Search for lepton flavour violation in B meson decays at LHCb

Lino FERREIRA LOPES

No mind to think.
No will to break.
No voice to cry suffering.
Born of God and Void.
— The Pale King

A mes parents et ma soeur…

Acknowledgements

After more than four years, I finally sit down and reflect about the journey that has unraveled towards the goal of adding "Dr." to my emails' signature. The path has been long and arduous and there have been a few moments of doubt but in the end a thesis was eventually written. I am not usually very good at expressing my feelings, except maybe for complaining, but I will try to make an effort.

First, I would like of course to thank the most experienced people in the lab, namely Aurelio, Olivier, Tatsuya, Guido and Fred for giving me the opportunity of doing a Ph.D. in this lab. It seems that my Master thesis work, done at LPHE in 2016–2017, was perhaps not so bad if they were persuaded to keep me for four more years. They are the backbone of the lab and it was always a treat seeing the discussions that were sparked during either the Monday meeting or the old reliable lunch at banane. While Aurelio and Tatsuya did not stick around until the end of my thesis, I always keep a thought about them in my mind and of course welcome the "replacement" Lesya.

I would like to thank our amazing secretaries Erika, Esther and Corinne, who are always so happy and upbeat, and who are always so helpful. Sometimes needing to take care of us like we are their children.

The LPHE truly is a great lab with an amazing work environment, where it is not shameful to ask the most trivial questions, which I did plenty of times. Everyone is very kind and ready to help. The various people of the lab, past and present, are really what makes the soul of the lab and what brought me a smile whenever I came to cubotron. There always were nice conversations around a coffee, at lunch, whenever there was cake or around a beer. Truly I would like to thank everyone that I've met in the lab as everyone was fantastic but I suppose that the list of names would be a bit too long and I would definitely forget someone. From the old lab, I would like to thank in particular the "quatuor du LPHE", *i.e.* Axel, Olivier Girard, Maxime and Brice and the fact that they introduced me to the wonders of the post-lunch soccer physics tournaments. I would like to also give a special mention to Guido Andreassi and Vincenzo. From the new blood, I would like to thank in particular Serhii and Surapat for introducing me on the path of the gainzz as well as Vlado, Sara, Carina, Veronica, Marie, Maria, Maarten and Ettore.

Naturally, I would like to give a huge thanks to my office mates Matthieu, Guillaume, Sonia and Sebastian (granted for a short time only) who have had the displeasure of listening to my

Acknowledgements

Abstract

The current limitations of the standard model of elementary particle physics drive the search for phenomenons forbidden within that framework with the goal to unveil a more universal theory. This thesis presents the search for two lepton flavour violating decays forbidden in the standard model: $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$. The search is performed by studying proton-proton collision data collected by the detector of the LHCb collaboration corresponding to $9.1 \, \text{fb}^{-1}$ of integrated luminosity. No significant signal is observed and as such the following upper limits at 90% confidence level are set: $\mathcal{B}(B^+ \to K^+ \tau^- \mu^+) < 1.3 \times 10^{-6}$ and $\mathcal{B}(B^+ \to K^+ \tau^+ \mu^-) < 1.2 \times 10^{-6}$. These limits correspond to one order of magnitude improvement relatively to the current world's best values for these decay modes.

Keywords: elementary particle physics, CERN, LHC, LHCb, flavour physics, lepton-flavour violation.

Résumé

Les limitations actuelles du modèle standard de la physique des particules élémentaires poussent à la recherche de phénomènes interdits dans ce contexte, avec pour objectif l'élaboration d'une théorie plus universelle. Cette thèse présente la recherche de deux modes de désintégration ne respectant pas la conservation des nombres leptoniques, à savoir : $B^+ \rightarrow K^+ \tau^- \mu^+$ et $B^+ \rightarrow K^+ \tau^+ \mu^-$. Cette recherche est effectuée en étudiant des données de collision proton-proton qui ont été collectées par le détecteur de la collaboration LHCb et qui correspondent à $9.1\,\mathrm{fb}^{-1}$ de luminosité intégrée. Aucun signal significatif n'est observé et en conséquence les limites supérieures suivantes, correspondant à un interval de confiance de 90%, sont déterminées : $\mathscr{B}(B^+ \rightarrow K^+ \tau^- \mu^+) < 1.3 \times 10^{-6}$ et $\mathscr{B}(B^+ \rightarrow K^+ \tau^+ \mu^-) < 1.2 \times 10^{-6}$. Ces limites correspondent à une amélioration d'un ordre de grandeur relativement aux meilleures limites actuelles sur ces modes de désintégrations.

Mots-clefs : physique des particules élémentaires, CERN, LHC, LHCb, physique de la saveur, violation de la saveur leptonique.

Contents

Contents

1 Introduction and theoretical motivation

In this chapter the current theoretical framework used in particle physics, known as the standard model is presented, and its limitations are introduced. These limitations drive the search for phenomenons which can only be described by so-called "new physics" *i.e.* by an extended theory with respect to the standard model. The core work of this thesis is a search for decays which can only be explained by new physics. The search for the decays $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ is presented in Chapters 4 and 5. Additionally, work has been performed in research and development of scintillating fibre mats used to upgrade the tracking system of the LHCb detector for the next periods of data taking. The current LHCb detector as well as its upgrade are presented in Chapters 2 and 3.

1.1   Standard model

The framework describing our current understanding of elementary particles and their interactions is called the standard model (SM) of particle physics. It is a quantum field theory describing the properties and behaviours of the elementary components of matter as well as three of the four fundamental forces of the universe. It describes the strong, electromagnetic and weak forces but does not describe gravity. Nevertheless, the fact that gravity is not described by the SM is not a problem from a practical point of view since its effect at the energy scales currently accessible experimentally is negligible. It is however an important shortcoming of the SM in the context of a search for a theory which fully describes elementary particles and their interactions. Despite this and other shortcomings, which will be mentioned later on, the SM is a very successful theory and most of its predictions agree well with the current experimental measurements. In particular, the existence of every single elementary particle that was postulated has been confirmed experimentally. The last particle being the Higgs boson, which was discovered in 2012 by the ATLAS and CMS collaborations [1, 2].

The SM describes the properties and interactions of elementary particles. A particle is considered elementary if it has no underlying sub-structure, otherwise it is called composite. The elementary particles present in the SM can be split into two categories according to their spin structure. Fermions have a half integer spin and follow Fermi-Dirac statistics while bosons have an integer spin and follow Bose-Einstein statistics. Within the fermion category we further distinguish between quarks and leptons, which both have spin 1/2. There are six different types, or flavours, of quarks and leptons, as well as their corresponding anti-particles. Amongst the quarks we distinguish between the up-type quarks which carry an electric charge of $+2/3$ and down-type quarks which carry a charge of $-1/3$, where the charge is expressed in term of the charge of the positron. The up-type quarks are the up, charm and top quarks, denoted u, c and t while the down-type quarks are the down, strange and bottom quarks, denoted d, s and b, respectively. The lepton sector is built analogously with the charged leptons carrying an electric charge of -1 and their corresponding massless neutral neutrinos. These are the electron, muon and tau denoted e, μ and τ and the corresponding electron, muon and tau neutrinos, denoted ν_e, ν_μ and ν_τ. Both the quarks and leptons are split into three families, or generations, where a generation is a doublet composed of one up-type and one down-type quarks or similarly one charged lepton and its corresponding neutrino. The first, second and third generations are the following : (u, d), (c, s) and (t, b) for the quarks, (e, ν_e), (μ, ν_μ) and (τ, ν_τ) for the leptons. The particles in the second generation have a mass greater than those in the first while particles in the third generation have a mass greater than those in both the first and second generation. As a result, the particles in the second and third generations are unstable and end-up decaying into particles of the lower generations. Each quark has an assigned quantum number called baryon number which has a value of $+1/3$ for quarks and $-1/3$ for anti-quarks. The baryon number is conserved by all the interactions of the SM and explains why a quark cannot decay into a lepton, since the latter carries a baryon number of 0. In the case of the leptons, a separate quantum number called lepton flavour is assigned to each of the generations : L_e for (e, ν_e), L_μ for (μ, ν_μ) and L_τ for (τ, ν_τ). The lepton flavour has a value of $+1$ for particles and -1 for anti-particles and it is individually conserved in every interaction of the SM. Both the baryon and lepton numbers are accidental symmetries which are not imposed directly by the gauge structure of the theory.

The bosons in the SM come from the gauge structure of the theory, with the exception of the Higgs, and are the mediators of the three fundamental forces. The SM is a quantum field theory based on the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The subscripts C, L and Y stand respectively for colour, left-handed and hypercharge. The strong interaction is described by the $SU(3)_C$ component while the weak and electromagnetic interactions are unified as the $SU(2)_L \times U(1)_Y$ electroweak component. In the strong sector the bosons mediating the interactions are the gluons. There are 8 of them, they are massless and each of them carries simultaneously a colour and an anti-colour charge. There are three colours, denoted red, green and blue, and their corresponding anti-colours. Quarks carry a single colour while leptons

are colour-less. This is why quarks are sensitive to the strong force while leptons are not. It should also be noted that neither quarks nor gluons can be observed as free particles. This is due to a property of the strong force known as colour confinement. As a result they can only be observed as hadrons which are colour-neutral combinations of quarks and gluons. The most common hadrons are mesons which carry a quark and an anti-quark and baryons which carry three quarks.

In the electroweak sector there are originally 4 massless gauge bosons W_μ^1, W_μ^2, W_μ^3 and B_μ from the gauge symmetry. A complex scalar Higgs doublet is added to generate mass through the mechanism of spontaneous symmetry breaking. After symmetry breaking, three of the four degrees of freedom from the Higgs doublet are absorbed by the electroweak gauge bosons which results in three massive weak bosons, which are the W^+, W^- and Z^0, and a massless photon γ which carries the electromagnetic interaction. The remaining degree of freedom results in the Higgs boson H. The Higgs boson gives mass to the fermions through the Yukawa interaction, where the coupling of the Higgs to the fermions is proportional to the fermion's mass. The Higgs boson is a spin 0 particle while the other gauge bosons are spin 1 particles. The photon mediates the electromagnetic force and as such only couples to electrically charged particles. It also conserves the flavour, as is the case with the Z^0. The W^+ and W^- on the other hand do not conserve flavour during interactions. An important property of the SM is that the coupling constants between gauge bosons and leptons of any flavour are equal for a given boson. This is known as lepton flavour universality and it is an accidental symmetry of the SM which does not come directly from the gauge structure of the SM. All the particles of the SM are shown schematically in Fig. 1.1.

1.2   Beyond the standard model

Currently, the standard model is the best description of elementary particles and their interactions. The most stringent test of the theory prediction is for the electron's anomalous magnetic moment where the predicted and measured values agree to more than 10 significant digits [4]. Nevertheless, the SM has known limitations and needs to be improved if one desires to have a theory which provides a full description of elementary particles and their interactions. The following is a list of some of the most prominent issues of the SM :

- Gravity is completely ignored in the SM. This is due to the difficulty in combining the theory of general relativity, which is our best description of gravity, with quantum field theories. Indeed, trying to apply a similar quantisation procedure as is done with the other three forces in order to produce the hypothetical graviton results in a non-renormalisable theory [5]. Additionally, current particle physics experiments do not have the precision necessary to probe graviton models since gravity is extremely weak at
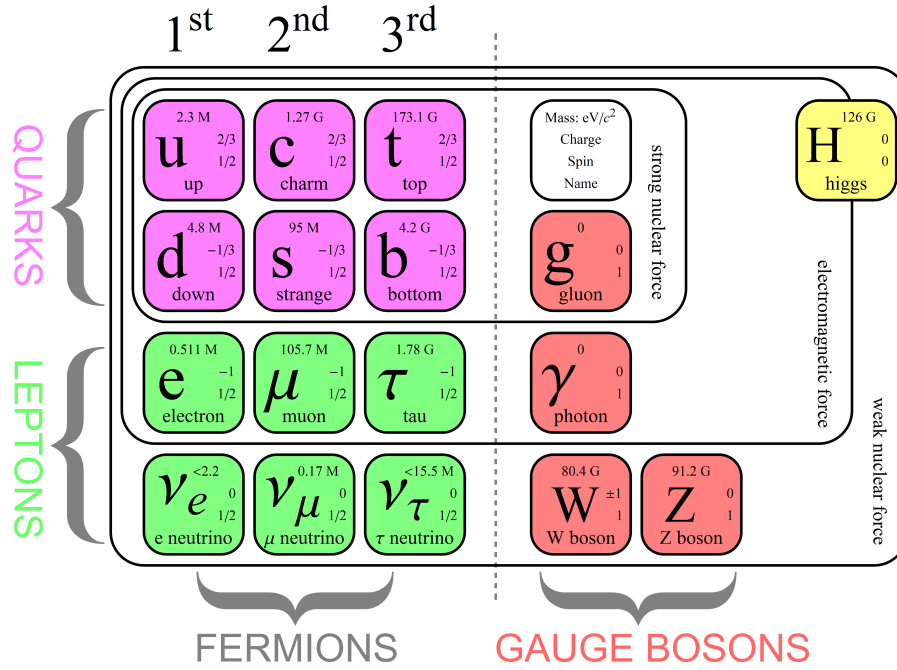
Figure 1.1 – Elementary particles of the SM. Image taken from Ref. [3].

the energy scales that are being probed.

- The existence of dark matter (DM) is now well established thanks to measurements from astrophysics and cosmology [6]. For example it has been observed that the rotational velocity of objects in a stable orbit around a galaxy is higher than what would be predicted by assuming only the existence of luminous matter. There is no candidate in the SM with properties and abundance which could correspond to those required by DM. From cosmology it is known that DM corresponds to about 25% of the energy density content of the universe with the usual baryonic matter described by the SM accounting only for ~ 5%. The remaining energy density comes from dark energy (see below).

- Cosmological observations show that the visible universe is expanding and that this expansion is accelerating [7, 8]. If the universe contained only matter, visible and dark, we would expect the expansion to slow down. Since this is not the case, there has to be a form of repulsive energy density which we call dark energy. This dark energy accounts for roughly 70% of the total energy density in the universe. Again, the SM is unable to provide any explanation for the existence of this dark energy.

- Charge-Parity (*CP*) violation is one of the three Sakharov conditions necessary to generate matter-antimatter asymmetry in the universe [9]. Even though *CP* violation is allowed in the SM, it is orders of magnitude below what is required to explain the

observed matter-antimatter asymmetry in the universe [10].

- The Lagrangian of the strong interaction allows for a *CP* violating term, yet experimentally the strong interaction is *CP* conserving to a high degree of precision [11, 12]. This is known as the strong *CP* problem.

- Why is there such a large difference between the scales of each interaction? In particular there is a difference of 33 orders of magnitude between the couplings of the gravitational and weak forces. This is known as the hierarchy problem. It is related to the issue as to why the Higgs mass of $\sim 100\,\text{GeV}/c^2$ is so light compared to the Planck scale of $\sim 10^{19}\,\text{GeV}/c^2$, where we expect to see the manifestations of the quantum effects of gravity. Indeed, from the point of view of an ultra-violet completion of the SM we expect that loop Feynman diagrams would make the Higgs mass diverge unless there is a great amount of fine-tuning [13].

- In the SM the neutrinos are all massless. However, measurements show that the neutrinos' flavour can oscillate, *i.e.* a neutrino of a given flavour can turn into a neutrino of a different flavour [14, 15, 16]. This implies that the neutrinos have mass and that the lepton flavour is not an exact symmetry. The origin of the mass of the neutrinos is still unclear.

Another property of the SM which is currently under scrutiny is lepton flavour universality, *i.e.* the assumption that the coupling between electroweak gauge bosons and leptons does not depend on the lepton's generation. Indeed, in the flavour physics sector, there have been many recent measurements [17, 18, 19, 20, 21] which display tensions with the hypothesis of lepton-flavour universality of the SM. There has been a lot of work from the theory community to provide extensions of the SM which could allow for lepton non-universality. It is often the case that models allowing lepton non-universality also violate lepton flavour [22]. Lepton-flavour violation (LFV) is the phenomenon in which a transition does not conserve the individual lepton flavour. As previously discussed, nowadays it is known that neutrinos can oscillate and thus lepton-flavour violating decays have been observed. However, the predicted branching fractions for lepton-flavour violating decays of charged leptons, and which proceed only through neutrino oscillations are extremely small, *e.g.* $\mathcal{B}(\mu^- \to e^- \gamma) \sim 10^{-54}$ [23]. As such, any significant measurement of a LFV transition would be a clear sign of new physics (NP) and would help constraining current NP models.

This thesis presents the search for two decay modes forbidden in the SM, $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$. The choice of these decay modes is motivated by the current anomalies observed in flavour physics. In particular, one of the main fields of research at the LHCb experiment currently is the study of the hypothesis of lepton flavour universality of the SM. As previously mentioned, there are various hints of lepton flavour non-universality which

have been observed already, however there is currently no single measurement which would disprove by itself this assumption of the SM. Perhaps the most convincing single measurement against lepton universality is the recently updated measurement of the R_K parameter, using the full data set available at LHCb, whose measured and predicted values display more than 3 standard deviations disagreement [24]. This parameter R_K represents the ratio of branching fractions of transitions of a B^+ to a K^+ and either a pair of oppositely charged muons or electrons. The measured value seems to suggest that the transition involving electrons is favoured with respect to the one with muons, while the SM predicts these transition rates to almost perfectly match. This type of ratio is a powerful tool to probe the existence of lepton non-universal effects. The LHCb experiment is currently performing measurements involving a variety of these ratios, both involving transitions at loop level in the SM, as is the case for R_K, as well as transitions occurring at tree level involving beauty to charm meson decays.

However, these ratios are not the only type of measurements being performed. Transitions involving electroweak decays proceeding only at loop-level in the SM are being thoroughly studied as well. The reason these decays are particularly interesting is that since they are forbidden at tree level in the SM, it is possible that transitions at tree level involving new heavy mediators would give sizeable contributions to various observables such as branching fractions or angular variables. Another type of transitions being studied is that of radiative decays, *i.e.* decays where a photon is emitted. In particular, the study of the photon polarisation is interesting since in the SM the photons are predominantly left-handed while NP could in principle allow new mediators to couple to photons with a right-handed chirality. Finally, the type of transitions that are being studied in this thesis, *i.e.* lepton-flavour violating decays. It is primordial that all of these various types of measurements are performed as they not only help in improving the chances of observing a strong anomaly in a given observable but also help in constraining the overall picture of NP phenomena. Indeed, given a single anomaly it is possible to build a plethora of NP models which could explain it. Naturally, increasing the number as well as diversifying the types of observables studied helps in constraining the various NP models and obtain an overall coherent picture of NP effects.

