-$



(c) $D^{*0}\ell^-$ (d) $D^{*+}\ell^-$

Figure 3.17: Distributions of the reconstructed charm meson three-momentum in the CM frame. The upper threshold value used for the candidates selection is shown with a vertical black line.

"Background" columns refer to the number of events that contain at least one signal, normalization or background candidate.

Table 3.9: Expected yields with efficiencies in parentheses for signal, normalization and background events, for all samples.

Sample	Signal	Normalization	Background
$D^0\ell^-$	991 (104×10^{-3})	13671 (86×10^{-3})	64218
$D^+\ell^-$	419 (56×10^{-3})	5850 (47×10^{-3})	12057
$D^{*0}\ell^-$	288 (15×10^{-3})	6962 (17×10^{-3})	2684
$D^{*+}\ell^-$	345 (23×10^{-3})	9362 (30×10^{-3})	1718

Table 3.10: Cut flow for data sample $D^0\ell^-$. For each selection criterion row, the relative change with respect to the previous row is shown in parantheses.

Selection	Signal	Normalization	Background	FoM signal	Signal/All (%)
	12698	277513	37235381	2.07	0.03
$E_{\rm ECL} < 1.2$ GeV	11202	246577	15501454	2.82	0.07
	(−11.7 %)	(−11.1 %)	(−58.4 %)	(+36.2 %)	(+133.3 %)
No extra tracks	7641	175614	1287558	6.30	0.59
	(−31.6 %)	(−28.7 %)	(−91.7 %)	(+123.4 %)	(+742.8 %)
No extra K_S^0 and	6815	157016	953499	6.45	0.71
no extra π^0	(−10.8 %)	(−10.6 %)	(−25.9 %)	(+0.2 %)	(+2.0 %)
$\lvert\boldsymbol{p}^*(D)\rvert < 2.0$ GeV/c	6807	99150	861965	6.92	0.79
	(−0.1 %)	(−36.8 %)	(−9.5 %)	(+6.3 %)	(+7.2 %)
$1.0 > \cos\theta_{B,D^{(*)}\ell} > -2.0$	5647	82792	542798	7.10	1.04
on tag side	(−17.0 %)	(−16.6 %)	(−37.0 %)	(+2.6 %)	(+31.6 %)
MVA output	1387	20984	103890	3.91	1.34
for $B_{\rm tag} > 10^{-1.5}$	(−75.4 %)	(−74.6 %)	(−80.8 %)	(−44.9 %)	(+28.8 %)
$\cos\theta_{B,D^{(*)}\ell} < 1.0$	1387	20553	102730	3.93	1.35
on signal side	(−0.0 %)	(−1.9 %)	(−1.0 %)	(+0.5 %)	(+0.7 %)
Best Candidate	992	13671	64218	3.59	1.63
Selection	(−30.4 %)	(−40.5 %)	(−42.3 %)	(−8.6 %)	(+22.2 %)

Table 3.11: Cut flow for data sample $D^+\ell^-$. For each selection criterion row, the relative change with respect to the previous row is shown in parantheses.

Selection	Signal	Normalization	Background	FoM signal	Signal/All (%)
	3936	87854	10920765	1.18	0.04
$E_{\mathrm{ECL}} < 1.2$ GeV	3029 (−22.7 %)	67521 (−22.6 %)	3536354 (−67.6 %)	1.59 (+33.8 %)	0.09 (+125.3 %)
No extra tracks	1569 (−46.2 %)	35098 (−47.3 %)	192689 (−94.5 %)	3.27 (+272.4 %)	0.81 (+777.7 %)
No extra K_S^0 and no extra π^0	1370 (−12.7 %)	30449 (−12.8 %)	124813 (−35.4 %)	3.42 (+4.5 %)	1.10 (+37.5 %)
$\lvert\boldsymbol{p}^*(D)\rvert < 2.0$ GeV/c	1370 (−0.0 %)	19234 (−36.6 %)	107336 (−14.0 %)	3.83 (+11.6 %)	1.28 (+16.3 %)
$1.0 > \cos\theta_{B,D^{(*)}\ell} > -2.0$ on tag side	1122 (−18.2 %)	15712 (−20.7 %)	54102 (−48.4 %)	4.21 (+10.4 %)	2.08 (+62.5 %)
MVA output for $B_{\mathrm{tag}} > 10^{-1.5}$	569 (−49.2 %)	8280 (−47.7 %)	16664 (−70.2 %)	3.56 (−14.2 %)	3.42 (+62.5 %)
$\cos\theta_{B,D^{(*)}\ell} < 1.0$ on signal side	569 (−0.0 %)	8090 (−2.4 %)	16384 (−1.8 %)	3.59 (+0.8 %)	3.48 (+1.7 %)
Best Candidate Selection	419 (−15.8 %)	5850 (−23.4 %)	12057 (−21.9 %)	3.40 (−5.5 %)	3.73 (+7.1 %)

Table 3.12: Cut flow for data sample $D^{*0}\ell^-$. For each selection criterion row, the relative change with respect to the previous row is shown in parantheses.

Selection	Signal	Normalization	Background	FoM signal	Signal/All (%)
	2934	71690	1275584	2.53	0.23
$E_{\rm ECL} < 1.2$ GeV	2638	65234	566643	3.31	0.47
	(−10.0 %)	(−9.0 %)	(−55.5 %)	(+31.6 %)	(+104.3 %)
No extra tracks	1971	50437	52291	6.09	3.77
	(−25.2 %)	(−22.7 %)	(−90.7 %)	(+81.5 %)	(+702.1 %)
No extra K_S^0 and	1742	45529	38721	5.94	4.50
no extra π^0	(−11.6 %)	(−9.7 %)	(−25.9 %)	(−2.4 %)	(+19.3 %)
$\lvert\boldsymbol{p}^*(D^*)\rvert < 2.0$ GeV/c	1742	39660	35398	6.28	4.92
	(−0.0 %)	(−12.8 %)	(−7.7 %)	(+5.7 %)	(+9.3 %)
$1.0 > \cos\theta_{B,D^{(*)}\ell} > -2.0$	1503	33996	22565	6.24	6.66
on tag side	(−13.7 %)	(−15.1 %)	(−36.2 %)	(−0.6 %)	(+34.6 %)
MVA output	423	11659	5004	3.24	8.46
for $B_{\rm tag} > 10^{-1.5}$	(−73.3 %)	(−64.7 %)	(−77.8 %)	(−48.0 %)	(+27.2 %)
$\cos\theta_{B,D^{(*)}\ell} < 1.0$	423	11395	4738	3.29	8.94
on signal side	(−0.0 %)	(−2.5 %)	(−4.9 %)	(+1.5 %)	(+5.9 %)
Best Candidate	290	6962	2684	2.82	10.0
Selection	(−28.3 %)	(−30.5 %)	(−37.2 %)	(−15.1 %)	(+11.2 %)

Table 3.13: Cut flow for data sample $D^{*+}\ell^-$. For each selection criterion row, the relative change with respect to the previous row is shown in parantheses.

Selection	Signal	Normalization	Background	FoM signal	Signal/All (%)
	2736	97833	836693	2.83	0.33
$E_{\rm ECL} < 1.2$ GeV	2142	76836	313682	3.41	0.68
	(−21.9 %)	(−20.4 %)	(−62.5 %)	(+21.2 %)	(+106.0 %)
No extra tracks	1185	42314	25525	4.50	4.64
	(−46.6 %)	(−44.9 %)	(−91.8 %)	(+32.3 %)	(+571.4 %)
No extra K_S^0 and	1041	36916	16971	4.44	6.14
no extra π^0	(−11.8 %)	(−9.7 %)	(−33.5 %)	(−1.3 %)	(+10.7 %)
$\lvert\boldsymbol{p}^*(D^*)\rvert < 2.0$ GeV/c	1041	30128	14845	4.85	7.01
	(−0.0 %)	(−18.9 %)	(−11.7 %)	(+9.0 %)	(+14.7 %)
$1.0 > \cos\theta_{B,D^{(*)}\ell} > -2.0$	861	24747	8232	4.68	10.46
on tag side	(−17.2 %)	(−15.6 %)	(−44.5 %)	(−3.5 %)	(+48.5 %)
MVA output	438	12706	2554	3.49	17.16
for $B_{\rm tag} > 10^{-1.5}$	(−49.9 %)	(−48.4 %)	(−49.3 %)	(−25.5 %)	(+64.0 %)
$\cos\theta_{B,D^{(*)}\ell} < 1.0$	438	12388	2486	3.54	17.63
on signal side	(−0.0 %)	(−2.3 %)	(−2.7 %)	(+1. %)	(+2.9 %)
Best Candidate	344	9362	1718	3.35	18.87
Selection	(−13.6 %)	(−24.4 %)	(−22.9 %)	(−5.6 %)	(+6.8 %)

Chapter 4

Signal Extraction

The goal of this analysis is to extract the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ from experimental data, to test the compatibility of the SM with measured data, or, in case of a discrepancy, the compatibility of other theories. We do this by building a model through Monte-Carlo (MC) simulations, according to the tested theory, but allowing some selected parameters of the model, which we want to measure, to be degrees of freedom of our model, such that their values will be determined to be the one that yield the best agreement between model the experimentally measured data. This defines the method called the maximum likelihood estimator, introduced in Section 4.1.3.

Firstly, we build probability distribution functions (PDFs) that represent what we expect from our model to be in the data, with one PDF for each component that makes the whole pool of reconstructed events, which include signals to be measured and the different sources of background.

We then fix some components' normalization values (or yields), and let some other yields float, which are the parameters of interest, and fit our PDFs to the data. We allow the fitting procedure determine which combination of determined floating parameters of interest is the most likely to describe the data.

4.1 Fit Procedure

The fit consists of a binned extended maximum likelihood fit computed by the `RooFit` package [62], and is performed on a 2D plane where each of the axes represents a variable with a specific purpose:

- E_{ECL} : the sum of the neutral energy deposited in the Electromagnetic Calorimeter, which is not associated to any of the particles used in the event reconstruction. Correctly reconstructed B-mesons on the signal side, which in our case correspond to signal and normalization events, have E_{ECL} values close to zero. On the other hand background events are characterized by missing some neutral particle (such as π^0) in the reconstruction, and peak at values higher than zero.

As a consequence this variable allows us to separate the signal and normalization components from background events, as shown in Figure 4.1;

- Multivariate classifier output: given that the reconstruction of signal and normalization events follows exactly the same procedure, and their kinematics are similar apart from the signal events having a softer lepton momentum spectrum, these two components show exactly the same distribution for the E_{ECL} variable. Hence we train a classifier to separate these signal events from normalization events, and we use the classifier output for the remaining axis of the fit 2D plane. A general introduction to multivariate classifiers is given in Section 4.1.1, while the details of the classifier used in this analysis is given in Section 4.1.2.

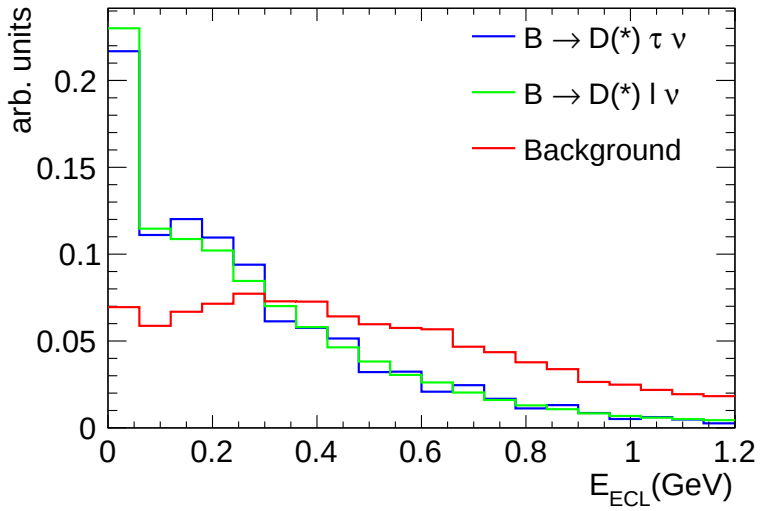


Figure 4.1: E_{ECL} distribution of signal, normalization and background events.

4.1.1 Gradient Boosted Decision Trees

Classifier introduction

We use a multivariate classifier to separate normalization events from signal events, allowing the fit to identify these two components and measure them precisely. The classifier learns how to identify each event as one of these two categories, when given enough features or input variables. A predictive model can thus be built by the classifier using MC simulations, and then applied to new data, such as those measured by the detector.

Formally, the classifier makes predictions depending on the input x, by using the model

$$F : \mathcal{X} \to \mathcal{A}, \tag{4.1}$$

mapping every input $x \in \mathcal{X}$ to a corresponding prediction $a \in \mathcal{A}$. Thus for a given x the classifier makes the prediction

$$a = f(X). \tag{4.2}$$

Since we are dealing with discrete categories, our problem is a classification problem, instead of a regression problem. The type of classifier we use is a decision tree, that learns how to divide the input space $\mathcal{X}$, in order to group together events of a same class. Specifically we use an algorithm called XGBoost [63], which uses a novel technique to exploit gradient boosted decision trees. To rest of this section is devoted to explain how GBDT work, and specifically what makes XGBoost an efficient and powerful algorithm.

Ensemble and Boosting

The type of classifier used in this analysis does not consist of just one tree, but it is made of an ensemble of trees, whose predictions will be then combined for a final event classification. The idea behind this concept is that we can combine multiple simple trees with limited predictive power, called weak or base-learners to produce an accurate, or strong, learner.

An historical way of using an ensemble of trees, is to have each of the trees randomly analyzing a different subset of the full data sample. By then averaging the response of all trees, this method ensure we also average out the *overfitting* of our predictor, i.e. how each tree learned statistical fluctuations of each subsample, which do not reflect any feature of the model we are trying to learn. Overfitting to a training data sample causes the response of the classifier to vary greatly when a new dataset is used, or in jargon, the classifier will have a high *variance*.

On the other hand there is also the risk of not capturing the features of the underlying model, i.e. underfitting, which means our predictions will always be far from the true labels, hence defining a classifier with high *bias*. Generally bias and variance can not be optimized at the same time, which introduces the bias-variance trade-off problem. This averaging ensemble method is used by the algorithm Random Forests [64].

A more recent method introduced an iterative way of adding trees to the ensemble, and was called *boosting*. After a first initial tree is used, we can choose a second tree such that the error of the first tree is minimized, i.e. we can perform a boost of the classifier. This can be generalized such that for any number N of trees that our ensemble contains, the $(N+1)^{\text{th}}$ tree will be chosen using the same method.

There are different ways choose the $(N+1)^{\text{th}}$ tree. One method is to train this tree on a dataset where the data points that suffer from large errors from the ensemble are reweighted to acquire a greater importance. This is the method used by AdaBoost which is regarded as the first practical boosting algorithm [65].

Another method is based on minimizing the so-called *loss function*, which is used to give a measure of the prediction accuracy of our classifier $F(x)$ when to compared

to the true labels y:

$$L(y_i, f(x_i)). \tag{4.3}$$

In this method we choose the $(N+1)^{\text{th}}$ tree by moving toward the direction that minimizes the loss function for the model built using N trees. This requires to identify this direction, which we can do by calculating the gradient of the loss function, and then moving the negative value of this gradient. Hence, the name of this method, gradient boosting [66].

The choice of the loss function is arbitrary, and different algorithms for gradient boosting are based on particular choices of the loss function. However if the loss function is chosen to be the classic squared-error loss, then what is actually is minimized is the errors or residuals from the ensemble predictions. The algorithm we have chosen, XGBoost, is based on a loss function that also includes a term that accounts for the complexity of the model, i.e. more complex models, which are likely to yield high variance, are penalized.

Gradient descent

After having initialized the model to a constant value

$$f_0(x) = \theta_0, \tag{4.4}$$

for a generic step m where $m = [0, M]$, we can compute the so-called pseudo residuals based on the ensemble build with $m-1$ base-learners:

$$-g_m(x_i) = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}. \tag{4.5}$$

These pseudo-residuals are then used as labels to train a base-learner $h_m(x)$ i.e. the training set is $\{(x_i, -g_m(x_i))\}_{i=1}^n$. The new base-learner is then added to the ensemble, with a certain weight, or step-length γ_m, determined by solving the following one-dimensional optimization problem:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, f_{m-1}(x_i) + \gamma h_m(x_i)) \tag{4.6}$$

A *shrinkage* factor $0 < \eta < 1$ is introduced to weight down the step length and therefore the contribution of this new learner to the ensemble, effectively slowing down the learning process (hence the other name learning rate) and making more the classifier more accurate. At the end of this iteration the ensemble model is updated:

$$f_m(x) = f_{m-1}(x) + \eta\gamma_m h_m(x). \tag{4.7}$$

These steps are repeated until a model is built in this iterative fashion.

$$f_M(x) = \sum_m f_m(x) \tag{4.8}$$

Hyperparameters and regularization

As explained in the previous section, classifiers can easily overfit and learn with too much precision the specific characteristics of a given dataset, and develop a very complex model that can classify perfectly all training data points. In order to avoid this outcome, a so-called *regularization* of the model needs to be used, so that the complexity of a model is penalized and minimized. In order to achieve this result the loss function, normally used to drive the classifier development, is modified to also include a term that accounts for the complexity of the model.

Several ways of accounting for the complexity of a boosted decision tree models are linked to the total number of trees in the ensemble, and also to the properties of the single trees, such as the depth of a tree, the number of terminal nodes, or the number of observations allowed in each terminal node. Furthermore, a measure of the complexity can be related also to the relative difference of the leaf weights, by assuming that each tree should not give such strong classification of the dataset to have widely different weights. The latter method is implemented by the classifier XGBoost used in this analysis, which stands out as the first algorithm to introduce complexity penalization for additive tree models. The penalization terms of the loss function can be written as:

$$\Omega(f) = \sum_{m=1}^{M} \left[\phi T_m + \frac{1}{2}\lambda ||w_m||_2^2 + \alpha ||w_m||_1 \right], \tag{4.9}$$

where the factor λ controls the contribution given to the penalization by the number of terminal nodes T_m and the following terms deal with the relative difference in leaf weights mentioned previously.

4.1.2 Signal vs Normalization Classifier

We train the XGBoost classifier using three input variables:

- m^2_{miss}, the missing four-momentum carried by the neutrino, introduced in Section 3.2.1.
- $\cos\theta_{B,D^{(*)}\ell}$, the cosine of the angle between the inferred direction of the reconstructed B-meson and that of the $D^{(*)}\,\ell$ system, introduced in Section 3.2.1.
- the sum of the energies of all the particles used in the $\Upsilon(4S)$ reconstruction, defined as:

  $$E_{\text{visible}} = \sum E_{i,\text{tag}} + \sum E_{i,\text{sig}}, \tag{4.10}$$

  where the sum extends over all final state particles over signal and tag sides.

The input variable distributions for the variables listed above are shown in Figures 4.2 and 4.3 for, respectively, B^+ and B^0 samples. Probability output distribution and relative ROC curve with Area Under Curve score for the trainings are shown in

Figures 4.9 and 4.8 for the B^+ and B^0 samples respectively. For the probability output distributions, the difference between the training and test distributions is also shown. The test dataset has not been used for the validation and training of the classifier. The AUC score is widely used as a measure of the classifier performance and is shown for the test dataset.

Given that the number of events in the signal class is approximately 20 times smaller than the one in the normalization class, if these two classes were fed in this proportion to the classifier, it would learn that it is always a safer choice to classify the event as normalization, defining a standard case of *imbalanced datasets*. For this reason, we resample the signal class using Synthetic Minority Over-sampling Technique (SMOTE) [67], to match the signal class and normalization class sizes.

4.1.3 Extended Maximum Likelihood

Assuming that a PDF normalized to one is being used to model the data, and it is a function of observables $\mathbf{x}$ and parameters $\mathbf{q}$

$$f(\mathbf{x};\mathbf{q}), \tag{4.11}$$

we can evaluate this PDF for a given data point $\mathbf{x_0}$ and call this value the *likelihood* of the PDF with respect to the observed data, as in

$$\mathcal{L}(\mathbf{q}) = f(\mathbf{x_0};\mathbf{q}). \tag{4.12}$$

We can also define the likelihood of a data set of experimental points, by combining the likelihood of all points

$$\mathcal{L}(\mathbf{q}) = \prod_i f(\mathbf{x_i};\mathbf{q}). \tag{4.13}$$

Clearly the likelihood function depends on the PDF parameters $\mathbf{q}$, which means that it can be maximized by an appropriate choice of these parameters.

This formalism assumes a PDF that is normalized to one, but if we wish to loosen this constraint, we can use extended maximum likelihood (EML) [68]. In this framework the likelihood is defined as

$$\mathcal{L}(\mathbf{q}) = \frac{e^{-N_f} N_f^{N_d}}{N_d!} \cdot \prod_i f(\mathbf{x}_i;\mathbf{q}), \tag{4.14}$$

where we have added the Poisson probability of finding the observed total number of event N_d with respect to the expectation values N_f, which is a function of the model parameters $\mathbf{q}$. This simple expression applies to cases where only one PDF is used.

However, it is often the case that the data sample is split into several model components, generally a signal component and one or more background components, where a PDF can be associated to each of them. The composite PDF can be written as

$$m(\mathbf{x};\mathbf{q}) = N_S \cdot s(\mathbf{x};\mathbf{q}) + \sum_i N_B^i b^i(\mathbf{x};\mathbf{q}^i), \tag{4.15}$$

where the subscript S refers to the signal component, and we generically summarize the sum of all background components running over the index i.

In our case the parameters of interest are the yields of selected components of the data sample, while the shape of the PDFs is fixed by the simulation expectations and does not depend on any parameter that has to be determined by the fit. Keeping in mind that the observables we use to build our PDFs are E_{ECL} and the MVA classifier output, we can therefore adapt equation 4.14 as

$$\mathcal{L}(N_1, N_2, ...) = \prod_{k=1}^{N_{\text{ch}}} \mathcal{P}_{\text{Poisson}}(n_k, q_k) \prod_{i=1}^{n_k} \left(\sum_{j=1}^{c_k} N_{j,k} \cdot m_{j,k}(\mathbf{x_{i,j,k}}) \right), \tag{4.16}$$

where N_{ch} denotes the number of channels simultaneously fit, n_k the number of experimental data points for channel k, q_k the total number of entries in this channel as determined by the fit, and c_k the number of PDF components always for channel k. The job of the fit algorithm is to find the combination of floating yields that maximises the likelihood function of Equation 4.16

4.2 Fit Components

We use RooFit [62] to extract Probability Distribution Functions (PDFs) from binned 2D distributions, to fit the data distributions using these PDFs and to extract the yields of the signal and normalization components. The components which make up the reconstructed events pool are:

- **Signal**: Decays in the channels $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ that are correctly reconstructed. The accuracy of the reconstruction is assessed through an algorithm that checks recursively if all the reconstructed particles in the decay tree of the B_{sig} candidate match the equivalent particles in the MC truth. When we find an error in this type of reconstruction to MC truth matching, we refer to such candidate as a *fake*. In the fit we measure the signal component by letting the R factors float, with the benefit of extracting the quantity which is the focus of this measurement, indeed R, directly from the fit. The reconstructed yields of signal and normalization are linked to the $R(D^{(*)})$ values using their reconstruction efficiencies and the branching ratio for the leptonic decay of the τ:

$$\begin{aligned} R(D^{(*)}) &= \frac{N_{\text{sig}}^{\text{gen}}}{N_{\text{norm}}^{\text{gen}}} \cdot \frac{1}{\mathcal{B}\left(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau\right)} \\ &= \frac{N_{\text{sig}}^{\text{fit}}}{N_{\text{norm}}^{\text{fit}}} \cdot \frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \cdot \frac{1}{\mathcal{B}\left(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau\right)} \\ &= \frac{N_{\text{sig}}^{\text{fit}}}{N_{\text{norm}}^{\text{fit}}} \cdot f_i \cdot \frac{1}{\mathcal{B}\left(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau\right)} \end{aligned} \tag{4.17}$$

where N^{fit} refers to the yield determined by the fit, N^{gen} refers to the generated events at MC truth level, and ϵ is their ratio. The factors f_i are calculated for each signal channel individually, using 10 streams of generic MC.

The calculation of the constrained signal yield can be expressed as

$$N_{D^{(*)}\ell,\,\text{sig}} = R(D^{(*)}) \cdot \frac{\mathcal{B}\left(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau\right)}{f_{D^{(*)}\ell}} \cdot N_{D^{(*)}\ell,\,\text{norm}} \tag{4.18}$$

where $N_{D^{(*)}\ell,\,\text{norm}}$ is the normalization yield.

- **Normalization**: Decays in the channels $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ that are correctly reconstructed, following the definition given above.

- **Feed-down, $\boldsymbol{\ell}$**: For the reconstructed events in the $\bar{B} \to D\ell^-\bar{\nu}_\ell$ signal channels, there is a contribution from the decays of the type $B \to (D^* \to D\overline{\pi_{\text{slow}},\gamma})\ell\nu$ where the line over the π indicates that the slow pion is missed by the reconstruction.

  More specifically, in the channels $B^+ \to D^0\ell\nu$ the feed down events will come from D^* decays of the type $D^{*0} \to D^0\overline{\pi^0/\gamma}$ and $D^{*+} \to D^0\overline{\pi^+}$, that arise respectively in events from charged and neutral (often denoted as mixed) B decays. Thus, the yield for the feed-down to $B^+ \to D^0\ell\nu$ is composed of two contributions

$$N(D^0\not{X}) = N(B^+ \to D^{*0} \to D^0\overline{\pi^0/\gamma}) + N(B^0 \to D^{*+} \to D^0\overline{\pi^+}). \tag{4.19}$$

  We refer to the former as **charged feed-down** and the latter as **mixed feed-down**. On the other hand, the feed-down events for the channels $B^0 \to D^+\ell\nu$ are only of one type, mixed feed-down, since they originate only from $D^{*+} \to D^+\overline{\pi^0}$ events

$$N(D^+\not{X}) = N(B^0 \to D^{*+} \to D^+\overline{\pi^0}). \tag{4.20}$$

  Mixed and charged feed-down have significantly different shapes in our 2D fit plane (as shown in Figure 4.10), and as a consequence they each have a separate PDF. Since these events differ from the normalization in the $D^*\ell$ channels only for the missing slow pion, their BDT classification will make them appear mostly in the normalization region of the 2D plane, and by let them float the statistical precision of the signal extraction will not be affected.

  Moreover, in this way the ratio of feed-down events in the $D\ell$ channels to the normalization events in the $D^*\ell$ channels will determined by the fit, and can be used to constrain the feed-down that come from the signal $B \to D^*\tau\nu$, as explained below.

- **Feed down, $\boldsymbol{\tau}$**: The number of events in the feed-down components $B \to (D^* \to D\overline{\pi_{\text{slow}}})\tau\nu$ can be correlated to the yield of the signal in the $D^*\ell$ channels using scale factors, which depend on the probability of reconstructing a slow pion, and

that are determined through MC studies. More specifically, for the charged feed-down to the channels $B^+ \to D^0 \ell\nu$ we can use the correlation

$$N(B^+ \to D^{*0} \to D^0 \cancel{\pi^0/\gamma}) = K_{\pi^0/\gamma} \cdot N(B^+ \to D^{*0} \to D^0 \pi^0/\gamma), \tag{4.21}$$

which can be written more concisely as

$$N_{D^0\ell,\ \text{charged fd-}\tau} = K_{\pi^0/\gamma} \cdot N_{D^{*0}\ell,\ \text{sig}}. \tag{4.22}$$

For the mixed feed-down to the channels $B^+ \to D^0 \ell\nu$ we can write

$$\begin{aligned} N(B^0 \to D^{*+} \to D^0 \cancel{\pi^+}) &= K_{\pi^+} \cdot N(B^0 \to D^{*+} \to D^0 \pi^+) \\ &= K_{\pi^+} \cdot S \cdot N(B^0 \to D^{*+} \to \text{all}), \end{aligned} \tag{4.23}$$

or equivalently

$$N_{D^0\ell,\ \text{mixed fd-}\tau} = K_{\pi^+} \cdot S \cdot N_{D^{*+}\ell,\ \text{sig}}. \tag{4.24}$$

On the other hand, for the definition of the feed-down to the channels $B^0 \to D^+ \ell\nu$ we only need to take into account the mixed feed-down, and the correlation can be written as

$$\begin{aligned} N(B^0 \to D^{*+} \to D^+ \cancel{\pi^0}) &= K_{\pi^0} \cdot N(B^0 \to D^{*+} \to D^+ \pi^0) \\ &= K_{\pi^0} \cdot (1 - S) \cdot N(B^0 \to D^{*+} \to \text{all}), \end{aligned} \tag{4.25}$$

or equivalently

$$N_{D^+\ell,\ \text{mixed fd-}\tau} = K_{\pi^0} \cdot (1 - S) \cdot N_{D^{*+}\ell,\ \text{sig}}. \tag{4.26}$$

By taking into account Eq. 4.18 that shows how the signal yield is itself expressed as a function of the floating parameters $R(D^*)$ and normalization yield, we can rewrite the feed-down yields as

$$\begin{aligned} N_{D^0\ell,\ \text{charged fd-}\tau} &= K_{\pi^0/\gamma} \cdot R(D^{*0}) \cdot \frac{\mathcal{B}(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau)}{f_{D^{*0}\ell}} \cdot N_{D^{*0}\ell,\ \text{norm}}, \\ N_{D^0\ell,\ \text{mixed fd-}\tau} &= K_{\pi^+} \cdot S \cdot R(D^{*+}) \cdot \frac{\mathcal{B}(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau)}{f_{D^{*+}\ell}} \cdot N_{D^{*+}\ell,\ \text{norm}}, \text{ and} \\ N_{D^+\ell,\ \text{mixed fd-}\tau} &= K_{\pi^0} \cdot (1 - S) \cdot R(D^{*+}) \cdot \frac{\mathcal{B}(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau)}{f_{D^{*+}\ell}} \cdot N_{D^{*+}\ell,\ \text{norm}}. \end{aligned} \tag{4.27}$$

Since the feed-down from $B \to D^* \ell\nu$ is a free parameter of the fit, we correct for any mismodeling of slow pion reconstruction efficiency by updating the correlation scale factors for the feed-down from $B \to D^* \tau\nu$ as

$$\left.\frac{N(D^*\tau\nu \to D\cancel{\pi})}{N(D^*\tau\nu \to D\pi)}\right|_{\text{Iter.}} = \left.\frac{N(D^*\tau\nu \to D\cancel{\pi})}{N(D^*\tau\nu \to D\pi)}\right|_{\text{MC}} \times \left.\frac{N(D^*\ell\nu \to D\cancel{\pi})}{N(D^*\ell\nu \to D\pi)}\right|_{\text{Fit}} \times \left.\frac{N(D^*\ell\nu \to D\pi)}{N(D^*\ell\nu \to D\cancel{\pi})}\right|_{\text{MC}} \tag{4.28}$$

- $\boldsymbol{\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell}$: The charm-meson ground states D and D^* have orbital angular momentum $L = 0$. Orbitally excited charm mesons with $L = 1$ are called D^{**} states. Events where the decay of the D^{**}-meson is not fully reconstructed, very often because a π^0 is missed, mimic a signal or normalization event. The experimental knowledge of $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ decays is still rather poor, especially for the broad D^{**} states. As a consequence, the modelling of this background suffers from large uncertainties and this yield is a free parameter of the fit. To model this component we use 4.6 MC streams of dedicated $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ events which feature more decay modes and updated branching ratios with respect to the older generic MC samples.

- **Fake $\boldsymbol{D^{(*)}}$-mesons**: Events where the $D^{(*)}$-meson candidate daughter of the B_{sig} is not correctly reconstructed. The yield of this component is calibrated using sidebands of the charm invariant mass distributions for the $D^{(*)}$, and this scale factor's uncertainty is taken into account in the systematics estimation.

- **Other**: All other categories such as, for example, continuum background, events with correctly reconstructed charm particles but with fake leptons, events that originate from $B \to D_s^{(*)}\ D^{(*)}$ decays, constitute small contributions and are grouped together under this component.

The yields of the fake $D^{(*)}$-meson backgrounds will be estimated using data-driven techniques, as explained in section 5.4. The PDFs for these components are shown in Figures 4.12 to 4.11 for the B^0 and B^+ samples respectively, while the expected yields for all PDFs components are detailed in tables 4.1 and 4.2 for B^0 and B^+ samples respectively.

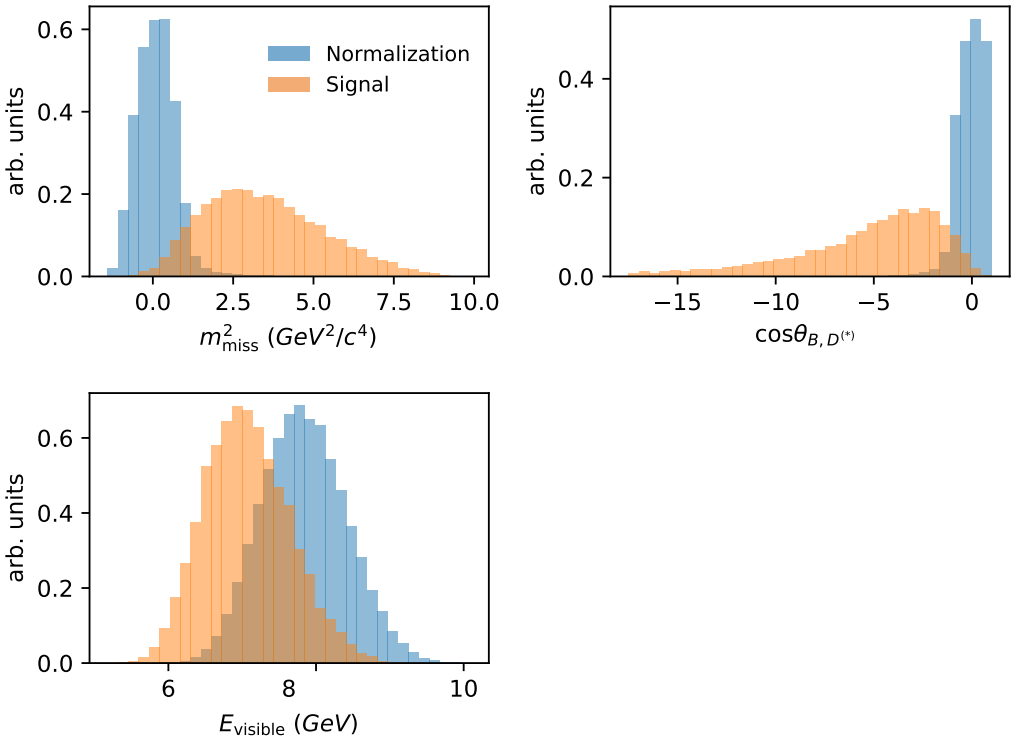
4.2.1 Fit Validation

In order to check the stability of the fit algorithm, we perform several tests. Firstly, we split our MC samples in 10 parts (which we label streams) and perform 10 different tests, where in each test we use 1 stream as pseudo-data and the remaining 9 streams to obtain the PDFs for the fit and determine the correlation factors. As expected, we find fit results in good agreement across the 10 streams.

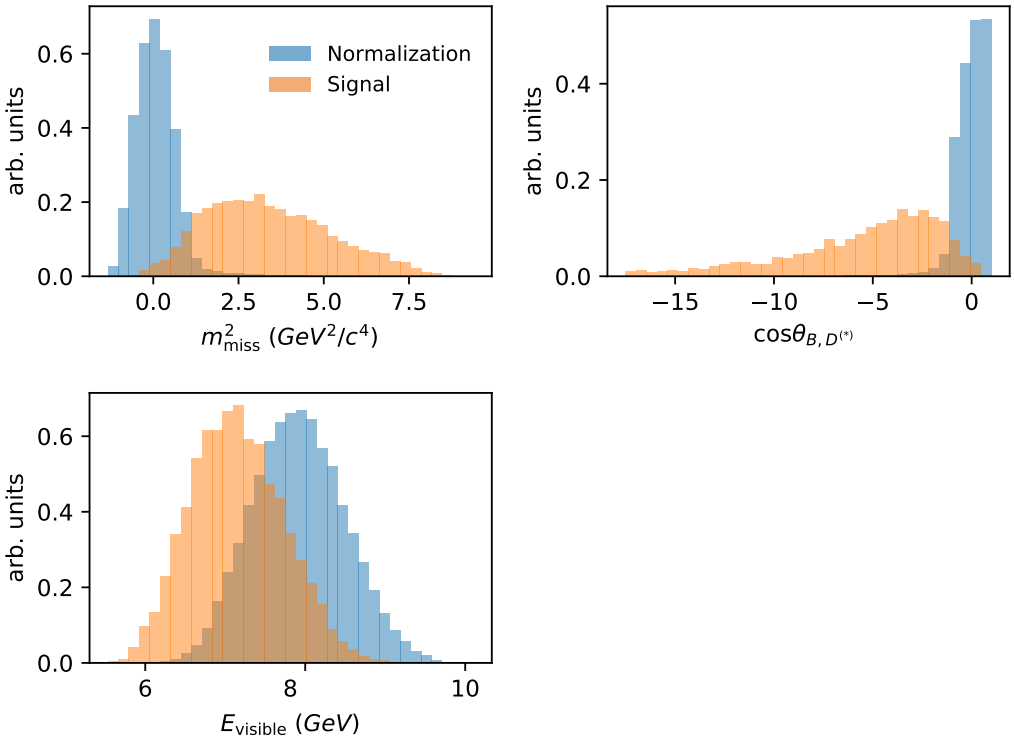
Since previous measurements of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ have found an excess in the signal yield with respect to the SM expectations, we test the capability of the fit to identify signal components with different size with respect to the one expected in our SM MC samples. We perform this test by introducing a signal component with a yield that varies from 0.25 to 2.00 times its nominal SM value, and verify that the fit algorithm is able to properly measure this enlarged signal component.

Lastly, we generate a large number of pseudo-datasets by varying the bin contents of the 2D classifier-E_{ECL} distributions built using the MC samples. We vary each bin content by sampling from a Poisson distribution built using the bin content and the bin error. For each pseudo-dataset built following this method, we repeat the fit procedure

and verify that the fit results follow into Gaussian distributions with central values equal to the nominal fit result values and standard deviations equal to the fit result errors.

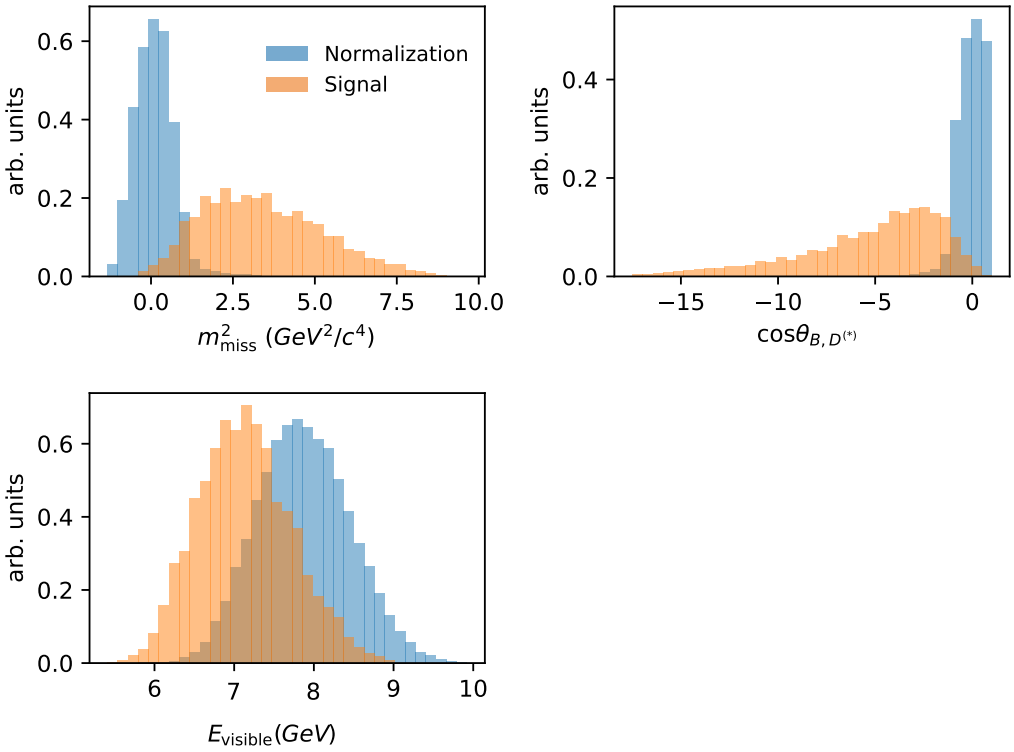


(a) Sample: $D^0\ell^-$.

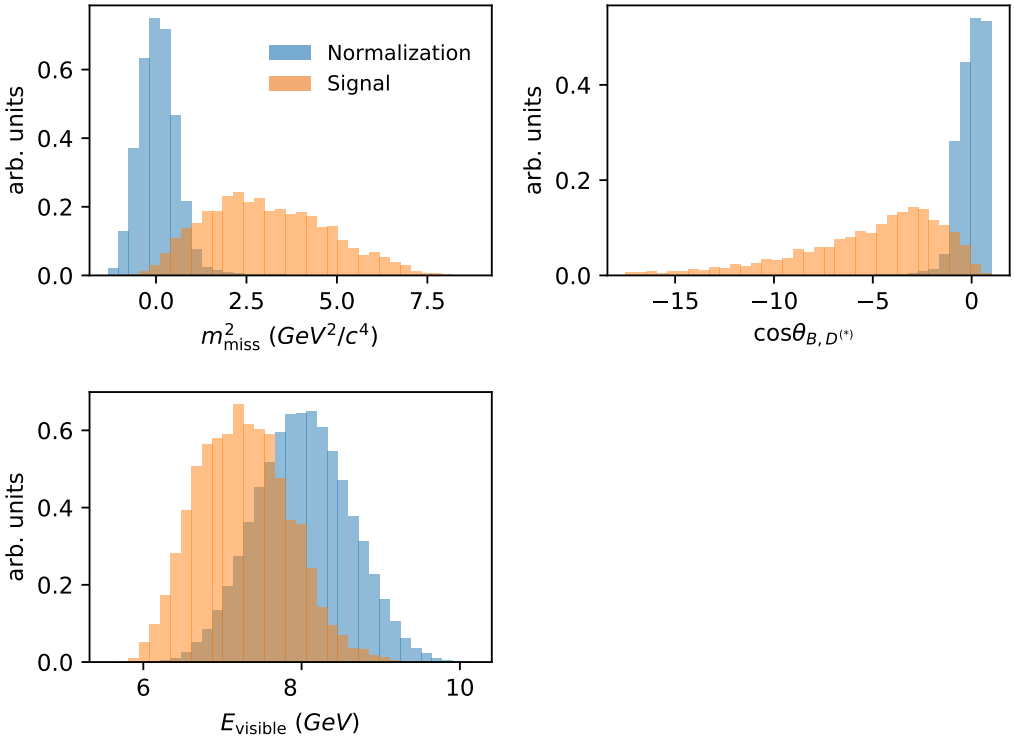


(b) Sample: $D^{*0}\ell^-$.

Figure 4.2: Input variable distributions for the B^+ samples.

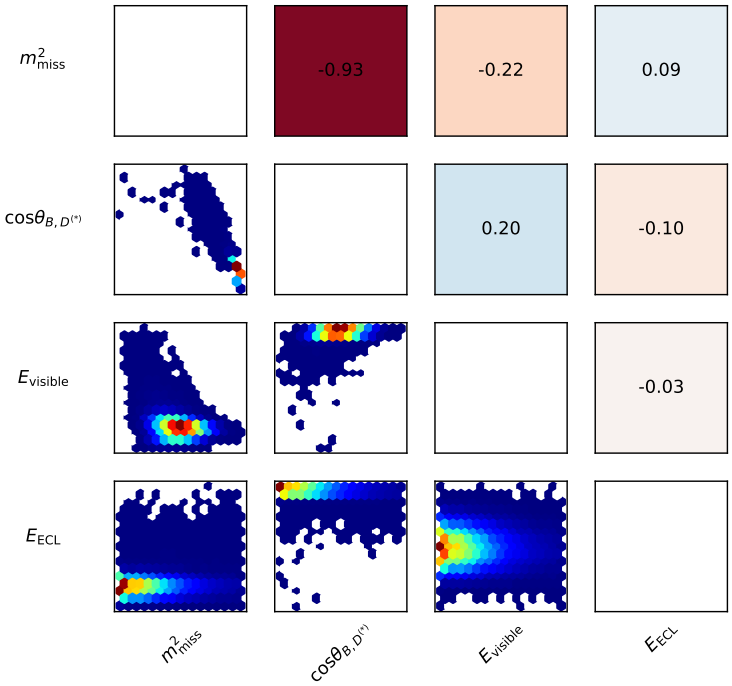


(a) Sample: $D^+\ell^-$.

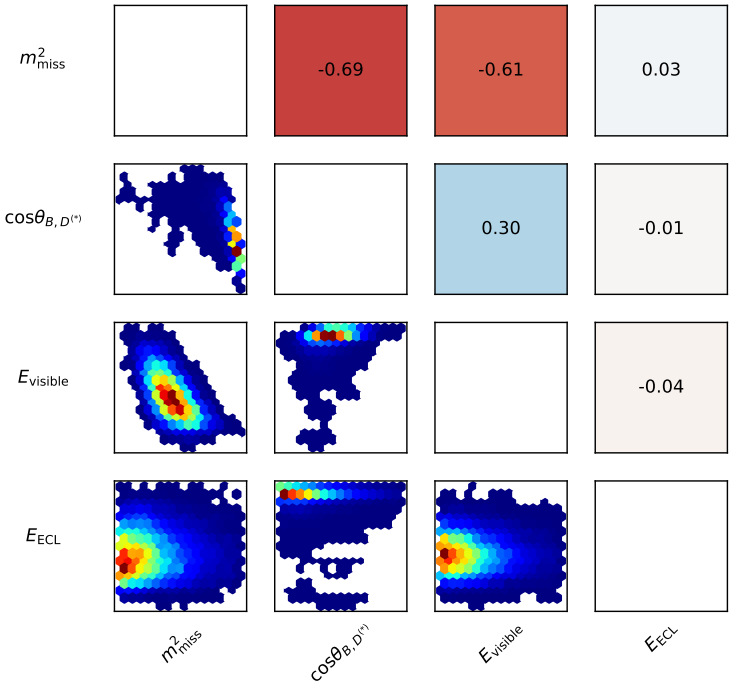


(b) Sample: $D^{*+}\ell^-$.

Figure 4.3: Input variable distributions for the B^0 samples.

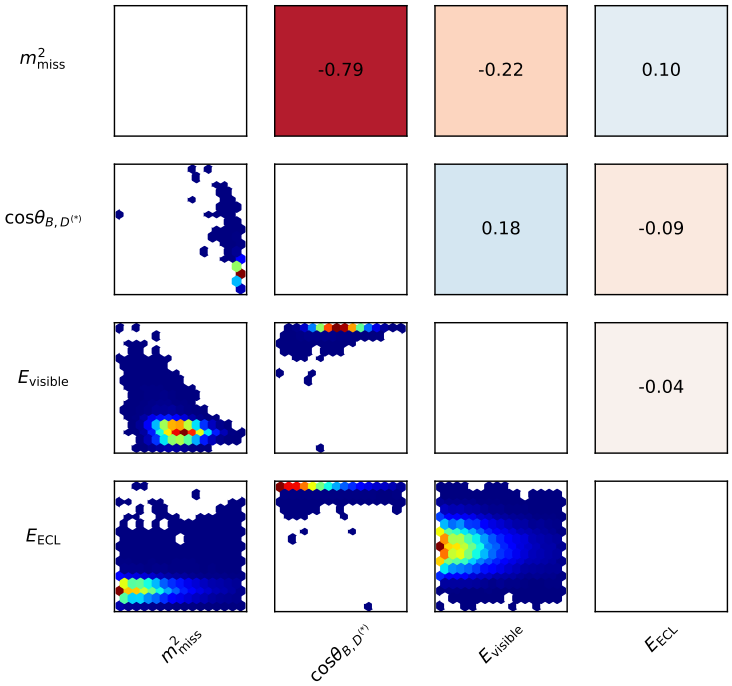


(a) Normalization

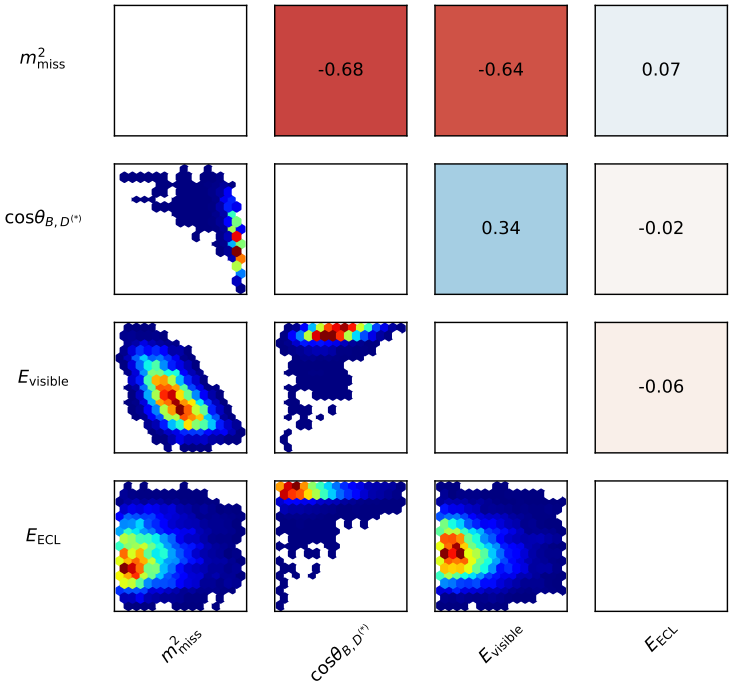


(b) Signal

Figure 4.4: Linear correlations between normalization and signal input variables for the sample $D^0\ell^-$.

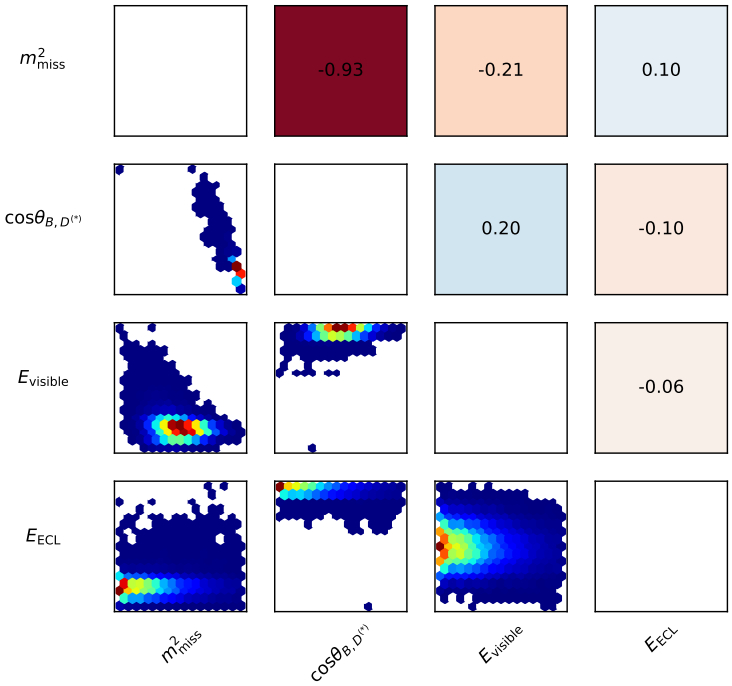


(a) Normalization

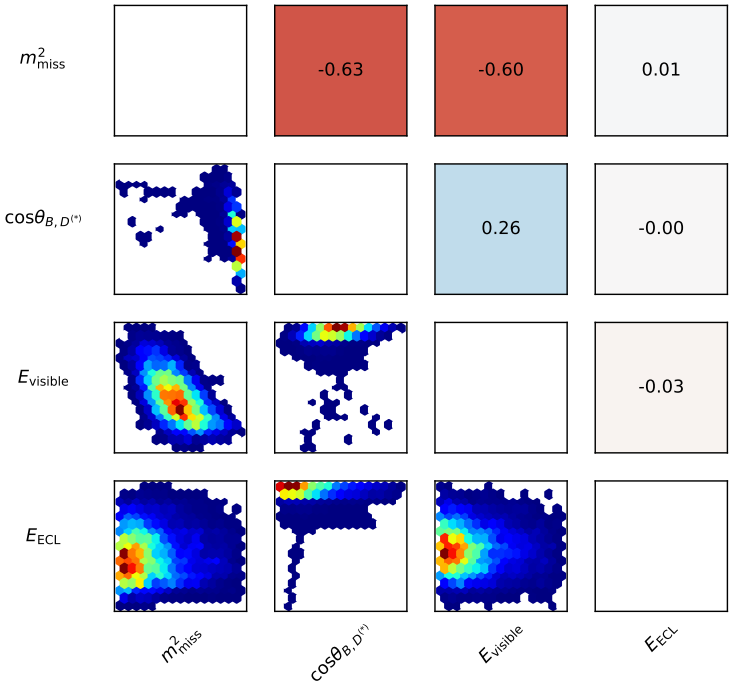


(b) Signal

Figure 4.5: Linear correlations between normalization and signal input variables for the sample $D^{*0}\ell^-$.

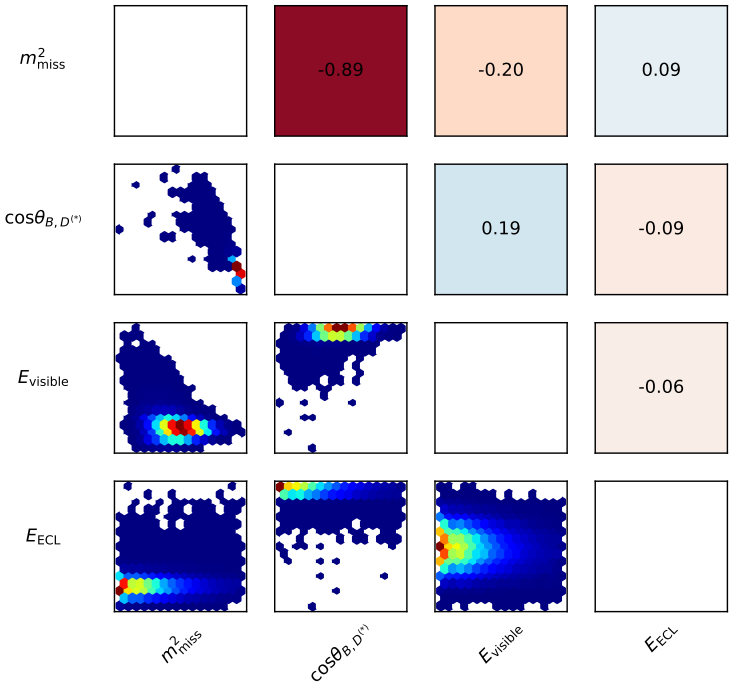


(a) Normalization

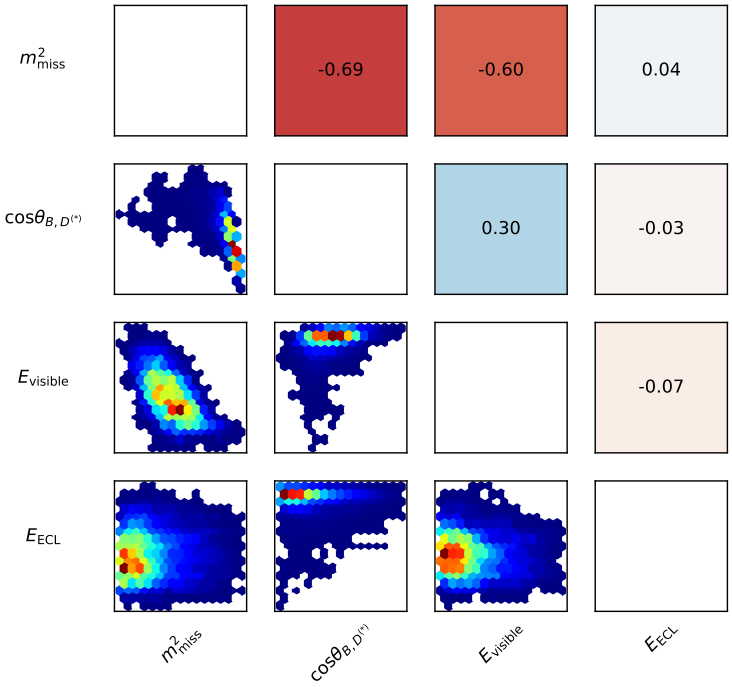


(b) Signal

Figure 4.6: Correlation between normalization and signal input variables for the sample $D^+\ell^-$.

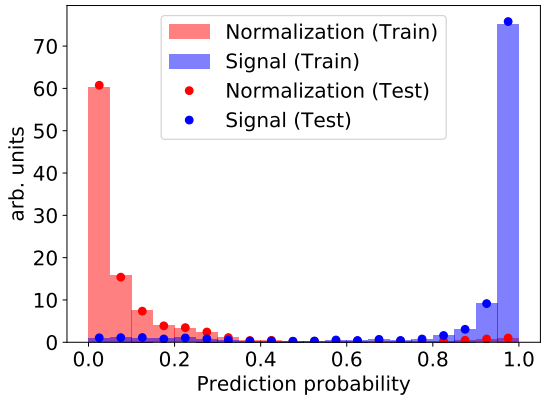


(a) Normalization

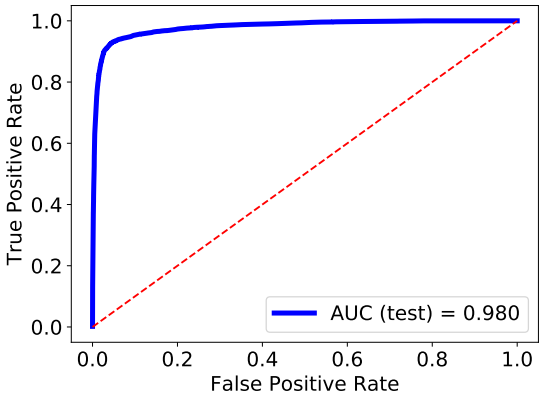


(b) Signal

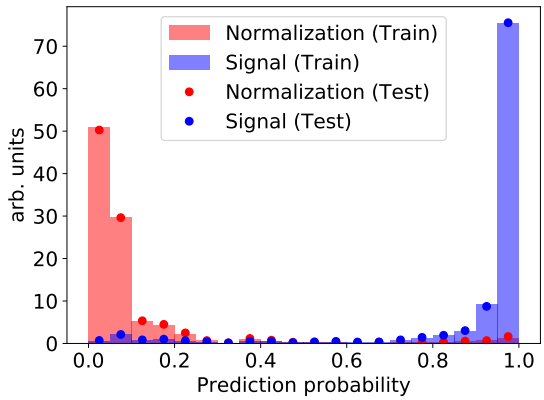
Figure 4.7: Correlation between normalization and signal input variables for the sample $D^{*+}\ell^-$.



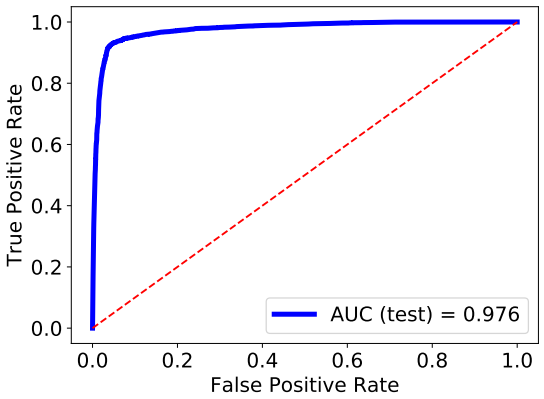
(a) Sample: $D^0\ell^-$



(b) Sample: $D^0\ell^-$

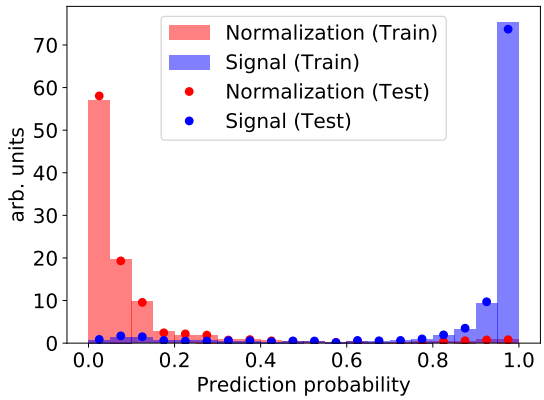


(c) Sample: $D^{*0}\ell^-$

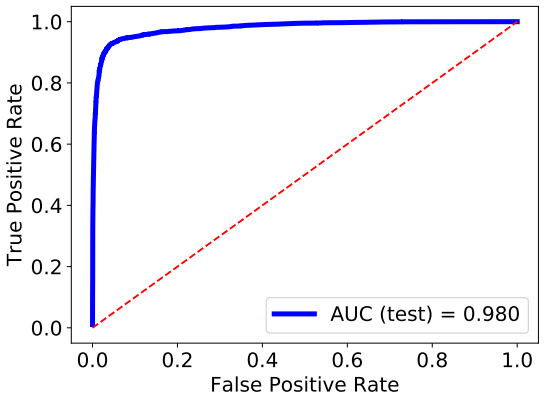


(d) Sample: $D^{*0}\ell^-$

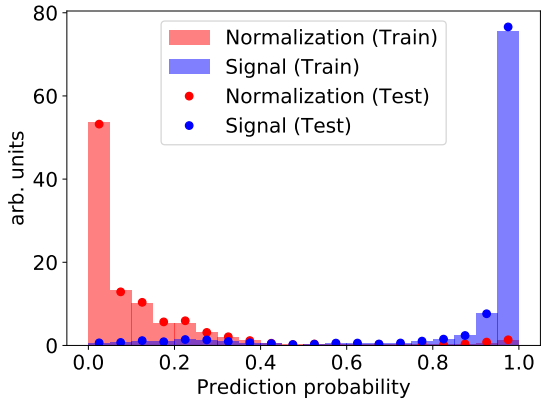
Figure 4.8: Probability output distribution (left) and ROC curve with Area Under Curve score (right) for the B^+ samples.



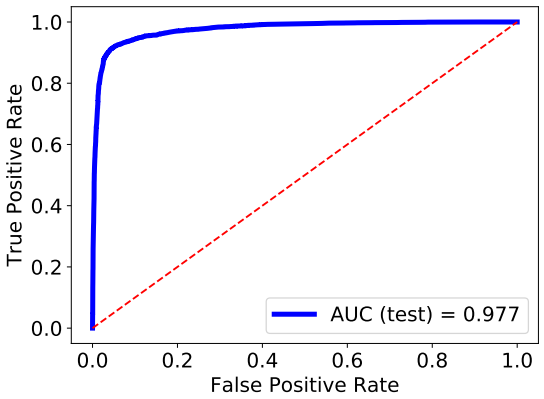
(a) Sample: $D^+\ell^-$



(b) Sample: $D^+\ell^-$

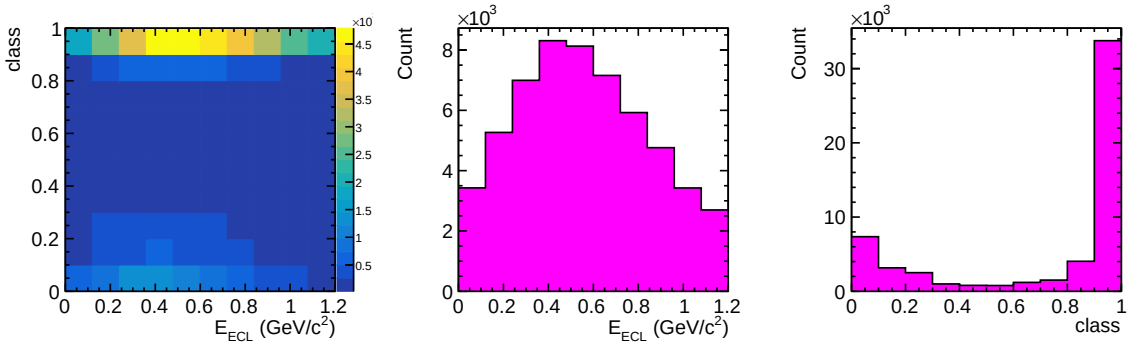


(c) Sample: $D^{*+}\ell^-$

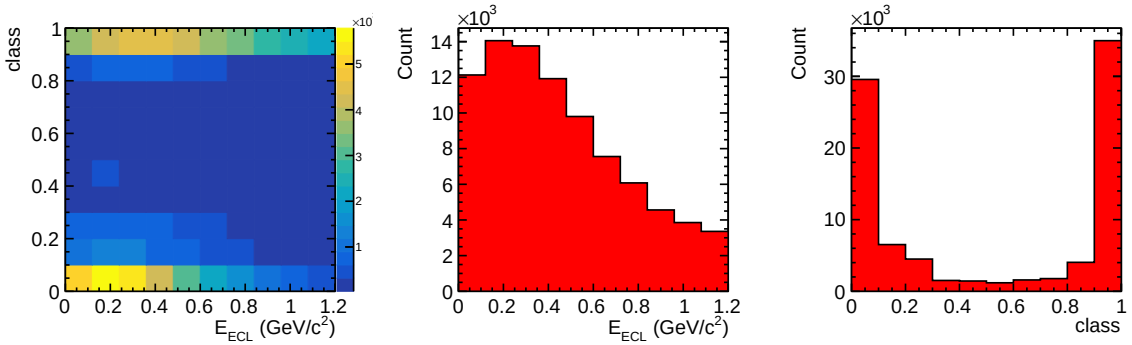


(d) Sample: $D^{*+}\ell^-$

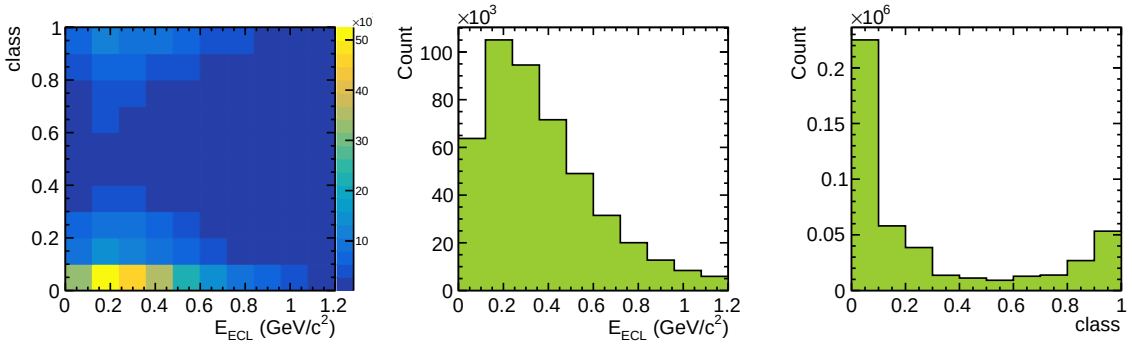
Figure 4.9: Probability output distribution (left) and ROC curve with Area Under Curve score (right) for the B^0 samples.