A search for lepton-flavour violating decays using data collected by the LHCb detector, is presented in this thesis. A search for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decays is performed, where $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ or $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$.[1] From a theoretical point of view, the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes can have different branching fractions while from the experimental side a different background composition is expected in data when reconstructing these modes and as such they need to be treated separately. Various NP models predict a higher branching fraction for decay modes containing leptons of the third generation compared to modes containing only leptons of the first and second generation [25]. There are multiple NP models which predict the existence of these two decay modes and the predicted branching

---

[1]The inclusion of charge-conjugated processes is implied throughout unless otherwise specified.

Figure 1.2 – Feynman diagrams describing the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decays on the left and right, respectively. These diagrams assume the presence of a vector leptoquark mediating the transition and denoted U_1 based on the model from Ref. [29].

fractions span a sizeable range up to $\mathcal{O}(10^{-5})$ [26, 27, 28, 29]. Hypothetical Feynman diagrams describing the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ transitions based on the vector leptoquark model from Ref. [29] are given for illustration in Fig. 1.2. Experimentally, upper limits have been set on these decays by the BaBar [30] and LHCb [31] collaborations. The limits at 90% confidence level (CL) from the BaBar collaboration are

$$\mathcal{B}(B^+ \to K^+ \tau^- \mu^+) < 4.5 \times 10^{-5} \text{ at } 90\% \text{ CL},$$
$$\mathcal{B}(B^+ \to K^+ \tau^+ \mu^-) < 2.8 \times 10^{-5} \text{ at } 90\% \text{ CL}.$$

The LHCb collaboration has set a limit on the $B^+ \to K^+ \tau^+ \mu^-$ mode, which is

$$\mathcal{B}(B^+ \to K^+ \tau^+ \mu^-) < 3.9 \times 10^{-5} \text{ at } 90\% \text{ CL}.$$

It should be noted that the result from the LHCb measurement is statistically independent from the one of this analysis. Indeed, in the previous LHCb analysis the candidates where the τ lepton decays into three charged pions plus one or more neutrals have been vetoed to allow the combination with the present analysis.

1.3 Interplay between lepton flavour non-universality and LFV

As mentioned previously, the tensions observed currently in various observables linked to lepton flavour universality are driving the search for NP models which could explain the observed anomalies. Generally, these models also imply the possibility of LFV. As a first step towards designing NP models, the measured values of the various observables are often interpreted in the context of an effective field theory framework. The idea is to interpret the SM as a low-energy effective field theory arising when integrating out the heavy mediators, *i.e.* with a mass greater than the accessible experimental energy scale, from an hypothetical ultra-violet complete quantum field theory. This idea is similar to that used by Fermi to describe the weak interactions when the existence of the $W^\pm$ and Z^0 bosons was unknown at the time [32]. By integrating out heavy mediators such as the $W^\pm$ bosons, the weak decay

of the neutron can be expressed in terms of a product of a dimension 6 operator involving the four spinors, *i.e.* the neutron, the proton, the electron and the neutrino, which essentially describes a local interaction and a dimensionful constant proportional to G_F with unit GeV^{-2}. The constant $G_F = \frac{\sqrt{2}g_2^2}{8m_W^2}$ is the Fermi constant, g_2 is the weak coupling constant and $m_W = 80.379 \pm 0.012\,\text{GeV}/c^2$ [33] is the mass of the $W^\pm$ boson. This idea can be generalised such as to perform a separation between low and high energy effects. The high energy effects arising from particles with a mass greater than the energy separation scale μ, including hypothetical heavy NP, can be described by so-called Wilson coefficient $C_i(\mu)$ which can be theoretically computed using perturbation theory. The low energy effects are described using a finite set of dimension 5, 6 and higher operators $O_i(\mu)$ respecting the gauge invariance of the SM [34].

When interpreting the results of the anomalies in the semileptonic decays involving $b \rightarrow s$ transitions, the relevant part of the weak effective Hamiltonian is given by [35, 36]

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} \frac{e^2}{16\pi^2} V_{tb} V_{ts}^* \sum_i (C_i O_i + C_i' O_i') + h.c., \tag{1.1}$$

where G_F is the Fermi constant, V_{tb} and V_{ts}^* are elements of the Cabibbo-Kobayashi-Maskawa matrix [37] describing the transition between flavour and mass eigenstates of the quarks. The list of operators in the sum is as follows :

$$\begin{aligned}
O_9^{l(')} &= (\bar{s}\gamma_\mu P_{L(R)} b)(\bar{l}\gamma^\mu l), \\
O_{10}^{l(')} &= (\bar{s}\gamma_\mu P_{L(R)} b)(\bar{l}\gamma^\mu \gamma_5 l), \\
O_S^{l(')} &= m_b(\bar{s} P_{R(L)} b)(\bar{l} l), \\
O_P^{l(')} &= m_b(\bar{s} P_{R(L)} b)(\bar{l}\gamma_5 l), \\
O_7^{(')} &= \frac{m_b}{e}(\bar{s}\sigma_{\mu\nu} P_{R(L)} b) F^{\mu\nu}, \\
O_8^{(')} &= \frac{g_c m_b}{e^2}(\bar{s}\sigma_{\mu\nu} P_{R(L)} T^a b) G^{\mu\nu a},
\end{aligned} \tag{1.2}$$

where $l = e, \mu, \tau$ corresponds to a given charged lepton and $P_{L(R)} = \frac{1}{2}(1 \mp \gamma_5)$ is the left (right) chirality projector. The Wilson coefficients can then be split into the contributions due to the SM and to NP as $C_i = C_i^{SM} + C_i^{NP}$. By doing so, the measurements of various observables related to lepton non-universality can be reinterpreted in terms of C_i^{NP} which is useful in understanding the structure of the hypothetical NP. An illustration of such combination of measurements is given in Fig. 1.3. When considering NP models that would be able to describe the effect of the various observables one often ends up predicting the possibility of LFV decays. There are currently many measurements of such decay modes which have been performed and that add further constraints on NP models. The current limits on LFV decay modes involving the decay of a beauty meson are shown in Fig. 1.4.
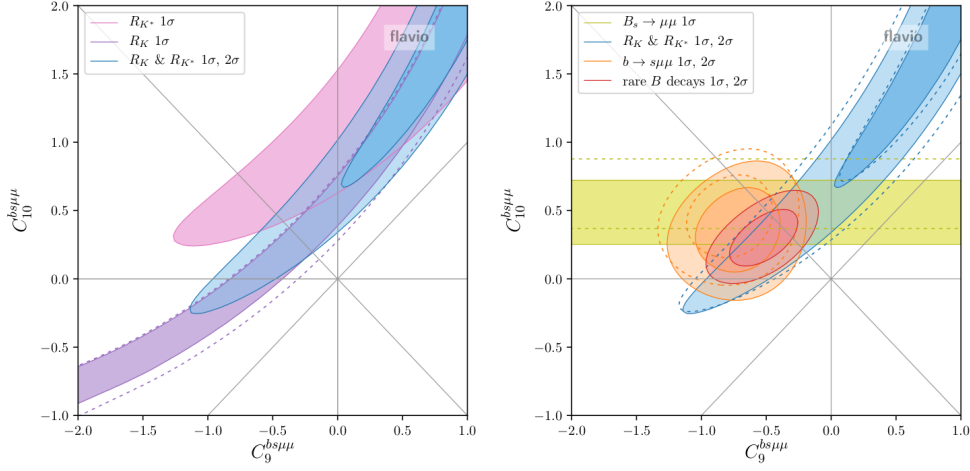
Figure 1.3 – The plot on the left shows the constraints on the $(C_9^\mu; C_{10}^\mu)$ plane given by the individual lepton universality ratios $R_{K^{(*)}} = \frac{\mathcal{B}(B^+ \to K^{(*)+}\mu^+\mu^-)}{\mathcal{B}(B^+ \to K^{(*)+}e^+e^-)}$ as well as the combination of both measurements while the plot on the right takes into consideration additional observables. Figure taken from Ref. [38].

There are various types of NP models which predict the existence of LFV such as those based on the two Higgs doublet model [40], heavy singlet Dirac neutrinos [41], the Pati-Salam model [42], Z' gauge bosons [43], leptoquarks [44], etc. Amongst these, models based on leptoquarks have received sizeable attention, in particular models based on the vector leptoquark U_1 with representation $(3, 1, 2/3)$ in the SM gauge group. This is because it seems possible to describe describe both tensions in the semileptonic $b \to s$ and $b \to c$ transitions by assuming the presence of this single mediator [25]. A specific U_1 leptoquark model with an ultra-violet completion assuming a non-vanishing right-handed coupling to SM fermions is given in Ref. [29]. This particular model also gives predictions for the branching fraction $\mathcal{B}(B^+ \to K^+\tau^+\mu^-)$ as can be seen in Fig. 1.5.
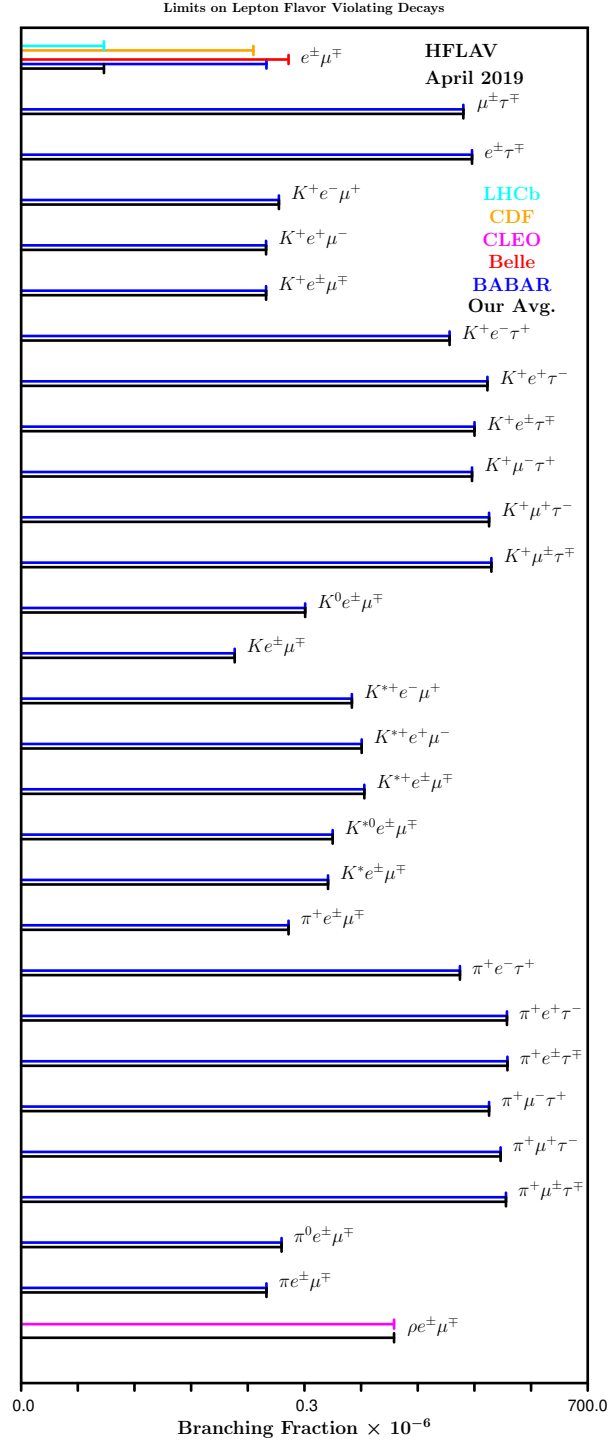
Figure 1.4 – Current upper limits at 90% confidence level on LFV decay modes involving the decay of a beauty meson. Figure taken from Ref. [39].
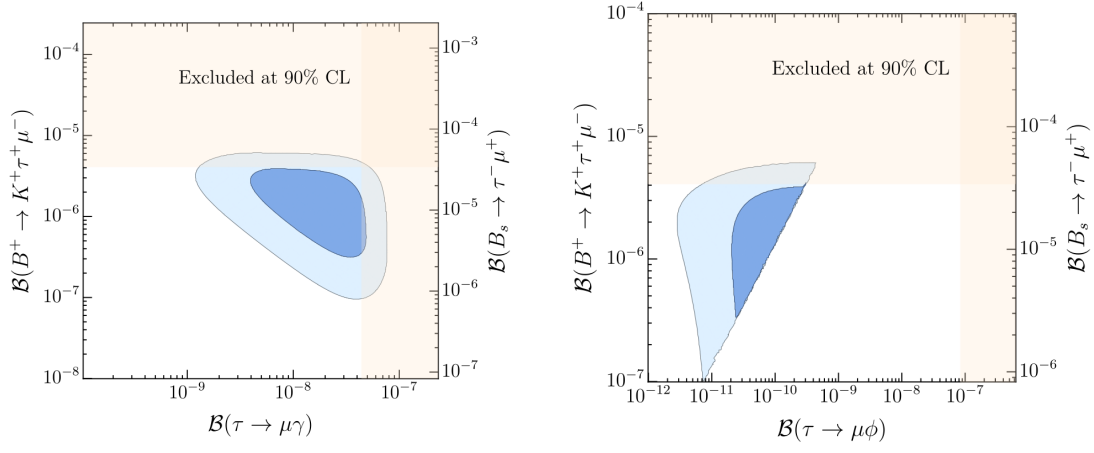
Figure 1.5 – The left plot shows the prediction of the value of $\mathscr{B}(B^+ \to K^+ \tau^+ \mu^-)$ as a function of $\mathscr{B}(B_s^0 \to \tau^- \mu^+)$ and $\mathscr{B}(\tau^- \to \mu^- \gamma)$ for the leptoquark model from Ref. [45]. In the plot on the right, $\mathscr{B}(\tau^- \to \mu^- \gamma)$ is replaced by $\mathscr{B}(\tau^- \to \mu^- \phi)$. The orange bands correspond to current experimental upper limits at 90% confidence level.

2 The experimental apparatus

In this chapter the LHCb detector used to acquire the data being analysed in this thesis is described. First, a general overview of the design of the detector and its purpose is given. Then, the different sub-detectors from which the LHCb detector is composed are described with more details.

2.1   The Large Hadron Collider

The data processed in this thesis was acquired by the LHCb detector. The LHCb detector is one of the four main detectors used to reconstruct data from proton-proton collisions at the Large Hadron Collider (LHC) [46]. The LHC is located at the European Organization for Nuclear Research (CERN) near Geneva, at the border between Switzerland and France. It is a particle accelerator and collider composed of a 27 km long ring with two vacuum pipes located approximately 100 m underground in the tunnel previously used for the Large Electron Positron (LEP) collider [47]. The LHC is primarily used to accelerate protons and perform $p + p$ collisions. However, it can also be operated with heavy ions which provide conditions different than the $p + p$ environment. So far the following collisions involving heavy ions have been performed : Pb + Pb, p + Pb and Xe + Xe. These are mainly used to study QCD at high energy and density regime as well as the behaviour of quark-gluon plasma [48].

The LHC has been designed to perform pp collisions at a centre-of-mass energy of $\sqrt{s} = 14$ TeV. Currently, the collisions were performed at $\sqrt{s} = 7$ TeV in 2010 and 2011, $\sqrt{s} = 8$ TeV in 2012 and $\sqrt{s} = 13$ TeV from 2015 to 2018. The period from 2010 to 2012 is referred to as Run 1 while the period from 2015 to 2018 is referred to as Run 2. In order to reach these energies, the protons have to be accelerated in successive steps. First, the protons are obtained by ionising hydrogen. They are then injected into a linear accelerator (Linac 2) which uses radio-frequency cavities to accelerate the protons up to an energy of 50 MeV. They then go through the proton
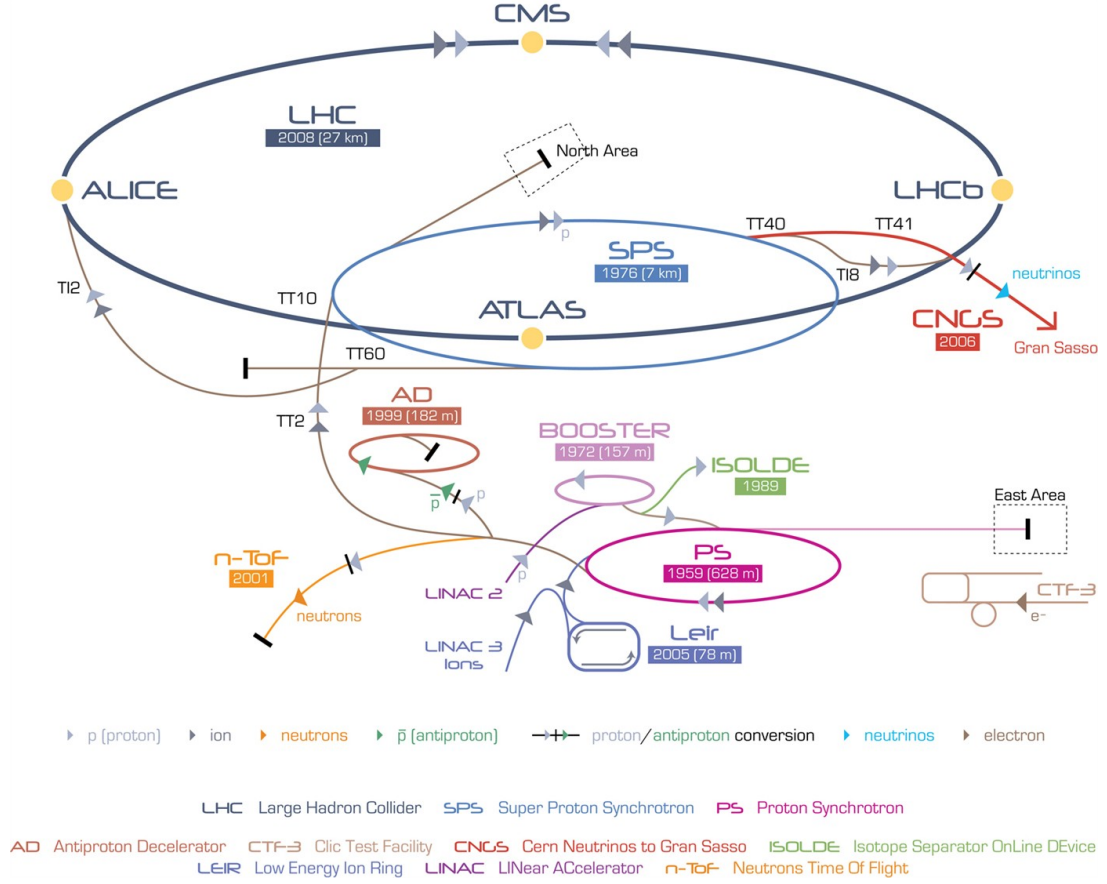
Figure 2.1 – Accelerator chain of the LHC. Figure from Ref. [52].

synchrotron booster (PSB) which raises their energy up to 1.4 GeV such that they can then be injected into the proton synchrotron (PS) where they reach an energy of 25 GeV. Finally, they go through the super proton synchrotron (SPS) where they reach 450 GeV before splitting the beam in two opposite directions in the LHC where the protons will reach their final energy and collide within the various detectors. The chain of accelerators is represented schematically in Fig. 2.1. There are 8 interaction points along the LHC rings where the detectors of the different experiments can potentially be placed. The four main detectors are ALICE [49], ATLAS [50], CMS [51] and LHCb [45].