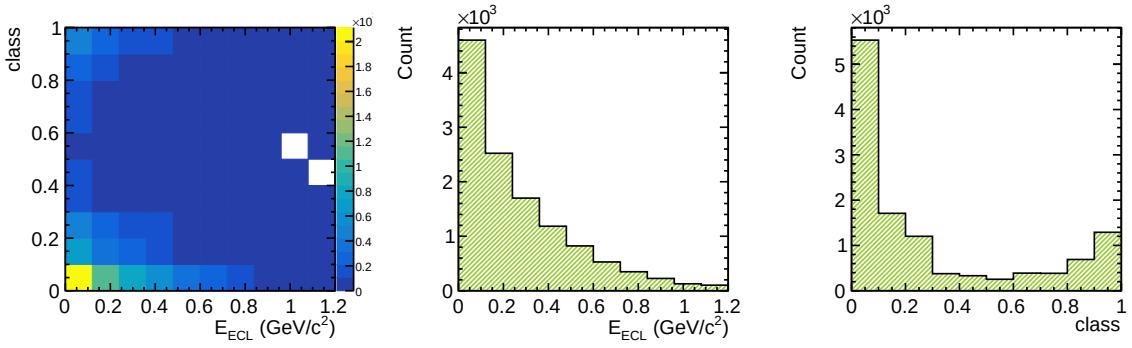
(a) $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ PDF component.



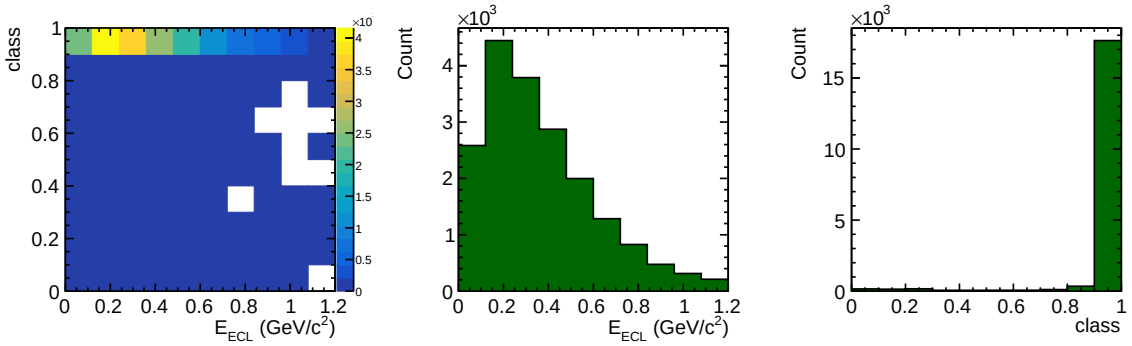
(b) Fake charm PDF component.



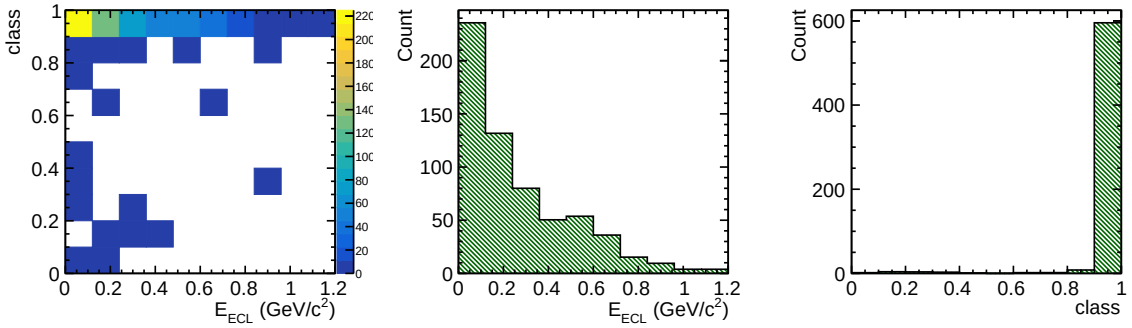
(c) Charged feed-down, ℓ PDF component.



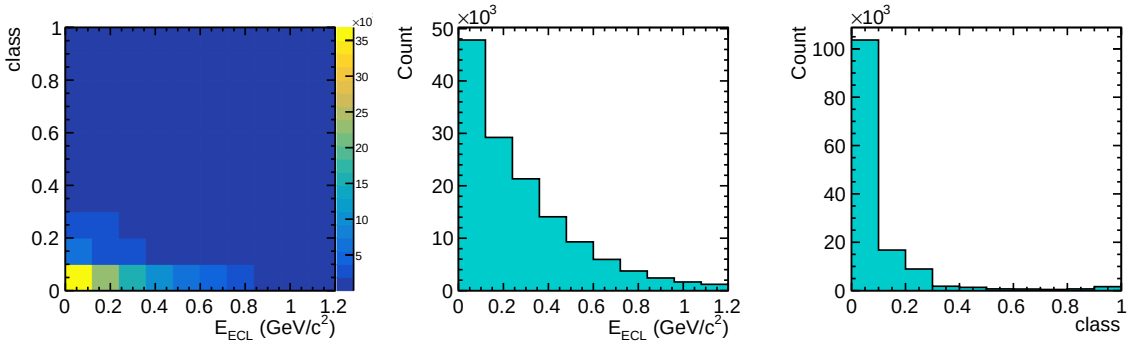
(d) Mixed feed-down, ℓ PDF component.



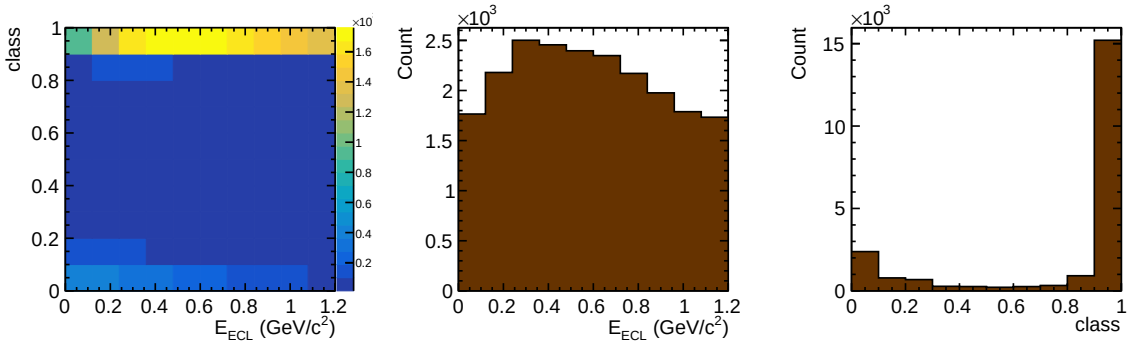
(e) Charged feed-down, τ PDF component.



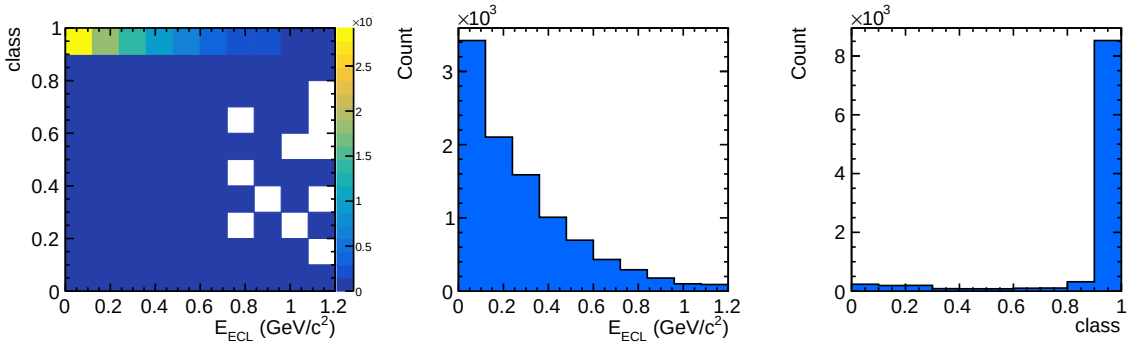
(f) Mixed feed-down, τ PDF component.



(g) Normalization PDF component.

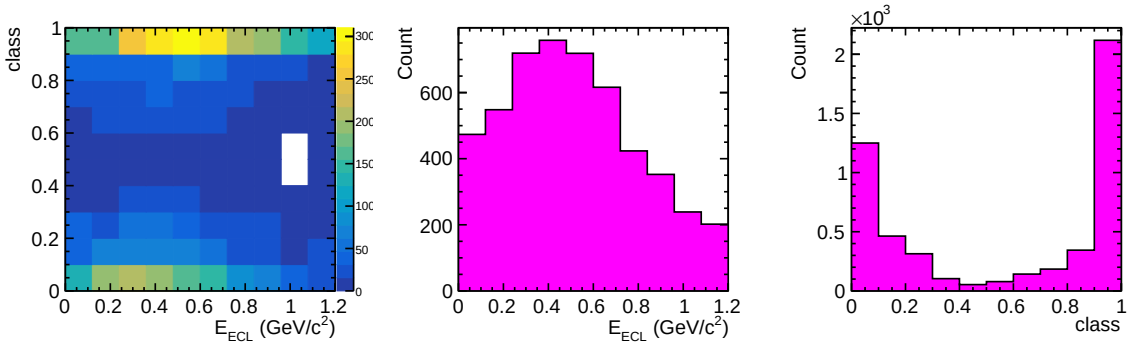


(h) Other PDF component.

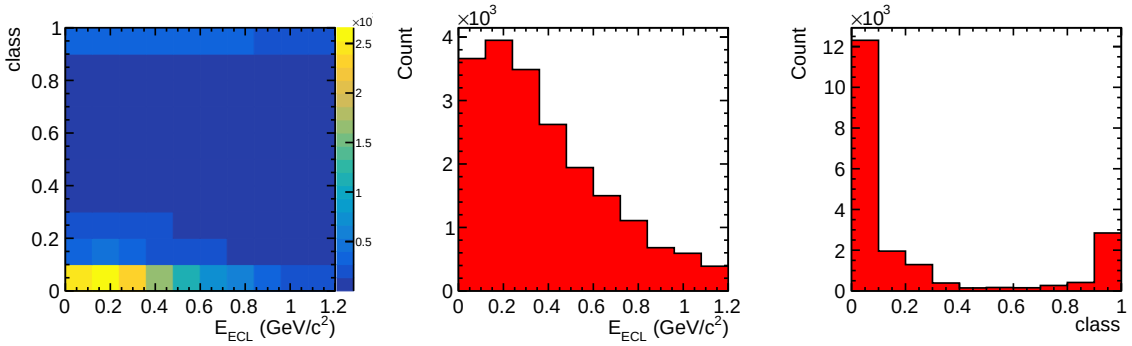


(i) Signal PDF component.

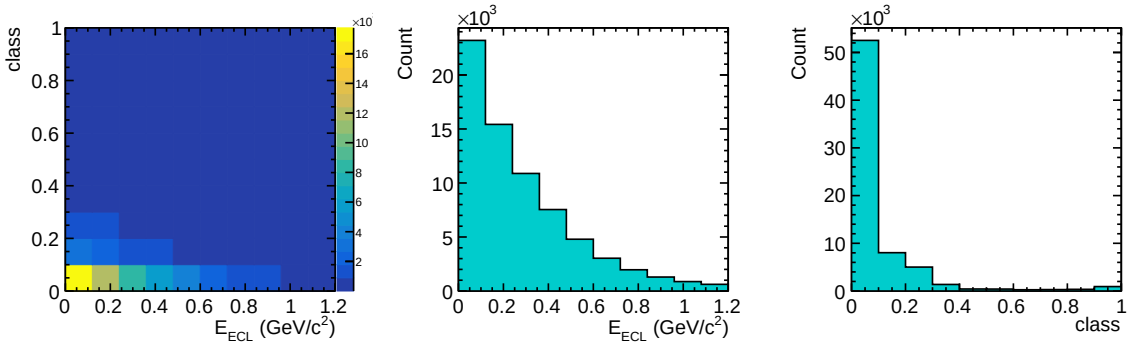
Figure 4.10: Fit PDFs for the sample $D^0\ell^-$.



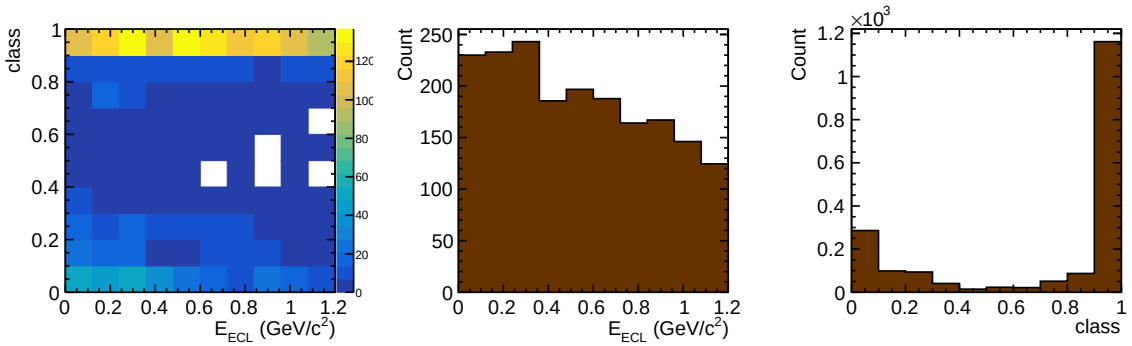
(a) $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ PDF component.



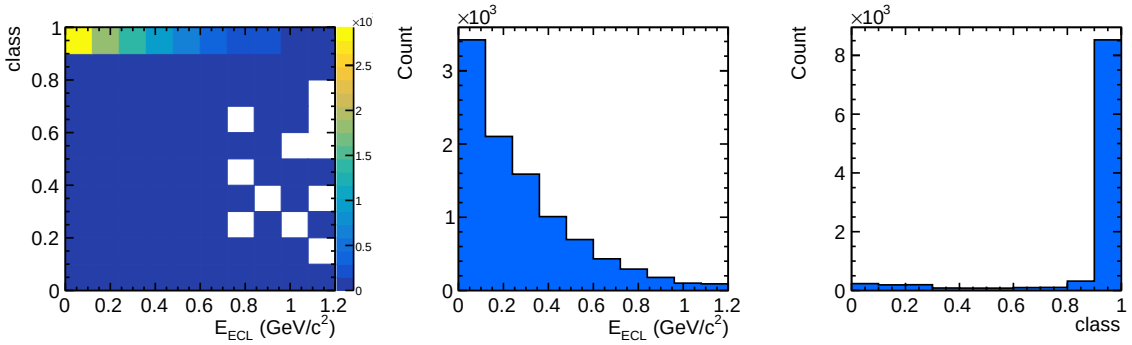
(b) Fake charm PDF component.



(c) Normalization PDF component.

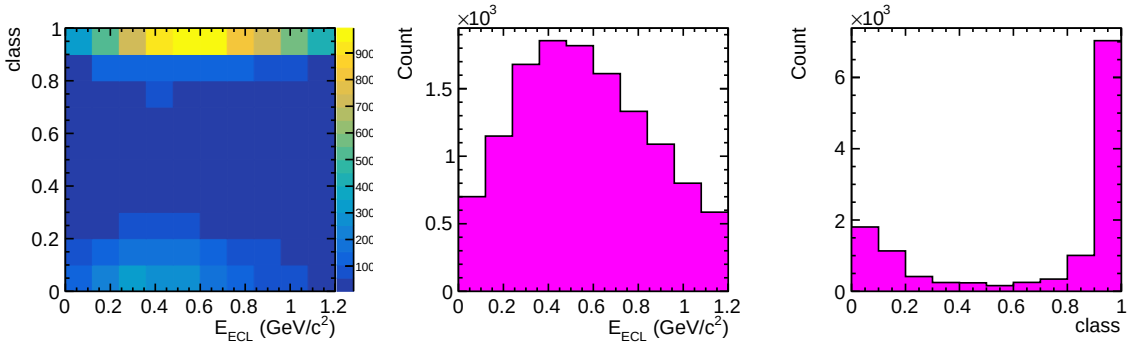


(d) Other PDF component.

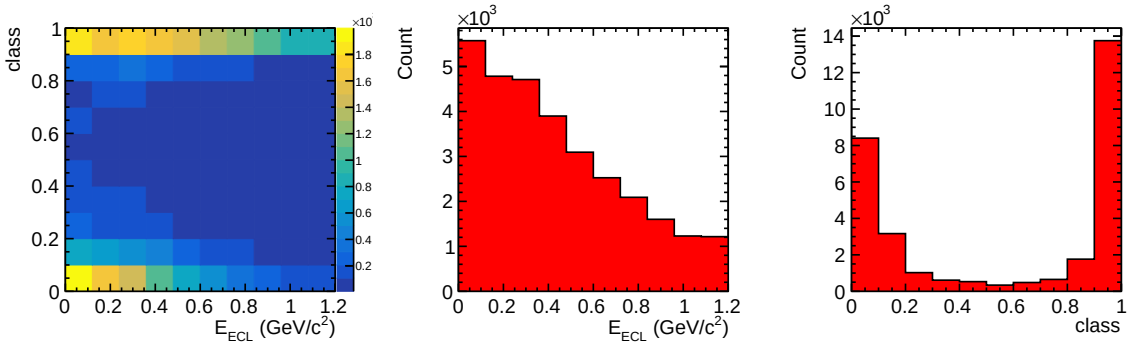


(e) Signal PDF component.

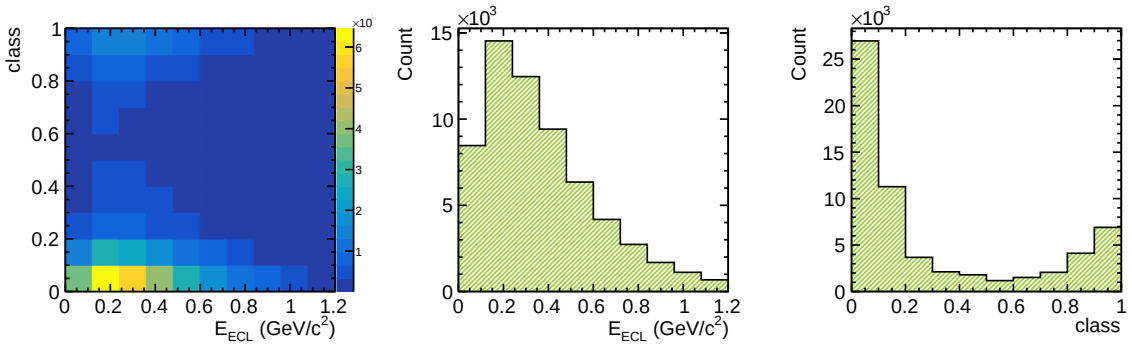
Figure 4.11: Fit PDFs for the sample $D^{*0}\ell^-$.



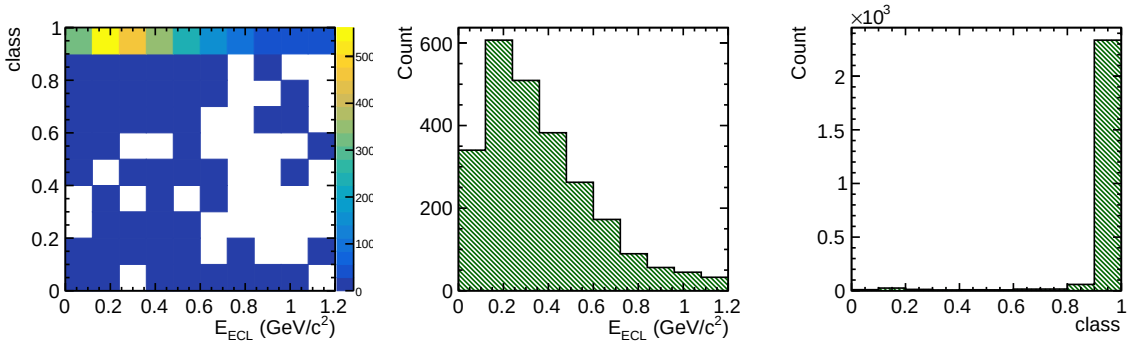
(a) $\bar{B} \to D^{**}\ell^{-}\bar{\nu}_{\ell}$ PDF component.



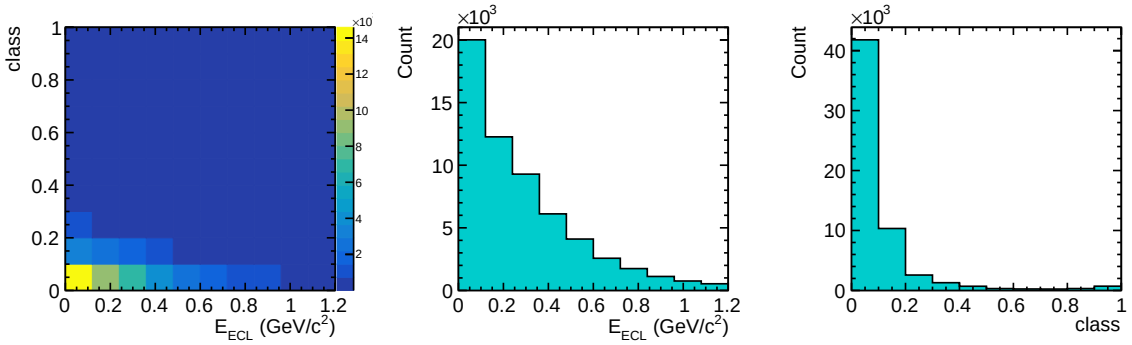
(b) Fake charm PDF component.



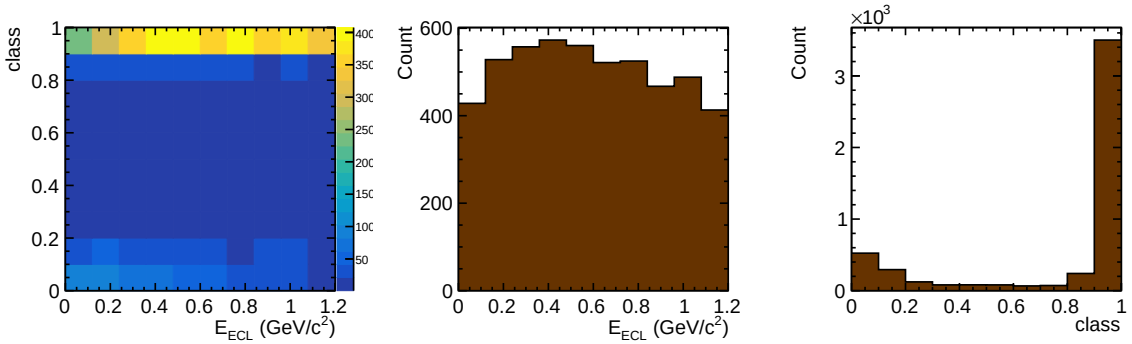
(c) Mixed feed-down, ℓ PDF component.



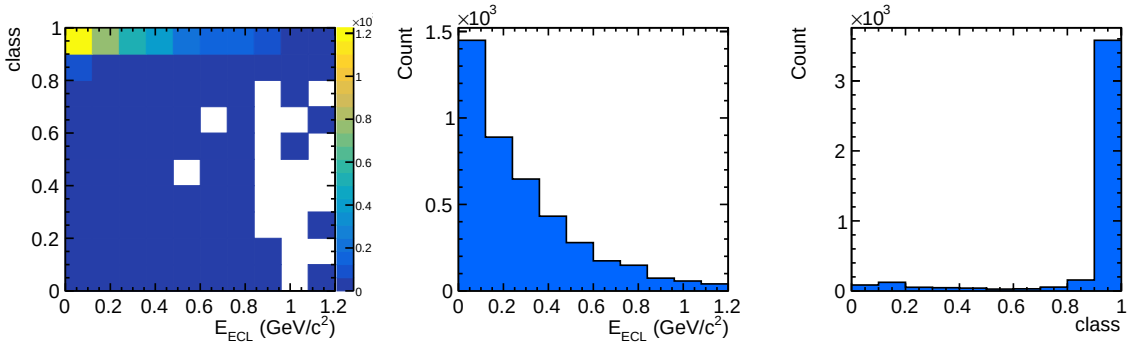
(d) Mixed feed-down, τ PDF component.



(e) Normalization PDF component.

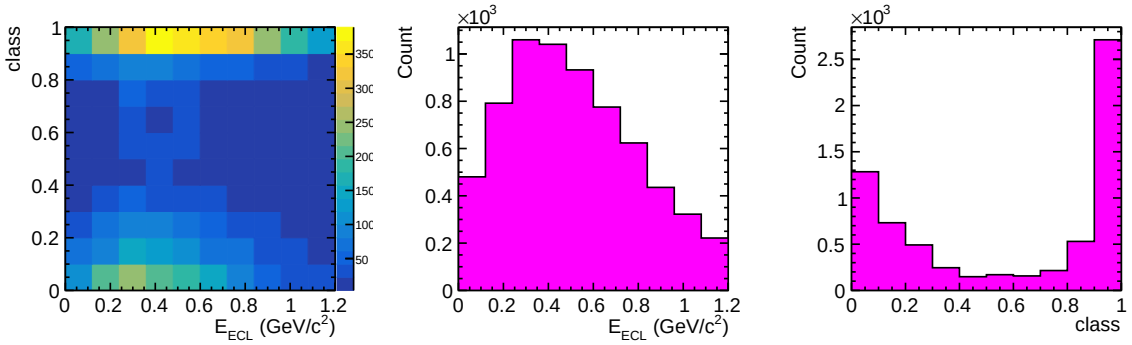


(f) Other PDF component.

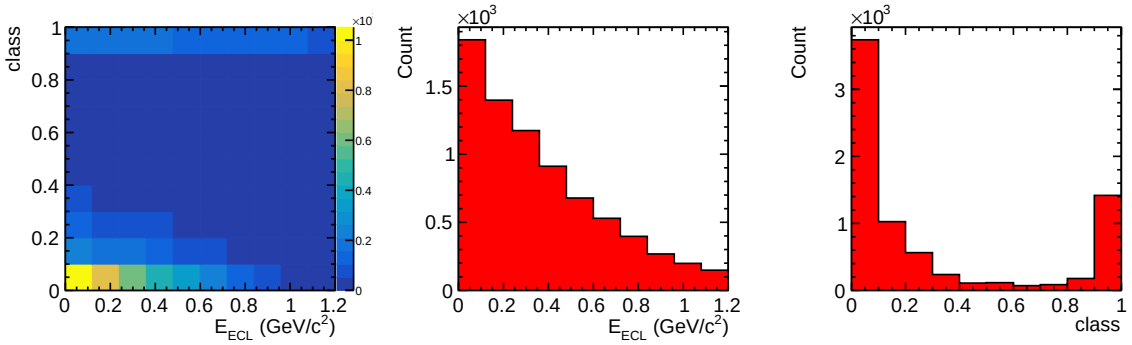


(g) Signal PDF component.

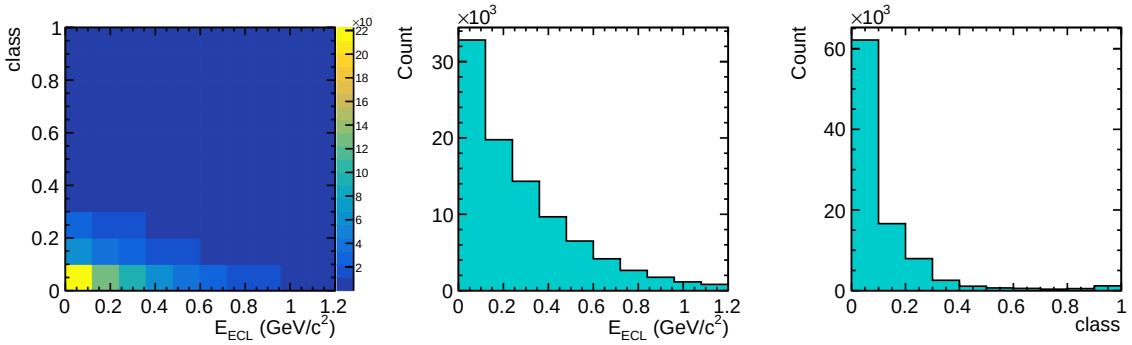
Figure 4.12: Fit PDFs for the sample $D^+\ell^-$.



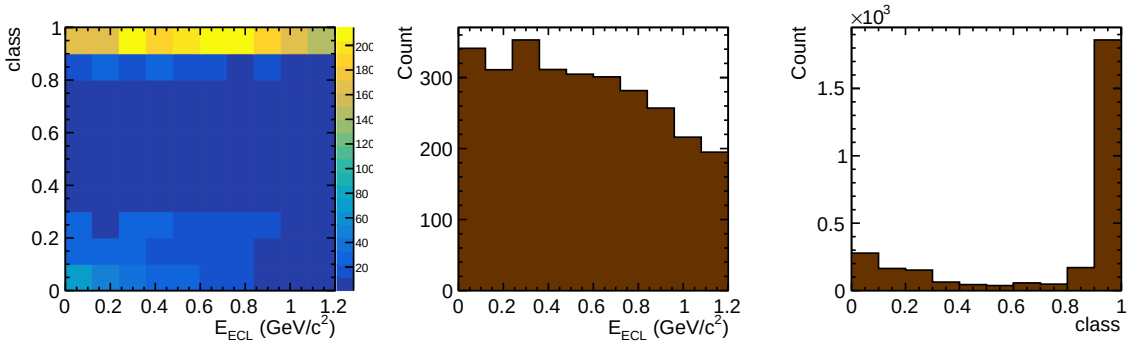
(a) $\bar{B} \to D^{**}\ell^{-}\bar{\nu}_{\ell}$ PDF component.



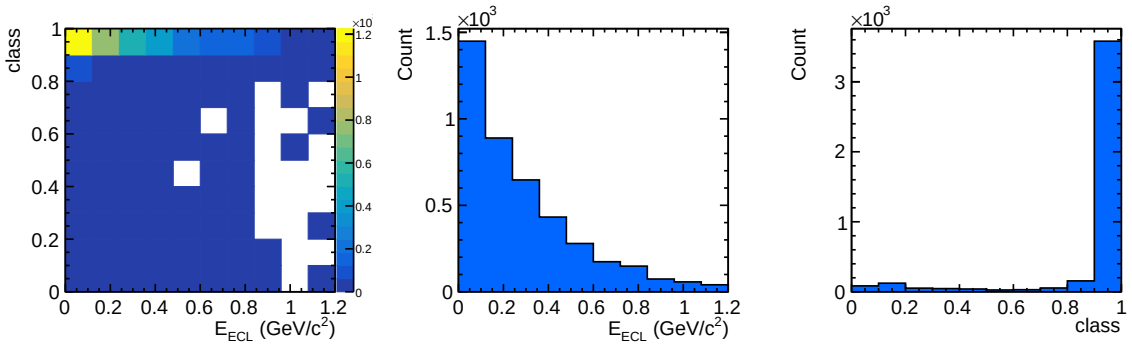
(b) Fake charm PDF component.



(c) Normalization PDF component.



(d) Other PDF component.



(e) Signal PDF component.

Figure 4.13: Fit PDFs for the sample $D^{*+}\ell^-$.