Other than the centre-of-mass energy $\sqrt{s}$ that was discussed above, the other important parameter for a collider is the integrated luminosity *i.e.* the number of events from the collisions for a given interaction cross-section. In particular, what is interesting at LHCb are events containing a b or a c quark. The number of events with a $b\bar{b}$ pair for a given period of time is given by :

$$N_{b\bar{b}} = L_{int} \cdot \sigma_{pp \to b\bar{b}} \quad , \tag{2.1}$$

where L_{int} is the integrated luminosity for the given period of time and $\sigma_{pp \to b\bar{b}}$ is the $b\bar{b}$ production cross-section from proton-proton collisions. Since the cross-section depends on $\sqrt{s}$, see Fig. 2.2, it is clear that the two important parameters for high-precision measurements are the instantaneous luminosity and the centre-of-mass energy. The instantaneous luminosity for a Gaussian beam distribution is given by :

$$L = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F \quad , \tag{2.2}$$

where n_b is the number of bunches, N_b is the number of protons per bunch, f_{rev} is the revolution frequency, γ is the Lorentz factor, ϵ_n is the normalised transverse beam emittance, β^* is the beta function at the interaction point and finally F is a factor that takes into account the reduction of luminosity due to a non-zero crossing angle. The LHC is designed to deliver a peak luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This is the luminosity used by the ATLAS and CMS experiments. However, for the LHCb experiment, the luminosity is lowered to $10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This is done to accommodate the specific requirements of the LHCb detector. In particular, reducing the luminosity results in a lower pile-up, *i.e.* the number of primary vertices per event, as well as a lower radiation damage to the detector. The integrated luminosity recorded by the LHCb experiment during Run 1 and Run 2 is shown in Fig. 2.3.

2.2 The LHCb detector

LHCb stands for Large Hadron Collider beauty. This reflects the fact that the main purpose of the experiment is to study the behaviour and properties of heavy hadrons, in particular those containing *b* or *c* quarks. As a result, the design of the detector is tailored for the study of such particles and therefore presents a geometry which differs significantly *e.g.* from the ATLAS and CMS detectors. Contrarily to these which are general purpose detectors with a 4π geometry, the LHCb detector is a single-arm forward spectrometer. The angular acceptance is from 10 to 300 mrad in the bending plane of the magnet and 10 to 250 mrad in the non-bending plane. In terms of particles' properties, this represents a pseudorapidity range $2 < \eta < 5$ where the pseudorapidity is expressed as :

$$\eta = -\ln\left(\tan\left(\theta/2\right)\right) = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right), \tag{2.3}$$

where $\vec{p}$ is the momentum of the particle, p_z is the momentum's component longitudinal to the beam axis and θ is the angle between the momentum and the beam axis. Generally, events are described using a right-handed Cartesian coordinate system where the *z*-axis points along the beam axis towards the muon chambers and the *y*-axis is orthogonal to the *z*-axis, in a vertical plane, pointing upwards. The origin of the coordinate system is set at the location

proton - (anti)proton cross sections
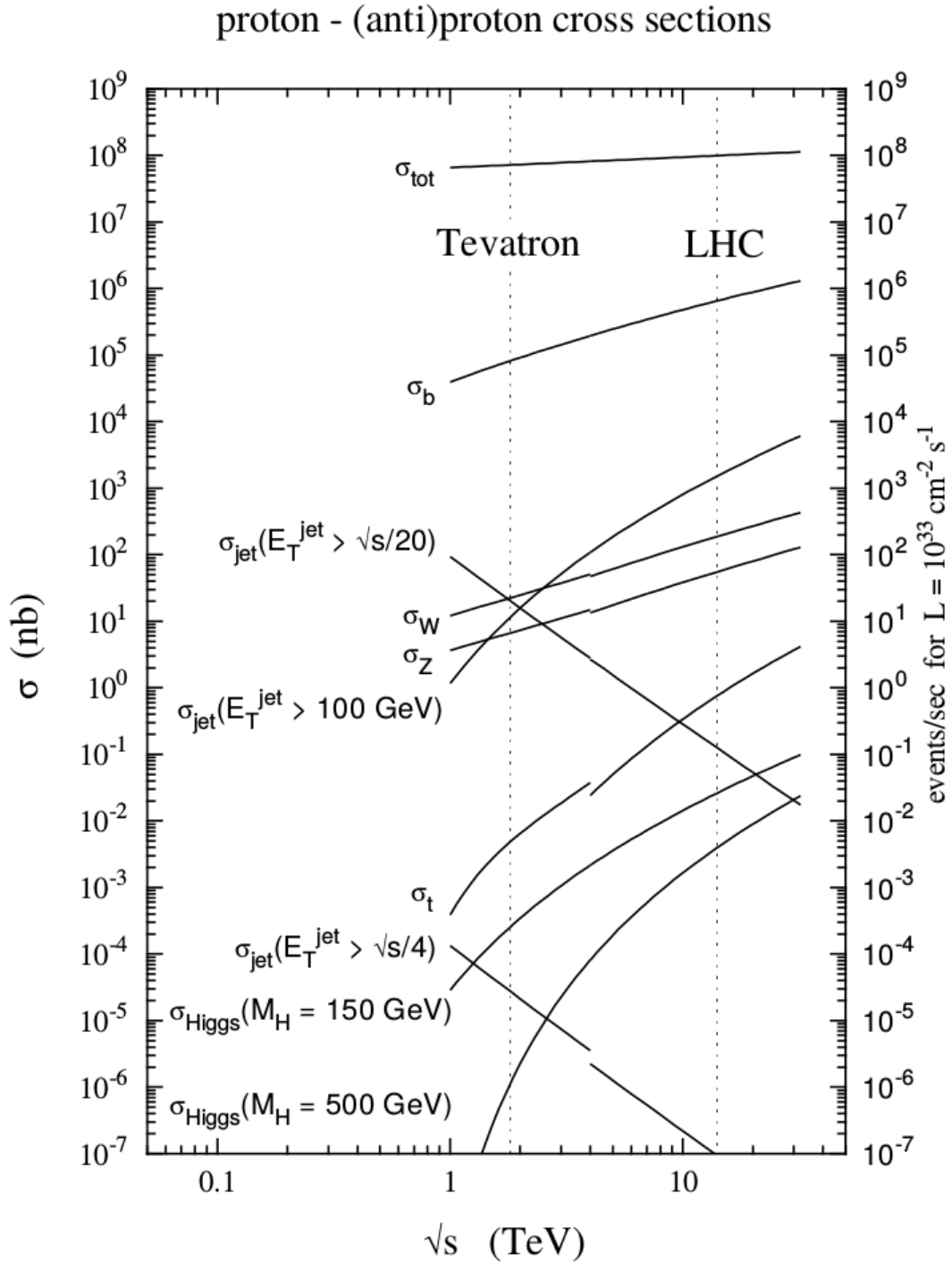
Figure 2.2 – Cross-sections of pp and $p\bar{p}$ collisions as a function of the centre-of-mass energy. The discontinuities are due to the switch from $p\bar{p}$ to pp cross-sections. At a centre-of-mass energy $\sqrt{s} = 7$ TeV there are approximately 10^{12} b-hadrons produced at LHCb for an integrated luminosity of 1 fb^{-1}. Figure from Ref. [53].
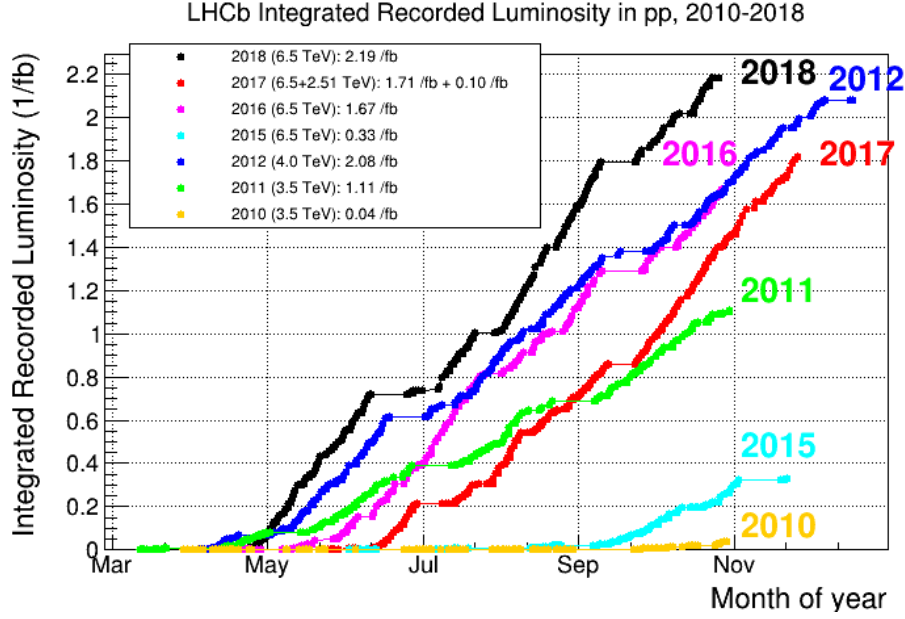
Figure 2.3 – Integrated luminosity collected by the LHCb experiment from 2010 to 2018. Figure from Ref. [54].

of the pp interaction point. The reason for this choice of geometry is due to the fact that at the LHC the $b\bar{b}$ pairs are predominantly produced in the forward and backward directions as shown in Fig. 2.4.

One of the interesting features of b-hadrons is that they have a relatively long lifetime of the order $\mathcal{O}(1\,\mathrm{ps})$. At the collision energies delivered by the LHC, they travel a distance of the order of $\mathcal{O}(1\,\mathrm{cm})$ thanks to the Lorentz time dilation. This means that with a sufficiently good vertex resolution it is possible to determine the position of the production vertex (PV) and decay vertex (SV) of a hadron with significant separation. This allows to measure its lifetime and is helpful in rejecting backgrounds coming from random combination of tracks, which is referred to as combinatorial background, as well as backgrounds coming from decays other than the one of interest. A precise measurement of the hadron lifetime is necessary in order to perform competitive measurements of CP violating observables, which is one of the main physics objectives of LHCb, while a very good background suppression is needed in searches for very rare or SM-forbidden decays. Since measuring the position of these vertices is primordial for the physics program of LHCb, it is desirable to have a low number of interactions per bunch crossing. This number, known as pile-up, needs to be low for the LHCb experiment in order to keep a good precision on the measurement of production and decay vertices. This is one of the reasons why the luminosity delivered by the LHC is lowered at LHCb, as mentioned earlier. The techniques used to modulate the luminosity are explained in Ref. [56]. As a result, the

17
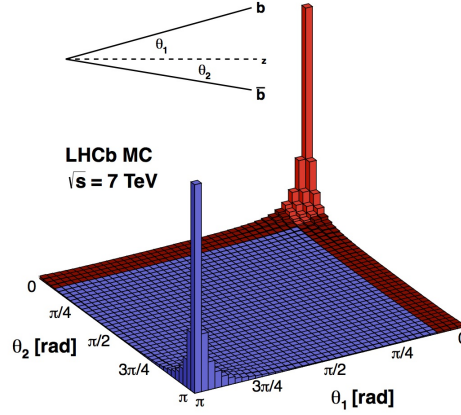
Figure 2.4 – Monte Carlo simulation of the $b\bar{b}$ quark pair angular distribution after a pp collision at an energy of $\sqrt{s} = 7\,\mathrm{TeV}$. The angles of the b and $\bar{b}$ quarks with respect to the beam axis are given by θ_1 and θ_2, respectively. Figure from Ref. [55].

average pile-up at the LHCb detector if around 1.7 while it is of the order of $\mathscr{O}(30)$ in the ATLAS and CMS experiments.

The LHCb detector is composed of multiple sub-detectors structured sequentially from the interaction region as shown in Fig. 2.5. The vertex locator (VELO) surrounds the pp interaction region where b or c hadrons are produced, then the decay products of these hadrons travel along the detector possibly up to the muon chambers. Each sub-detector plays a particular role in the reconstruction process of an event. However, they can be divided into two main categories depending on their purpose. They are either part of the tracking or the particle identification systems. These systems and the associated sub-detectors are described in more details in the next sections.

2.2.1 Tracking system

The tracking system is composed of the VELO, the Tracker Turicensis (TT), the magnet and the tracking stations. It is designed to determine the trajectories of charged particles as well as their production and decay vertices. The PV and SV positions are measured in the VELO, as explained earlier, while the trajectories are extrapolated from hits in the multiple tracking stations positioned upstream and downstream from the bending magnet. Additionally, the curvature of the trajectory of charged particles within the magnetic field allows the determination of their momentum and charge. The resulting precision on the measurement of the particles momenta varies from 0.4% to 0.6% for momenta $p = 5\,\mathrm{GeV}/c$ and $p = 100\,\mathrm{GeV}/c$, respectively.
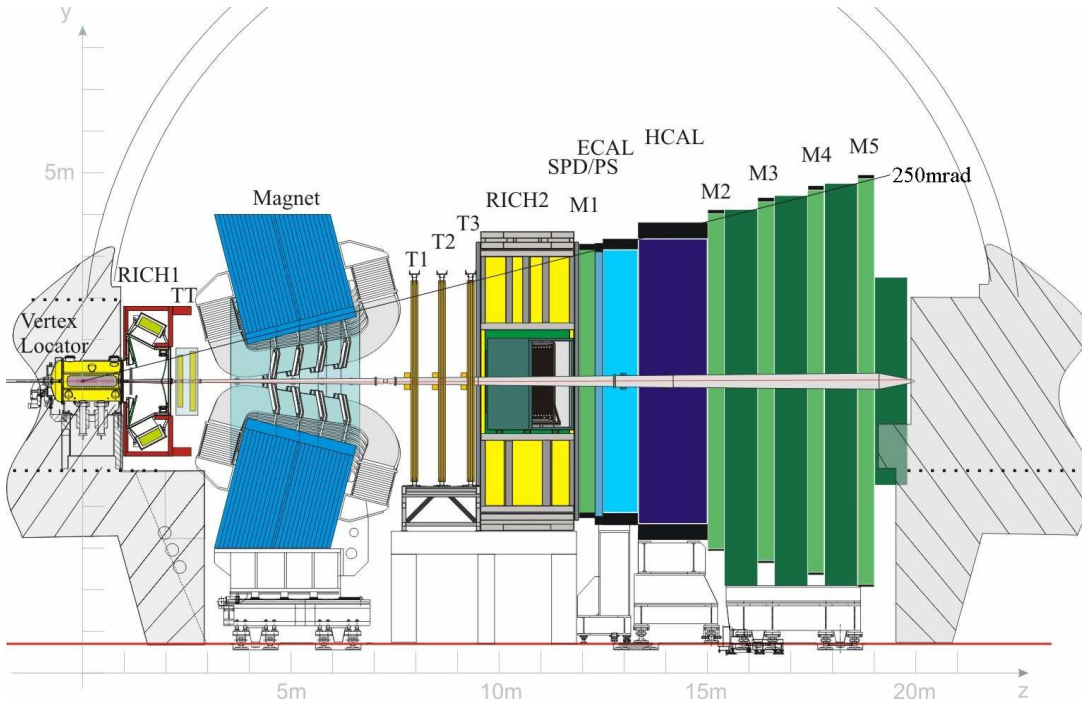
Figure 2.5 – LHCb detector with its associated coordinate system. Figure from Ref. [45].

Vertex Locator

The VELO uses silicon microstrip technology in order to accurately measure the positions of the PV and SV of the particles produced during a pp collision. As such, this detector surrounds the interaction region and consists of multiple modules placed along the beam line *i.e.* along the z-axis. Each module provides a measurement of the r and ϕ coordinates of a hit using so-called R-sensors and ϕ-sensors. The choice of a cylindrical coordinate system instead of a Cartesian one is natural given the geometry at hand. Additionally, the cylindrical system allows for a faster real-time reconstruction at the software trigger level (see below). The geometry of the VELO is motivated by the fact that it has to cover the angular acceptance of the overall detector, *i.e.* it needs to cover the range $2 < \eta < 5$. Additionally, the constraint along the beam axis is $|z| < 10.6\,\text{cm}$ and tracks have to pass through at least 3 stations. There are 42 modules made out of pairs of half detectors. The halves within a pair are separated by $1.5\,\text{cm}$. This shift is required in order for the two detector halves to overlap when operating the detector. Indeed, the modules are retractable since they need to be as close as possible to the beam during operation in order to maximise the precision of the measurements but the distance from the modules to the beam is smaller than the opening needed to operate the LHC during the injection. This means that the VELO detector is open during the LHC injection phase and is closed only during the data-taking period when the LHC beams are in a stable condition. A
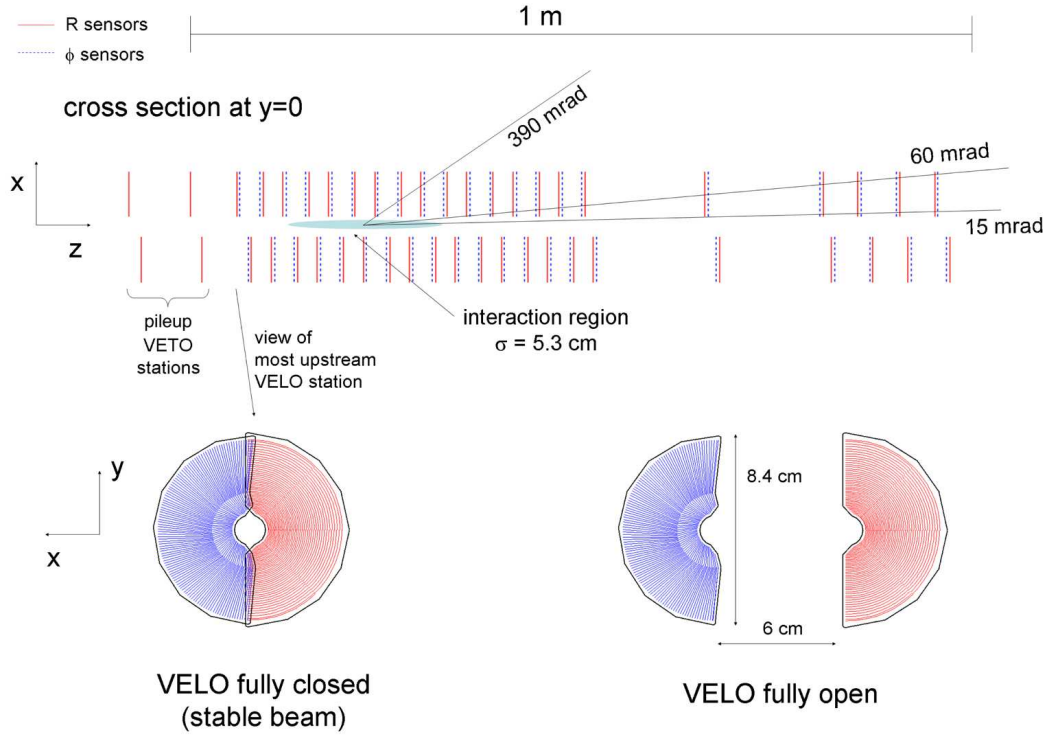
Figure 2.6 – The top part of the figure shows a cross-section of the VELO in (x, z) plane while the bottom part shows a VELO module in open and closed positions in the (x, y) plane. Figure from Ref. [45].

sketch of the VELO detector is given in Fig. 2.6.

In order to maximise the resolution of the measured vertices, the pitch is the smallest close to the beam and increases further away from the beam. The minimal distance between the sensors and the beam's centre is 8 mm. The R-sensors are shaped as concentric circles centered around the beam with an inner-most pitch of 38 µm and an outer-most pitch of 101.6 µm. The structure of the ϕ-sensors is slightly more complicated. It is divided in two regions separated at a radius of 17.25 mm. In the inner region the pitch goes from 39 µm to 78 µm while in the outer region it goes from 39 µm to 97 µm. An additional feature of the ϕ-sensors is that they are inclined with respect to the radial direction. Two consecutive modules are inclined in opposite directions. This pattern is used to reject ghost hits, *i.e.* fake hits originating when at least two particles hit two planes of crossing strips. A detailed sketch of a module showing both the R-sensors and ϕ-sensors is shown in Fig. 2.7. By combining the measurements of the r, ϕ and z coordinates, it is possible to fully reconstruct the three-dimensional position of the primary and secondary vertices.

Figure 2.7 – Sketch of a VELO module showing the R-sensors and ϕ-sensors. Two consecutive modules are shown in for the ϕ-sensor in order to display the stereo angle. Figure from Ref. [45].

Tracker turicensis

The tracker turicensis (TT) detector is the first set of tracking stations placed downstream from the VELO. Its dimensions are 150 cm × 130 cm and cover the full LHCb acceptance. It uses silicon microstrip sensors with a pitch of 200 μm. The TT is composed of four detection layers. The outer layers have vertical strips while the two middle layers are inclined by ±5° in order to provide a measurement of the y coordinate of the hits and reduce the number of fake tracks. These four layers are split into two pairs, each with two layers, separated by 27 cm along the z axis. The readout strip lengths have been adjusted to account for the fact that hit occupancy is significantly greater in the inner part of the TT than in the outer part. The resulting hit resolution is 50 μm. A sketch of the TT is given in Fig. 2.8.

Dipole magnet

In order to measure the momentum of particles traveling through the detector, a magnetic field is applied to bend their trajectory. From the curvature of the track it is possible to infer the particle momentum. An integrated magnetic field of 4 Tm, for tracks of 10 m length, is produced by a warm dipole magnet placed between the TT and the tracking stations. The magnet is made out of two coils with a conical saddle shape and placed within a magnet yoke. The magnetic field is mainly oriented along the y axis and thus the trajectory of charged particles is bent in the (x, z) plane. In order to provide a sufficiently good resolution on the measurement of the particles' momentum it is necessary to measure the integrated field within the magnet's volume with a relative precision of the order of $\mathcal{O}(10^{-4})$. Additionally, the magnetic field is

Figure 2.8 – Layout of the TT. Figure from Ref. [57].

Figure 2.9 – LHCb dipole magnet. Figure from Ref. [58].

Figure 2.10 – Sketch of a layer of the inner tracker. Figure from Ref. [45].

periodically inverted between MagUp (MU) and MagDown (MD) configurations, *i.e.* when the field is in the positive or negative *y* direction, in order to check for any systematic effect arising from a geometric asymmetry in the detector. This is mainly important for measurement of *CP* violating observables since any discrepancy between particles and anti-particles due to the experimental setup has to be under control. A sketch of the magnet is given in Fig. 2.9.

Tracking stations

The final sub-detectors of the tracking system are the tracking stations. There are three tracking stations, T1, T2 and T3, located downstream of the magnet. Each of these tracking stations is made of four layers which are positioned in the same way as for the TT, *i.e.* with the central two layers rotated by ±5°. An interesting feature of these stations is that they are essentially divided into two sub-detectors using different technologies. There is the inner tracker (IT) in the centre of the stations at high pseudorapidities, while the rest of the stations compose the outer tracker (OT). The variation of the hit occupancy as a function of the distance from the beam pipe is the main reason why a single technology was not selected.

Similarly to the TT, the IT uses silicon microstrip technology and is able to provide a spatial resolution of 50 µm. It is positioned close to the beam pipe where the occupancy is the greatest and where the highest track resolution is required. The IT stations are made of four detector boxes positioned in a cross-like shape around the beam pipe, see Fig. 2.10. Each box contains four detection layers and each layer is made of seven modules. The modules which are in the boxes positioned on the top or bottom of the beam pipe contain a single silicon sensor while the modules in the boxes positioned on the side of the pipe contain two sensors.

The OT is made of modules containing a double layer of drift tubes. These straw tubes have a diameter of 4.9 mm and contain a mixture of gases which is ionised when a charged particle traverses the detector. The mixture is made of 70% of argon and 30% of CO_2. This mixture of gases provides a drift time below 50 ns and a spatial resolution of 200 µm. Compared to the IT
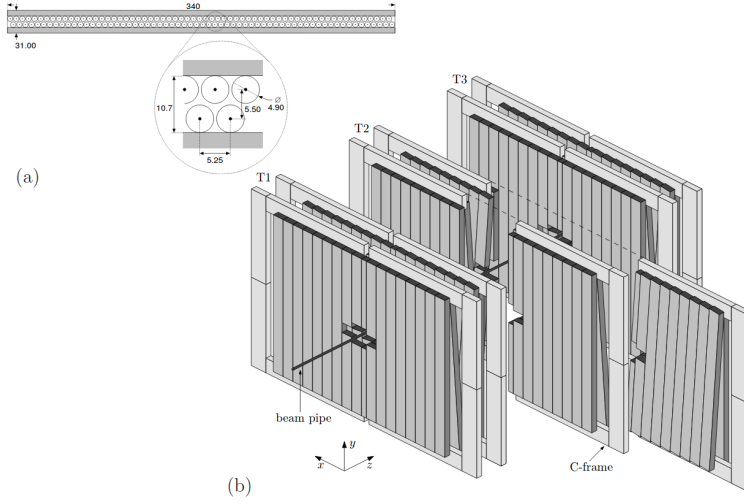
Figure 2.11 – The top left part of the figure shows the double layered structure of the straw tubes within a module while the bottom part shows the OT within the tracking stations. Figure from Ref. [59].

which uses silicon microstrips, this choice of technology provides a lower resolution of 200 μm which is sufficient considering the lower hit occupancy in the OT. A sketch of the OT as well as the straw tubes within a module is given in Fig. 2.11.

2.2.2 Particle identification system

The particle identification (PID) system is composed of the Cherenkov detectors, the electromagnetic and hadronic calorimeters as well as the muon chambers. The purpose of this system is to distinguish between the different kinds of particles produced in the pp collisions and in particular to assign an identity to the most basic particles passing through the LHCb detector. The particles which are considered basic in the LHCb detector, *i.e.* which are considered stable, are the following: $\pi, K, p, e, \mu, \gamma$. The different sub-detectors of the PID system are tailored to distinguish between these different kinds of particles. For each final state particle an identity is assigned corresponding to one of these 6 particles. In practice, this means that the known value of the mass of one of these particles is assigned to the track. The rest of the decay chain is then reconstructed from the bottom up through energy-momentum conservation and topological considerations. It is thus primordial to correctly identify these particles in order to reject various sources of background. Additionally, since most analyses use the reconstructed invariant mass to measure the number of signal decays, if one or more particles in the final state are misidentified this result in a total reconstructed invariant mass biased with respect to the true mass of the hadron under consideration. For the decays considered in this thesis the final state is composed of pions, kaons and muons. The overall PID system results in an

identification efficiency of roughly 95% for kaons and 97% for muons, and mis-identification rates from pion to kaon and pion to muon of 5% and 3%, respectively.

Cherenkov detectors

The identity of the different hadrons produced in LHCb, *i.e.* pions, kaons and protons, is distinguished through the use of ring imaging Cherenkov detectors (RICH). These detectors exploit the Cherenkov effect in which a charged particle passing through a dielectric medium with a speed greater than c/n will emit a cone of light centred around the particle's trajectory and whose opening angle depends on the particle's velocity. More precisely the opening angle θ follows the relation :

$$\cos(\theta) = \frac{1}{n\beta},\tag{2.4}$$

where n is the refractive index of the medium and $\beta = v/c$ is velocity of the particle normalised to the speed of light in vacuum. Thus, by measuring the opening angle of the light cones produced by an incoming particle its velocity can be estimated. Combining this information with the measurement of the momentum of that particle given by the tracking system, its mass, and therefore its identity, can be determined.

The choice of the medium called "radiator" used in the RICH dictates the momentum range in which a significant separation power can be achieved on the identity of the different hadrons. It would be impossible to cover the full momentum spectrum of the hadrons traversing the LHCb detector by using a single RICH. For this reason, the LHCb detector uses two RICH detectors abbreviated RICH1 and RICH2. The RICH1 is positioned between the VELO and the TT, with a radiator made of C_4F_{10} and covers the momentum range $1-60\,\mathrm{GeV}/c$, while RICH2 is positioned between the tracking stations and the first muon station, with a radiator made of CF_4 and covers the momentum range $15-100\,\mathrm{GeV}/c$. The RICH1 covers the full LHCb acceptance while RICH2 only covers a limited solid angle, from $\pm 15\,\mathrm{mrad}$ to $\pm 120\,\mathrm{mrad}$ ($\pm 100\,\mathrm{mrad}$) in the horizontal (vertical) plane. Nonetheless, this limited angular is acceptance is not an issue since it covers the region containing the high momentum particles. The Cherenkov light is guided outside of the LHCb acceptance where it is collected by hybrid photon detectors (HPDs) which are protected from the LHCb magnet's field with an iron shielding. A sketch of both RICH detectors is given in Fig. 2.12. The opening angle distribution as a function of the momentum is given in Fig. 2.13. It illustrates the excellent discriminating power of the RICH.

Calorimeter system

The calorimeter subsystem is designed to fulfill multiple purposes. Not only is it used as part of the identification system but it is also used to measure the direction of incoming electron,
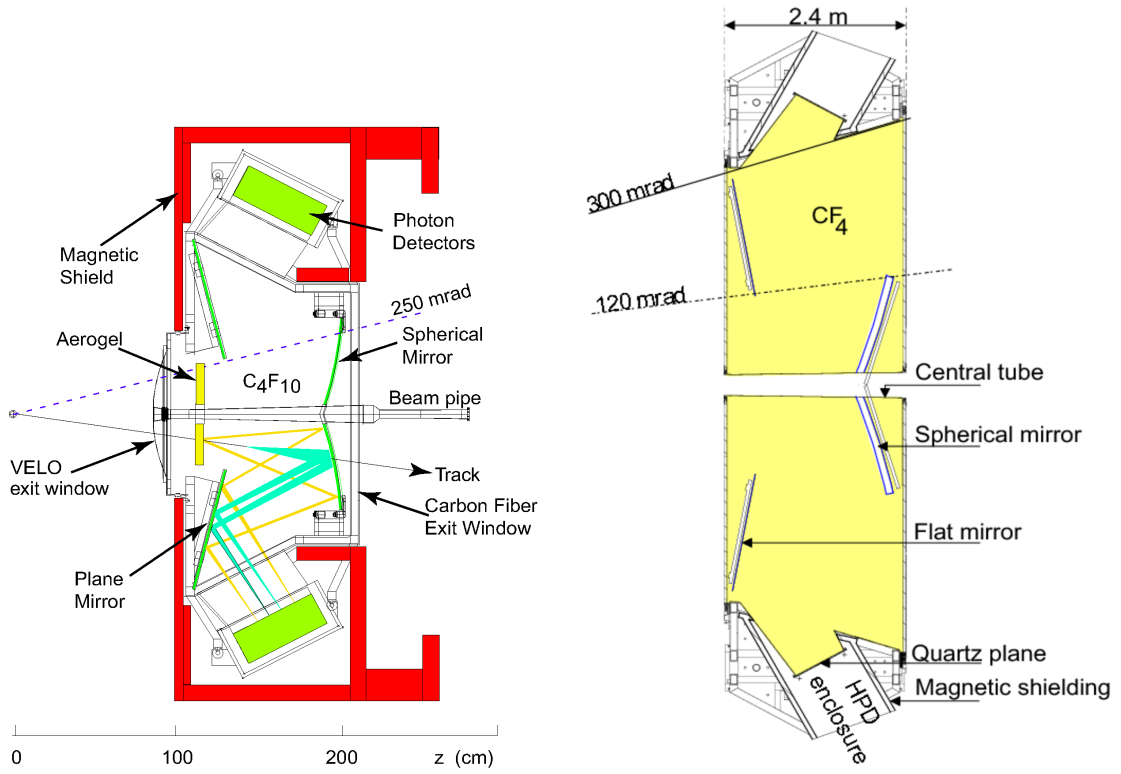
Figure 2.12 – Layout of the RICH1 (left) and RICH2 (right) detectors. Figure taken from Ref. [45].
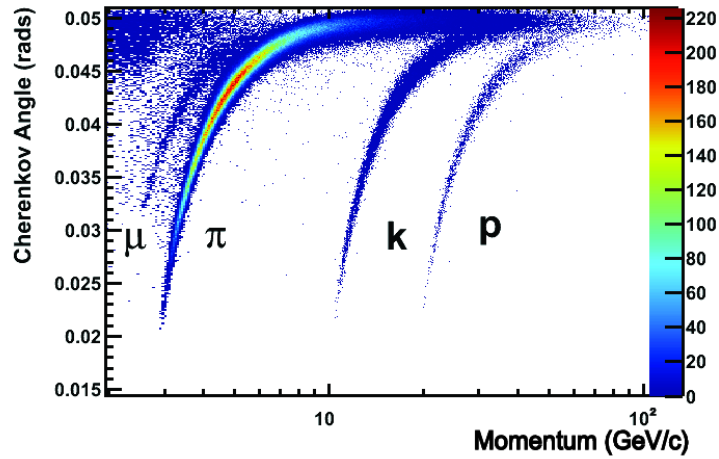
Figure 2.13 – Measured distribution of Cherenkov opening angle as a function of the particle's momentum for the RICH1. Figure from Ref. [60].

photon and hadrons, as well as the transverse energy $E_T \equiv \sqrt{m^2 + p_T^2}$, where m and p_T are the particle's mass and transverse momentum with respect to the beam direction. The transverse energy is a primordial input to the hardware trigger decision as it carries valuable information about the potential presence of b-hadron decays. The calorimeter system is located between the first and second muon stations. It is composed of four elements, the scintillating pad detector (SPD), the preshower detector (PS), and the electromagnetic (ECAL) and hadronic (HCAL) calorimeters.

The SPD and PS sub-detectors are mainly used to distinguish electrons from neutral pions and from photons. It is made of a 1.5 cm thick lead sheet used to initiate electromagnetic showers whose properties are then measured using scintillators positioned before and after the lead plates. The light from the scintillators is guided by wave-length shifting fibres up to photo-multiplier detectors. The width of 1.5 cm of lead corresponds to 2.5 radiation lengths.

Both calorimeters are based on the same principle. They are made of alternating layers of heavy material and scintillators. The heavy material is used to produce an interaction between the incoming particle and the layer thus creating either electromagnetic or hadronic showers, while scintillators are used to produce light from the interaction with the secondary particles. The amount of collected light by the photo-multipliers is a function of incoming particle's energy. The ECAL uses lead as heavy material while the HCAL uses iron instead. In the ECAL, the lead and scintillating layers are 2 mm and 4 mm thick, respectively. The full composition of the ECAL corresponds to 25 radiation lengths and fully absorbs incoming electrons or photons. The corresponding energy resolution, where the energy E is expressed in GeV, is :

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%. \tag{2.5}$$

In the case of the HCAL, the full length of the detector corresponds to 5.6 hadronic interaction lengths to which are added the 1.2 interaction lengths from the ECAL. In general, the size of the HCAL is not sufficient to fully absorb incoming hadrons and as such the resulting energy resolution is lower than for the ECAL. The energy resolution is :

$$\frac{\sigma_E}{E} = \frac{69\%}{\sqrt{E}} \oplus 9\%. \tag{2.6}$$

As a result, the HCAL is mainly used as a particle identification tool.

Both calorimeters are segmented laterally in order to provide a measurement of the direction of incoming particles. The cells' density distribution is non-uniform. This reflects the fact that the occupancy varies by two orders of magnitude along the radial direction. The ECAL is divided in three sections while there are only two sections in the HCAL. A sketch of this segmentation for both the ECAL and HCAL is given in Fig. 2.14.
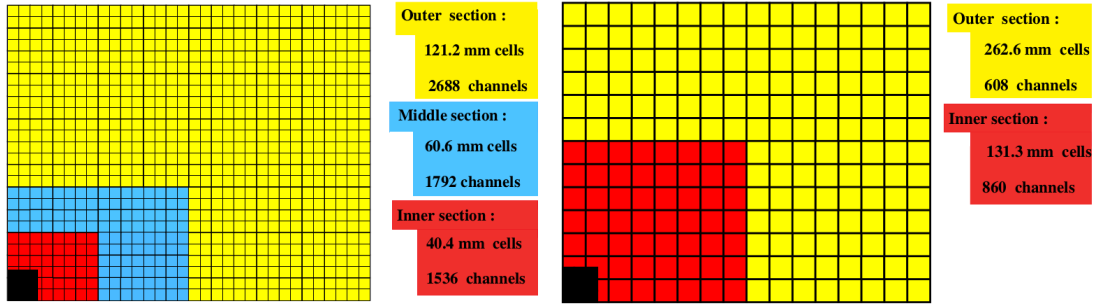
Figure 2.14 – Transverse segmentation of the ECAL on the left and HCAL on the right. Each sketch represents solely one quarter of the front face of the detector. Figure from Ref. [45].

Muon system

The final piece of the PID system is the muon system. It is composed of five stations labelled M1 to M5. These are used to bring information to the hardware trigger as well as to identify muons. The first station is located before the PS while the four remaining stations are located downstream of the calorimeters. The complete system covers the angular acceptance $20 - 306$ mrad and $16 - 258$ mrad in bending and non-bending planes, respectively. The stations are located at the end of the LHCb detector since muons have a sizeable lifetime and are more difficult to absorb compared to the rest of the final state particles studied at LHCb.