Table 4.1: Expected yields for all fit PDFs for the B^0 samples are given for the whole 2D fit plane, signal region and the complement to the signal region for $E_{\mathrm{ECL}} < 0.5$ GeV. For each column the relative contributions of each cell are given in parantheses. A fit PDF can either freely float (Free), be fixed and not float (Fixed), be correlated to another PDF (Correlated) or be estimated through data sidebands (Sidebands)

PDF component	Type	Whole region	`class` > 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV	`class` < 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV
Signal	Free	419.0 (2.4 %)	296.5 (13.4 %)	51.1 (0.5 %)
Normalization	Free	5850.1 (33.2 %)	34.8 (1.6 %)	4812.2 (46.1 %)
$B \to D^{**}\ell\nu$	Free	1262.2 (7.2 %)	268.9 (12.2 %)	300.5 (2.9 %)
Fake $D^{(*)}$	Sidebands	3158.8 (17.9 %)	817.7 (37.0 %)	1060.3 (10.1 %)
Mixed feed-down (ℓ)	Free	6161.4 (35.0 %)	484.6 (21.9 %)	4127.9 (39.5 %)
Mixed feed-down (τ)	Correlated	249.7 (1.4 %)	175.6 (7.9 %)	12.8 (0.1 %)
Other backgrounds	Fixed	506.2 (2.9 %)	134.1 (6.1 %)	83.8 (0.8 %)
All		17607.4 (100.0 %)	2212.2 (100.0 %)	10448.6 (100.0 %)

(a) Sample: $D^+\ell^-$.

PDF component	Type	Whole region	`class` > 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV	`class` < 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV
Signal	Free	345.0 (3.0 %)	243.0 (41.8 %)	45.5 (0.5 %)
Normalization	Free	9362.1 (82.0 %)	58.9 (10.1 %)	7729.0 (90.9 %)
$B \to D^{**}\ell\nu$	Free	668.4 (5.9 %)	119.5 (20.6 %)	235.9 (2.8 %)
Fake $D^{(*)}$	Sidebands	718.5 (6.3 %)	76.4 (13.1 %)	421.7 (5.0 %)
Other backgrounds	Fixed	323.4 (2.8 %)	83.5 (14.4 %)	73.1 (0.9 %)
All		11417.3 (100.0 %)	581.3 (100.0 %)	8505.4 (100.0 %)

(b) Sample: $D^{*+}\ell^-$.

Table 4.2: Expected yields for all fit PDFs for the B^+ samples are given for the whole 2D fit plane, signal region and the complement to the signal region for $E_{\mathrm{ECL}} < 0.5$ GeV. For each column the relative contributions of each cell are given in parantheses. A fit PDF can either freely float (Free), be fixed and not float (Fixed), be correlated to another PDF (Correlated) or be estimated through data sidebands (Sidebands)

PDF component	Type	Whole region	`class` > 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV	`class` < 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV
Signal	Free	991.0 (1.2 %)	707.1 (7.3 %)	117.8 (0.2 %)
Normalization	Free	13671.0 (17.0 %)	82.0 (0.8 %)	11347.6 (23.6 %)
$B \to D^{**}\ell\nu$	Free	5610.1 (7.0 %)	1367.5 (14.1 %)	1165.8 (2.4 %)
Fake $D^{(*)}$	Sidebands	8650.6 (10.7 %)	1813.6 (18.7 %)	3303.1 (6.9 %)
Mixed feed-down (ℓ)	Free	1215.8 (1.5 %)	102.0 (1.1 %)	914.9 (1.9 %)
Charged feed-down (ℓ)	Free	46260.0 (57.5 %)	3672.5 (37.8 %)	30780.4 (64.0 %)
Mixed feed-down (τ)	Correlated	62.0 (0.1 %)	48.5 (0.5 %)	2.0 (0.0 %)
Charged feed-down (τ)	Correlated	1880.5 (2.3 %)	1321.3 (13.6 %)	89.7 (0.2 %)
Other backgrounds	Fixed	2131.5 (2.6 %)	593.9 (6.1 %)	340.7 (0.7 %)
All		80472.4 (100.0 %)	9708.4 (100.0 %)	48062.0 (100.0 %)

(a) Sample: $D^0\ell^-$.

PDF component	Type	Whole region	`class` > 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV	`class` < 0.9 and $E_{\mathrm{ECL}} < 0.5$ GeV
Signal	Free	288.3 (2.9 %)	203.3 (37.7 %)	38.2 (0.5 %)
Normalization	Free	6962.3 (70.2 %)	46.8 (8.7 %)	5750.2 (79.7 %)
$B \to D^{**}\ell\nu$	Free	505.1 (5.1 %)	90.7 (16.8 %)	170.1 (2.4 %)
Fake $D^{(*)}$	Sidebands	1865.9 (18.8 %)	134.4 (24.9 %)	1178.1 (16.3 %)
Other backgrounds	Fixed	298.2 (3.0 %)	64.3 (11.9 %)	80.0 (1.1 %)
All		9919.9 (100.0 %)	539.6 (100.0 %)	7216.5 (100.0 %)

(b) Sample: $D^{*0}\ell^-$.

Chapter 5

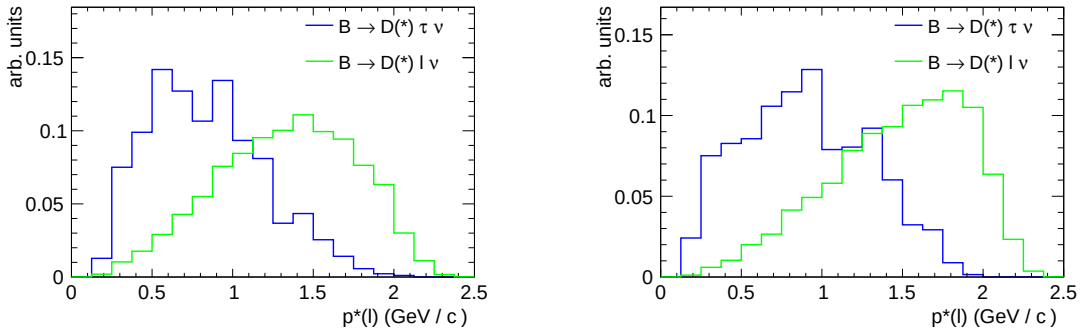
Correction and Validation of Simulated Data

The event selection detailed in the Chapter 3 is based on MC simulations. Moreover, we extract the signal and normalization yields with a fit, taking the Probability Density Functions (PDFs) from these simulations. Hence, it is fundamental to have good agreement between data and MC. Since these MC samples have been produced using theoretical and experimental parameters which may be outdated or inaccurate, we apply a set of corrections to the MC simulation. A list of these corrections is given in this Chapter. We then perform several tests to confirm that the data is well represented by MC samples in the control regions, to confirm the effect of the corrections and in general the good level of agreement between data and MC.

5.1 Lepton Reconstruction

5.1.1 Lepton ID Efficiency

As shown in Figure 5.1, the momentum spectrum of electron and muons varies greatly from signal mode to normalization mode, as the leptons produced by τ decays (as in the signal mode) have a lower average momentum with respect to the leptons that come directly from the B (as in the normalization mode). For this reason any mis-modelling in lepton reconstruction may affect the two modes in different ways, and the overall effect on the final measurement needs to be taken into account. Differences between data and MC are measured using calibration studies and a correction determined in ten bins of track momentum and seven bins of azimuthal angle, both measured in the detector's rest frame [69]. The efficiency to identify a charged track as either an electron or a muon has been studied in low-multiplicity $e^+e^- \to e^+e^-\ell^+\ell^-$ events and average-multiplicity inclusive $e^+e^- \to J/\psi X \to \ell^+\ell^- X$ events. The uncertainties on the correction factors are statistical and systematic, the latter are derived from the differences between low- and average-multiplicity studies, and are a source of sytematic uncertainty for the final result of this analysis.



(a) Sample $D\ell$. (b) Sample $D^*\ell$.

Figure 5.1: The difference in lepton momentum spectrum between the signal mode and the normalization mode is shown in the subfigures, for B decays to both D and D^* mesons.

5.1.2 Lepton Fake Rate

Similar to above, the probability of identifying hadrons as leptons is different in data and MC. To estimate corrections to the MC, a dedicated $D^* \to D\pi$ sample is studied. The lepton mis-identification corrections are obtained for the four combinations of π, K being misreconstructed as e, μ separately.

The correction depends on the track momentum (eleven bins) and azimuthal detector angle (six bins). Due to the four categories of mis-identification ($\pi, K \to e, \mu$), the sample sizes of the fake rate study are low and dominate the uncertainty of the correction factors, whereas the systematic uncertainties are negligible. The correction is applied to all fake lepton candidates and the uncertainties are propagated into the final results of this analysis.

5.2 Form Factors for Semileptonic B-mesons Decays

In the MC samples used for this analysis, $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays are generated with the `HQET2` EvtGen package, based on the CLN parametrization discussed in Section 1.2.1. Since the model's parameters have been updated since the MC generation, we apply an event-by-event reweighting using correction factors obtained by the ratio of the distributions of the new and old model. This reweighting procedure is described in an internal Belle note [70], and is based on 2-dimensional binning in the momentum transfer, q^2, and the lepton momentum in the CM-frame of the decaying B meson, p^*_ℓ.

On the other hand, the $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ decays are generated in the MC samples with the outdated `ISGW2` EvtGen package, based on the quark model described in Ref. [71]. The more accurate LLSW model [72], is used to obtain correction factors

based on the ratio of the analytic predictions of LLSW and MC distributions generated with `ISGW2`. For these decays the reweighting is based on variables such as the recoil $w = m_B^2 + m_{D^*}^2 - q^2/2m_B m_{D^*}$ (where p_B and p_{D^*} are respectively the 4-momentum of B and $D^{(*)}$ mesons) and $\cos\theta_\ell$, the angle between the momenta of D meson and lepton, in the rest frame of the W-boson [70]. The uncertainties on the model parameters are subject of systematic uncertainties, detailed in Chapter 6.

5.3 Branching Ratios and Luminosity

The most recent branching ratios reported by the PDG [10] are used to obtain the updated branching ratios which are then used to correct the B, D, D^* and D^{**} meson decays used in this analysis. Tables 5.1 and 5.2 contain a summary of these corrections. We expect the signal and normalization modes to have approximately the same channel composition for the D and D^* mesons, so that the effect of the uncertainties on these branching ratios carry an effect that cancels out almost completely in the final $\mathcal{R}(D^{(*)})$ measurement.

The $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ background affects signal and normalization modes in a different way, and for this reason the uncertainty on the $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ branching ratios needs to be assessed as a systematic uncertainty of the analysis, as shown in Section 6.3.3.

Table 5.1: Correction factors for the branching ratios of $B \to D^{(*)}\ell\nu$ and $D^{(*)}$ decays.

Channel	$\mathcal{B}_{\rm MC}$	$\mathcal{B}_{\rm PDG}$	Correction
$\bar{B}^0 \to D^{*+} e^- \bar{\nu}_e$	0.053	0.049 ± 0.001	0.925 ± 0.021
$B^- \to D^{*0} e^- \bar{\nu}_e$	0.058	0.057 ± 0.002	0.983 ± 0.033
$\bar{B}^0 \to D^{+} e^- \bar{\nu}_e$	0.023	0.022 ± 0.001	0.948 ± 0.052
$B^- \to D^{0} e^- \bar{\nu}_e$	0.023	0.023 ± 0.001	0.983 ± 0.048
$D^{*+} \to D^0 \pi^+$	0.677	0.677 ± 0.005	1.000 ± 0.007
$D^{*+} \to D^+ \pi^0$	0.307	0.307 ± 0.005	1.000 ± 0.016
$D^{*0} \to D^0 \pi^0$	0.619	0.647 ± 0.009	1.045 ± 0.015
$D^+ \to K^- \pi^+ \pi^+$	0.095	0.090 ± 0.003	0.944 ± 0.029
$D^0 \to K^- \pi^+$	0.038	0.039 ± 0.000	1.018 ± 0.010
$D^0 \to K^- \pi^+ \pi^0$	0.131	0.142 ± 0.005	1.086 ± 0.038
$D^0 \to K^- \pi^+ \pi^+ \pi^-$	0.071	0.081 ± 0.002	1.144 ± 0.021

The number of events present in the MC samples differs from the number of $B\bar{B}$ pairs recorded in the Belle data sample, and the generic MC on average misses 2.4% of generated luminosity compared to the collision data. The corresponding correction is

Table 5.2: Correction factors for the branching ratios of $B \to D^{**}\ell\nu$ decays. The label A refers to experiment numbers 7 to 27 and the label B for experiment numbers 31 to 65.

B channel	D^{**} channel	$\mathcal{B}_{\mathrm{MC}}(B)$ (A)	$\mathcal{B}_{\mathrm{MC}}(B)$ (B)	$\mathcal{B}_{\mathrm{MC}}(D^{**})$	$\mathcal{B}_{\mathrm{PDG}}(B \to D^{**} \to X)$	Correction (A)	Correction (B)
$\bar{B^0} \to D_1^+ e^- \bar{\nu}_e$	$D^{*0}\pi^-$	0.006	0.007	0.667	0.00280 ± 0.00028	0.69 ± 0.06	0.56 ± 0.05
$\bar{B^0} \to D_0^{*+} e^- \bar{\nu}_e$	$D^0\pi^-$	0.008	0.008	0.667	0.00300 ± 0.00120	0.57 ± 0.23	0.53 ± 0.21
$\bar{B^0} \to D_1'^+ e^- \bar{\nu}_e$	$D^{*0}\pi^-$	0.012	0.007	0.667	0.00310 ± 0.00090	0.38 ± 0.11	0.62 ± 0.18
$\bar{B^0} \to D_2^{*+} e^- \bar{\nu}_e$	$D^{*0}\pi^-$	0.002	0.004	0.459	0.00068 ± 0.00012	0.67 ± 0.11	0.41 ± 0.07
$B^- \to D_1^0 e^- \bar{\nu}_e$	$D^{*+}\pi^-$	0.007	0.008	0.667	0.00303 ± 0.00020	0.68 ± 0.04	0.56 ± 0.03
$B^- \to D_0^{*0} e^- \bar{\nu}_e$	$D^+\pi^-$	0.009	0.009	0.667	0.00250 ± 0.00050	0.44 ± 0.08	0.41 ± 0.08
$B^- \to D_1'^0 e^- \bar{\nu}_e$	$D^{*+}\pi^-$	0.013	0.008	0.667	0.00270 ± 0.00060	0.31 ± 0.06	0.49 ± 0.11
$B^- \to D_2^{*0} e^- \bar{\nu}_e$	$D^+\pi^-$	0.002	0.004	0.459	0.00153 ± 0.00016	1.38 ± 0.14	0.85 ± 0.08

based on the experiment number and the type of $B\bar{B}$ pair that is being reconstructed (neutral or charged), and the correction factors are displayed in Figure 5.2.

Using the lmost recent world average values from the PDG [10] we also update the branching ratio for the decay $\Upsilon(4S) \rightarrow B^0\bar{B}^0(B^+B^-)$ to 48.6% (51.4%), since in the MC generation it was set to 50% for both decays.

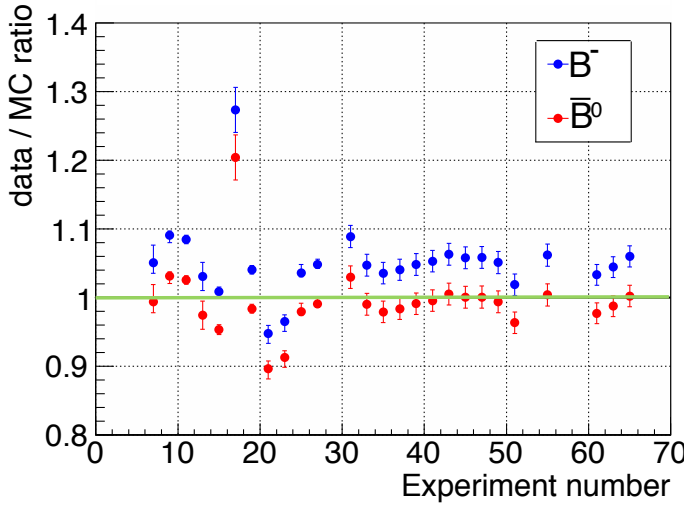


Figure 5.2: Correction factors for the number of $B\bar{B}$ pairs as a function of the experiment number for the experimental data and MC samples. High correction values are caused by MC samples that are no longer available.

5.4 Fake $D^{(*)}$ -meson Calibration

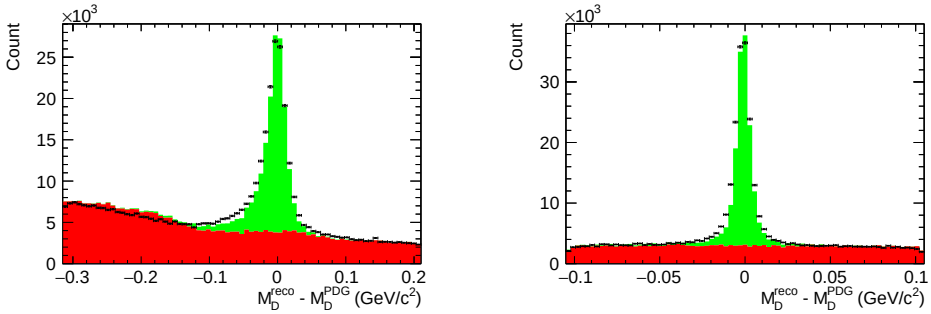
One of the main background sources for this analysis are the B_{sig} candidates that contain fake charm mesons, that is a D or D^* -meson that is incorrectly reconstructed.

As explained in Chapter 3, correctly reconstructed $D^{(*)}$ mesons will peak for the distribution of the variable $M_{\text{reco}}(D^{(*)}) - M_{PDG}(D^{(*)})$. For these variables, the combinatorial background will lie in a continuous band that stretches along the full variable axis, and we define the region away from the signal peak, as *sidebands*. The comparison of data and MC for the sidebands for all charm channels are shown in Figures 5.3, 5.4 and 5.5 for D^0, D^+ and D^* mesons respectively.

To calibrate the amount of fake $D^{(*)}$ events in our model, we compare a sideband region of the data sample, which contains only fake charm background events, to the equivalent region in the MC samples. The comparison is performed with respect to two variables, E_{ECL} and BDT classifier output, which define the 2D fit plane for the signal extraction and are introduced in Chapter 4.

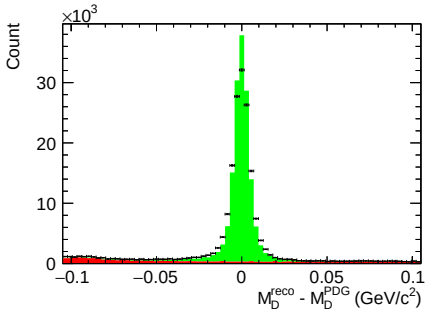
We follow the following procedure:

1. Select events in the sidebands of the mass distributions;

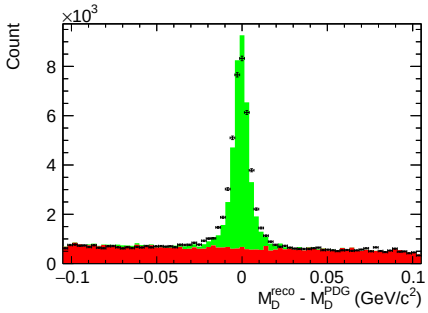


(a) $D^0 \to K^- \pi^+ \pi^0$

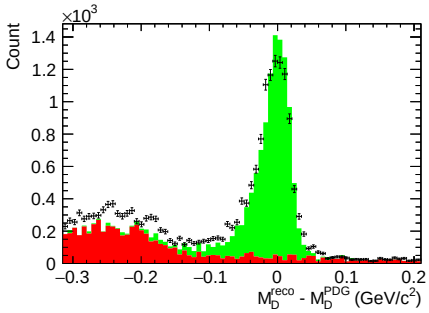
(b) $D^0 \to K^- \pi^+ \pi^+ \pi^-$



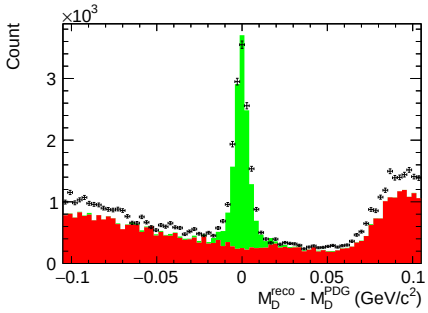
(c) $D^0 \to K^- \pi^+$



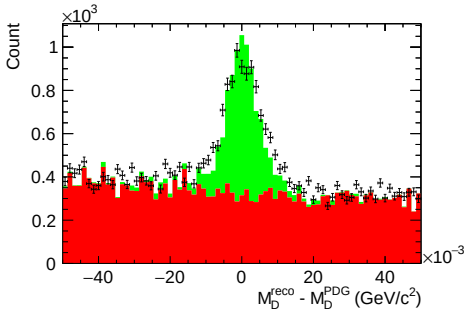
(d) $D^0 \to K^0_S \pi^+ \pi^-$



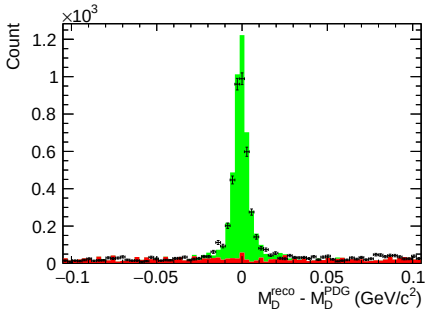
(e) $D^0 \to K^0_S \pi^0$



(f) $D^0 \to K^- K^+$

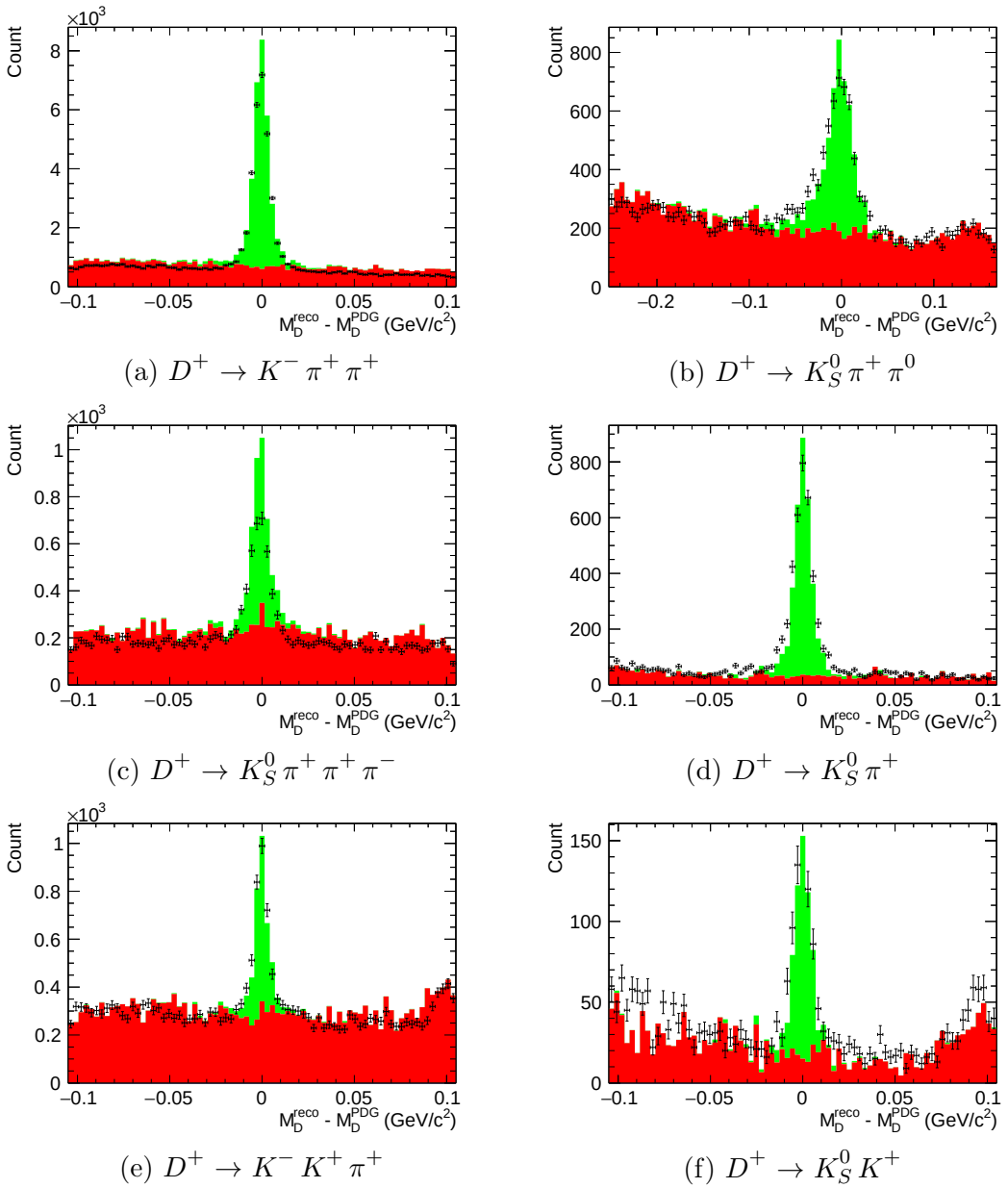


(g) $D^0 \to \pi^- \pi^+$



(h) $D^0 \to K^0_S K^+ K^-$

Figure 5.3: Difference between the reconstructed D^0-meson invariant mass and its PDG value. Correctly reconstructed candidates are shown in green, and mis-reconstructed candidates in red. Experimental data points are overlayed on the MC distribution.

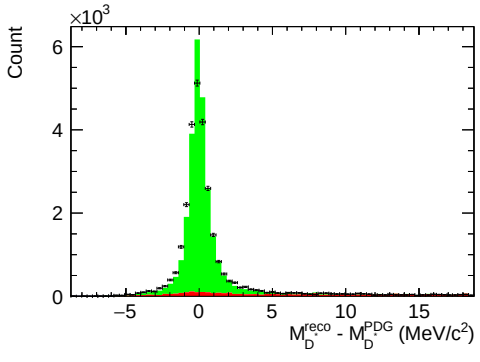


(a) $D^+ \to K^- \pi^+ \pi^+$ (b) $D^+ \to K^0_S \pi^+ \pi^0$

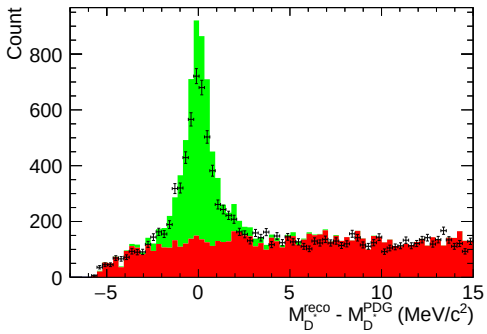
(c) $D^+ \to K^0_S \pi^+ \pi^+ \pi^-$ (d) $D^+ \to K^0_S \pi^+$

(e) $D^+ \to K^- K^+ \pi^+$ (f) $D^+ \to K^0_S K^+$

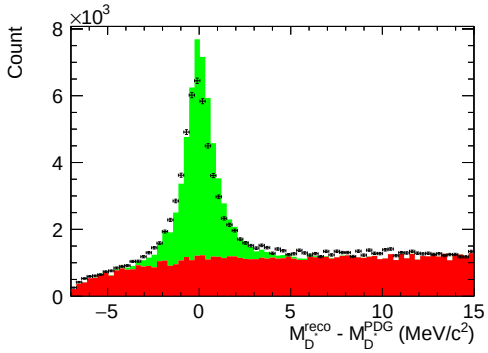
Figure 5.4: Difference between the reconstructed D^+-meson invariant mass and its PDG value. Correctly reconstructed candidates are shown in green, and mis-reconstructed candidates in red. Experimental data points are overlayed on the MC distribution.



(a) $D^{*+} \to D^0 \pi^+$



(b) $D^{*+} \to D^+ \pi^0$



(c) $D^{*0} \to D^0 \pi^0$

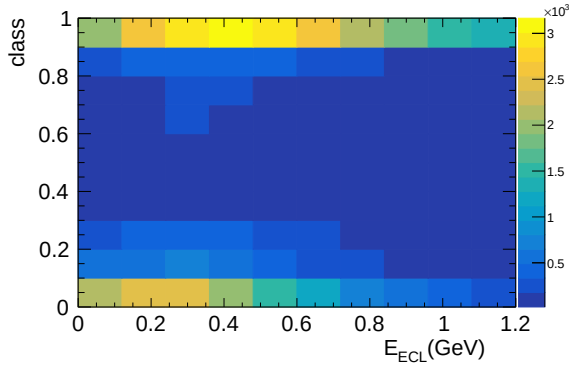
Figure 5.5: Difference between the reconstructed D^*-meson invariant mass and its PDG value. Correctly reconstructed candidates are shown in green, and mis-reconstructed candidates in red. Experimental data points are overlayed on the MC distribution.

2. Fill 2D histograms E_{ECL} -classifier using these events for the data and MC samples;

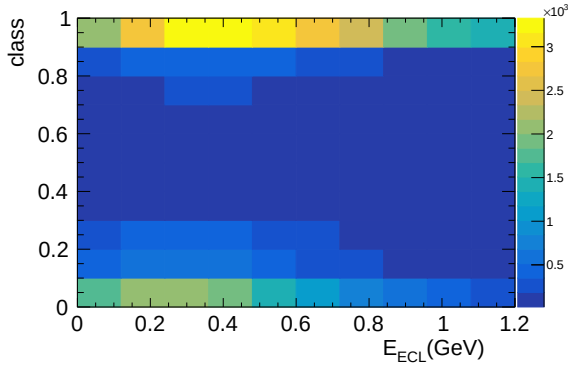
3. Divide the 2D histogram for the data sample by the 2D histogram for the MC sample;

4. For each bin of the 2D histogram extract a calibration factor $r_{\text{bin}} = \frac{N_{\text{bin, data}}}{N_{\text{bin, MC}}}$.

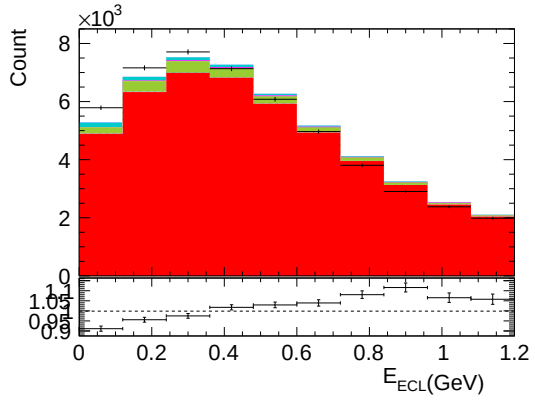
The results of the calibration procedure are shown in Figures 5.6 to 5.9. Given that the $B \to D^*$ channels contain a lower number of events than the $B \to D$ channels, the systematic uncertainty caused by this calibration will be larger for the former channels.



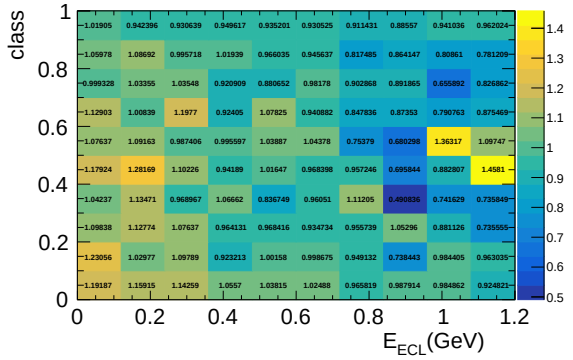
(a) Data



(b) MC

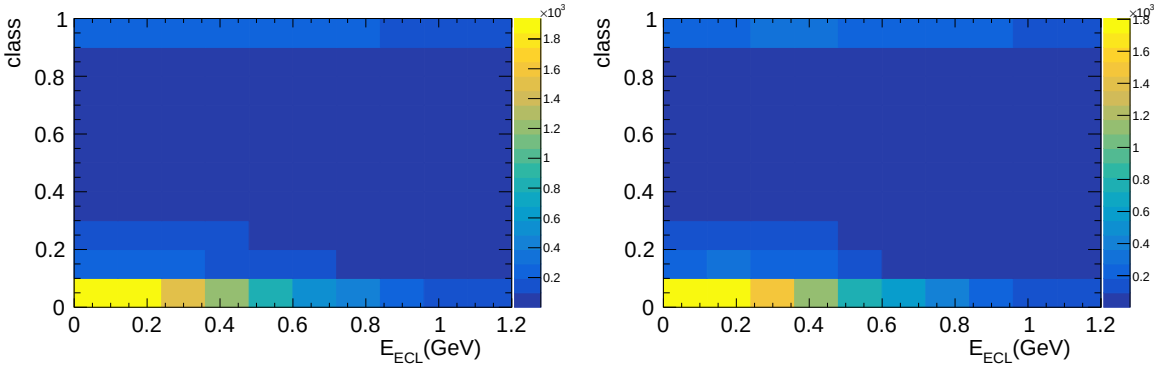


(c) E_{ECL} - projection of data and MC.



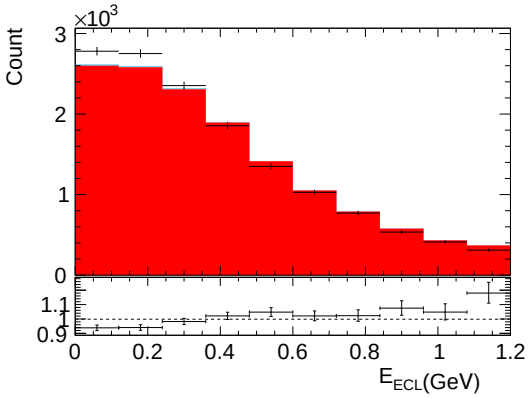
(d) Ratio of data and MC.

Figure 5.6: Fake $D^{(*)}$ calibration for sample: $D^0\ell^-$. The output of the BDT classifier (introduced in Chapter 4) is denoted as `class`.

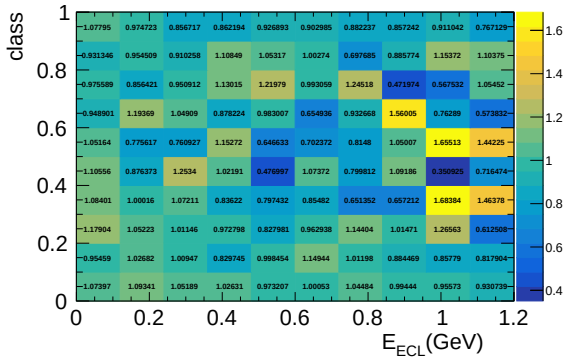


(a) Data

(b) MC

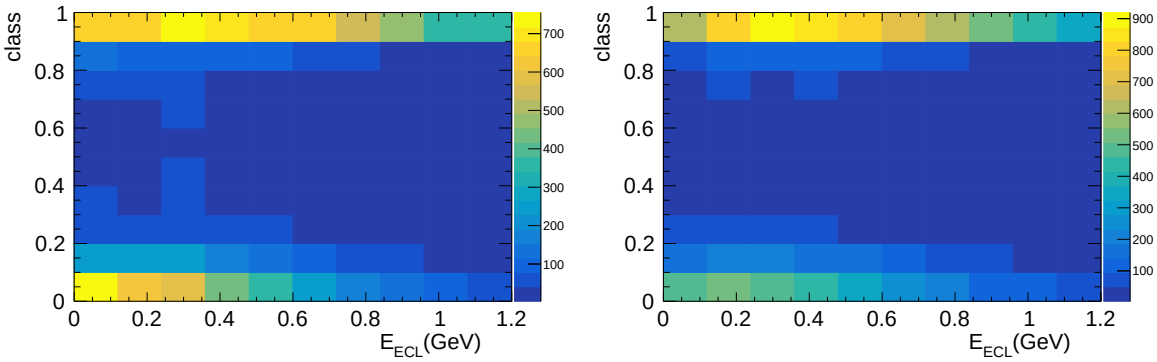


(c) E_{ECL} - projection of data and MC.



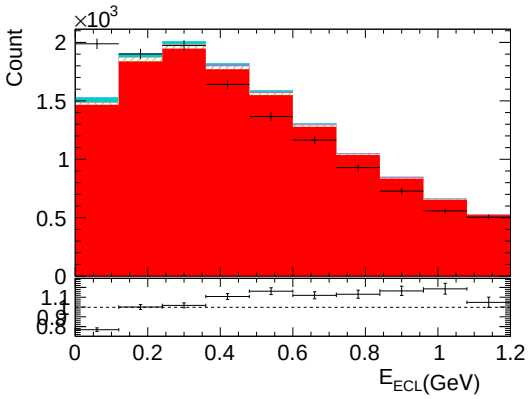
(d) Ratio of data and MC.

Figure 5.7: Fake $D^{(*)}$ calibration for sample: $D^{*0}\ell^{-}$. The output of the BDT classifier (introduced in Chapter 4) is denoted as `class`.

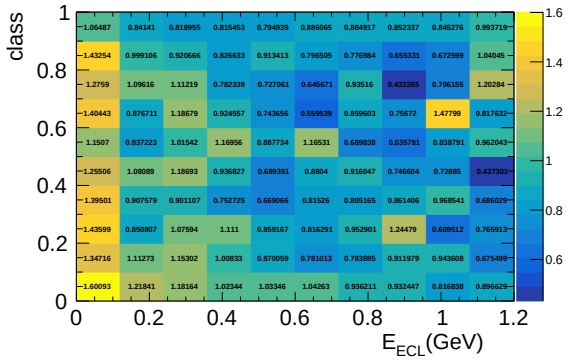


(a) Data

(b) MC

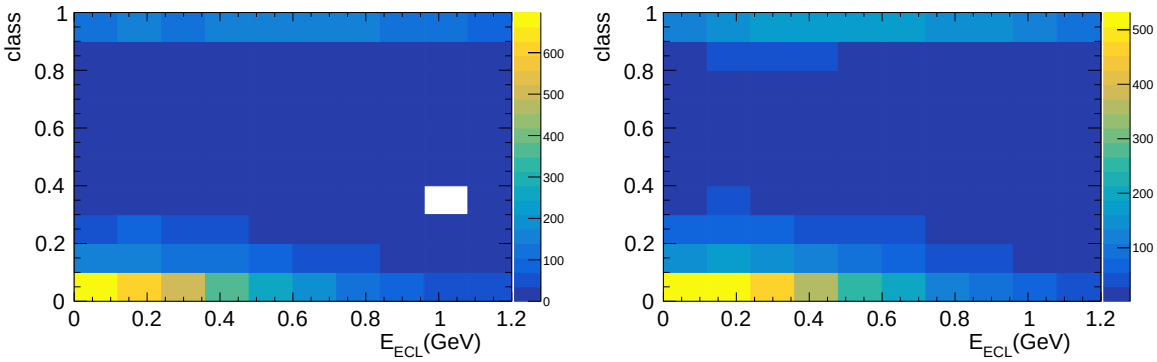


(c) E_{ECL} - projection of data and MC.



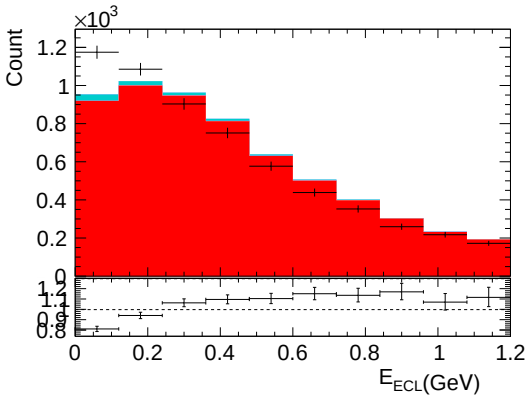
(d) Ratio of data and MC.

Figure 5.8: Fake $D^{(*)}$ calibration for sample: $D^{+}\ell^{-}$. The output of the BDT classifier (introduced in Chapter 4) is denoted as `class`.

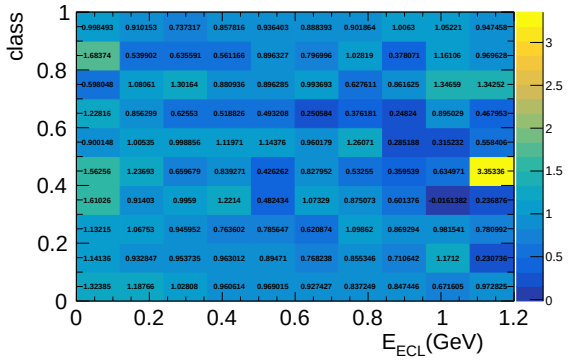


(a) Data

(b) MC



(c) E_{ECL} - projection of data and MC.



(d) Ratio of data and MC.

Figure 5.9: Fake $D^{(*)}$ calibration for sample: $D^{*+}\ell^{-}$. The output of the BDT classifier (introduced in Chapter 4) is denoted as `class`.

5.5 *B*-tag Calibration

The number of B_{tag} reconstructed by the tagging algorithm depends on the modelling of its input variables in the MC samples, which differs from the measured distributions found in the data samples. As a consequence, we find a discrepancy not only between the number of B_{tag} candidates in the MC samples and in the data samples, but also in their classifier output shape and thus we apply a B_{tag} calibration. This calibration will affect the expected yields of the B_{sig} candidates that accompany the B_{tag} .

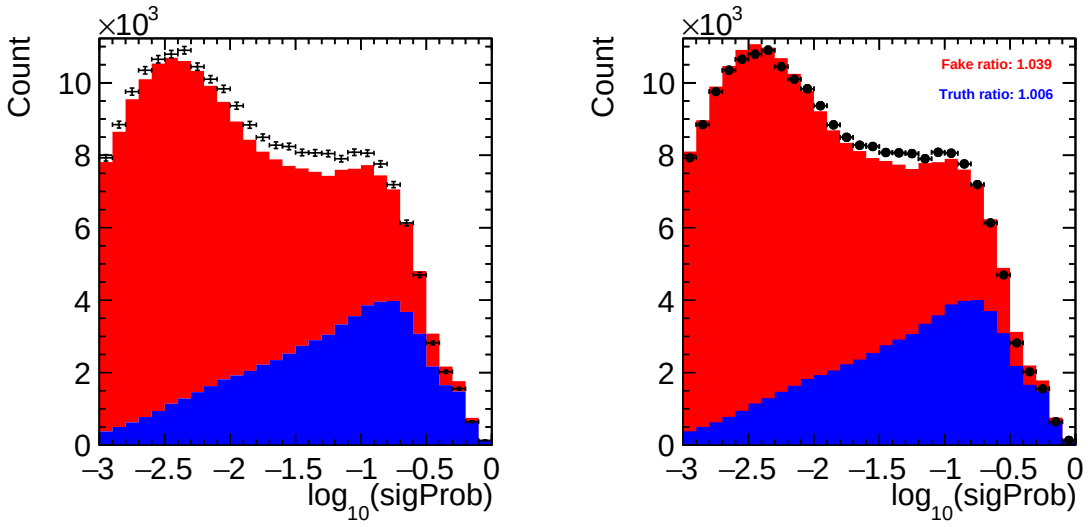
The calibration procedure we follow is to use MC simulations to build *fake tag* (mis-reconstructed B_{tag}) and *true tag* (correctly reconstructed B_{tag}) probability distribution functions (PDFs) for the classifier output distributions, and let the size of these PDF float in a fit to the same distribution measured in the data sample. The scale factor is defined as

$$s = N_i^{\text{fit}}/N_i^{\text{exp}} \tag{5.1}$$

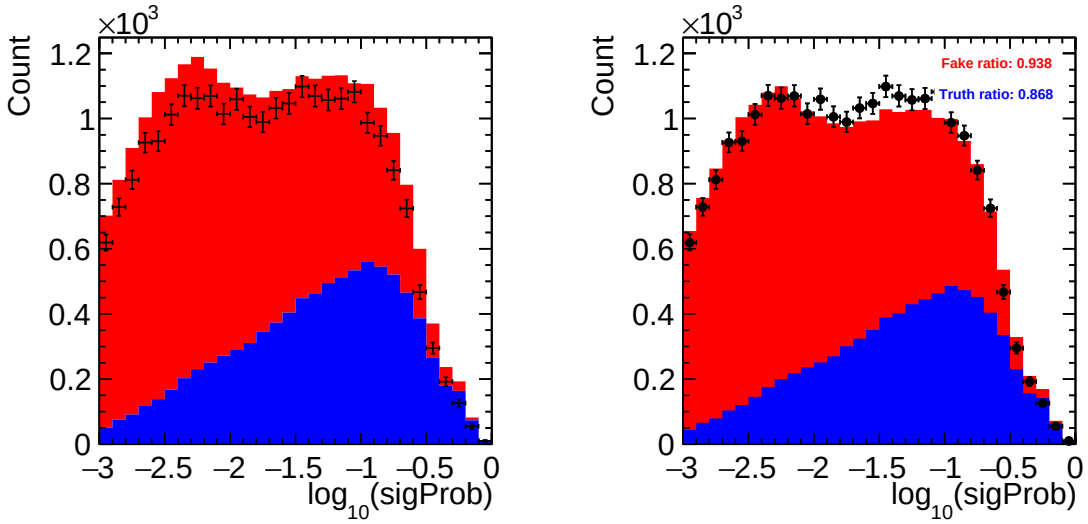
where the index i refers to fake and truth PDFs, N^{fit} is the number of events in the PDF as determined by the fit, and N^{exp} is the number of events of events in the PDF before the fit, is used to re-weight the relevant events in the PDFs. The tag calibration procedure is shown visually in Figure 5.10 and the numerical results are summarized in Table 5.3.

Table 5.3: Numerical values for the B_{tag} calibration for all signal channels.

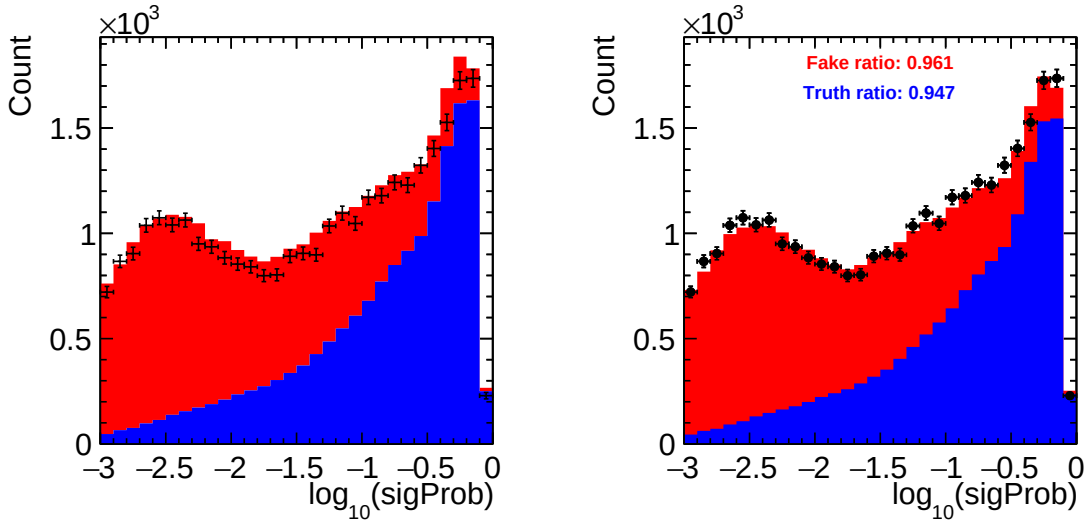
B_{tag} channel	Fake B_{tag} ratio	True B_{tag} ratio
$B^- \to D^0 \ell^- \nu$	1.04 ± 0.01	1.00 ± 0.01
$B^- \to D^{*0} \ell^- \nu$	0.94 ± 0.01	0.87 ± 0.02
$B^0 \to D^+ \ell^- \nu$	0.96 ± 0.01	0.95 ± 0.01
$B^0 \to D^{*+} \ell^- \nu$	0.93 ± 0.02	1.02 ± 0.02



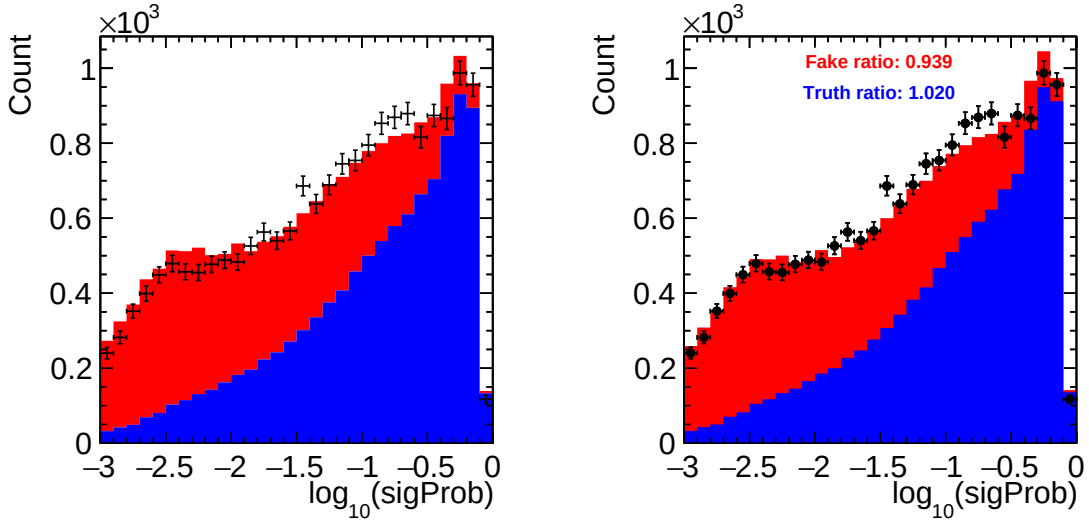
(a) Sample: $D^0\ell^-$.



(b) Sample: $D^{*0}\ell^-$.



(c) Sample: $D^{+}\ell^{-}$.



(d) Sample: $D^{*+}\ell^{-}$.

Figure 5.10: Pre-fit (left) and post-fit plots (right) used to determine the tag calibration factors are shown for all samples. The classifier output of the tagging algorithm is denoted by `sigProb`.

5.6 Normalization Control Region Study

In order to assess the quality of the agreement between data and MC samples, without biasing the signal extraction procedure, we define a normalization control region by requiring events with a value of the classifier smaller than 0.2. Given that the normalization events $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ are more likely to be classified with values close to 0, and the signal events $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ with values close to 1, this requirement effectively defines a signal-depleted sample.

Our goal in this section is to extract from this control sample the ratios of branching ratios

$$R^{+/0}_{\text{ctrl}} = \frac{\mathcal{B}\left(B^{+/0} \to D^*\ell^-\bar{\nu}_\ell\right)}{\mathcal{B}\left(B^{+/0} \to D\ell^-\bar{\nu}_\ell\right)}. \tag{5.2}$$

which does not depend on the yields of the signal region. We do so by setting up a fit procedure where the only PDFs that are floating are the normalization, feed-down (ℓ) and $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ components.

We obtain the results summarized in Tables 5.5, where the quoted uncertainties refer only to statistical ones, while the expectations are extracted from the HFLAV averaging group [40]. The branching ratios used for the calculation of the expectation values are shown in Table 5.4. The projections for the fit results are shown in pre-fit and post-fit in figures from 5.11 to 5.14.

We obtain a good agreement between measured values of the ratios $R^{+/0}_{\text{ctrl}}$, and their world-average values. It is important to notice that, as our results show, the expected values of these ratios should be the same, according to isospin symmetry. Moreover, we notice that, in most cases, the fitted yields of the $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ component obtain low values but are always compatible with 0 within uncertainties.

Table 5.4: Branching ratios used for the calculation of the expectation values for the results of the normalization control region study [40].

Channel	Branching ratio
$B^- \to D^0\ell^-\nu$	$2.33 \pm 0.04 \pm 0.09$
$B^- \to D^{*0}\ell^-\nu$	$5.59 \pm 0.02 \pm 0.19$
$\bar{B}^0 \to D^+\ell^-\nu$	$2.20 \pm 0.04 \pm 0.09$
$\bar{B}^0 \to D^{*+}\ell^-\nu$	$4.88 \pm 0.01 \pm 0.10$

Table 5.5: Fit results for the normalization control region test are compared to expectations from the HFLAV averaging group [40].

	Fit result	Expectaction
R^{+}_{ctrl}	2.53 ± 0.11	2.39 ± 0.06
R^{0}_{ctrl}	2.49 ± 0.10	2.29 ± 0.06

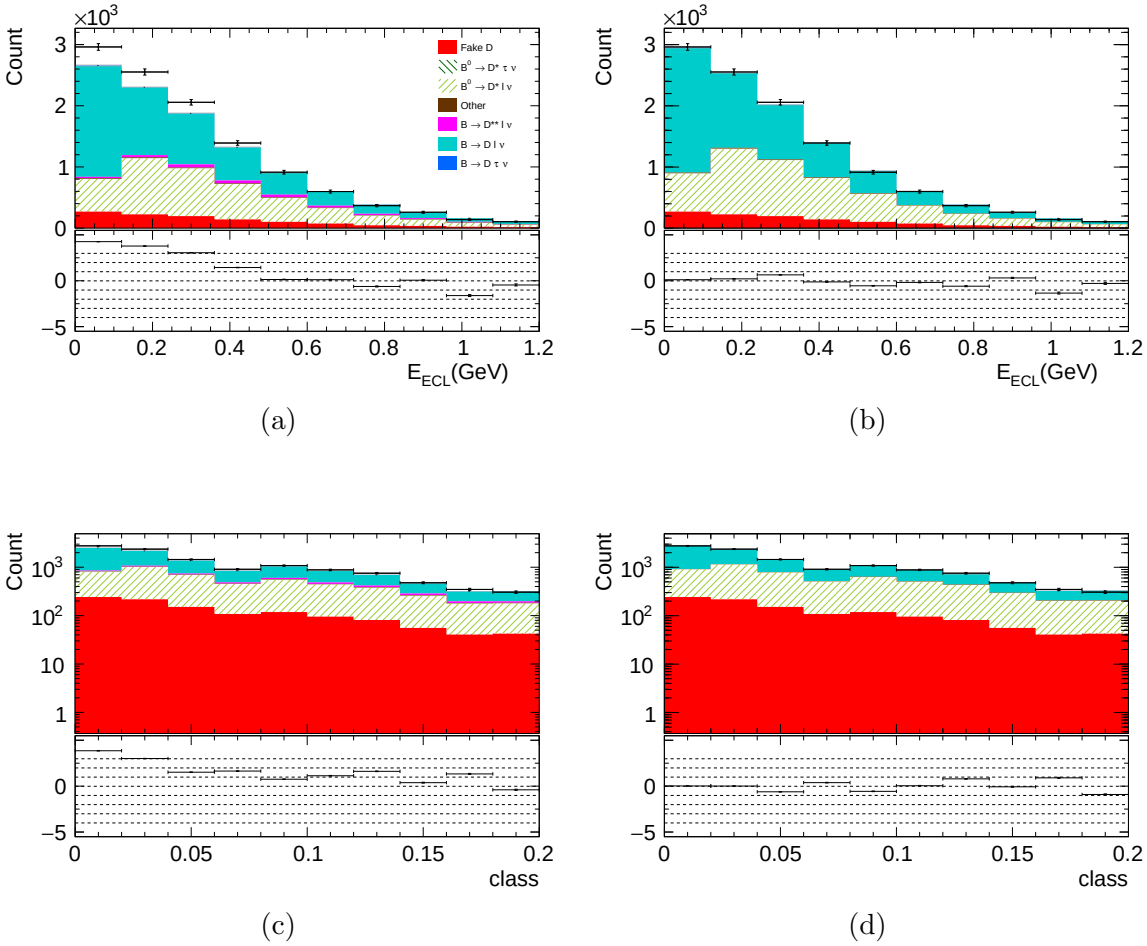


Figure 5.11: Projections of the fit results for the normalization control region test for the channel $B^0 \to D^+ \ell^- \nu$. Pre-fit projections are shown on the left and post-fit projections are shown on the right. Figures (a) and (b) show the E_{ECL} projection , figures (c) and (d) show the classifier projection .

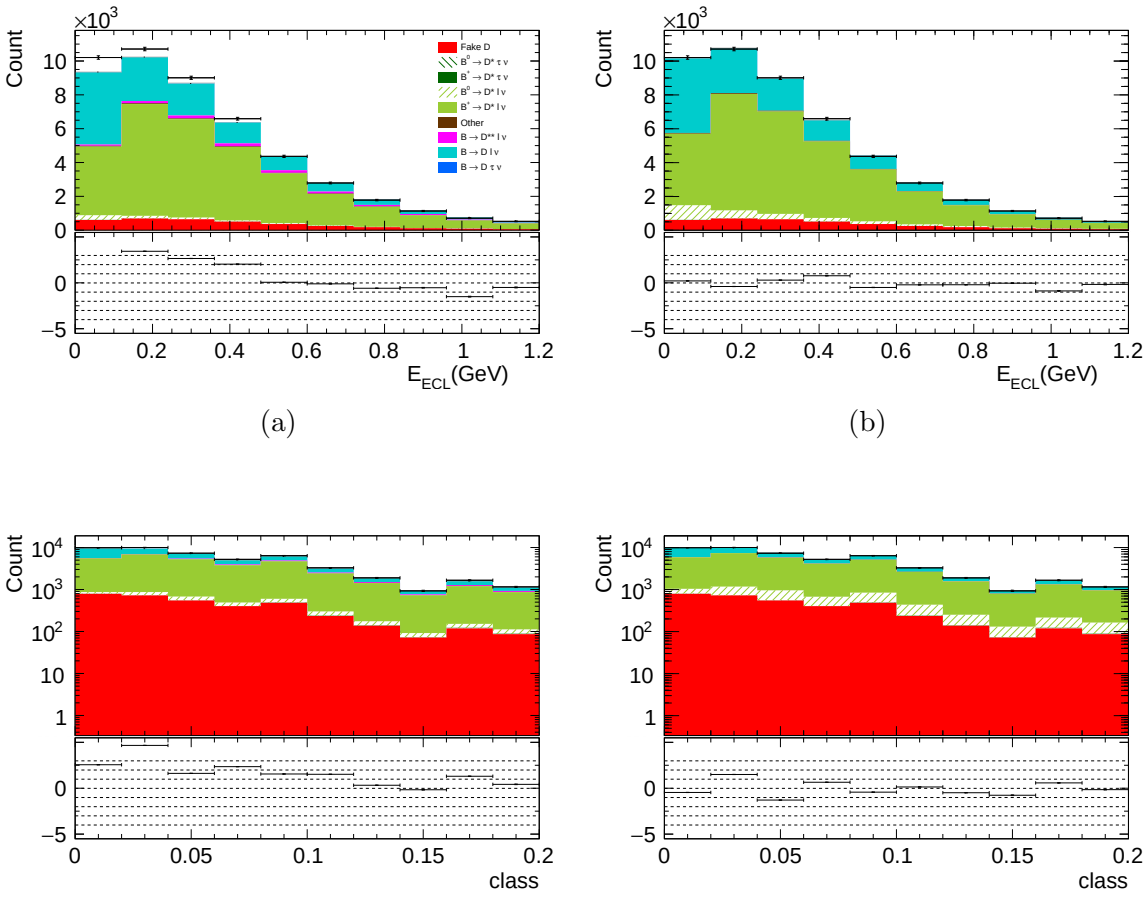


Figure 5.12: Projections of the fit results for the normalization control region test for the channel $B^- \to D^0 \ell^- \nu$. Pre-fit projections are shown on the left and post-fit projections are shown on the right. Figures (a) and (b) show the E_{ECL} projection , figures (c) and (d) show the classifier projection .

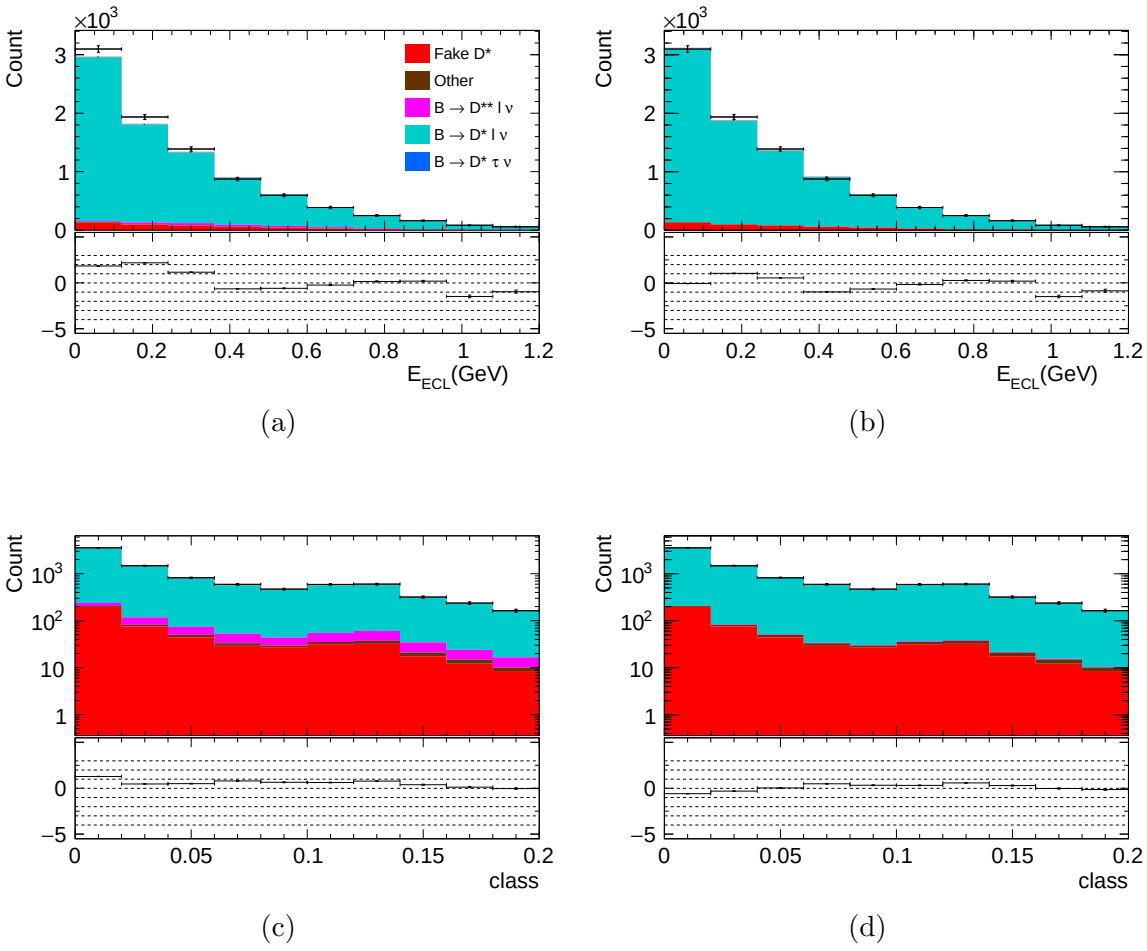
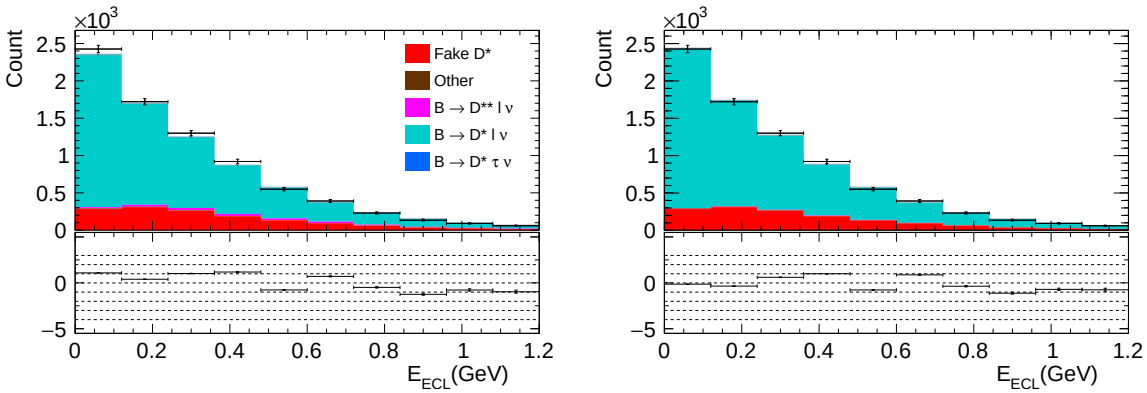
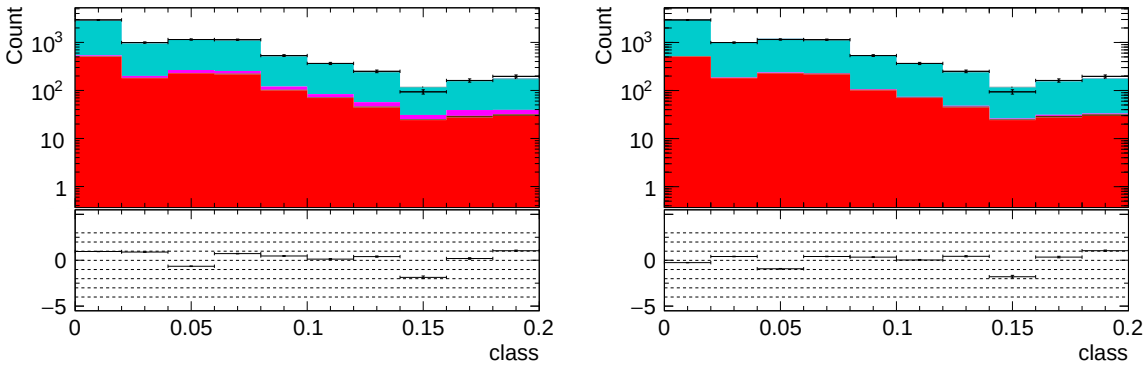


Figure 5.13: Projections of the fit results for the normalization control region test for the channel $B^0 \to D^{*+}\ell^-\nu$. Pre-fit projections are shown on the left and post-fit projections are shown on the right. Figures (a) and (b) show the $E_{\rm ECL}$ projection , figures (c) and (d) show the classifier projection .