The stations M1-M3 are used to measure the transverse momentum of the muon track candidates and as such have a good spatial resolution while the stations M4 and M5 have a lower spatial resolution as they are mainly used to identify the most penetrating particles. In the inner part of the station M1 three gas electron multiplier (GEM) foils are used while the rest of the stations use multi wire proportional chambers (MWPC). The M1 station uses GEM in order to cope with the high occupancy in this region of the detector. The MWPC use a mixture of Ar, CO_2 and CF_4 in the proportion 40%, 55% and 5%, respectively while the GEM uses the proportions 45%, 15% and 40%. These gas mixtures are chosen to comply with the readout speed requirement imposed by the 25 ns bunch crossing time. To act as filters, 80 cm wide iron absorbers are placed between the MWPC of stations M2-M5. These are used to select only the most penetrating muons. In order to go through the full muon system, a muon would require a minimal momentum of 6 GeV/c. A sketch of the muon chambers is given in Fig. 2.15.

2.2.3 Particle identification variables

As described previously, each type of particle interacts in a specific way with the detector as illustrated in Fig. 2.16. The various sub-detectors have been designed to exploit the traces left by these interactions such that by combining the overall information from every sub-detector
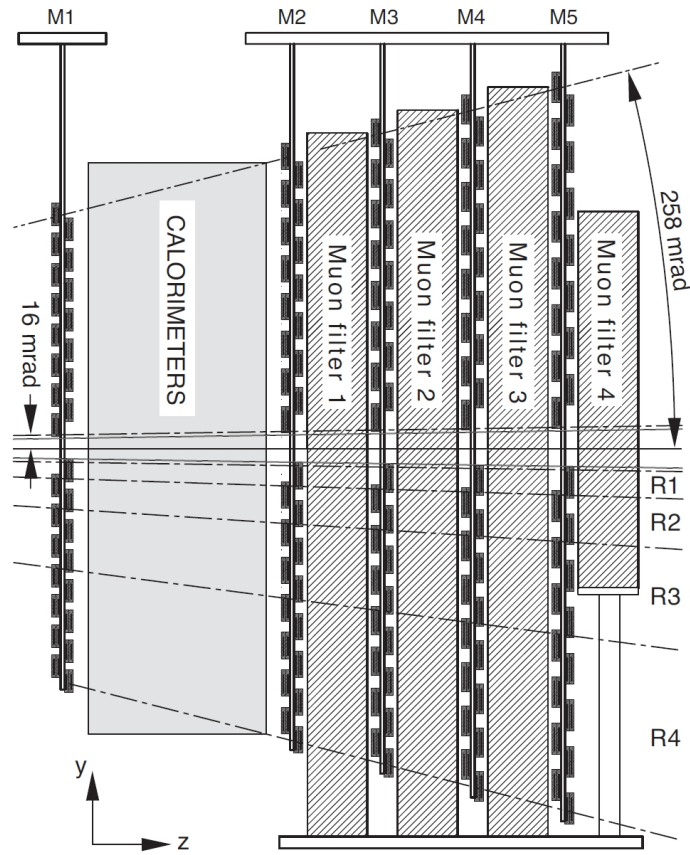
Figure 2.15 – Side view of the muon chambers. Figure from Ref. [45].

it is possible to assign an identity to a track with a low misidentification rate. In LHCb there are three different algorithms used to combine this information which results in three different types of PID variables.

- `isMuon` : A binary variable indicating whether a track is likely or not to be a muon. It is based on the number of muon stations where a hit can be associated to the track.

- Delta log likelihood (DLL) : The identification information from all the relevant sub-detectors is combined into a likelihood variable $\mathscr{L}_X(Y)$ for a given PID hypothesis X and reconstructed track Y. Then, instead of using directly the likelihood for a given hypothesis, the difference of log-likelihood between hypothesis X and the pion hypothesis is used : $DLL_X(Y) = \ln \mathscr{L}_X(Y) - \ln \mathscr{L}_\pi(Y) = \ln\left(\frac{\mathscr{L}_X(Y)}{\mathscr{L}_\pi(Y)}\right)$.

- `ProbNN` : Output of a neural network using as input information from all sub-detectors, including those associated to tracking. It gives a value $ProbNN_X(Y)$ between 0 and 1 describing how likely a given track Y corresponds to an identity X, where 1 is very likely and 0 is very unlikely.

By using the distributions of these different variables, it is possible to further tailor the PID selection criteria for a given decay of interest and reject mis-identified backgrounds.

2.2.4 The trigger system

The nominal LHC bunch crossing rate is 40 MHz. However, it is impossible to process and record data at this rate with the current technology. For this reason, it is necessary to devise a trigger system whose role is to reduce this nominal rate to a more manageable rate of a few kHz. Additionally, there is a large proportion of events which are not particularly interesting in terms of potential physics output. As such, the trigger system is designed to reduce the event rate while maximising the proportion of interesting physics events stored for further study. In order to achieve this, the idea is to identify key signatures of heavy hadron decays such as displaced vertices from the interaction point or candidates with high transverse momenta. The trigger system is split into a hardware and a software stage. The Level 0 (L0) stage is implemented in hardware while the High Level Trigger (HLT) is implemented in software. The HLT trigger is further split into two stages called HLT1 and HLT2.

The L0 trigger reduces the event rate from 40 MHz down to 1 MHz. To do so, the decisions need to be performed quickly and as such they can only be based on relatively simple information. Multiple sets of selection requirements, known as trigger lines, are set-up to search for specific physics signatures and allow to separate the stored events in different streams depending on the expected particle content. The decisions are based on the information provided

innermost layer ——————→ outermost layer

| tracking system | electromagnetic calorimeter | hadronic calorimeter | muon system |

photons →

electrons →

muons →

protons
Kaons
pions →

neutrons
K_L^0 →

C. Lippmann − 2003
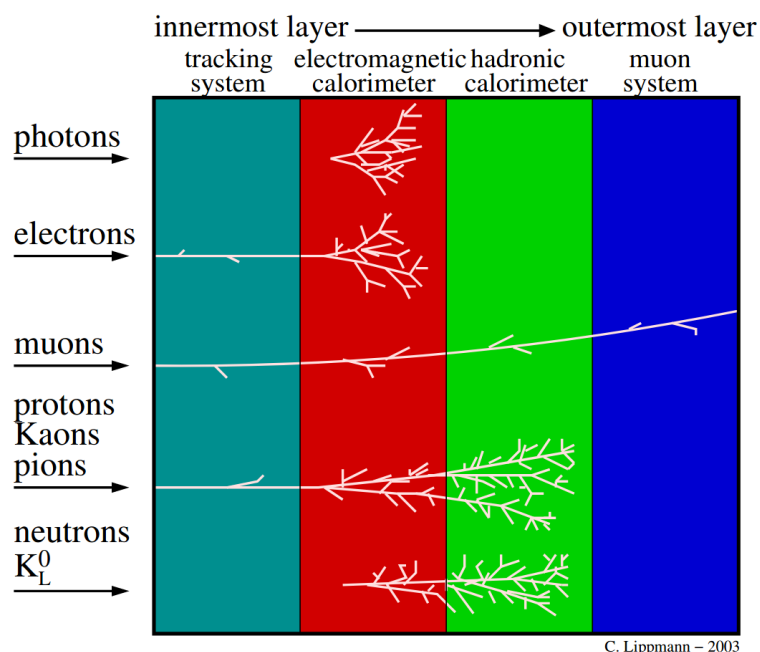
Figure 2.16 – Sketch showing the traces left by different types of particles inside a generic particle detector. Figure from Ref. [61].

by the muon system, the calorimeters as well as the pile-up system. It tries to reconstruct the muon pair with the highest transverse momentum as well as the hadron, electron and photon clusters with the highest transverse energy deposits in the calorimeters. Additionally, the pile-up system estimates the number of primary pp collisions per bunch crossing and uses this information in order to reject events with a high multiplicity. Based on all this information, the L0 trigger is able to split the stored events into a muon/dimuon, a hadron and an electron/photon stream. It is possible for an event to be stored in more that one stream at a time.

The HLT trigger runs a C++ application using 29000 CPU cores in order to reduce the event rate from 1 MHz down to a few kHz. The aim of the HLT1 stage is to first reduce to rate to roughly 30 kHz where a full reconstruction can be performed. In order to do so, the HLT1 stage performs a partial reconstruction of the events where the decisions from the L0 level are further investigated using the additional information from the VELO and tracking stations. By using these sub-detectors it is possible to request that the tracks possess either a high impact parameter or a high transverse momentum. During the Run 1 data period, the events were then sent immediately to the HLT2 stage to be processed. For the Run 2 data, the trigger scheme was slightly altered such that events which passed the HLT1 stage would first be stored on disk for about 150 hours. This buffer allows to perform online alignment and calibration in

LHCb 2012 Trigger Diagram

40 MHz bunch crossing rate

L0 Hardware Trigger : 1 MHz readout, high E_T/P_T signatures

| 450 kHz h± | 400 kHz μ/μμ | 150 kHz e/γ |

Software High Level Trigger

29000 Logical CPU cores

Offline reconstruction tuned to trigger time constraints

Mixture of exclusive and inclusive selection algorithms

5 kHz (0.3 GB/s) to storage

| 2 kHz Inclusive Topological | 2 kHz Inclusive/ Exclusive Charm | 1 kHz Muon and DiMuon |

LHCb 2015 Trigger Diagram

40 MHz bunch crossing rate

L0 Hardware Trigger : 1 MHz readout, high E_T/P_T signatures

| 450 kHz h± | 400 kHz μ/μμ | 150 kHz e/γ |

Software High Level Trigger

Partial event reconstruction, select displaced tracks/vertices and dimuons

Buffer events to disk, perform online detector calibration and alignment

Full offline-like event selection, mixture of inclusive and exclusive triggers

12.5 kHz (0.6 GB/s) to storage

Figure 2.17 – Trigger schemes during Run 1 (left) and Run 2 (right) data taking periods. Figures taken from Ref. [62].

order to increase the data quality. At the HLT2 stage, the rate is sufficiently low such that it is possible to fully reconstruct an event. It uses sets of inclusive (*i.e.* not tailored for a specific decay) and exclusive lines in order to further reduce the event rate. An important type of inclusive trigger lines is the one based on topological information *e.g.* requesting vertices produced by 2, 3 or 4 tracks with high invariant mass. This stage reduces the event rate to approximately 10 kHz at which point the data is stored on disk in order to be further analysed offline. The schemes of the trigger algorithm for the Run 1 and Run 2 data taking periods are shown in Fig. 2.17.

There are three trigger decision categories used to describe the properties of an event where a signal decay candidate is recorded. The candidate is classified as Triggered On Signal (TOS) if the requirements of the trigger lines are passed by particles belonging exclusively to the signal candidate. On the contrary, if the decision is positive based solely on particles belonging to the rest of the underlying event, the candidate is classified as Triggered Independently of Signal (TIS). In the case where both particles from the signal candidate and the rest of the event are required in order to obtain a positive trigger decision, the candidate is classified as Triggered

On Both (TOB).

2.2.5   The LHCb software

The LHCb software is designed around the GAUDI [63] framework, which allows the use of C++ applications that handle specialised tasks related to the data flow. This framework is versatile and allows to treat both real and simulated data in a similar manner. Most analyses also use simulated data based on Monte Carlo (MC) techniques for various purposes such as measuring efficiencies associated to requirements or determine the mass distribution of a given signal candidate decay. The applications which are commonly used in data analysis, where the first few are specific to the simulated data, are described below.

- Gauss is the application that is used to simulate data created within the LHCb environment. The simulation of the pp collision as well as the hadronisation of the various quarks produced in the collision are performed by the Pythia [64, 65] package which has been tuned to reproduce the features of the data collected by the LHCb detector. The decay of the particles produced in the event is described by the EvtGen [66] package where the final state radiation is generated using Photos [67]. The interaction of the particles with the detector is then simulated using the Geant4 [68, 69] toolkit.

- Boole is used to digitise the response of the LHCb detector to the interaction from the particles produced in the simulation. Boole also simulates the hardware trigger's response.

- Moore performs the software trigger's evaluation. After this stage, both real and simulated data are being treated in the same manner.

- Brunel realises the full track reconstruction and stores the information in data storage tape (DST) files which can then be further used for offline analysis.

- DaVinci is the application which combines the information from the final state particles in order to reconstruct a given signal decay of interest. As such, it is the application used for offline studies. When reconstructing a given decay mode, a set of loose requirements, called a stripping line, is applied to make a preliminary filtering of the events. Additionally, this step converts the data from the DST file to a ROOT [70] ntuple which is used to perform the analysis offline.

3 The LHCb detector upgrade

The LHCb experiment has been running successfully during the Run 1 and Run 2 data taking periods and has managed to obtain several interesting results. As explained in Chapter 2, the integrated luminosity is a key element in the context of the physics in the LHC environment. It essentially drives the statistical precision of the measurements performed at LHCb as well as the potential physics reach in the context of searches for rare or SM-forbidden decays. However, the LHCb experiment in Run 1 and 2 was not using the full luminosity delivered by the LHC. The main limiting factor hindering the full use of the luminosity delivered by the LHC is the readout rate from the L0 trigger level. In this context, an upgrade of the detector is being performed during the long shutdown 2 period (2018–2021) in order to maximise the physics output of the LHCb experiment during the Run 3 (2022–2024) and Run 4 (2028–2030) data taking periods. During these periods, the collision centre-of-mass energy will be $\sqrt{s} = 14\,\text{TeV}$ and the instantaneous luminosity will be $\mathcal{L} = 2 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$. The aim is to collect an integrated luminosity of $50\,\text{fb}^{-1}$ with an expected pile-up of 5. The upgrade relies on two major modifications. The first is that the L0 trigger system is replaced by a full software trigger able to read out data at a 40 MHz rate. The second is a modification of multiple detector sub-systems in order to be able to cope with the increased pile-up and sustain the expected radiation damage during the following data taking periods. Amongst these sub-systems, the upgrade of the tracking stations is of particular interest in the context of this thesis since a sizeable part of my work has been towards this particular upgrade and will be discussed in more details in Sec. 3.2.

As stated previously, the upgraded trigger will be fully software-based and will be able to operate at a 40 MHz rate. More precisely, it needs to be able to operate at a 30 MHz rate since this is the inelastic collision rate. It will perform a full event reconstruction and separate the data according to either inclusive or exclusive selection requirements. Similarly to the Run 2 trigger scheme, the data is then buffered such as to be able to perform online calibrations. Once this is done, various track quality and particle identification information will be added to

LHCb Upgrade Trigger Diagram

30 MHz inelastic event rate (full rate event building)

Software High Level Trigger

Full event reconstruction, inclusive and exclusive kinematic/geometric selections

Buffer events to disk, perform online detector calibration and alignment

Add offline precision particle identification and track quality information to selections

Output full event information for inclusive triggers, trigger candidates and related primary vertices for exclusive triggers

2-5 GB/s to storage

Figure 3.1 – Trigger scheme for the detector upgrade. Figure taken from Ref. [62].

the recorded events. Finally, the full event information will be saved in the case of the exclusive selection lines while only the triggered inclusive candidates will be saved. This new trigger scheme is shown in Fig. 3.1. Further details on the upgraded trigger are given in Ref. [71].

3.1 Tracking detectors

The tracking system will be almost entirely modified and there will be some improvements made to the PID system as well. Concerning the latter, the readout electronics of the Cherenkov detectors, the calorimeters and the muon stations are modified in order to meet the 40 MHz readout requirement. Additionally, the mirrors of the RICH detectors are replaced and the M1 station will be removed altogether since its main use was to provide information to the L0-hardware trigger. Further information about the upgraded PID system is given in Ref. [72]. Concerning the tracking system, the VELO technology will use hybrid pixel sensors which will improve both the track reconstruction speed and precision. Additionally, the position of this upgraded VELO will be closer to the beam which will improve the resolution on the measured impact parameter. A detailed description of the VELO upgrade can be found in Ref. [73]. The
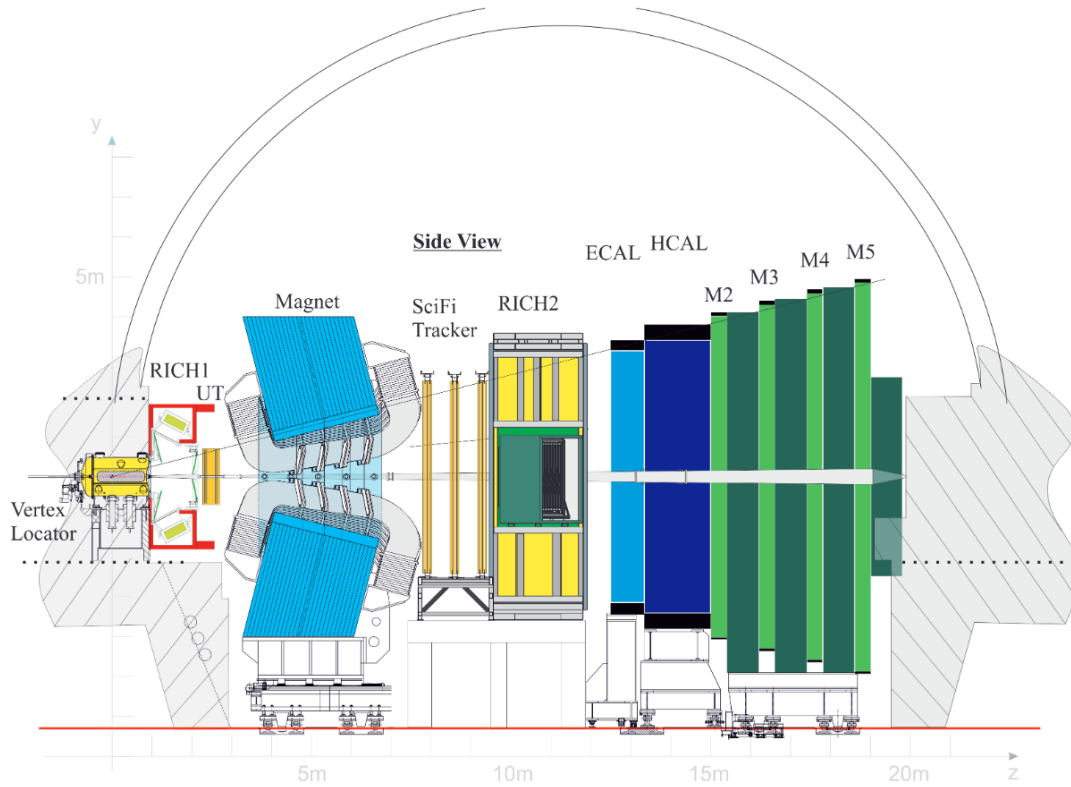
Figure 3.2 – Upgraded LHCb detector. Figure from Ref. [74].