(a) (b)



(c) (d)

Figure 5.14: Projections of the fit results for the normalization control region test for the channel $B^- \to D^{*0}\ell^-\nu$. Pre-fit projections are shown on the left and post-fit projections are shown on the right. Figures (a) and (b) show the E_{ECL} projection , figures (c) and (d) show the classifier projection .

5.7 D^{**} Control Region Study

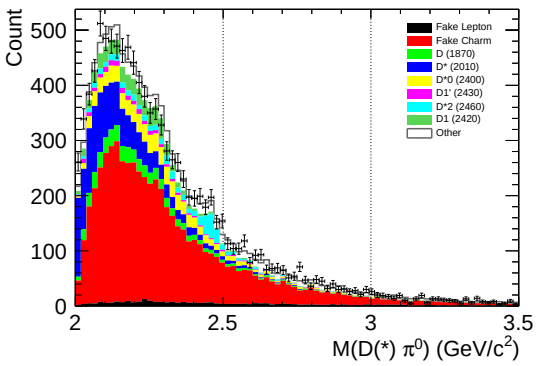
In order to assess the usability of our $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ PDF, we define a D^{**} control region by adding a π^0 in the reconstruction of our nominal signal channels, i.e. we reconstruct $\bar{B} \to D^{(*)}\ell^-\pi^0\bar{\nu}_\ell$. We do not change any of the selection criteria for the signal and tag reconstruction and we apply a tag calibration based on this sample, which differs from the one used for the signal extraction sample.

We then compare the MC expectations provided by the dedicated D^{**} sample to the experimental data, by plotting the comparisons of E_{ECL} distribution and of the invariant mass of the D^{**} meson, which is defined by the sum of the charm meson and the π^0 four momenta, shown in Figure 5.15 to 5.18.

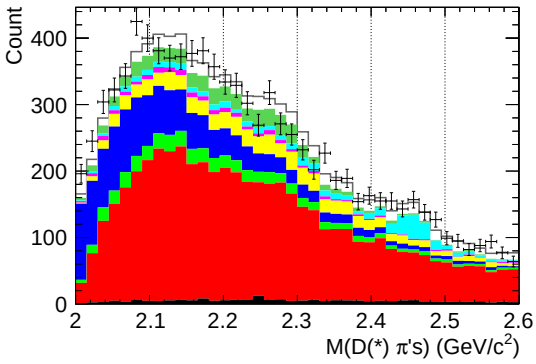
We observe good agreement between MC and data samples, and conclude that our $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ PDF can be used for the signal extraction as it is. Notably, in the invariant mass distributions we can clearly see the peaks of the D_1 and D_2^* states, which have small decay widths. We will assess the impact of the uncertainty of the D^{**} PDF composition in the evaluation of the systematics uncertainty, as explained in Section 6.3.3.

We further study the $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ decays by listing the D^{**} decay modes for each D^{**} type, for each signal channel, shown in Tables 5.9 to 5.21. In the tables we only show decay modes with a relative contribution larger than 2%, and the + superscript refers to both positively and negatively charged particles. This study is performed with the dedicated D^{**} MC samples used for the nominal signal study, and not on the π^0 enriched sample.

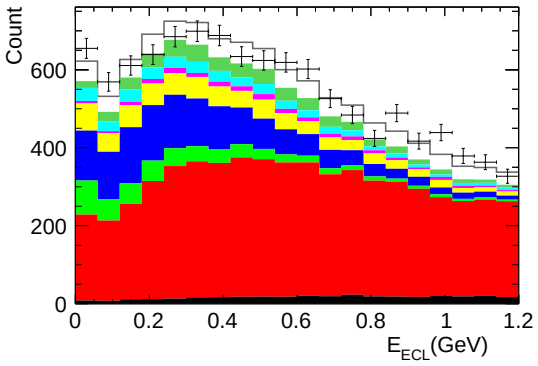
As expected the greatest contributions for D^{**} decay modes comes from missing a π^0. We only notice a significant contribution from D^{**} decay modes with charge that is not compatible with the charge of the expected charm particles in the signal mode. This contribution comes from fake tags with a wrong charge, and might suffer from mismodeling. For this reason we will change the relative contributions to the D^{**} decay modes and verify the effect of this change on the final result.



(a)



(b)



(c)

Figure 5.15: Plots for the D^{**} control region for the sample $D^{+}\pi^{0}\ell^{-}$. The invariant mass of the D^{**} is shown in figure (a), a zoomed in version of the same distribution is shown in figure (b), while figure (c) shows the E_{ECL} distribution.

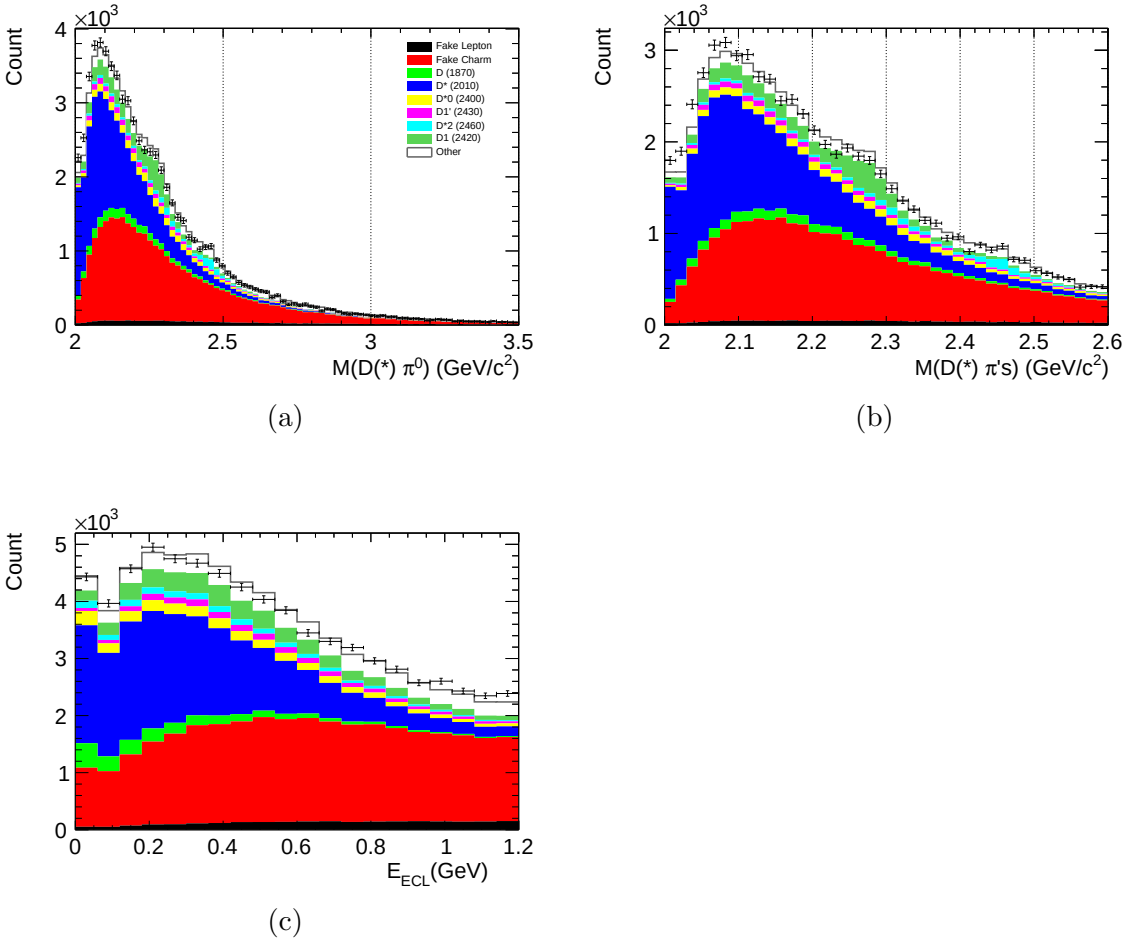
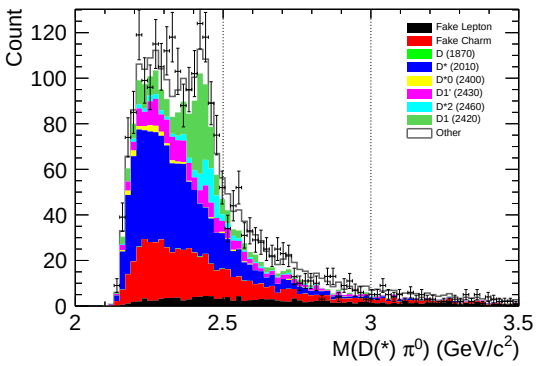
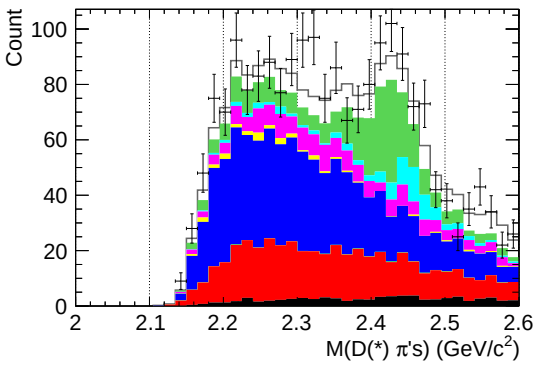


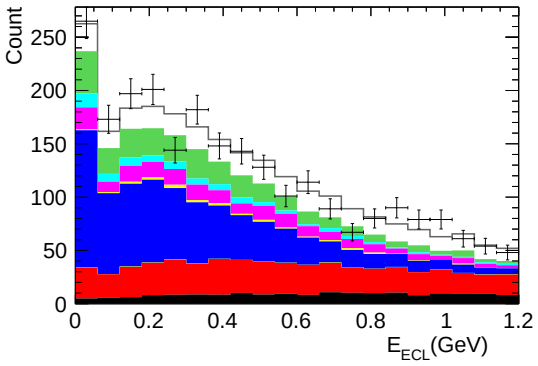
Figure 5.16: Plots for the D^{**} control region for the sample $D^0\pi^0\ell^-$. The invariant mass of the D^{**} is shown in figure (a), a zoomed in version of the same distribution is shown in figure (b), while figure (c) shows the $E_{\rm ECL}$ distribution.



(a)

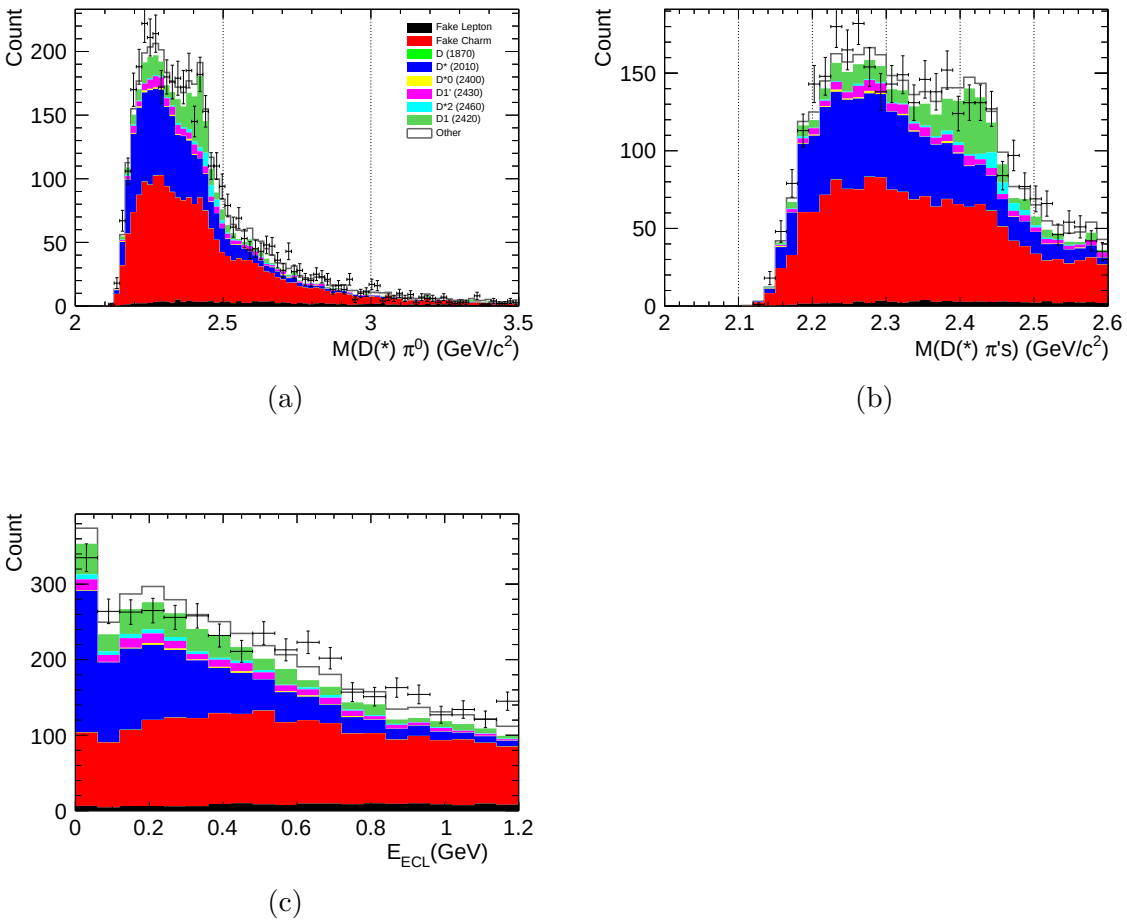


(b)



(c)

Figure 5.17: Plots for the D^{**} control region for the sample $D^{*+}\pi^0\ell^-$. The invariant mass of the D^{**} is shown in figure (a), a zoomed in version of the same distribution is shown in figure (b), while figure (c) shows the E_{ECL} distribution.



(a) (b)

(c)

Figure 5.18: Plots for the D^{**} control region for the sample $D^{*0}\pi^0\ell^-$. The invariant mass of the D^{**} is shown in figure (a), a zoomed in version of the same distribution is shown in figure (b), while figure (c) shows the $E_{\rm ECL}$ distribution.

Table 5.6: D_1 (39.4 %) channels for the sample $D^0\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*0}\pi^0$	434	38.8
$D^{*0}\pi^+$	315	28.1
$D^0\pi^0\pi^0$	154	13.7
$D^{*0}\pi^0\pi^0$	70	6.3
$D^{*+}\pi^+$	41	3.6
$D^{*+}\pi^0$	33	2.9

	Count	Rel count (%)
$D^{*0}\pi^0$	366	58.6
$D^0\pi^0\pi^0$	130	20.8
$D^{*0}\pi^0\pi^0$	52	8.4
$D^{*+}\pi^+$	30	4.7
$D^0\pi^+\pi^+$	26	4.2
$D^{*0}\pi^+\pi^+$	16	2.4

Table 5.7: D_2^* (14.8 %) channels for the sample $D^0\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^0\pi^0$	149	43.1
$D^{*0}\pi^0$	107	30.9
$D^0\eta$	44	12.5
$D^0\pi^+$	14	4.2
$D^{*+}\pi^+$	13	3.8
$D^{*+}\pi^0$	9	2.6

	Count	Rel count (%)
$D^0\pi^0$	149	48.2
$D^{*0}\pi^0$	96	31.3
$D^0\eta$	53	17.0
$D^{*+}\pi^+$	11	3.4

Table 5.8: D_0^* (31.5 %) channels for the sample $D^0\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^0\pi^0$	462	60.0
$D^0\eta$	201	28.1
$D^0\pi^+$	32	4.5
$D^{*0}\pi^0\pi^0$	30	4.2

	Count	Rel count (%)
$D^0\pi^0$	406	59.9
$D^0\eta$	226	33.3
$D^{*0}\pi^0\pi^0$	33	4.9

Table 5.9: D_1' (14.3 %) channels for the sample $D^0\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*0}\pi^0$	187	50.7
$D^{*0}\eta$	79	21.2
$D^{*0}\pi^+$	35	9.4
$D^{*+}\pi^+$	28	7.3
$D^{*+}\pi^0$	16	4.1

	Count	Rel count (%)
$D^{*0}\pi^0$	159	60.4
$D^{*0}\eta$	68	25.7
$D^{*+}\pi^+$	26	9.8

Table 5.10: D_1 (56.0 %) channels for the sample $D^{*0}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*0}\pi^0$	93	51.3
$D^{*0}\pi^+$	66	37.0
$D^{*0}\pi^0\pi^0$	18	9.6

	Count	Rel count (%)
$D^{*0}\pi^0$	79	85.4
$D^{*0}\pi^0\pi^0$	11	10.8
$D^{*0}\pi^+\pi^+$	4	3.9

Table 5.11: D_2^* (10.7 %) channels for the sample $D^{*0}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*0}\pi^0$	25	94.2
$D^{*0}\pi^+$	2	5.8

	Count	Rel count (%)
$D^{*0}\pi^0$	26	100.0

Table 5.12: D_0^* (4.2 %) channels for the sample $D^{*0}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*0}\pi^0\pi^0$	6	73.0
$D^{*0}\pi^+\pi^+$	1	15.4
$D^{*0}\pi^+\pi^0$	1	11.6

	Count	Rel count (%)
$D^{*0}\pi^0\pi^0$	10	78.3
$D^{*0}\pi^+\pi^+$	2	21.7

Table 5.13: D_1' (29.0 %) channels for the sample $D^{*0}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*0}\pi^0$	48	62.0
$D^{*0}\eta$	21	27.1
$D^{*0}\pi^+$	7	8.8
$D^{*0}\rho^+$	2	2.0

	Count	Rel count (%)
$D^{*0}\pi^0$	40	63.7
$D^{*0}\eta$	21	33.1
$D^{*0}\rho^+$	1	2.3

Table 5.14: D_1 (24.8 %) channels for the sample $D^+\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^+\pi^0\pi^0$	42	37.4
$D^{*+}\pi^0$	37	32.2
$D^+\pi^+\pi^0$	10	8.6
$D^+\pi^+\pi^+$	9	7.3
$D^{*+}\pi^+$	8	6.7
$D^{*+}\pi^0\pi^0$	5	4.9

	Count	Rel count (%)
$D^+\pi^0\pi^0$	89	48.8
$D^{*+}\pi^0$	70	37.7
$D^+\pi^+\pi^+$	14	7.6
$D^{*+}\pi^0\pi^0$	9	4.6

Table 5.15: D_2^* (18.5 %) channels for the sample $D^+\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^+\pi^0$	40	47.9
$D^+\pi^+$	19	23.1
$D^+\eta$	14	17.2
$D^{*+}\pi^0$	8	9.7

	Count	Rel count (%)
$D^+\pi^0$	94	68.7
$D^+\eta$	25	17.9
$D^{*+}\pi^0$	18	13.0

Table 5.16: D_0^* (50.1 %) channels for the sample $D^+\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^+\pi^0$	119	56.3
$D^+\eta$	61	28.9
$D^+\pi^+$	23	10.8
$D^{*+}\pi^0\pi^0$	5	2.7

	Count	Rel count (%)
$D^+\pi^0$	245	62.8
$D^+\eta$	138	35.4

Table 5.17: D_1' (6.7 %) channels for the sample $D^+\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*+}\pi^0$	16	46.5
$D^{*+}\pi^+$	10	30.7
$D^{*+}\eta$	4	12.3
$D^+\rho^+$	3	8.0

	Count	Rel count (%)
$D^{*+}\pi^0$	28	60.2
$D^{*+}\eta$	14	29.5
$D^+\rho^+$	4	7.5

Table 5.18: D_1 (51.8 %) channels for the sample $D^{*+}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*+}\pi^0$	82	69.7
$D^{*+}\pi^+$	18	14.6
$D^{*+}\pi^0\pi^0$	12	10.9
$D^{*+}\pi^+\pi^0$	3	2.5
$D^{*+}\pi^+\pi^+$	3	2.3

	Count	Rel count (%)
$D^{*+}\pi^0$	184	86.7
$D^{*+}\pi^0\pi^0$	25	11.3

Table 5.19: D_2^* (11.3 %) channels for the sample $D^{*+}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*+}\pi^0$	19	71.5
$D^{*+}\pi^+$	7	28.5

	Count	Rel count (%)
$D^{*+}\pi^0$	46	100.0

Table 5.20: D_0^* (4.7 %) channels for the sample $D^{*+}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*+}\pi^0\pi^0$	6	63.0
$D^{*+}\pi^+\pi^+$	2	25.9
$D^{*+}\pi^+\pi^0$	1	11.1

	Count	Rel count (%)
$D^{*+}\pi^0\pi^0$	16	79.4
$D^{*+}\pi^+\pi^+$	4	20.6

Table 5.21: D_1' (32.1 %) channels for the sample $D^{*+}\ell^-$ for the left (right) and fake (true) B-tags.

	Count	Rel count (%)
$D^{*+}\pi^0$	39	55.5
$D^{*+}\pi^+$	22	30.9
$D^{*+}\eta$	9	12.6

	Count	Rel count (%)
$D^{*+}\pi^0$	86	66.3
$D^{*+}\eta$	40	30.3
$D^{*+}\rho^+$	4	3.0

Chapter 6

Systematic Uncertainties

6.1 Introduction

The PDFs we use for the signal extraction, are built from Monte Carlo (MC) simulations by combining the theoretical SM modelling of the physics event we want to measure, and the detector response to the products of these events. Our best knowledge about theoretical and experimental parameters has been used to produce these MC samples, however this knowledge is not immune from uncertainties.

As a consequence, the final result of the analysis can be potentially biased by the model we use to extract this result, thus we aim at quantifying the impact of the known uncertainties that affect our MC model. The goal of this section is to describe the methods used to estimate the magnitude of the systematic uncertainty of the measurement, the sources that contribute to this uncertainty, and its numerical values.

6.2 Systematic Uncertainty Estimation Procedure

Generally speaking, we can separate the sources of systematic uncertainties in two categories: the ones generated by uncertainties on the model's parameters and the ones generated by the limited size of the MC sample that we use to build our model. The latter can be seen as a statistical source of systematic uncertainties, but to avoid confusion we will label the statistical uncertainty only the one related to the size of the data sample, which is given as an output of the fit.

There is, however, a difference in the treatment of these two sources of uncertainties, since the former is assigned a Gaussian error profile, while the latter a Poisson error profile. For the Gaussian-like sources of systematic uncertainties, we produce new MC models with model parameters that have been varied from the ones used in the standard fit. For each parameter, we build a gaussian distribution that has as mean the nominal parameter value, and as sigma the parameter's uncertainty. From this distribution we draw N new parameters values, which will be used to repeat the fit procedure N times.

Following this method we obtain a Gaussian distribution of fit results, whose standard deviation is used to define the uncertainty error on the final result introduce by the uncertainty of the parameter. The treatment of Poisson-like sources of systematic uncertainties is explained in Section 6.3.1.

6.3 Systematic Uncertainty Sources

6.3.1 Monte Carlo Statistics

The shapes of the fit PDFs are fixed and only their yield is a floating parameter of the fit. We assume our knowledge of these shapes is dependent on the number of events belonging to each histogram PDF. Therefore, in order to estimate the effect on the final results introduced by the uncertainty on the PDFs shapes we produce toy MC samples by varying each bin's content of the PDFs histogram, according to a Poisson distribution with the bin content as event rate, and repeat the fit procedure. As shown in Table 6.1 this type is the dominant source for the total systematic uncertainty of the measurement.

6.3.2 Feed-down Correlations and Signal/Normalization Efficiency

All the fixed factors that are used in the fitting procedure detailed in Section 4.2) are evaluated using Monte Carlo samples. They are only known up to a certain precision due to statistical fluctuations. The error on these parameters reflect the size of the event sample used for their determination. Therefore, to evaluate the effects of this imprecision to the final result, we produce MC toy samples with gaussian error variations of these parameters.

6.3.3 $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ Branching Fractions

The branching fractions of the $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ decay modes and the decays of the D^{**}-mesons are not well known and therefore contribute significantly to the total PDF uncertainty for the $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ component. The branching fraction of each $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ decay is varied within its uncertainty. The uncertainties depend on the D^{**} state and are assumed to be:

- D_1 : $\pm 6\%$
  The D_1 state contributes through the decay modes $D^*\pi^0$, $D^*\pi^+$ and $D\pi^0\pi^0$. We estimate the relative accuracy of $\mathcal{B}(B \to D_1\ell\nu)$ to be 6% from the HFLAV value of $(0.427 \pm 0.027)\%$.

- D_2^* : $\pm 10\%$
  The D_1 state contributes through the decay modes $D^*\pi^0$ and $D\pi^0$. We estimate

the relative accuracy of $\mathcal{B}(B \to D_2^*\ell\nu)$ to be 10% from the HFLAV value of $(0.117 \pm 0.012)\%$.

- D_1' : $\pm 83\%$
  There are large discrepancy among experiments for the branching ratio $\mathcal{B}(B \to D_1'\ell\nu)$, which are reflected in the PDG average of $(0.195 \pm 0.054)\%$. We found the main contributions from this state to be through the decay modes $D^*\pi^0$ and $D^*\eta$. take a conservative approach and estimate this uncertainty to be 83%.

- D_0^* : $\pm 100\%$
  The D_0^* state contributes through the decay modes $D\pi^0$ and $D\eta$. We conservatively estimate the relative accuracy of $\mathcal{B}(B \to D_2^*\ell\nu)$ to be 100%.

By reweighting our PDFs with gaussian variation of these branching ratios, we produce new PDFs and repeat the fit procedure.

6.3.4 *B* Decay Form Factors

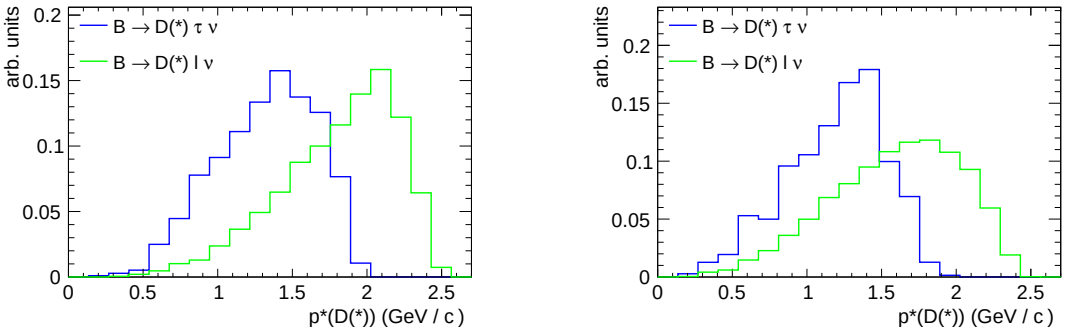
As described in Section 5.2 we reweight our MC sample to account for updated values of the form factors parameters describing the$\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ and $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ decays. Using Gaussian error distributions we generate 500 variations of the form factor parameters. The simulated events are reweighted, the PDFs re-calculated, and the fit repeated for each set of form factors. We estimate the uncertainty as the standard deviation of the $\mathcal{R}(D^{(*)})$ distribution.

6.3.5 Slow Pion Efficiency

The slow pion efficiency does not completely cancel in the ratio as the D^* mesons coming from $\bar{B} \to D^*\ell^-\bar{\nu}_\ell$ decays have slightly higher momenta than D^* mesons from $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ decays, as shown in Figure 6.1. As a result, the momentum of the soft pion coming from D^* decays also differs. The slow pion reconstruction efficiency affects the normalization and signal PDF shapes, as a consequence we estimate the systematic uncertainty introduced by the slow pion efficiency corrections.

6.3.6 Lepton Efficiency and Fake Rate

The momentum spectra of electrons and muons vary greatly from signal mode to normalization mode. Indeed, leptons produced by τ decays (as in the signal mode) having a softer spectrum than leptons that come from the W decay (as in the normalization mode). As described in Sections 5.1.1 and 5.1.2 the lepton reconstruction efficiency and fake rates in the MC samples are reweighted using experimental results. These factors affect the normalization and signal PDF shapes, as a consequence we estimate the systematic uncertainty introduced by these corrections.



(a) Sample $D\ell$. (b) Sample $D^*\ell$.

Figure 6.1: The difference in charm meson momentum spectrum between the signal mode and the normalization mode is shown in the subfigures, for B decays to both D and D^* mesons.

6.3.7 Tag and Fake Charm Calibration

Both the efficiencies for the tag reconstruction and the fake charm reconstruction are calibrated using collision data, as explained in Sections 5.5 and 5.4, respectively. The calibration factors are generally extracted by taking the ratio of events in the data samples to events in the MC samples. The errors on these factors is affected by the size of the samples used in the calibration. By using gaussian variations of these parameters we build reweighted PDFs and repeat the fit procedure, in order to assess the systematic uncertainty introduced by these calibrations.

6.4 Summary of Systematic Uncertainties

An overview of all the contributions to the systematics uncertainties for the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ results is given in Table 6.1, with a breakdown of the contribution to the MC statistics component provided in Table 6.2. We add all contributions in quadrature since we treat them as uncorrelated.

The size of the MC samples used to build the PDFs is among the dominant contributions to the total systematic uncertainty of the measurement. The factors used to correlate the feed-down components, together with the efficiency factors used for the $R^{(*)}$, which both have a statistical origin, also have high contributions. The overall correlation ρ_{tot} of the systematic uncertainties is calculated using the expression:

$$\sum_{i=1}^{n} \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \sigma_i^* \\ \rho_i \sigma_i \sigma_i^* & \sigma_i^{*2} \end{pmatrix} = \begin{pmatrix} \sigma_{\text{tot}}^2 & \rho_{\text{tot}} \sigma_{\text{tot}} \sigma_{\text{tot}}^* \\ \rho_{\text{tot}} \sigma_{\text{tot}} \sigma_{\text{tot}}^* & \sigma_{\text{tot}}^{*2} \end{pmatrix}, \tag{6.1}$$

where σ refers to uncertainties on $\mathcal{R}(D)$, σ^* to uncertainties on $\mathcal{R}(D^*)$ and i runs over

Table 6.1: Overview of all the contributions to the systematic uncertainties on the $\mathcal{R}(D^{(*)})$results. The total systematic uncertainty is calculated by summing each uncertainty in quadrature.

Source	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)	Correlation
D^{**} composition	0.76	1.41	−0.41
PDF shapes	4.39	2.25	−0.55
Feed-down factors	1.69	0.44	0.53
Efficiency factors	1.93	4.12	−0.57
Fake $D^{(*)}$ calibration	0.19	0.11	−0.76
B_{tag} calibration	0.07	0.05	−0.76
Lepton efficiency and fake rate	0.36	0.33	−0.83
Slow pion efficiency	0.08	0.08	−0.98
B decay form factors	0.55	0.28	−0.60
Luminosity and $\mathcal{B}(\Upsilon(4S))$	0.10	0.04	−0.58
$\mathcal{B}(B \to D^{(*)}\ell\nu)$	0.05	0.02	−0.69
$\mathcal{B}(D)$	0.35	0.13	−0.65
$\mathcal{B}(D^*)$	0.04	0.02	−0.51
$\mathcal{B}(\tau^- \to \ell^- \bar{\nu}_\ell \nu_\tau)$	0.15	0.14	−0.11
Total	5.21	4.94	−0.52

Table 6.2: Breakdown of systematic uncertainty due to the limited MC size sample. Each line in the table corresponds to a PDF used in the fit. The total systematic uncertainty is calculated by summing each uncertainty in quadrature.

PDF	$\Delta R(D)$ (%)	$\Delta R(D^*)$ (%)	Correlation
Signal	2.18	1.38	-0.82
Normalization	1.08	0.75	-0.74
$\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$	0.91	0.67	-0.59
Other	1.07	0.78	-0.60
Fake $D^{(*)}$-mesons	1.99	0.93	-0.48
Mixed Feed-down, ℓ	1.86	0.34	-0.59
Charged Feed-down, ℓ	1.63	0.62	0.02
Mixed Feed-down, τ	0.77	0.24	-0.63
Charged Feed-down, τ	0.85	0.33	-0.45
Total	4.39	2.25	-0.55

each source of uncertainty. From this expression we finally extract

$$\rho_{\text{tot}} = \frac{\sum_{i=1}^{n} \rho_i \sigma_i \sigma_i^*}{\sigma_{\text{tot}} \sigma_{\text{tot}}^*}. \tag{6.2}$$

We obtain a total correlation of the systematic uncertainty equal to -0.52, driven mostly by the Monte Carlo statistics source of systematic uncertainty.

Chapter 7

Results and Discussion

7.1 Results

After performing the fit and evaluating the systematic uncertainty, we extract the results:

$$\mathcal{R}(D) = 0.307 \pm 0.037 \pm 0.016 \tag{7.1}$$
$$\mathcal{R}(D^*) = 0.283 \pm 0.018 \pm 0.014, \tag{7.2}$$

where the first uncertainty is statistical and provided by the fit, and the second error is systematic. A break-down of electron and muon channel results is given in Table 7.1. We exploit the isospin symmetry between B^0 and B^+ to impose the relationship

Table 7.1: Fit results for the electron, muon and sum of electron and muon channels.

$R(D, e)$	$0.281 \pm 0.042 \pm 0.017$
$R(D, \mu)$	$0.373 \pm 0.068 \pm 0.030$
$R(D, \ell)$	$0.307 \pm 0.037 \pm 0.016$
$R(D^*, e)$	$0.304 \pm 0.022 \pm 0.016$
$R(D^*, \mu)$	$0.245 \pm 0.035 \pm 0.020$
$R(D^*, \ell)$	$0.283 \pm 0.018 \pm 0.014$

$R(D^{(*)}) = R(D^{(*)+}) = R(D^{(*)0})$ in the fit. Results for all fit parameters of the four samples are given in Table 7.2.

The fit projection on the E_{ECL} axis and on the classifier axis, for both the whole 2D fit region and for the signal region defined by `class` > 0.9, are shown in Figures 7.1 to 7.8. The correlation matrix for all floating parameters of the fit is shown in Figure 7.9. As expected, we find a statistical correlation factor of -0.53 between $\mathcal{R}(D^*)$ and $\mathcal{R}(D)$, while the systematic correlation factor is equal to -0.52.

Table 7.2: Fitted yields in the 4 channels for the partial box opening. N_{exp} is the expected number of events, and N_{fit} the number of events obtained in the fit. Yields with 0 error are fixed.

Channel	Component	N_{exp}	N_{fit}
$D^+\ell^-$	Other	506	506 ± 0
	$B \to D^{**}\ell\nu$	1262	413 ± 110
	Fake D	3072	3072 ± 0
	$B^0 \to D^*\ell\nu$	6161	6370 ± 225
	$B^0 \to D^*\tau\nu$	249	269 ± 24
	$B \to D\ell\nu$	5850	6800 ± 179
	$B \to D\tau\nu$	419	307 ± 65
$D^{*+}\ell^-$	Other	287	287 ± 0
	$B \to D^{**}\ell\nu$	668	314 ± 65
	Fake D^*	754	754 ± 0
	$B \to D^*\ell\nu$	9362	9794 ± 109
	$B \to D^*\tau\nu$	344	376 ± 36
$D^0\ell^-$	Other	2131	2131 ± 0
	$B \to D^{**}\ell\nu$	5610	3595 ± 252
	Fake D	8708	8708 ± 0
	$B^+ \to D^*\ell\nu$	46259	45042 ± 563
	$B^0 \to D^*\ell\nu$	1215	2302 ± 531
	$B^+ \to D^*\tau\nu$	1891	1704 ± 177
	$B^0 \to D^*\tau\nu$	61	123 ± 11
	$B \to D\ell\nu$	13671	16096 ± 436
	$B \to D\tau\nu$	992	1471 ± 193
$D^{*0}\ell^-$	Other	187	187 ± 0
	$B \to D^{**}\ell\nu$	505	406 ± 64
	Fake D^*	1993	1993 ± 0
	$B \to D^*\ell\nu$	6962	7148 ± 100
	$B \to D^*\tau\nu$	290	275 ± 29

The 2D combination of the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ results, together with their correlation and the SM expectation is shown in Figure 7.10 for $R(D,\ell) - R(D^*,\ell)$ with contours up to 5σ, while the 1σ contour results for $R(D,e) - R(D^*,e)$, $R(D,\mu) - R(D^*,\mu)$ and $R(D,\ell) - R(D^*,\ell)$ are shown in Figure 7.11.

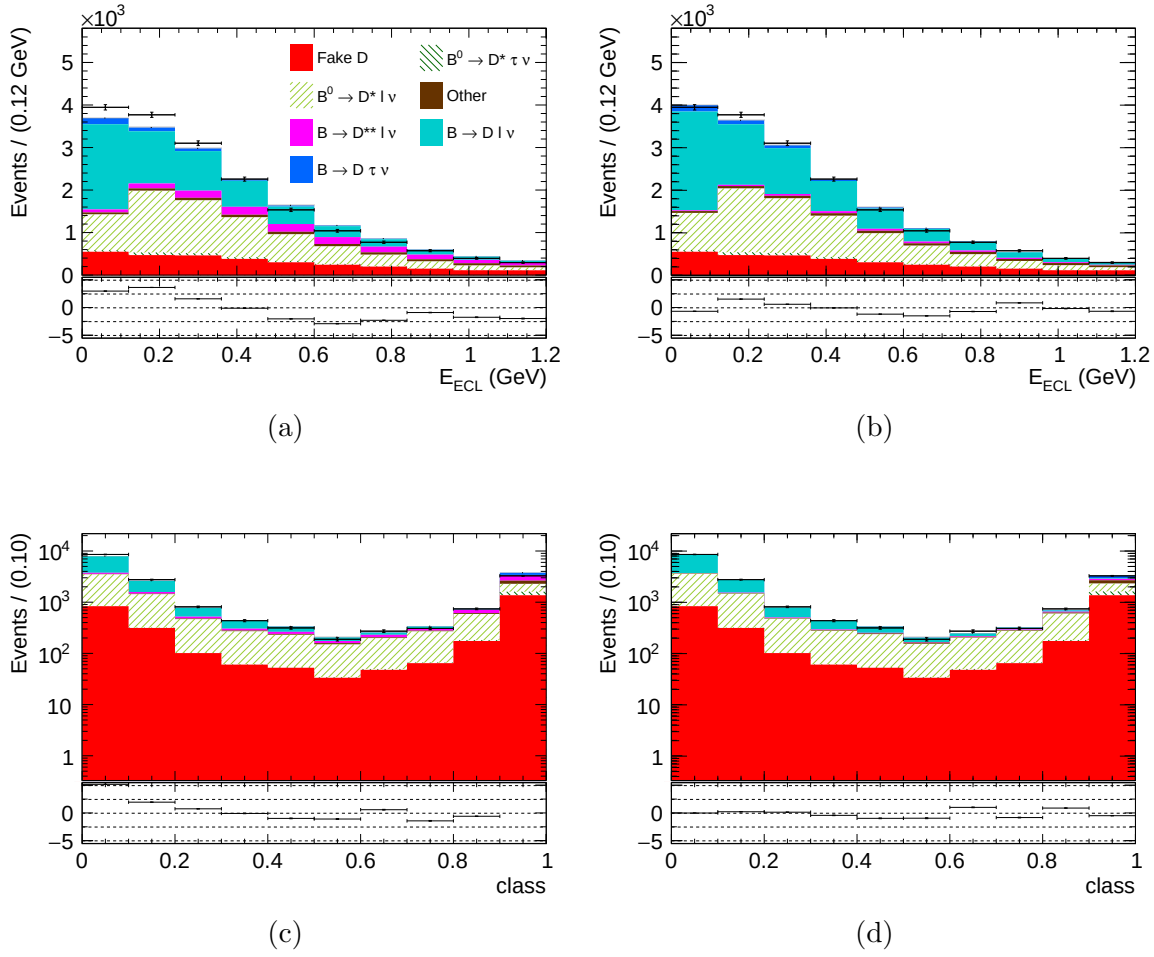
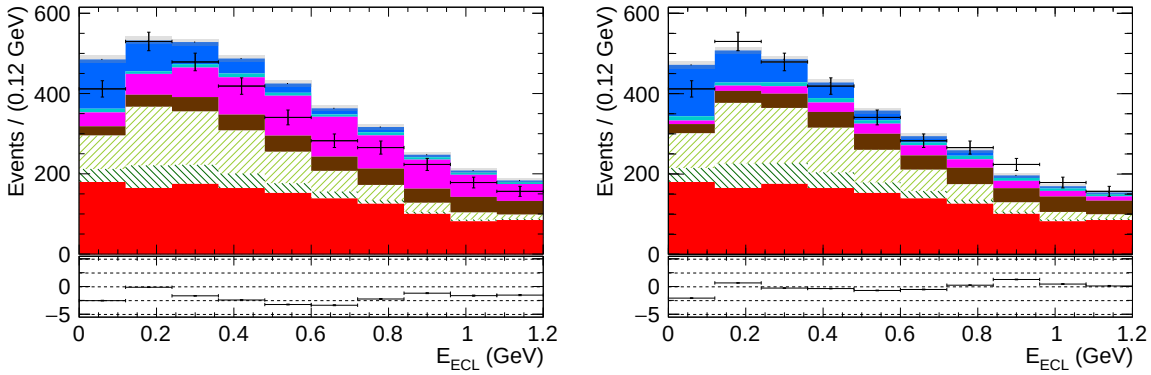
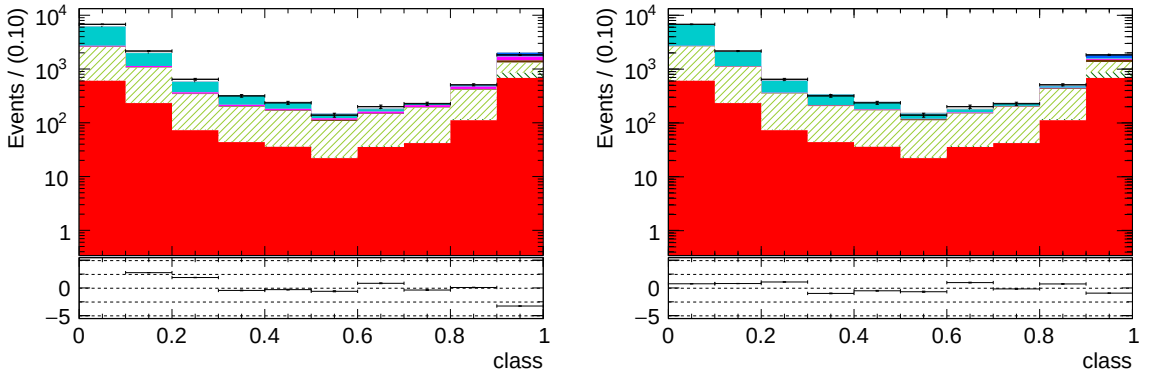


Figure 7.1: Projections of the fit results for the sample $D^+\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the E_{ECL} projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\mathrm{ECL}} < 0.48$ GeV.



(a) (b)



(c) (d)

Figure 7.2: Signal region projections for the fit results for the sample $D^+\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the $E_{\rm ECL}$ projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\rm ECL} < 0.48$ GeV.

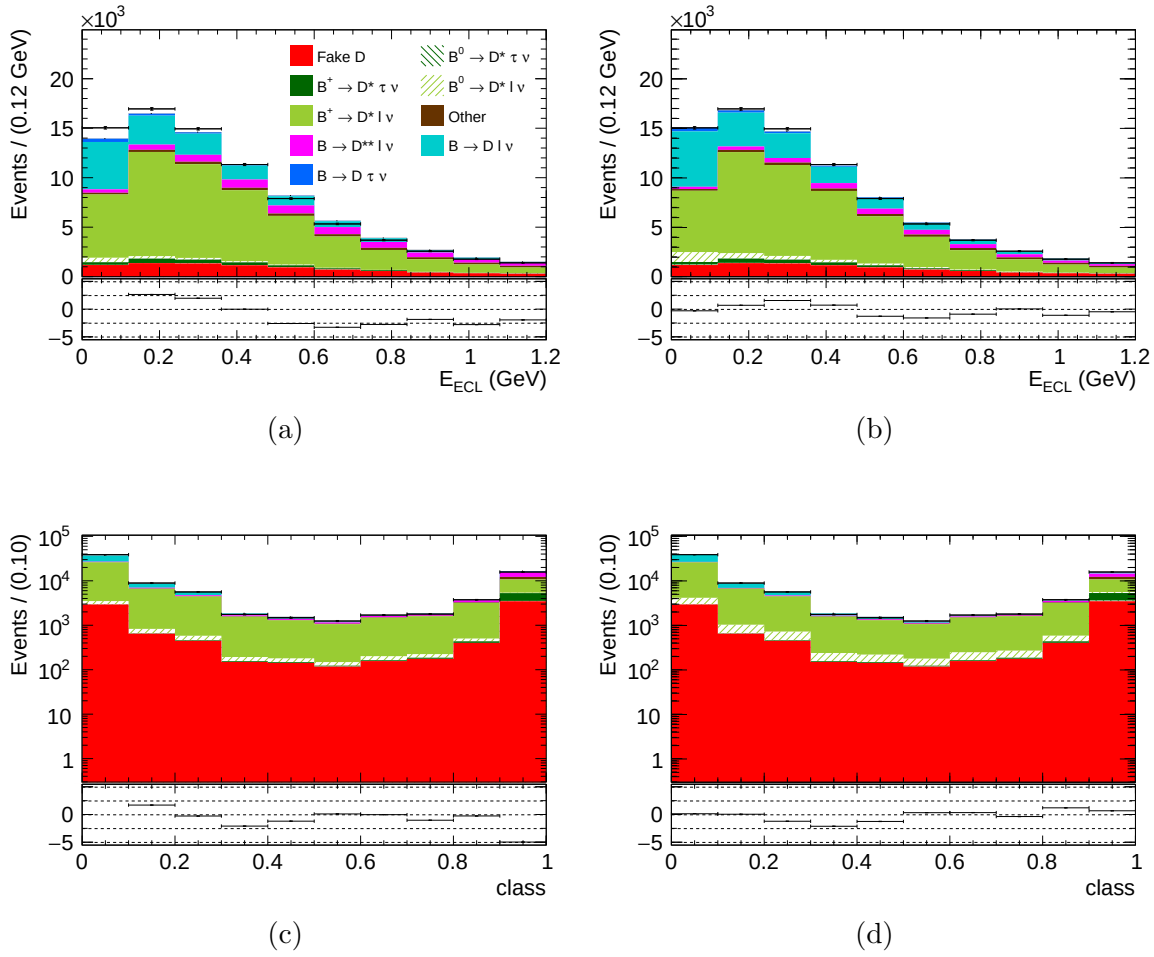
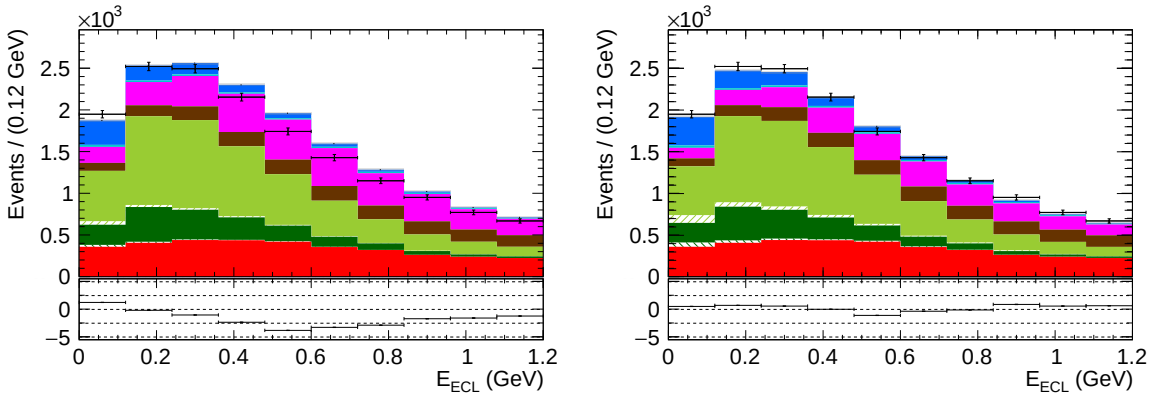
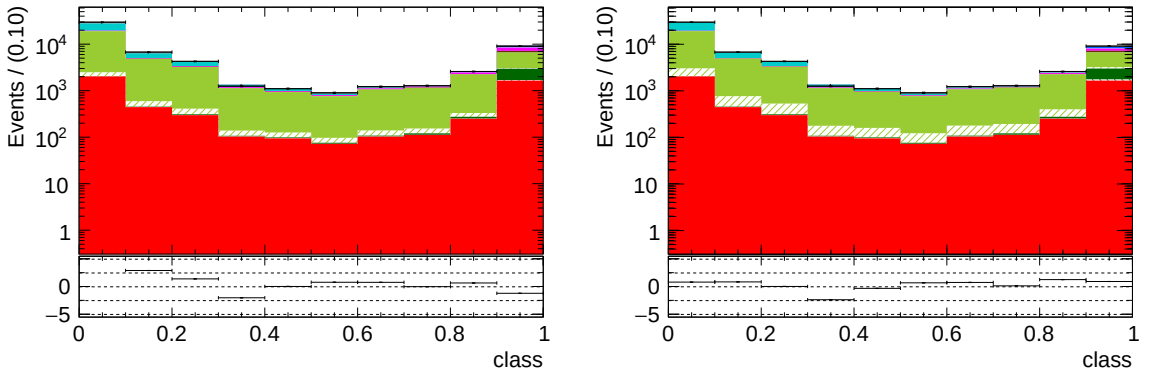


Figure 7.3: Projections of the fit results for the sample $D^0\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the E_{ECL} projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\mathrm{ECL}} < 0.48$ GeV.



(a) (b)



(c) (d)

Figure 7.4: Signal region projections for the fit results for the sample $D^0\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the E_{ECL} projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\mathrm{ECL}} < 0.48$ GeV.

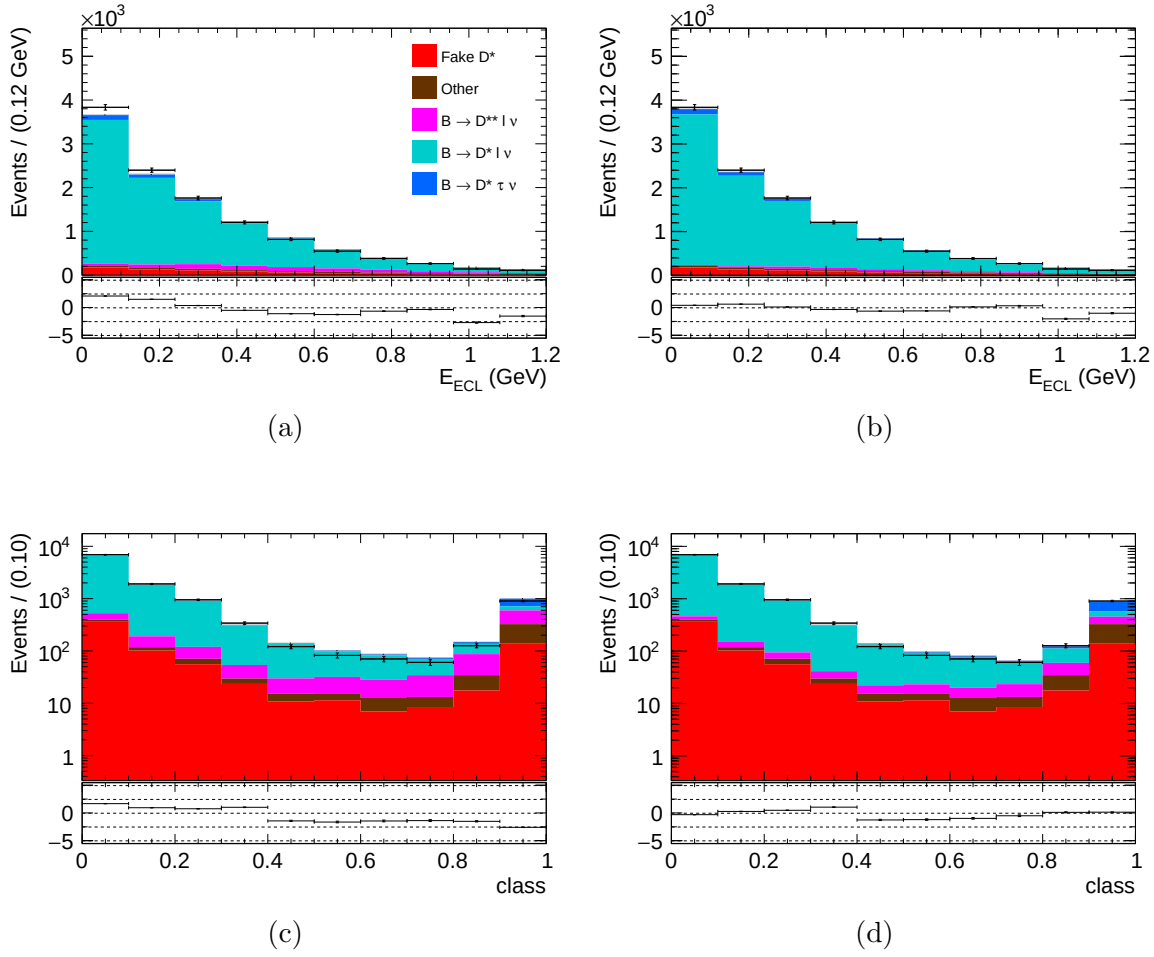


Figure 7.5: Projections for the fit results for the sample $D^{*+}\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the E_{ECL} projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\mathrm{ECL}} < 0.48$ GeV.

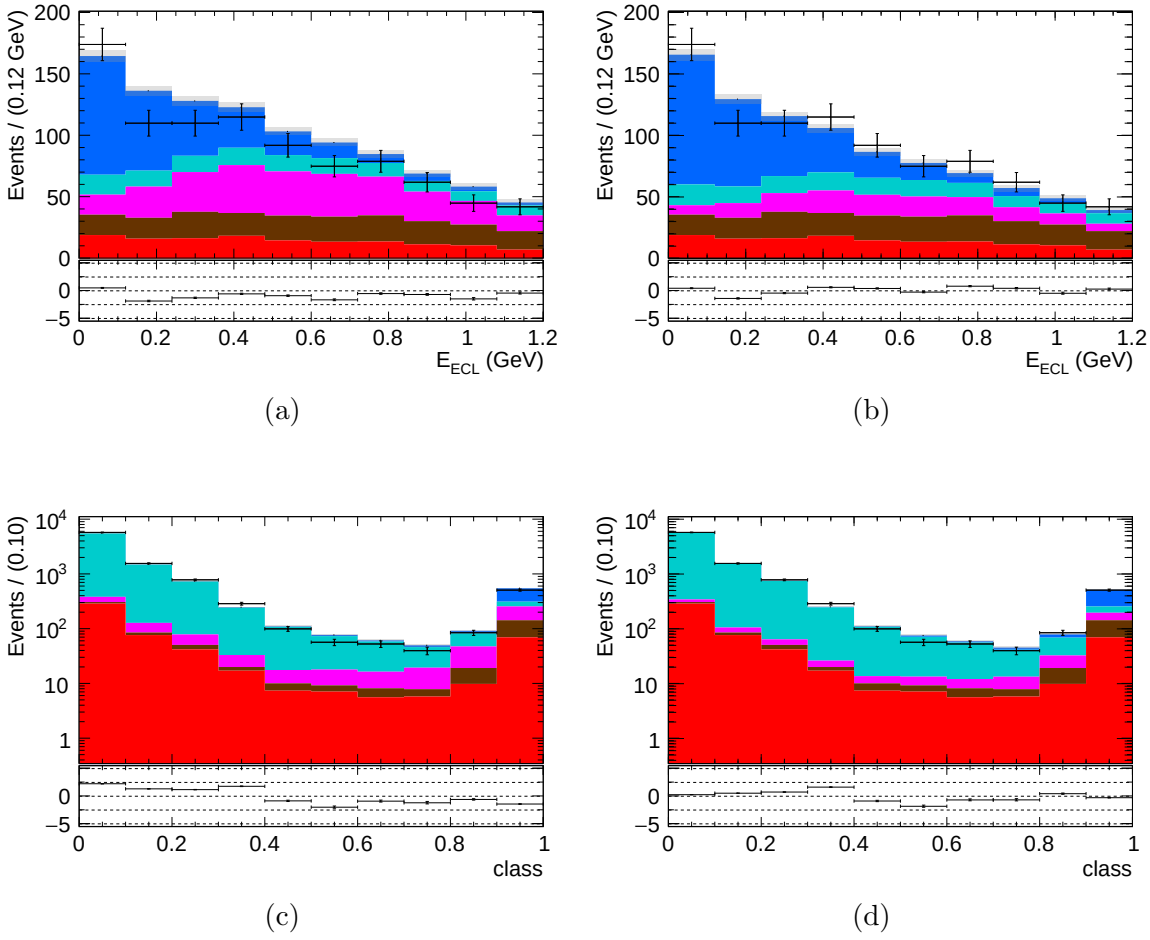


Figure 7.6: Signal region projections for the fit results for the sample $D^{*+}\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the E_{ECL} projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\mathrm{ECL}} < 0.48$ GeV.

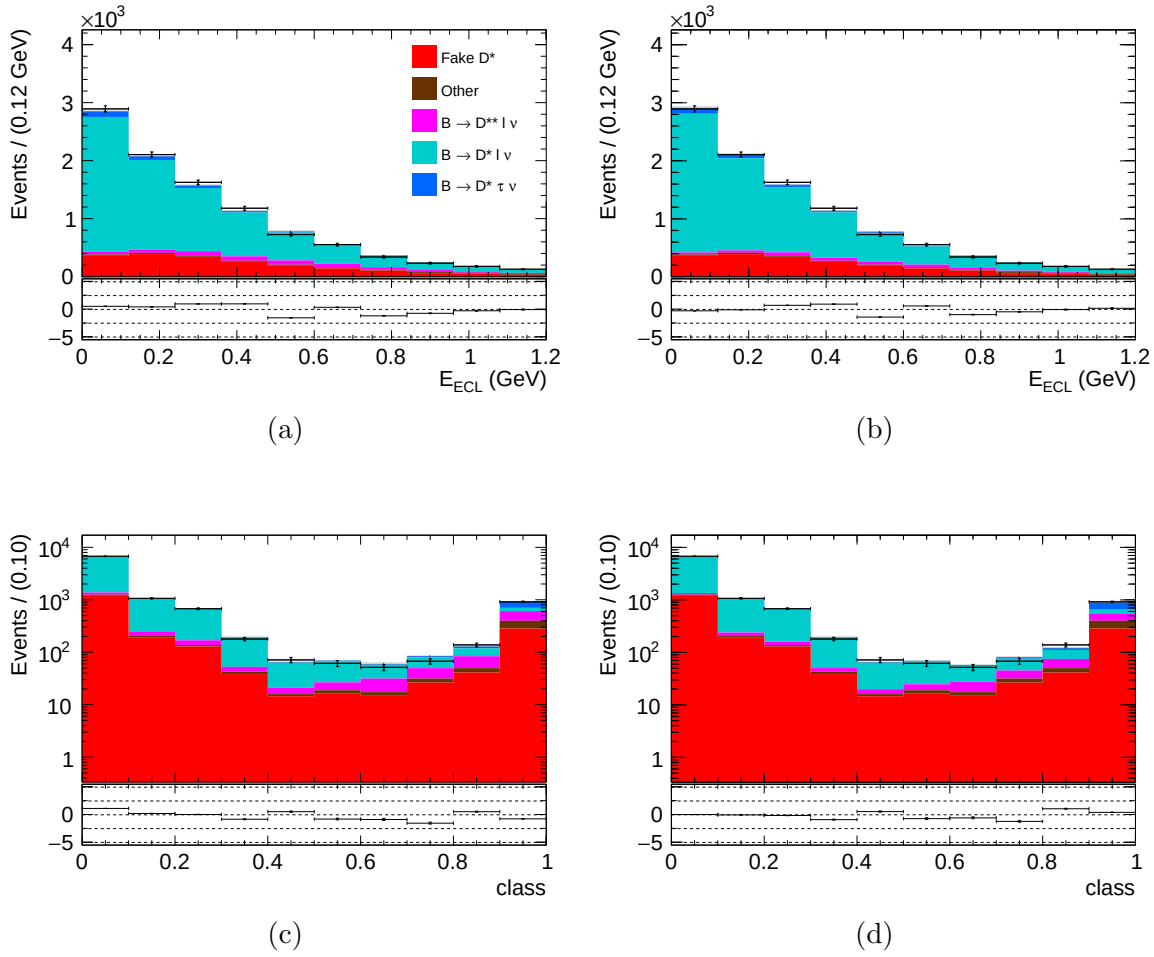
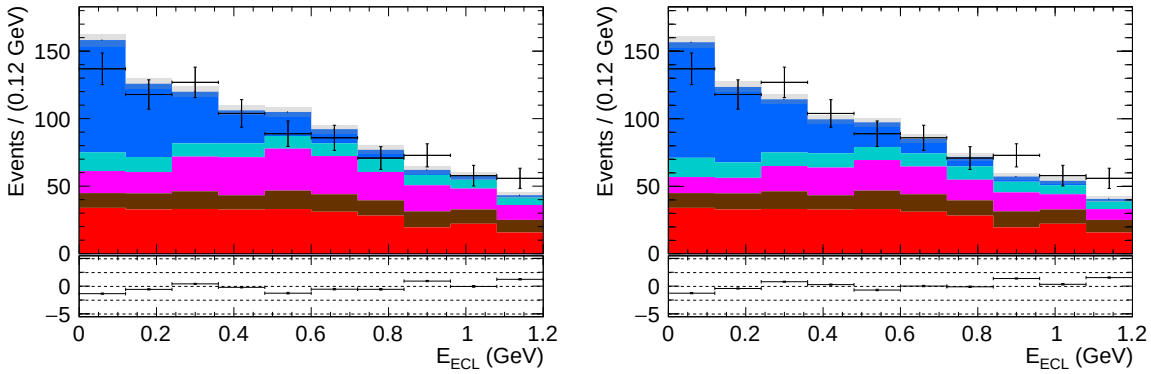
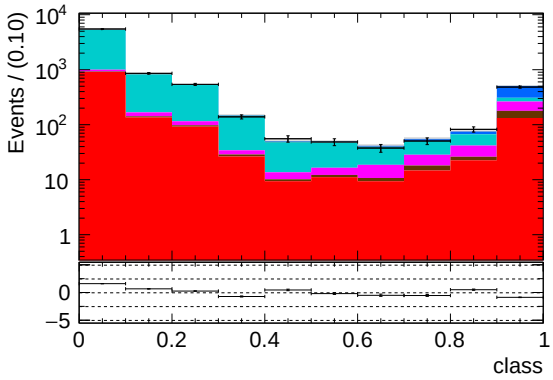


Figure 7.7: Projections for the fit results for the sample $D^{*0}\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the E_{ECL} projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\mathrm{ECL}} < 0.48$ GeV.

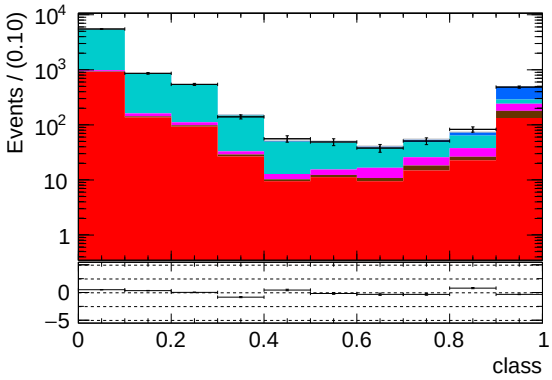


(a)

(b)



(c)



(d)

Figure 7.8: Signal region projections for the fit results for the sample $D^{*0}\ell^-$. Pre-fit results are shown on the left and post-fit plots are shown on the right. Figures (a) and (b) show the $E_{\rm ECL}$ projection after applying the criterion `class` > 0.9, figures (c) and (d) show the classifier projection after applying the criterion $E_{\rm ECL} < 0.48$ GeV.