TT is replaced by the Upstream Tracker (UT). The UT is a silicon-microstrip detector made of four detection layers. The first and last detection layers are parallel to the y-axis while the two middle layers are rotated by $\pm 5°$. The detectors granularity is adapted to the expected geometric occupancy by having a smaller granularity close to the beam pipe where the highest activity is expected. Additional details can be found in Ref. [74]. The three tracking stations will be replaced by the Scintillating Fibre (SciFi) tracker which, unlike the current IT and OT, uses a single technology, which is based on scintillating fibres. The SciFi is described in further details in Sec. 3.2. A schema of the upgraded detector is shown in Fig. 3.2.

3.2 The scintillating fibre tracker

As stated previously, the current tracking stations T1–T3 are being replaced by the scintillating fibre tracker. The core principle of this tracker is to collect, using silicon photomultipliers (SiPMs), the scintillation light produced whenever charged particles pass through the scintillating fibres. The choice of this particular technology is dictated by the following requirements:

- The electronics need to be operated at a 40 MHz rate.

- The single hit detection efficiency needs to be as large as possible, ideally above 99%.

- The single hit resolution in the magnet's bending plane needs to be better than 100 μm.

- The radiation length of a single detection layer needs to be sufficiently low ($X/X_0 < 1\%$) in order to minimise the impact of multiple scattering.

- The detector's designed performance needs to sustain the irradiation effects corresponding to an integrated luminosity of 50 fb^{-1}.

On top of these physics related constraints, the design of the detector has to account for the geometric constraints imposed by the position of the sub-detector within the whole LHCb detector. The design of the detector is then as follows. The 250 μm diameter scintillating fibre is used to produce 2.5 m-long densely packed arrays from which are built 0.5 m × 5 m modules. By joining side-by-side 10 or 12 of these modules, a single detection layer is built. Similarly to the Run 1 and 2 detector, the SciFi is made of three tracking stations with four detection layers per station. Once again, the stations are positioned in the usual x-u-v-x geometry, meaning that the two layers on the outside of a station are parallel to the y-axis while the two inner layers are rotated by ±5°. Both extremities of the modules are equipped with SiPM arrays to collect the scintillation light. Each mat is associated to 4 SiPM arrays with 128 channels each. A schematic of a single station is given in Fig. 3.3 while the layout of all three stations is shown in Fig. 3.4. The various components of this detector are described in great detail in Ref. [74]. In the following, some selected aspects of this detector will be discussed, with a particular emphasis on the production process of the fibre mats since a sizeable part of the work of this thesis has been dedicated to the said production.

3.2.1 Silicon photomultiplier

The SiPMs are solid state photon detectors designed for an optimal synergy with the tracking characteristics of the SciFi. In particular, their geometric structure has been tuned to match with geometric specifications of the fibre mats and they are designed to have a large photon detection efficiency which is necessary due to the low light yield generated by the scintillating fibres. A single SiPM detector is made of 128 channels split over two 64 channels silicon dies separated by 250 μm. A single channel is 250 μm wide and 1.5 mm long and is composed of 96 pixels where each pixel has dimensions 57.5 μm × 62.5 μm. The SiPM channels and fibres are not exactly aligned since the fibre mat thread is 275 μm wide. Each fibre mat, made of 500 fibres, is connected to four SiPM detectors. Each pixel is made of an avalanche photodiode operated in Geiger-Müller mode with its own quenching resistor whose role is to terminate the avalanche produced by the initial ionisation. In order to keep the noise to a minimum, the
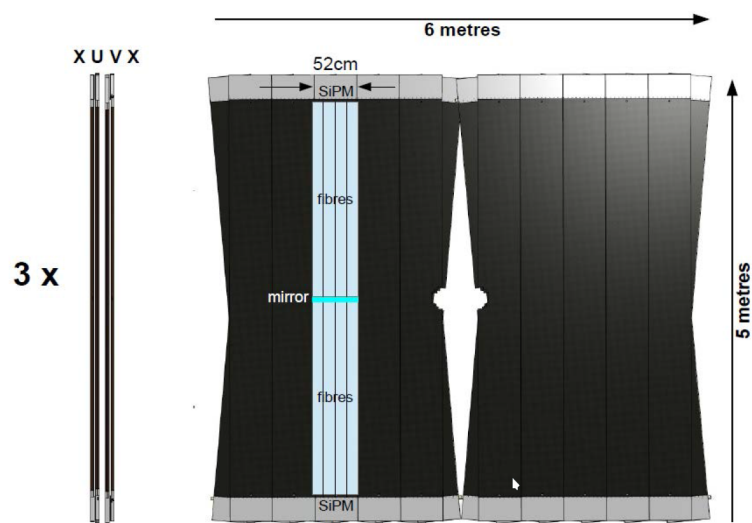
Figure 3.3 – Schematic drawing of a single SciFi station. Figure from Ref. [75].
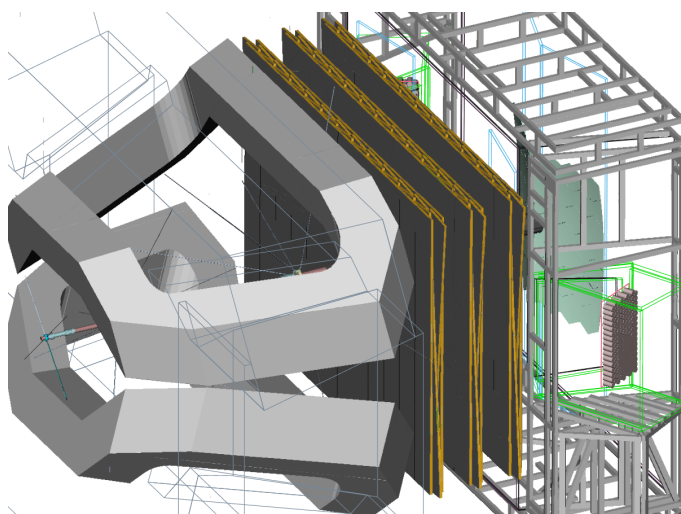
Figure 3.4 – Layout of the three SciFi stations positioned between the dipole magnet and the RICH2 detector. Figure from Ref. [74].
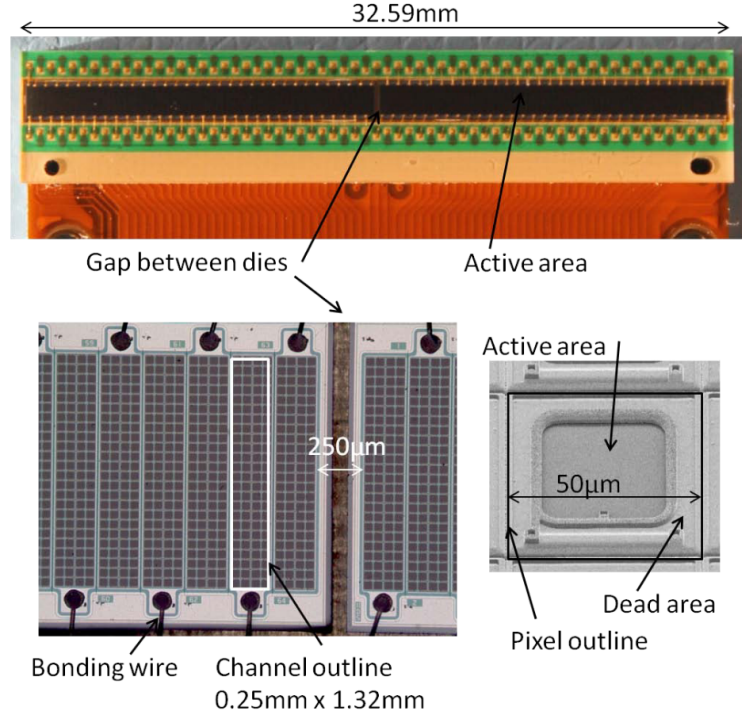
Figure 3.5 – Top part of the figure shows a single SiPM where the two 64 channel silicon dies can be seen. The bottom left of the figure shows various channels in greater detail while the bottom right of the figure shows a microscope image of a single pixel. Figure from Ref. [74].

SiPMs need to be cooled at $-40°C$. Pictures of a SiPM are given in Fig. 3.5. Further technical details are given in Refs. [76, 77].

3.2.2 Scintillating fibre

The active material in the SciFi tracker is the scintillating plastic fibre. The fibre is made by the company Kuraray in Japan. It consists of a polystyrene core, where the primary scintillation light is produced, doped with two dyes. The primary dye consists of p-terphenyl and is used to absorb the light produced by the core of the fibre and transmit the associated excitation energy to the second dye. The second dye is made of tetraphenyl-butadiene and acts as a wavelength shifter whose purpose is to modify the emission energy spectrum such as to minimise the light re-absorption by the polymer. The resulting decay time constant for the scintillation light is 2.8 ns. The fibre possesses a circular geometry with a diameter of 250 µm. In order to increase the amount of light collected by the fibre, a double cladding with decreasing refractive indices is used. This results in a photon collection efficiency of 5.34%. A sketch of the fibre is shown in Fig. 3.6. The emission wavelength's spectrum peaks at 450 nm when measuring
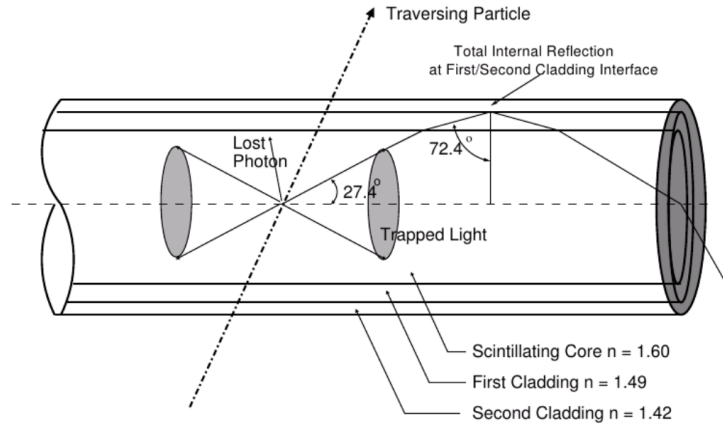
Figure 3.6 – Schematic of a scintillating fibre. Figure from Ref. [74].

close to the emission source and the absorption length is greater than 3.5 m for most of the emitted photons. The average propagation time along the fibre for the scintillation light is 6 ns/ m. As previously mentioned, an important constraint when designing the detectors for the LHCb upgrade is that they should be able to sustain the radiation's impact associated to the collection of 50 fb^{-1} of data. In this particular case, the radiation damage will cause a reduction of the attenuation length of the fibre. In the worst case scenario, it is expected that the signal loss in the central region of the detector, where the highest fluency is expected, will be smaller than 40% after the end of Run 4. The light yield produced by the fibres is further discussed in Sec. 3.4.

3.2.3 Fibre mats

The scintillating fibre is used to produce fibre mats. These mats are then arranged into modules. A single module is made of eight mats, with the top four mats being oriented such that their output light goes towards the positive y direction while the bottom four mats are oriented in the opposite direction. The organisation of the production chain was as follows. The fibre itself is first delivered to CERN where quality assurance tests are performed. It is then delivered to the winding centers which will produce and test the fibre mats. The four winding centers are situated in Aachen, Dortmund, Lausanne and Moscow. Once the mats are ready, they are delivered to the module assembly centers which are situated in Amsterdam and Heidelberg.

In the context of this thesis, a lot of work has been done for the production and quality assurance testing of the fibre mats. In the following, the various production and quality assurance steps required to make a fibre mat are explained in further details with an emphasis on the steps under my responsibility.

3.3 Fibre mat productions steps

As described previously, the nominal diameter of the fibre is 250 μm. In general however, the fibre has some imperfections and as such the diameter is not constant. For this reason, the fibre's quality is firstly tested at CERN where a machine is used to flatten the regions of the fibre where the diameter is too large, *i.e.* > 300 μm. Once the fibre has been treated at CERN it is sent to the winding institutes. The first step in the fabrication of a mat is the winding of the fibre. For this purpose, a winding machine specifically developed for the SciFi project is used. The machine keeps tension in the fibre while it is being laid onto a threaded wheel with a pitch size of 275 μm. Additionally, the machine monitors precisely the position of the fibre by using a laser system in order to avoid that the fibre accidentally skips over a channel. The wheel is threaded such as to build mats with 500 channels. After laying the fibre over the full 500 channels width, a layer of epoxy glue is laid on top of the fibres. The process is then repeated until 6 layers of fibres have been deposited on the wheel. Once the mat has been wound, it is cured in a controlled environment for 20 hours. The next step is to temper the mat for 2 hours at $40°C$ in order to reduce bending. Kapton foils are glued on both sides of the mat for optical and mechanical protection.

In order to fulfill the requirement of 100 μm single hit resolution, it is primordial to align the mats with great accuracy. Indeed, the fibre mats are quite flexible. In order to align the mats, the winding wheel possesses 11 dedicated pin holes. These holes are filled with glue during the winding process and as such leave solid bumps on the mat which can be used for the alignment process. Naturally, the positions of the pin holes on the winding wheel have to be precisely aligned. Four winding wheels are used at EPFL and a study of the alignment's precision of these pin holes has been performed in order to make sure that the 100 μm resolution requirement is met. For each wheel the position of the pin holes with respect to the thread is measured by taking pictures of each hole and using a software ruler. Assuming that the nominal thread pitch of 275 μm has negligible uncertainty, the distance between the left border of the pin hole and the second thread on the right of the hole is measured. This distance is measured for each hole and the difference between the two extrema for a given wheel gives an estimate of the alignment precision. The result of the study shows that two wheels out of four have a relatively poor alignment. As such, the decision was taken that the mats should be built in priority using the two wheels with a good alignment and that mats wound in the other two wheels would see their final quality rating systematically downgraded to "B". The quality assurance process based on which a grade is assigned is explained in detail later on. The measured distance of the pin holes to the thread is shown in Fig. 3.7 for a grade "A" and a grade "B" wheel.

After tempering, polycarbonate endpieces are glued to the extremities of the mat. These are used to provide pins which allow to reliably couple the mat to the SiPMs. They are also used to fix the mat inside a module. The gluing procedure is performed on a precision jig
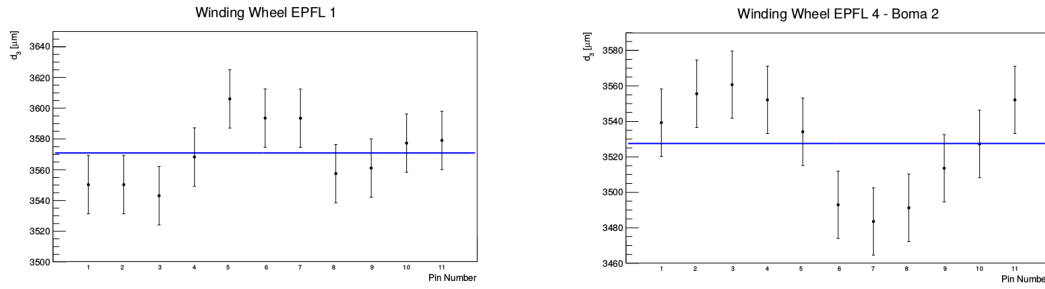
Figure 3.7 – Measured distance between the left border of a pin hole and the second thread on the right of the hole for each hole of a good (left) and bad (right) quality wheel. In the first case, the difference between the two extrema is roughly $70\,\mu$m while on the other it is $100\,\mu$m.
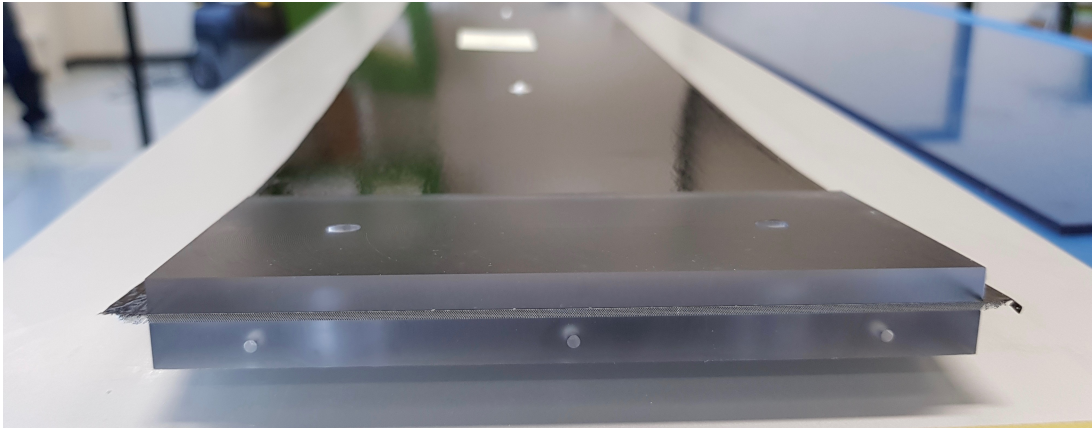
Figure 3.8 – Picture of the endpiece on the readout end of a mat. The layers of scintillating fibres which are then coupled to the SiPMs can be observed.

in a temperature controlled environment such as to cancel any possible thermal expansion of the mat. The size of the endpiece depends on which side of the mat it is fixed to. In the non-readout end of the mat, the size of the endpiece is smaller as it is designed to facilitate the application of a mirror. A picture of a mat with the endpiece glued on the SiPM end is shown in Fig. 3.8.

The SiPM end of the mat is then cut using a diamond-milling tool in order to have a clean contact surface between the mat and the SiPMs and minimise signal loss when transmitting the light from the mat to the SiPMs. Additionally, a longitudinal cut is performed over the full length of the mat in order to match the width of the mat with the required geometric tolerances. At this stage, a series of quality assurance tests is performed in order to assess the performance of the mat and check for any possible defect.
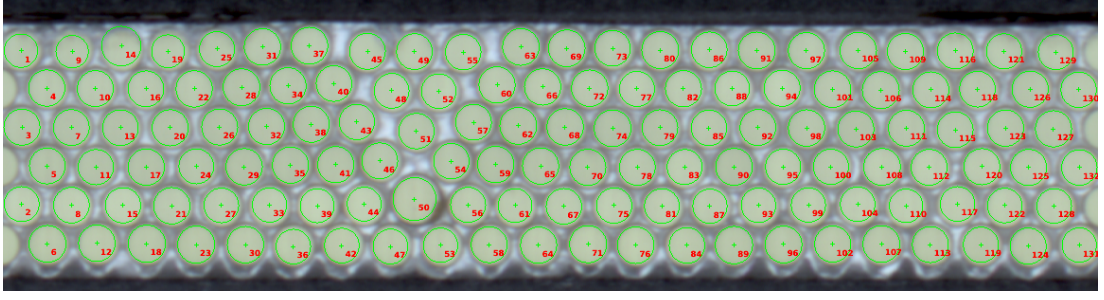
Figure 3.9 – Optical scan of a mat. It can be observed that the fibre No. 50 has an oversized diameter which causes a local disturbance to the structure of the adjacent fibre layers.

3.4 Fibre mat quality assurance

Various quality assurance tests are performed in order to assess the quality of each individual mat. The idea is to place the mats of highest quality close to the beam pipe where the occupancy is the highest and where it is critical to have a good resolution. The first quality assurance test is the optical scan. A high resolution scanner is used on both sides of the mat in order to control the transparency of the fibres as well as the geometric properties of the mat. In particular, the software used in the scanner is able to identify each individual fibre in the mat and check from the fibre's cross-section whether or not its shape is circular with the nominal 250 µm diameter or if there are deviations. When the diameter of a fibre is not conform to the nominal specifications, it can impact the overall geometry of the mat. In the most extreme cases, the upper most layer of the mat is not straight and does not fit within the geometric tolerances required to match with the SiPM channels. The scan counts the number of fibers from the top and bottom layers which fall outside the SiPM detection region. Based on this result, a grade is assigned to the mat from 1 (best) to 4 (worst). The scan of a mat with a defective fibre is shown in Fig. 3.9.

Once the optical scan has been performed, a light yield (LY) measurement of the mat is made. The purpose of this test is to measure the overall amount of light produced by the mat as well as check the uniformity in the channels' light output. The LY is defined as the average of the light output in photo-electrons measured by a SiPM array whenever a charged particle deposits energy in the fibre mat. In order to perform this measurement, a ^{90}Sr radioactive source emitting β^- particles is placed on the mirror-end of the mat. The following transitions occur with this source :

$$^{90}_{38}\text{Sr} \quad \rightarrow \quad ^{90}_{39}\text{Y}\, e^- \, \overline{\nu}_e \,, \tag{3.1}$$

$$^{90}_{39}\text{Y} \quad \rightarrow \quad ^{90}_{40}\text{Zr}\, e^- \, \overline{\nu}_e \,, \tag{3.2}$$

where the strontium and yttrium half-lives are roughly 29 years and 64 hours, respectively.

The released energy, distributed between the three decay products, is roughly 0.55 MeV for the ^{90}Sr decay and 2.28 MeV for the ^{90}Y decay. A scintillating bar is positioned at the level of the radioactive source under the fibre mat and is used as a trigger during the data acquisition. On the opposite side of the mat four SiPM arrays, totalling 512 channels, are used to measure the LY. In order to have the highest possible detection efficiency, the SiPMs are placed on a movable cart which is positioned as close as possible to the mat's extremity. Additionally, the vertical position of the SiPMs with respect to the mat can be adjusted using micrometer screw gauges. These are used to align the SiPM geometric acceptance with the fibre mat's cross-section. The mat is placed inside a dark box in order to shield it from ambient light. Indeed, the amount of light produced by the mat is many orders of magnitude below the amount of ambient light.

In order to have sufficiently high statistics, each scan is performed until 500×10^3 events are collected. After performing the scan, a report is created, which indicates the average cluster and channel LY as well as the cluster width as a function of the position across the fibre mat. When an electron traverses the mat, light is produced by each fibre along the particle's trajectory. The resulting light is typically detected by more than a single SiPM channel. As a result, it is necessary to devise an algorithm which is able to form a cluster from the light produced by a single particle. The properties of this cluster are then used to infer precisely the trajectory of the charged particle that traversed the mat. A sketch describing this multi-channel effect is given in Fig. 3.10. A report showing the results of a measurement is given in Fig. 3.11. The data corresponding to the measured mat is given in black while the red distributions correspond to the values measured on a reference mat. Additionally, for each measurement the ratio between the currently measured mat and the reference mat is shown. This ratio is used to check for possible defects such as LY non-uniformity over the 500 channels and is also used to help in the calibration of the position of the SiPM when using the micro-metric screws to perform the vertical alignment.

For each mat two scans are performed, one with and one without absorber. The absorber is a 3 mm-thick rectangular piece of plastic which is positioned between the mat and the trigger. It is used to absorb the lower energy electrons such as to reproduce the effect of measuring the signal given by minimally ionising particles, as only this type of particles would be able to activate the trigger. Usually, the effects caused by defects in the mat are more pronounced when performing the measurement with the absorber. Once these two scans have been performed, a mirror foil is glued to the non-readout extremity of the mat such as to increase the overall LY output. Once again, a scan with and without absorber is performed. This is used to ensure that the mirror gluing was properly performed. Additionally, the LY increase from the mirror can be measured. Typically, it consists of a 70% LY increase, depending on the position of the ^{90}Sr source. The distribution of the average cluster LY for a sizeable number of mats before and after mirror gluing is shown in Fig. 3.12.
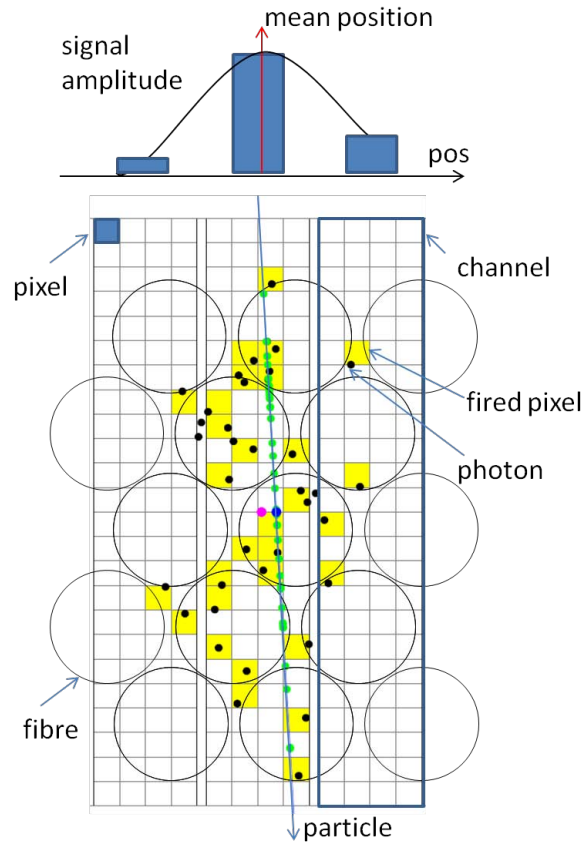
Figure 3.10 – Schematic of signal LY clustering. A charged particle goes through multiple fibres which produce light. The photons produced in the fibres are detected by the SiPMs. Each yellow square representing a SiPM pixel which has detected a photon. The total signal amplitude per channel is computed by summing over all fired pixels within that channel, as illustrated in the top part of the figure. By using a weighted mean of the signal over a few neighbouring channels, the particle's hit position can be determined. An algorithm studies the LY of the neighbouring channels to the one showing the highest signal in order to form a cluster. Figure from Ref. [74].
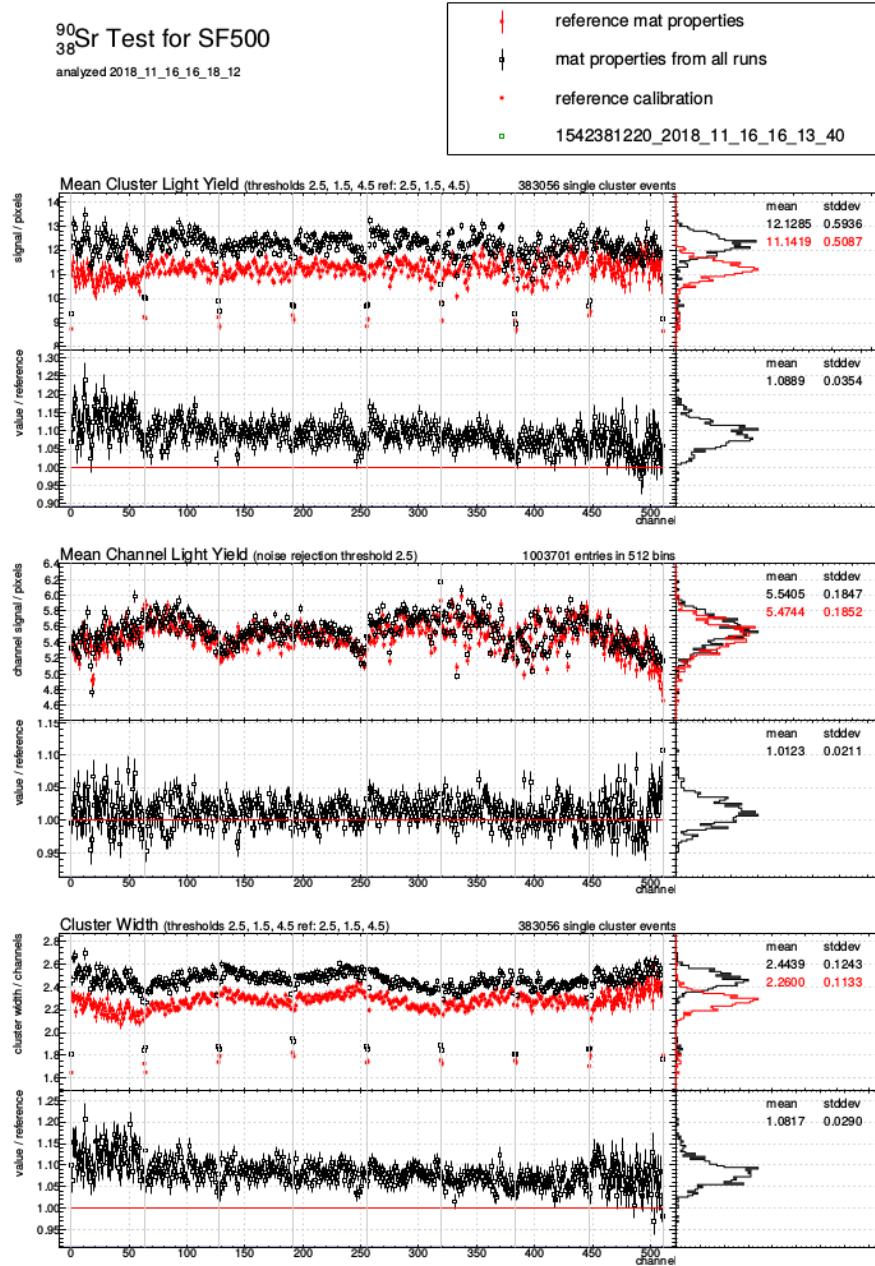
Figure 3.11 – Report from the LY measurement of a mat. The data corresponding to the measured mat is given in black while the red distributions correspond to the values measured on a reference mat. The top, middle and bottom plots show the mean cluster LY, the mean channel LY and the cluster width, respectively. Additionally, for each measurement the ratio between the currently measured mat and the reference mat is shown. The periodic drops in the measured values are due to the gaps between each 64-channels array of SiPMs.
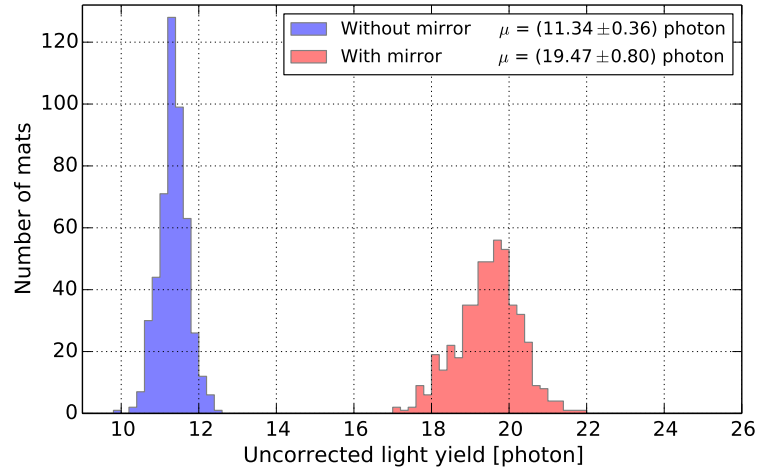
Figure 3.12 – Histogram of the average cluster LY before (blue) and after (red) mirror gluing. The two distributions correspond to the same mats. A 70% increase in the average LY after mirror gluing can be observed.

An issue which did arise during the production and had to be investigated is the dependence of the measured LY on the SiPM temperature. The LY measurement was not performed in a temperature-controlled environment but rather was done at room temperature. However, it is known that the breakdown voltage of the SiPMs depends on the temperature, which is one of the reasons why their temperature is controlled in the final detector. Since the applied bias voltage is kept constant for all the LY measurements, a change in breakdown voltage implies a change in the detector gain. This variation of the gain is not taken into account by default and as such the resulting measured LY is temperature dependent. In order to quantify this effect, a series of LY measurements on the same fibre mat was performed for varying SiPM temperatures. The resulting distribution of cluster LY as a function of SiPM temperature is then fitted using a linear function. The resulting slope of this function is $-0.153 \pm 0.004°C^{-1}$, as can be seen in Fig. 3.13. The result of this study was then used to improve the reliability of the LY measurements by systematically correcting the raw measured LY using this correction factor of $-0.153 \pm 0.004°C^{-1}$. This was shown to improve the overall LY measurement procedure's quality as the distribution of cluster LY per mat had a better resolution after applying this correction factor.

Finally, once both LY measurements, with and without mirror foil, have been completed, a final inspection of the fibre mat is performed. In particular, it is checked that the mat fits the overall geometric constraints, the transverse bending is measured and any local disturbance of the fibre matrix structure is registered in a database shared by all the winding institutes. Additionally, a final grade is assigned to the mat in order to assess its overall quality. The grade is composed of two contributions which go in decreasing quality from A to D, based on the
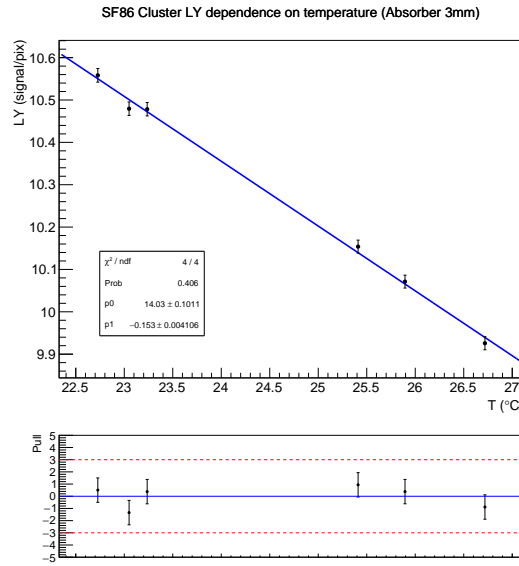
Figure 3.13 – Distribution of the average cluster LY as a function of the SiPM temperature. The measured LY is shown in black while the blue curve is a linear fit. The uncertainty on the measured LY values was rescaled to produce $\chi^2/\text{ndf} = 1$.

optical scan's result (as well as on the winding wheel which was used) and from 1 to 4, based on the LY scan's result. The final grade is then the combination of the letter and the number such that a mat of top quality will be graded A1 while a mat of lowest quality will receive the grade D4. In the final detector, mats of the highest quality will be placed at the centre, in the region of highest fluency, while mats of lower quality will be used in the regions of lower occupancy or be kept as spare parts.

4 Event selection for the $B^+ \rightarrow K^+\tau^-\mu^+$ and $B^+ \rightarrow K^+\tau^+\mu^-$ decay modes

The goal of this search is to measure the branching fractions $\mathscr{B}(B^+ \rightarrow K^+\tau^-\mu^+)$ and $\mathscr{B}(B^+ \rightarrow K^+\tau^+\mu^-)$. In order to do so, an invariant mass fit will be performed. In the case where the signal yield is significant, the branching fraction can be directly measured, otherwise an upper limit will be set instead. The $B^+ \rightarrow K^+\tau^\pm\mu^\mp$ decay candidates are reconstructed through the $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ decay mode, but also include the $\tau^- \rightarrow \pi^-\pi^+\pi^-\pi^0\nu_\tau$ decay mode. Both of these sub-decays account for approximately 14% of all τ^- decays. In comparison, the decay mode $\tau^- \rightarrow \mu^-\overline{\nu}_\mu\nu_\tau$ accounts for approximately 17% of all decays, but has the disadvantage of containing two neutrinos which cannot be reconstructed. Additionally, the decay modes with three pions have the advantage that they produce a fairly good decay vertex for the tauon which can be exploited to improve the reconstruction of the signal candidates and the background rejection.

Instead of measuring directly the branching fractions of the $B^+ \rightarrow K^+\tau^\pm\mu^\mp$ decay modes, a normalisation channel is used and the ratios of branching fractions are determined from the ratio of yields of the signal and normalisation decay modes. Performing the measurement in this way removes the need to use the $b\overline{b}$ production cross-section nor the integrated luminosity, which are two parameters known with relatively poor precision. Additionally, the use of a ratio of yields and efficiencies is useful in suppressing systematic effects which affect similarly both channels. Furthermore, the normalisation channel can be used to cross-check the sanity of the analysis procedure through various tests which cannot be done using the signal channel. For these reasons it is important to select a decay mode with features similar to that of the signal decay. Additionally, it should be a relatively abundant decay, easy to measure and which does not introduce peculiar systematic effects. As a result, the decay $B^+ \rightarrow \overline{D}^0(\rightarrow K^+\pi^-)D_s^+(\rightarrow K^+K^-\pi^+)$ is chosen as the normalisation channel. The topology of this decay mode is similar to that of the $B^+ \rightarrow K^+\tau^\pm\mu^\mp$ decays. In particular, the D_s^+ and $\overline{D}^0$ decay vertices mimic respectively the τ^- and $K\mu$ vertices. Furthermore, the number of charged tracks is the same for both modes. The topologies of the signal and normalisation
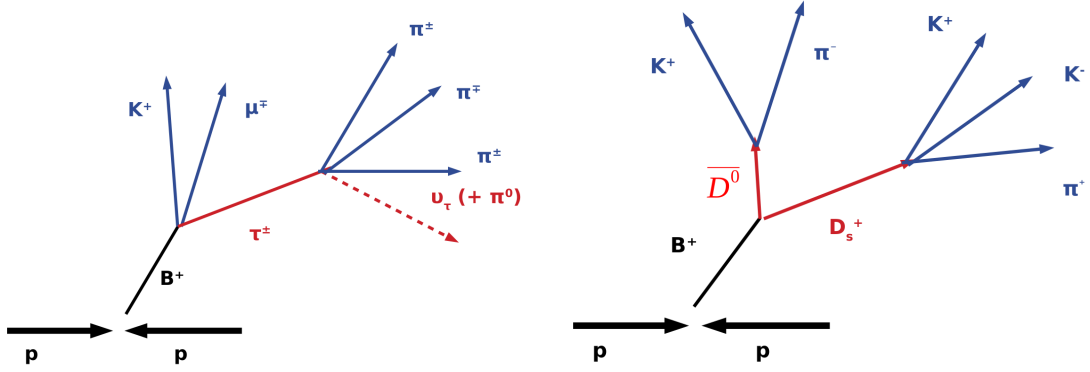
Figure 4.1 – Decay topologies of the $B^+ \to K^+\tau^\pm\mu^\mp$ (left) and $B^+ \to \overline{D}^0 D_s^+$ (right) decay modes.

decay modes are illustrated in Fig. 4.1.