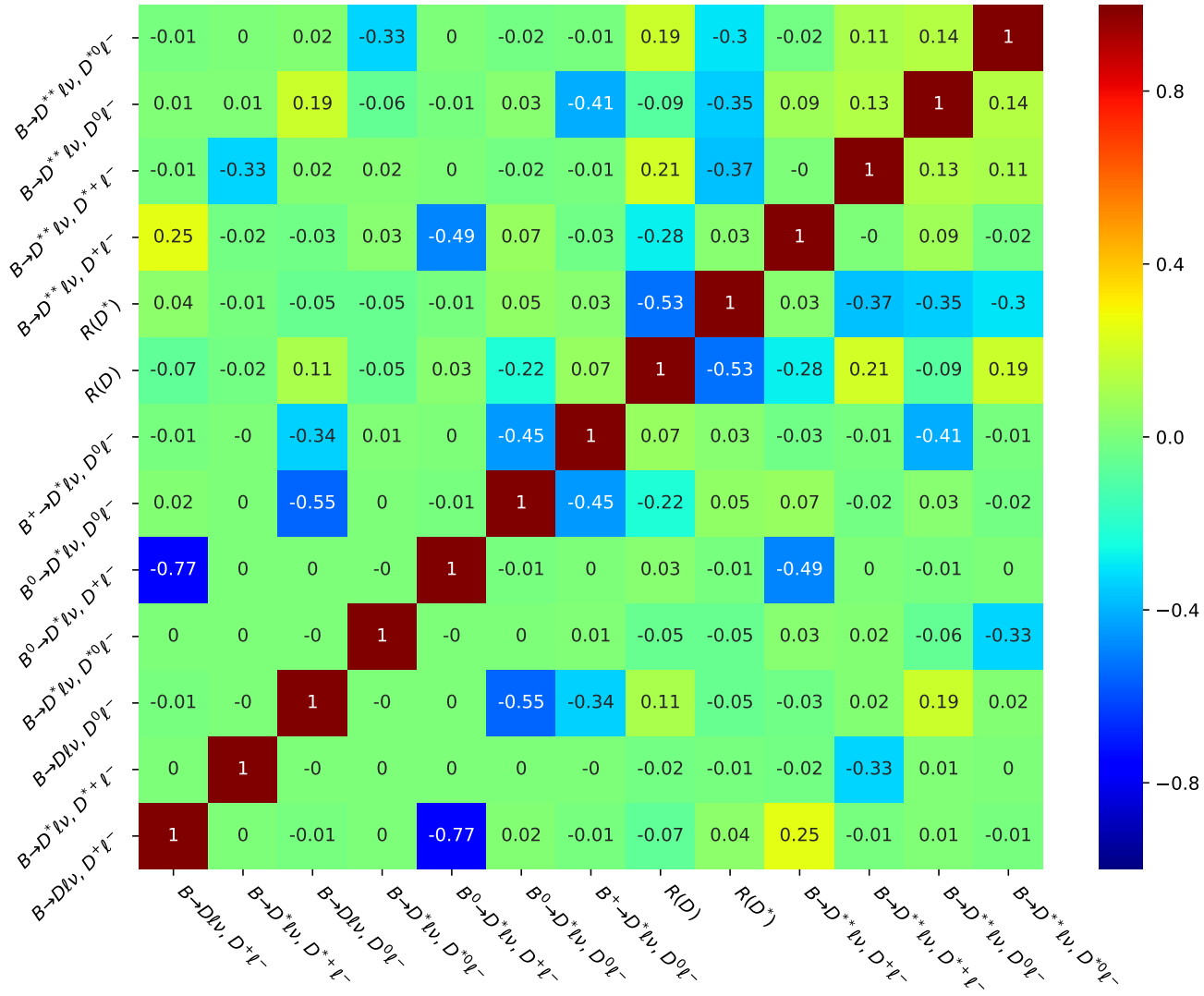


Figure 7.9: Correlation matrix for the floating fit parameters.

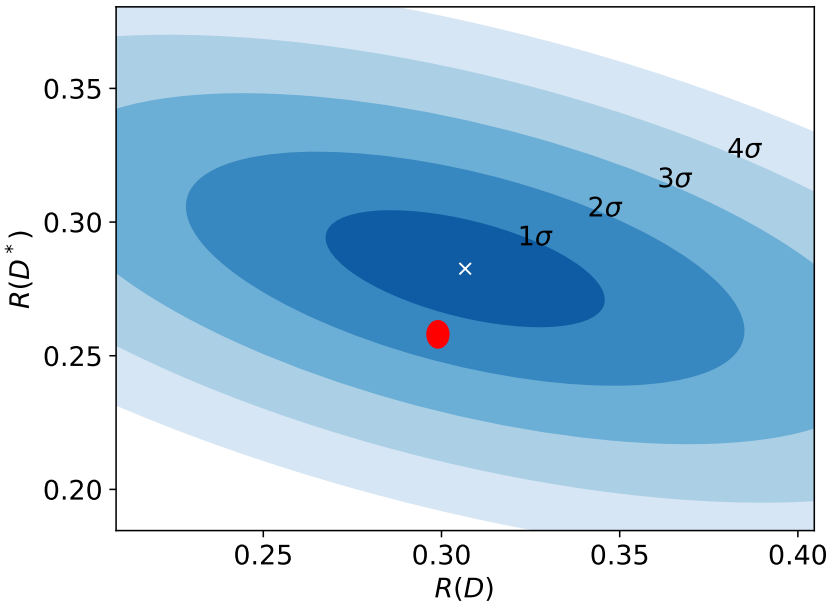


Figure 7.10: Fit results are shown on a 2D $R(D) - R(D^*)$ plane. The red ellipse represents the SM expectations, with values $\mathcal{R}(D)_{\text{SM}} = 0.299 \pm 0.003$ and $\mathcal{R}(D^*)_{\text{SM}} = 0.258 \pm 0.005$ [40].

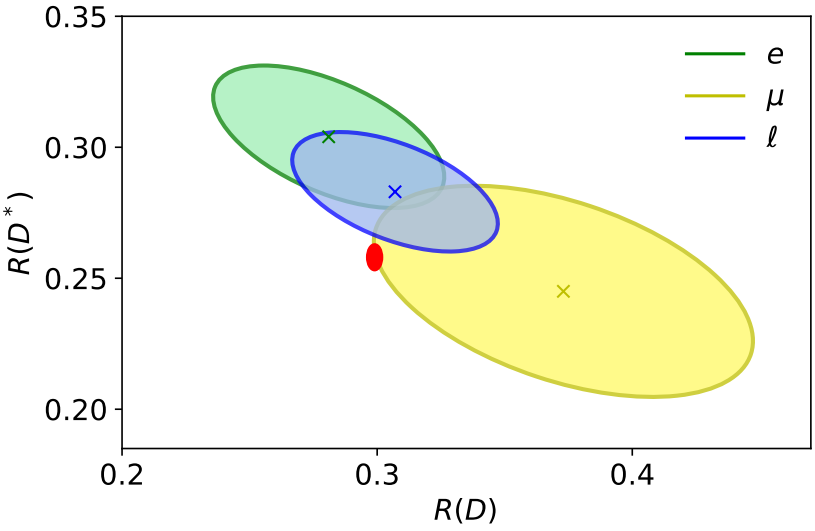


Figure 7.11: Fit results are shown for the 1σ contour of the combinations $R(D, e) - R(D^*, e)$, $R(D, \mu) - R(D^*, \mu)$ and $R(D, \ell) - R(D^*, \ell)$. The red ellipse represents the SM expectations, with values $\mathcal{R}(D)_{\text{SM}} = 0.299 \pm 0.003$ and $\mathcal{R}(D^*)_{\text{SM}} = 0.258 \pm 0.005$ [40].

7.2 Discussion

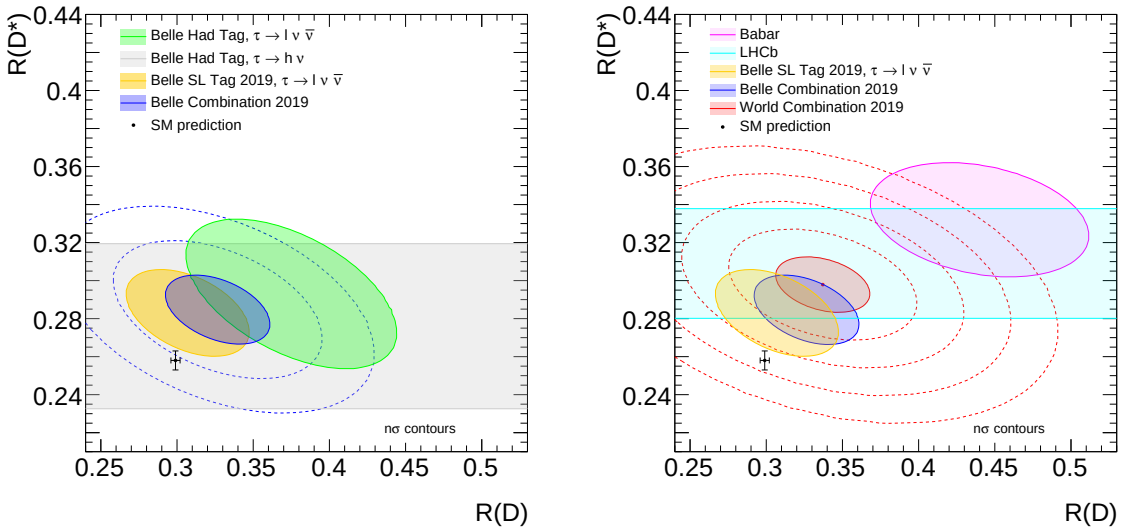
Our results for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are in agreement with the SM predictions within 0.2σ and 1.1σ respectively, and provide the most precise measurement of these quantities to date. The combination of the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ results is compatible with the SM within 1.3σ. The world average for the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ experimental results before this work showed a discrepancy of approximately 3.4σ with the SM expectation. This deviation is mostly driven by the BaBar analysis which found a 3σ discrepancy with the SM, with following results showing either a considerably smaller discrepancy, or compatibility with the SM. This work goes in the direction of reconciliating the experimental results with the SM expectation.

The combination of the Belle results on $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$, which include this work and the hadronic tag measurement [36], is shown in Figure 7.12a. The improvement in precision with respect to the previous analysis is large, particularly for $\mathcal{R}(D)$. The world average for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ updated with these results, together with the BaBar and Belle results is shown in Figure 7.12b. The discrepancy of the experimental world average with the SM expectations decreases from 4 to 3σ when including these latest results, pulled mostly by the BaBar result.

Our results favour new physics models with values of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ compatible with the SM values. For example, the 2HDM-Type II introduced in Section 1.2.2 can produce SM-like values of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ simultaneously for small values of its $\tan\beta/m_{H^\pm}$ parameter ratio. This is shown in Figure 7.13, where our results which have not been corrected for efficiency effects, are compared to 2HDM-Type II predictions for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$.

As for the 2HDM-Type III (where both Higgs doublets couple to up quarks and down quarks as well) we follow the study done in Reference [73] to constrain the allowed regions for the Wilson coefficients C_R^{cb} and C_R^{cb}, associated to the operators $\mathcal{O}_R^{cb} = \bar{c}P_R b\,\bar{\tau}P_L\nu_\tau$ and $\mathcal{O}_L^{cb} = \bar{c}P_L b\,\bar{\tau}P_L\nu_\tau$ respectively. The allowed 1σ regions for the coefficients C_L^{cb} and C_R^{cb} from the combination of our $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ results are shown in Figure 7.14.

The precision of this measurement is limited by the size of the available dataset. This fact is reflected by the statistical uncertainty dominating over the systematic uncertainty. The Belle II experiment benefits from an accelerator with an instantaneous luminosity 40 times larger that its predecessor, which will eventually produce a dataset that is expected to bring the total uncertainty of these measurements to the 2% level. Hence, Belle II will probably set once and for all the $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ anomaly.



(a) (b)

Figure 7.12: On the left: combination of the Belle results on $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$. On the right: new world average for the experimental results on $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$, shown together with the BaBar and Belle results. For both plots, the SM expectations taken from Ref. [40] are shown.

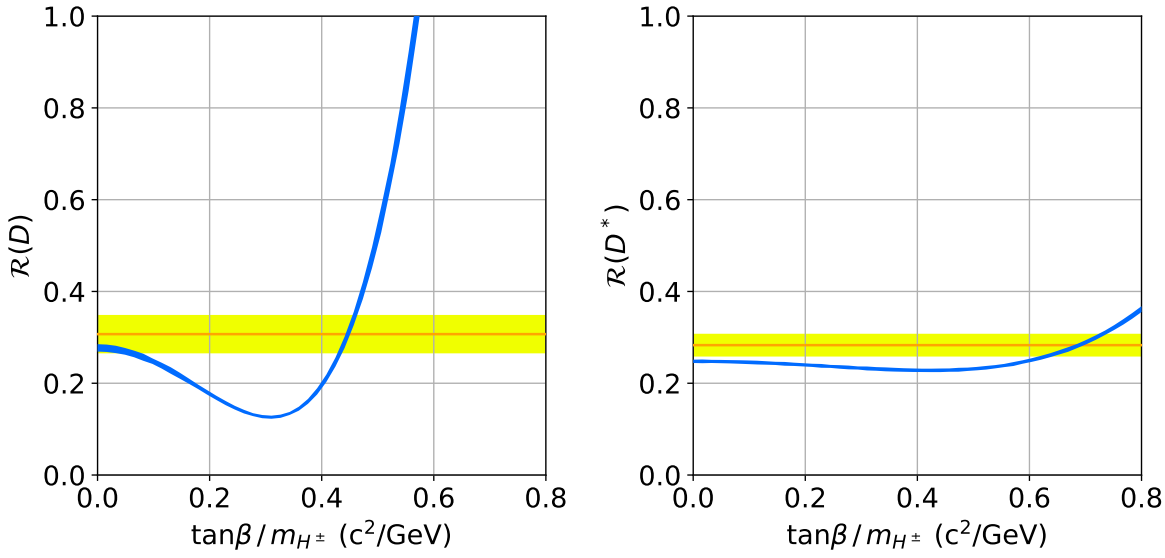


Figure 7.13: Compatibility of our results with 2HDM-II. The predictions for the 2HDM-II are shown in blue with a 1σ band and have been obtained with the `flavio` package [34]. Our results are shown in yellow and orange with a 1σ band that refers to the sum in quadrature of statistical and systematic uncertainties.

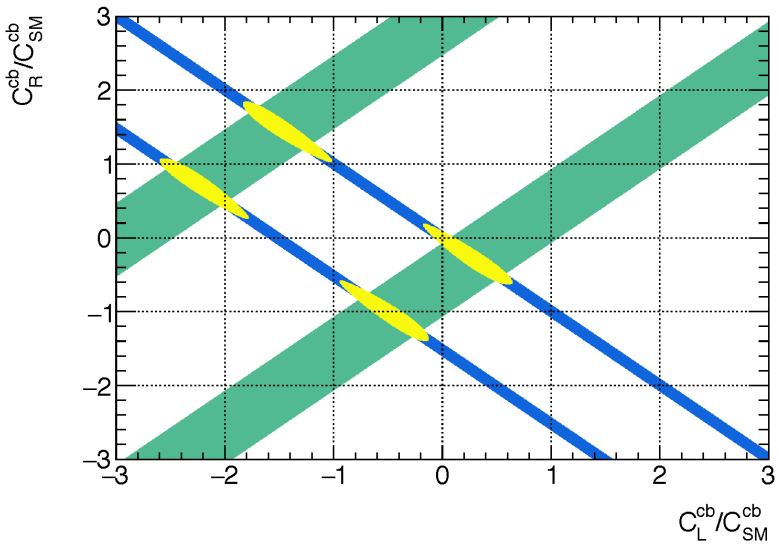


Figure 7.14: $C_L^{cb}C_{\rm SM}^{cb}$-$C_R^{cb}C_{\rm SM}^{cb}$ plane for real values of $C_L^{cb}C_{\rm SM}^{cb}$ and $C_R^{cb}C_{\rm SM}^{cb}$.

Chapter 8

Conclusion

This thesis presents the measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays – where ℓ is either e or μ – using semileptonic tagging channels and leptonic τ decays exclusively. It is performed on the full dataset on the $\Upsilon(4S)$ resonance of the Belle experiment.

In the past these measurements have been carried out using hadronic tags, and this work is the first analysis that uses a semileptonic tag for a combined measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$. Furthermore, with respect to the previous semileptonic measurement of $R(D^{*+})$ by Belle [44], this analysis uses a larger number of B_{tag} channels, which directly translates to a larger analysis dataset.

Our results are

$$\mathcal{R}(D) = 0.307 \pm 0.037 \pm 0.016 \tag{8.1}$$

$$\mathcal{R}(D^*) = 0.283 \pm 0.018 \pm 0.014, \tag{8.2}$$

where the first uncertainty is statistical and provided by the fit, and the second error is systematic. This is the single most precise measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ ever performed. The results are in agreement with the previous Belle measurement of $\mathcal{R}(D^*)$ performed with a semileptonic tag, which is now superseded.

The goal was to test the compatibility of this experimental data with the SM, whose expectation values are

$$\mathcal{R}(D)_{\text{SM}} = 0.299 \pm 0.003 \tag{8.3}$$

$$\mathcal{R}(D^*)_{\text{SM}} = 0.258 \pm 0.005. \tag{8.4}$$

Our results for $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are in agreement with the SM predictions within 0.2σ and 1.1σ respectively. The combination of our $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ results is compatible with the SM within 1.2σ. Before these results, the experimental $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ world average showed a discrepancy of approximately 4σ with the SM expectations. However, given the compatibility of our results with the SM and their high precision, this discrepancy is reduced to 3σ when including these latest results.

Bibliography

[1] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[2] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[3] ATLAS collaboration, G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[4] CMS collaboration, S. Chatrchyan et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[5] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652. eprint: /oup/backfile/content_public/journal/ptp/49/2/10.1143/ptp.49.652/2/49-2-652.pdf. URL: http://dx.doi.org/10.1143/PTP.49.652.

[6] BABAR collaboration, B. Aubert et al., "Observation of *CP* Violation in the B^0 Meson System". In: *Phys. Rev. Lett.* 87 (9 Aug. 2001), p. 091801. DOI: 10.1103/PhysRevLett.87.091801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.091801.

[7] Belle collaboration, K. Abe et al., "Observation of Large *CP* Violation in the Neutral *B* Meson System". In: *Phys. Rev. Lett.* 87 (9 Aug. 2001), p. 091802. DOI: 10.1103/PhysRevLett.87.091802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.091802.

[8] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967). [Usp. Fiz. Nauk161,no.5,61(1991)], pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[9] J. F. Donoghue, E. Golowich, and B. R. Holstein. "Dynamics of the standard model". In: *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 2 (1992). [Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.35(2014)], pp. 1–540. DOI: 10.1017/CBO9780511524370.

[10] Particle Data Group collaboration, M. Tanabashi et al., "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[11] S. Bifani et al. "Review of Lepton Universality tests in B decays". In: (2018). arXiv: 1809.06229 [hep-ex].

[12] K. Hagiwara, A. Martin, and M. Wade. "Exclusive semileptonic B-meson decays". In: *Nuclear Physics B* 327.3 (1989), pp. 569–594. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(89)90306-4. URL: http://www.sciencedirect.com/science/article/pii/0550321389903064.

[13] C. G. Boyd, B. Grinstein, and R. F. Lebed. "Constraints on form-factors for exclusive semileptonic heavy to light meson decays". In: *Phys. Rev. Lett.* 74 (1995), pp. 4603–4606. DOI: 10.1103/PhysRevLett.74.4603. arXiv: hep-ph/9412324 [hep-ph].

[14] I. Caprini, L. Lellouch, and M. Neubert. "Dispersive bounds on the shape of anti-B to D(*) lepton anti-neutrino form-factors". In: *Nucl. Phys.* B530 (1998), pp. 153–181. DOI: 10.1016/S0550-3213(98)00350-2. arXiv: hep-ph/9712417 [hep-ph].

[15] C. Bourrely, I. Caprini, and L. Lellouch. "Model-independent description of $B \to \pi\ell\nu$ decays and a determination of $|V(ub)|$". In: *Phys. Rev.* D79 (2009). [Erratum: Phys. Rev.D82,099902(2010)], p. 013008. DOI: 10.1103/PhysRevD.82.099902, 10.1103/PhysRevD.79.013008. arXiv: 0807.2722 [hep-ph].

[16] M. Neubert. "Heavy quark effective theory". In: *Subnucl. Ser.* 34 (1997), pp. 98–165. arXiv: hep-ph/9610266 [hep-ph].

[17] CLEO collaboration, A. Anastassov et al., "Investigation of Semileptonic B Meson Decays to p-Wave Charm Mesons". In: *Phys. Rev. Lett.* 80 (19 May 1998), pp. 4127–4131. DOI: 10.1103/PhysRevLett.80.4127. URL: http://link.aps.org/doi/10.1103/PhysRevLett.80.4127.

[18] D0 collaboration, V. M. Abazov et al., "Measurement of semileptonic branching fractions of B mesons to narrow D^{**} states". In: *Phys. Rev. Lett.* 95 (2005), p. 171803. DOI: 10.1103/PhysRevLett.95.171803. arXiv: hep-ex/0507046 [hep-ex].

[19] J. Link et al. "Measurement of masses and widths of excited charm mesons D_2^* and evidence for broad states". In: *Physics Letters B* 586.1–2 (2004), pp. 11–20. ISSN: 0370-2693. DOI: http://dx.doi.org/10.1016/j.physletb.2004.02.017. URL: http://www.sciencedirect.com/science/article/pii/S0370269304003016.

[20] Belle collaboration, D. Liventsev et al., "Study of $\bar{B} \to D^{**}\ell^-\bar{\nu}_\ell$ with full reconstruction tagging". In: *Phys. Rev.* D77 (2008), p. 091503. DOI: 10.1103/PhysRevD.77.091503. arXiv: 0711.3252 [hep-ex].

[21] BaBar collaboration, B. Aubert et al., "Measurement of the Branching Fractions of $\bar{B} \to D^{(**)}\ell^-\bar{\nu}$ Decays in Events Tagged by a Fully Reconstructed B Meson". In: *Phys. Rev. Lett.* 101 (2008), p. 261802. DOI: 10.1103/PhysRevLett.101.261802. arXiv: 0808.0528 [hep-ex].

[22] BaBar collaboration, B. Aubert et al., "Measurement of Semileptonic B Decays into Orbitally-Excited Charmed Mesons". In: *Phys. Rev. Lett.* 103 (2009), p. 051803. DOI: 10.1103/PhysRevLett.103.051803. arXiv: 0808.0333 [hep-ex].

[23] LHCb collaboration, R. Aaij et al., "Study of D_J meson decays to $D^+\pi^-$, $D^0\pi^+$ and $D^{*+}\pi^-$ final states in pp collision". In: *JHEP* 09 (2013), p. 145. DOI: 10.1007/JHEP09(2013)145. arXiv: 1307.4556.

[24] CDF collaboration, A. Abulencia et al., "Measurement of mass and width of the excited charmed meson states D0(1) and D*0(2) at CDF". In: *Phys. Rev.* D73 (2006), p. 051104. DOI: 10.1103/PhysRevD.73.051104. arXiv: hep-ex/0512069 [hep-ex].

[25] DELPHI collaboration, J. Abdallah et al., "Determination of heavy quark nonperturbative parameters from spectral moments in semileptonic B decays". In: *Eur. Phys. J.* C45 (2006), pp. 35–59. DOI: 10.1140/epjc/s2005-02406-7. arXiv: hep-ex/0510024 [hep-ex].

[26] Belle collaboration, K. Abe et al., "Study of $B^- \to D^{**0}\pi^-(D^{**0} \to D^{(*)+}\pi^-)$ decays". In: *Phys. Rev.* D69 (11 June 2004), p. 112002. DOI: 10.1103/PhysRevD.69.112002. URL: http://link.aps.org/doi/10.1103/PhysRevD.69.112002.

[27] BaBar collaboration, B. Aubert et al., "Dalitz Plot Analysis of $B^- \to D^+\pi^-\pi^-$". In: *Phys. Rev.* D79 (2009), p. 112004. DOI: 10.1103/PhysRevD.79.112004. arXiv: 0901.1291 [hep-ex].

[28] K. G. Wilson and W. Zimmermann. "Operator product expansions and composite field operators in the general framework of quantum field theory". In: *Communications in Mathematical Physics* 24.2 (June 1972), pp. 87–106. ISSN: 1432-0916. DOI: 10.1007/BF01878448. URL: https://doi.org/10.1007/BF01878448.

[29] J. F. Gunion et al. "The Higgs Hunter's Guide". In: *Front. Phys.* 80 (2000), pp. 1–404.

[30] S. P. Martin. "A Supersymmetry primer". In: (1997). [Adv. Ser. Direct. High Energy Phys.18,1(1998)], pp. 1–98. DOI: 10.1142/9789812839657_0001, 10.1142/9789814307505_0001. arXiv: hep-ph/9709356 [hep-ph].

[31] M. Trodden. "Electroweak baryogenesis: A Brief review". In: *Proceedings, 33rd Rencontres de Moriond 98 electrowek interactions and unified theories: Les Arcs, France, Mar 14-21, 1998*. 1998, pp. 471–480. arXiv: hep-ph/9805252 [hep-ph].

[32] Belle collaboration, A. Limosani et al., "Measurement of Inclusive Radiative B-meson Decays with a Photon Energy Threshold of 1.7-GeV". In: *Phys. Rev. Lett.* 103 (2009), p. 241801. DOI: 10.1103/PhysRevLett.103.241801. arXiv: 0907.1384 [hep-ex].

[33] BaBar collaboration, J. P. Lees et al., "Measurement of B($B \to X_s\gamma$), the $B \to X_s\gamma$ photon energy spectrum, and the direct CP asymmetry in $B \to X_{s+d}\gamma$ decays". In: *Phys. Rev.* D86 (2012), p. 112008. DOI: 10.1103/PhysRevD.86.112008. arXiv: 1207.5772 [hep-ex].

[34] D. M. Straub. "flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond". In: (2018). arXiv: 1810.08132 [hep-ph].

[35] BaBar collaboration, J. P. Lees et al., "Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons". In: *Phys. Rev.* D88.7 (2013), p. 072012. DOI: 10.1103/PhysRevD.88.072012. arXiv: 1303.0571 [hep-ex].

[36] Belle collaboration, M. Huschle et al., "Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle". In: *Phys. Rev.* D92.7 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. arXiv: 1507.03233 [hep-ex].

[37] W. Buchmuller, R. Ruckl, and D. Wyler. "Leptoquarks in Lepton - Quark Collisions". In: *Phys. Lett.* B191 (1987). [Erratum: Phys. Lett.B448,320(1999)], pp. 442–448. DOI: 10.1016/S0370-2693(99)00014-3,10.1016/0370-2693(87)90637-X.

[38] Y. Sakaki et al. "Testing leptoquark models in $\bar{B} \to D^{(*)}\tau\bar{\nu}$". In: *Phys. Rev.* D88.9 (2013), p. 094012. DOI: 10.1103/PhysRevD.88.094012. arXiv: 1309.0301 [hep-ph].

[39] M. Tanaka and R. Watanabe. "New physics in the weak interaction of $\bar{B} \to D^{(*)}\tau\bar{\nu}$". In: *Phys. Rev.* D87.3 (2013), p. 034028. DOI: 10.1103/PhysRevD.87.034028. arXiv: 1212.1878 [hep-ph].

[40] HFLAV. URL: http://www.slac.stanford.edu/xorg/hflav/semi/index.html.

[41] BaBar collaboration, J. P. Lees et al., "Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays". In: *Phys. Rev. Lett.* 109 (2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802. arXiv: 1205.5442 [hep-ex].

[42] LHCb collaboration, R. Aaij et al., "Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$". In: *Phys. Rev. Lett.* 115 (11 Sept. 2015), p. 111803. DOI: 10.1103/PhysRevLett.115.111803. URL: http://link.aps.org/doi/10.1103/PhysRevLett.115.111803.

[43] LHCb collaboration, R. Aaij et al., "Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays". In: *Phys. Rev.* D97.7 (2018), p. 072013. DOI: 10.1103/PhysRevD.97.072013. arXiv: 1711.02505 [hep-ex].

[44] Belle collaboration, Y. Sato et al., "Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method". In: *Phys. Rev.* D94.7 (2016), p. 072007. DOI: 10.1103/PhysRevD.94.072007. arXiv: 1607.07923 [hep-ex].

[45] Belle collaboration, S. Hirose et al., "Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle". In: *Phys. Rev.* D97.1 (2018), p. 012004. DOI: 10.1103/PhysRevD.97.012004. arXiv: 1709.00129 [hep-ex].

[46] T. Abe et al. "Achievements of KEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (2013), 03A001. DOI: 10.1093/ptep/pts102. eprint: /oup/backfile/content_public/journal/ptep/2013/3/10.1093/ptep/pts102/2/pts102.pdf. URL: http://dx.doi.org/10.1093/ptep/pts102.

[47] S. Kurokawa and E. Kikutani. "Overview of the KEKB accelerators". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 499.1 (2003). KEK-B: The KEK B-factory, pp. 1–7. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01771-0. URL: http://www.sciencedirect.com/science/article/pii/S0168900202017710.

[48] A. Abashian et al. "The Belle Detector". In: *Nucl. Instrum. Meth.* A479 (2002), pp. 117–232. DOI: 10.1016/S0168-9002(01)02013-7.

[49] Belle collaboration, A. J. Bevan et al., "The Physics of the B Factories". In: *Eur. Phys. J.* C74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9. arXiv: 1406.6311 [hep-ex].

[50] Y. Ushiroda. "Belle silicon vertex detectors". In: *Nucl. Instrum. Meth.* A511 (2003), pp. 6–10. DOI: 10.1016/S0168-9002(03)01739-X.

[51] E. Nakano. "Belle PID". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 494.1 (2002). Proceedings of the 8th International Conference on Instrumentatio n for Colliding Beam Physics, pp. 402–408. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(02)01510-3. URL: http://www.sciencedirect.com/science/article/pii/S0168900202015103.

[52] J. Brodzicka, T. Browder, and P. Chang. "Physics achievements from the Belle experiment". In: *Progress of Theoretical and Experimental Physics* 2012.1 (2012), p. 04D001. DOI: 10.1093/ptep/pts072. eprint: /oup/backfile/content_public/journal/ptep/2012/1/10.1093_ptep_pts072/1/pts072.pdf. URL: http://dx.doi.org/10.1093/ptep/pts072.

[53] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4. URL: http://www.sciencedirect.com/science/article/pii/S0168900201000894.

[54] T. Sjostrand et al. "High-energy physics event generation with PYTHIA 6.1". In: *Comput. Phys. Commun.* 135 (2001), pp. 238–259. DOI: 10.1016/S0010-4655(00)00236-8. arXiv: hep-ph/0010017 [hep-ph].

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

[56] E. Barberio, B. van Eijk, and Z. Was. "Photos — a universal Monte Carlo for QED radiative corrections in decays". In: *Computer Physics Communications* 66.1 (1991), pp. 115–128. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(91)90012-A. URL: http://www.sciencedirect.com/science/article/pii/001046559190012A.

[57] T. Keck. "Machine learning algorithms for the Belle II experiment and their validation on Belle data". PhD thesis. Karlsruher Institut für Technologie (KIT), 2017. 240 pp. DOI: 10.5445/IR/1000078149.

[58] G. Ciezarek et al. "A Challenge to Lepton Universality in B Meson Decays". In: *Nature* 546 (2017), pp. 227–233. DOI: 10.1038/nature22346. arXiv: 1703.01766 [hep-ex].

[59] H. Nakano, A. Ishikawa, and K. Sumisawa. *K^0_S selection with NeuroBayes and nisKsFinder class.* Belle note no 1253. 2012.

[60] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth.* A262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[61] G. C. Fox and S. S.Wolfram. "Observables for the Analysis of Event Shapes ine+e-Annihilation and Other Processes". In: *Physical Review Letters* 41.23 (Dec. 1978), pp. 1581–1585. DOI: 10.1103/physrevlett.41.1581.

[62] W. Verkerke and D. P. Kirkby. "The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[63] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *ArXiv e-prints* (Mar. 2016). arXiv: 1603.02754 [cs.LG].

[64] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324.

[65] Y. Freund and R. E. Schapire. "Experiments with a New Boosting Algorithm". In: *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. ICML'96. Bari, Italy: Morgan Kaufmann Publishers Inc., 1996, pp. 148–156. ISBN: 1-55860-419-7. URL: http://dl.acm.org/citation.cfm?id=3091696.3091715.

[66] J. H. Friedman. "Greedy function approximation: A gradient boosting machine." In: *Ann. Statist.* 29.5 (Oct. 2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451.

[67] N. V. Chawla et al. "SMOTE: Synthetic Minority Over-sampling Technique". In: *ArXiv e-prints* (June 2011). arXiv: 1106.1813 [cs.AI].

[68] R. Barlow. "Extended maximum likelihood". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 297.3 (1990), pp. 496–506. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(90)91334-8. URL: http://www.sciencedirect.com/science/article/pii/0168900290913348.

[69] L. Hinz. *Lepton ID efficiency correction and systematic error*. Belle note no 954. 2006.

[70] C. Oswald et al. *"Modelling of semileptonic $B_{(s)}$ decays"*. Belle note no 1335. 2014.

[71] D. Scora and N. Isgur. "Semileptonic meson decays in the quark model: An update". In: *Phys. Rev. D* 52 (5 Sept. 1995), pp. 2783–2812. DOI: 10.1103/PhysRevD.52.2783. URL: https://link.aps.org/doi/10.1103/PhysRevD.52.2783.

[72] A. K. Leibovich et al. "Semileptonic B decays to excited charmed mesons". In: *Phys. Rev. D* 57 (1 Jan. 1998), pp. 308–330. DOI: 10.1103/PhysRevD.57.308. URL: https://link.aps.org/doi/10.1103/PhysRevD.57.308.

[73] A. Crivellin, C. Greub, and A. Kokulu. "Explaining $B \to D\tau\nu$, $B \to D^*\tau\nu$ and $B \to \tau\nu$ in a 2HDM of type III". In: *Phys. Rev.* D86 (2012), p. 054014. DOI: 10.1103/PhysRevD.86.054014. arXiv: 1206.2634 [hep-ph].