One of the main difficulties in this kind of searches is being able to distinguish the events which contain the decay of interest from the events where the reconstructed candidate merely displays features similar to those of the signal decay but actually originates from one or multiple other decay modes unrelated to the signal. The content of the data samples under study is by far dominated by these so-called background decays. It is primordial to be able to reliably distinguish between the real and fake signal decays. If background decays are wrongly assumed to be signal decays, the signal yield will end up being overestimated, which in turn results in overestimated branching fractions. However, correctly identifying signal and background decays is not sufficient for a good measurement. Even if the signal and background yields are correctly measured, as long as the proportion of background decays far outweighs that of the signal, the resulting measurement will have a poor statistical precision. For these reasons, a sizeable part of the analysis consists in improving the separation between the signal and the background. This is done in various steps as there is a variety of background decays, each with its own features, which need to be suppressed by using different strategies. Once that the background contribution has been sufficiently well reduced, the yields of the $B^+ \to K^+\tau^\pm\mu^\mp$ and $B^+ \to \overline{D}^0 D_s^+$ decays can be measured through the use of an invariant mass fit. In order to not bias the analysis procedure, the measurement is performed blind, meaning that the signal region in data is not observed until the analysis procedure has been finalised and all the sanity cross-checks performed. This way of proceeding is to ensure that no human bias is introduced. Since the analysis is blind, it is necessary to define the limit setting procedure in case the significance of the measured branching fraction is too small. Once that this has been done and that the sources of systematic uncertainty have been taken into account, the final measurement can be performed.

The structure of this chapter is as follows. The various data and simulated samples used in

the analysis are described in Sec. 4.1. The issue of the undetected neutrino from the τ decay in the signal reconstruction is addressed in Sec. 4.2. The invariant mass obtained from an over-constrained kinematic fit is shown to provide the best resolution when compared to other standard mass reconstruction variables. The first step of the selection procedure is then defined in Sec. 4.3 and is based on offline requirements as well as on a combination of requirements imposed on the outputs of trigger and stripping lines. The next step in background suppression, described in Sec. 4.4, is through the use of multivariate analysis techniques. The remaining sources of background are even further suppressed through the use of a combination of particle identification variables. The distributions of these variables are corrected using a data-driven method in order to improve the agreement between data and simulation. This procedure is described in Sec. 4.5. The requirements on the multivariate analysis techniques and particle identification distributions are optimised to obtain the best expected limits on the signal branching fractions in the absence of signal. Section 4.6 gives the value of selection efficiencies needed to compute the signal branching fraction. Additionally, in this section the effect of the selection on various sources of background is studied. The fitting and limit setting procedures are explained in Sec. 5.1 and 5.2. A simultaneous maximum likelihood fit of mass distributions of the signal and normalisation candidates is performed in order to measure the signal branching fractions relative to the normalisation branching fraction in Sec. 5.3. The study of the sources of systematic uncertainties is discussed extensively in Sec. 5.4. Finally, the results are given in Sec. 5.5.

4.1   Data and simulation samples

For this search, the full data set collected by the LHCb experiment is being used, which contains approximately $9\,\mathrm{fb}^{-1}$ of integrated luminosity. This corresponds to the data acquired during 2011 and 2012, referred to collectively as Run 1, as well as the data acquired during 2015, 2016, 2017 and 2018 referred to as Run 2. Additionally, various simulated samples which reproduce the conditions of the real data have been requested. These samples simulate the pp collisions taking into account the specific conditions that were present in a given year. In each simulated event, one of the b quarks produced in the collision is forced to hadronise into a B^+ and subsequently decay into the signal decay mode of interest while the rest of the event evolves as it would in the real data. The decay chain is configured such as to reproduce accurately the real physics of each particle's sub-decay according to kinematics and the spin density matrix. In the case of the $B^+ \rightarrow K^+ \tau^\pm \mu^\mp$ decay modes, the decay of the B^+ meson is described by a phase-space model while the decay of the τ^- lepton is described by the `TAUOLA` model [78]. More precisely, the `TAUOLA 5` and `TAUOLA 8` decay models are used to describe the $\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \rightarrow \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decays, respectively. For the $B^+ \rightarrow \overline{D}^0(\rightarrow K^+ \pi^-)D_s^+(\rightarrow K^+ K^- \pi^+)$ decay mode, the D_s^+ decay follows the `D_Dalitz` model which is based on a measurement from the BaBar collaboration [79] while both the B^+ and $\overline{D}^0$

meson decays are described by a phase-space model.

When producing simulated samples it is possible to apply a set of requirements in order to filter out candidates which would be uninteresting in the analysis. This also has the advantage of saving CPU time. For these reasons, every simulated sample requires the signal candidate's final state charged tracks to be located within the detector's geometric acceptance. Additionally, the following requirements are imposed on the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes: $p(\mu^-) > 2.5\,\text{GeV/}c$, $p_\text{T}(K^+) > 650\,\text{MeV/}c$, $p_\text{T}(\pi^+) > 220\,\text{MeV/}c$ and $p_\text{T}(B^+) > 2.5\,\text{GeV/}c$. These requirements are less stringent versions of requirements which are later applied in the stripping line, as discussed in Sec. 4.3. The requirements should not be exactly the same in order to avoid resolution effects. Indeed, the requirements applied at generator level, *i.e.* during the simulation, are imposed on generated quantities while the stripping line imposes requirements on reconstructed quantities. For the $B^+ \to \overline{D}^0(\to K^+ \pi^-) D_s^+(\to K^+ K^- \pi^+)$ decay mode, a minimum momentum of $1.6\,\text{GeV/}c$ is requested for each daughter of the charm mesons.

For the data samples the DAVINCI reconstruction software applies a standard momentum scaling correction while for the simulated samples a momentum smearing correction is applied. This is to account for the fact that the scale used to compute the momentum from the detector readings is slightly off in data while the momentum's resolution is overestimated in simulation. The number of events requested after generator level in simulation is the same for the signal and normalisation decay modes and is equal to 500k for 2011, 1M for 2012, 2016, 2017 and 2018 and 300k for 2015 configurations. The fact that these numbers differ depending on the year reflects the fact that the integrated luminosity in data has changed depending on the year. In the case of the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes, the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decays are produced in equal numbers. These quoted numbers of generated events are the sum of both the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decays. The generator level efficiencies associated with the simulated samples are given in given in Sec. 4.7.

4.2 Signal reconstruction

The $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes are reconstructed through the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ decay mode but in principle the events also contain $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decays. As a result, the final state contains either one or two undetected particles. This implies that using the $m(B^+)$ invariant mass of the visible decay products will yield a distribution whose mean value is not centred at the B^+ mass of $5279.34 \pm 0.12\,\text{MeV/}c^2$ [33] but rather at a lower value due to the missing energy. Additionally, the signal peak in the invariant mass distribution will have a relatively poor resolution. This is a common issue for semi-leptonic analyses with the presence of a neutrino in the final state. As a result, various correction methods have been devised in order to mitigate this effect of partial reconstruction. A standard correction technique is the

so-called minimally corrected mass which is given by [80],

$$m_{\mathrm{MCM}}(B^+) = \sqrt{m(B^+)^2 + p_\perp^2} + p_\perp, \tag{4.1}$$

where $m(B^+)$ is the B^+ visible invariant mass and $p_\perp$ is the component of the visible momentum of the B^+ orthogonal to the B^+ flight direction, which is determined from the B^+ and $K^\pm \mu^\mp$ production vertices.

Instead of using the minimally corrected mass to correct for the missing neutrino energy, a different approach based on the "DecayTreeFitter" (DTF) algorithm [81] is chosen. The advantage of this method compared to the minimally corrected mass is that the DTF correction can fully retrieve the missing energy in the decay. The idea is to fit the decay chain with a given decay mode hypothesis while imposing various constraints to reduce the number of degrees of freedom. Assuming an event in data contains a $B^+ \to K^+ \tau^\pm \mu^\mp$ decay, it is impossible to know a priori whether the decay proceeded through a $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ or a $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ transition. As a result, it is necessary to make an assumption for the signal decay chain when applying the DTF algorithm. Since the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ decay mode is expected to be the dominant source of signal, based simply on branching fraction arguments, the algorithm reconstructs the decay chain $B^+ \to K^+ \tau^\pm \mu^\mp$ with $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$. Even when simulating $B^+ \to K^+ \tau^\pm \mu^\mp$ with $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ the algorithm assumes that the only missing neutral particle is a neutrino.

The DTF algorithm applies a constrained fit to the decay chain including the neutrino, taking into account the uncertainties and imposing constraints on the τ^- and neutrino masses as well as the production and decay vertices of the B^+ and $\tau^\mp$ candidates. The number of constraints can be computed as follows. In the decay chain there are 8 particles whose four-momenta have to be determined, these being the B^+, K^+, τ^-, μ^+, ν_τ and the three pions. Out of these, the three-momenta of the final state charged tracks are measured and the world average measurement of their mass by the Particle Data Group (PDG) [33] for their corresponding identity is assigned to them. This means that the four-momenta of the three pions, the kaon and the muon are fully determined. This leaves 12 unknowns corresponding to the components of the B^+, τ^- and ν_τ four-momenta. The energy-momentum conservation equations corresponding to the B^+ and τ^- decays remove an additional 8 unknowns. The flight directions of these last two particles are constrained by using the information on their decay vertices, which removes two degrees of freedom per particle. Finally, the masses of the τ^- and ν_τ are constrained to their known masses, adding two extra constraints. The result is a fit that is over-constrained by 2 equations, sometimes referred to as a 2C fit.

The mass distribution obtained from the DTF algorithm is studied using signal simulation candidates and is compared to the minimally corrected mass distribution from Eq. (4.1). The usual visible invariant mass distribution is also added in the comparison to provide a reference.
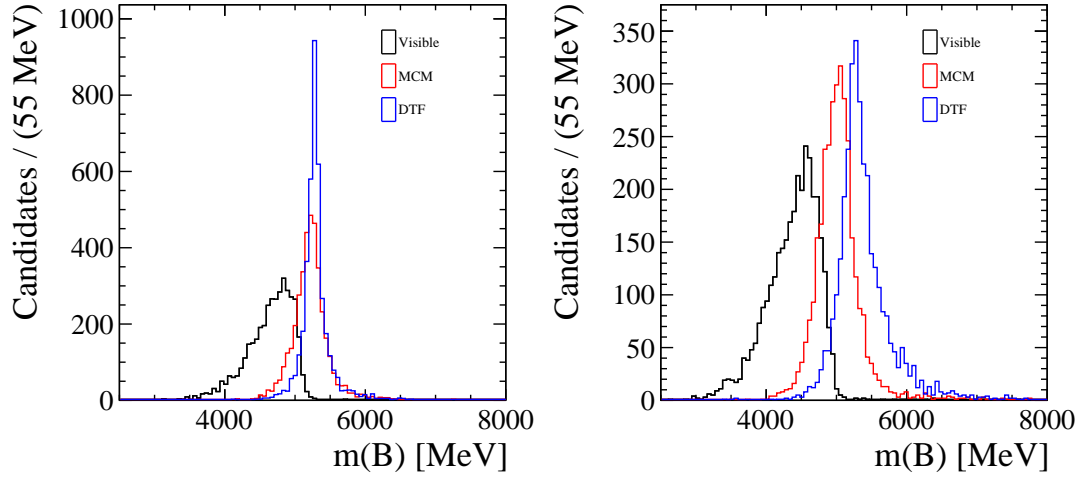
Figure 4.2 – Comparison of different mass variables for 2016 $B^+ \to K^+ \tau^- \mu^+$ signal simulation candidates after imposing the stripping and trigger selections. The left and right plots show respectively the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay modes. The different masses variables are: visible in black, minimally corrected in red and DecayTreeFitter in blue. Natural units, where $c = 1$, are used.

The comparison between these three distributions for the simulated $B^+ \to K^+ \tau^- \mu^+$ decay candidates is shown in Fig. 4.2, for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay modes. Similarly, Fig. 4.3 shows this comparison for signal $B^+ \to K^+ \tau^+ \mu^-$ decay candidates. These datasets correspond to 2016 data conditions and have the stripping and trigger requirements applied (the selection is discussed in further detail in Sec. 4.3). As can be observed in these figures, the DTF distribution outperforms both the visible as well as the minimally corrected mass distributions. Indeed, the DTF distributions are centred more closely to the known value of the B^+ meson mass and also have a better resolution, with a core resolution for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ decay mode of approximately $50 \, \text{MeV}/c^2$ for the DTF distributions. Naturally, the resolution for the $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay mode is worse than for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ mode. This behaviour is expected since the decay chain reconstructed by the DTF algorithm in the case of the $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay mode is different from the real decay because of the missing neutral pion. Nevertheless, the core resolution of both decay modes is remarkably good considering the presence of undetected neutral particles in the final state.

While the study of the signal simulation shows that the DTF mass distribution performs well in the signal reconstruction, this variable could in principle prove to be problematic when applied to background candidates. The main worry is that by using the DTF mass variable, peaking structures, especially in the signal region, would be created for background candidates whose standard invariant mass distribution is smoothly decaying. This could be quite problematic as
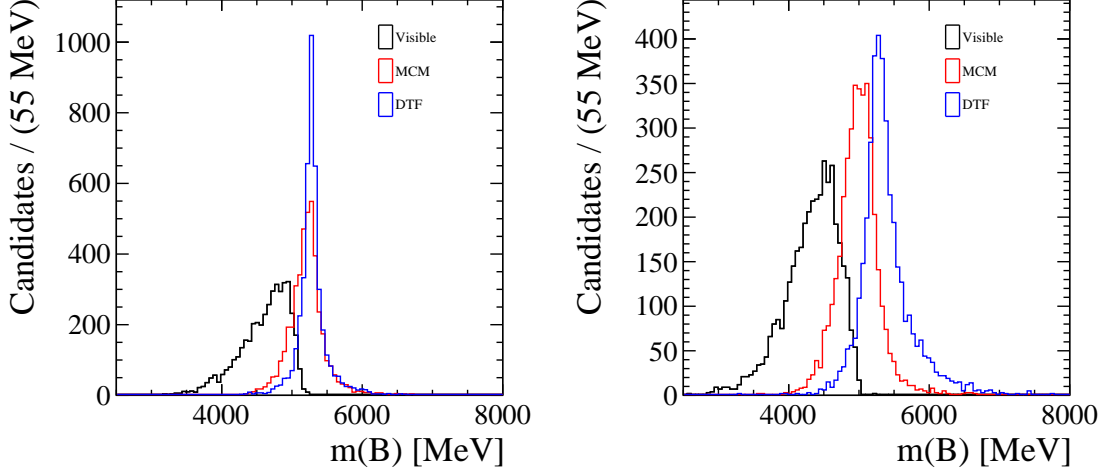
Figure 4.3 – Comparison of different mass variables for 2016 signal simulation after stripping and trigger selection for $B^+ \to K^+ \tau^+ \mu^-$ candidates. The left plot shows the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ mode while the right plot shows $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$. The different masses are : visible in black, minimally corrected in red and DecayTreeFitter in blue. Natural units, where $c = 1$, are used.

these background candidates could prove difficult to distinguish from signal candidates. In order to make sure that this is not the case, the DTF mass distribution is inspected for two data samples with 2016 conditions where the candidates are reconstructed as $B^- \to K^+ \tau^- \mu^-$ and $B^+ \to K^+ \tau^+ \mu^+$. Using these data samples, where the two leptons carry the same electric charge ensures that there is no signal since the $B^+ \to K^+ \tau^+ \mu^+$ decay is completely forbidden due to non-conservation of the electric charge while the $B^- \to K^+ \tau^- \mu^-$ decay is either forbidden or further suppressed than the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes in NP models which predict these decays. Since there is no signal, the distribution does not need to be blinded and can be shown over the full mass range, including in the signal region. Additionally, the various background modes contained in these samples share similar properties to those in the real data samples such that their invariant mass distribution, in this case the DTF mass, approximates well the behaviour of the invariant mass distribution of the background in the real data samples. The mass distributions for these two samples are shown in Fig. 4.4. No peaking structure is observed and the DTF mass distribution is smoothly decaying for large mass values. This test confirms that the DTF mass variable is well-behaved and can be used to measure the signal yield in an invariant mass fit. The DTF mass variable will be referred hereafter as $m_{\text{corr}}(B^+)$.

The study of the signal simulated samples shows that the region $4.8 < m_{\text{corr}}(B^+) < 5.8 \, \text{GeV}/c^2$ contains approximately 93% (85%) of all signal $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ ($\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$) decays. This region is chosen as the blinding window in data. The B^+ corrected mass distributions for data and simulation are shown in Fig. 4.5. An interesting feature, which can be observed
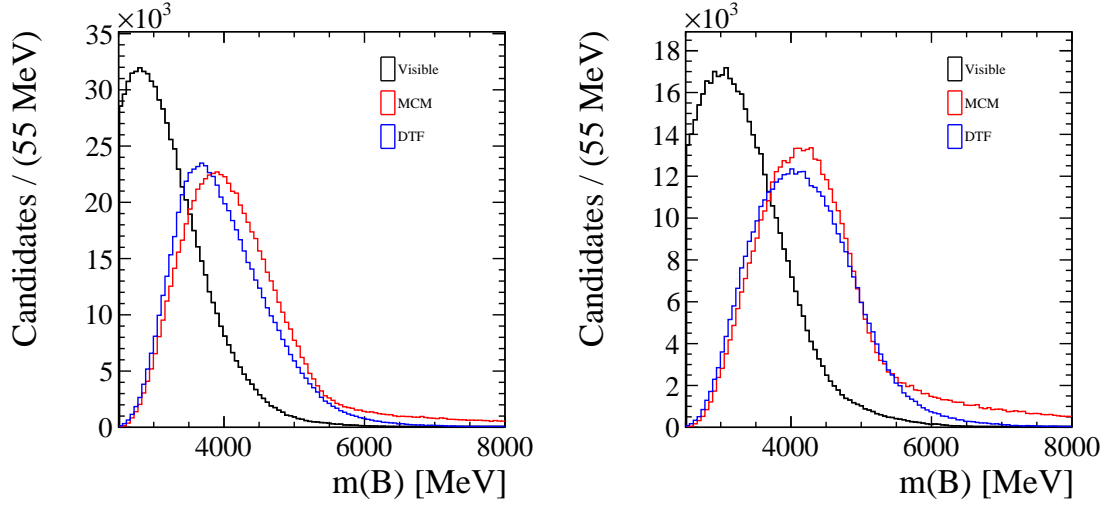
Figure 4.4 – Comparison of different mass variables in 2016 data after stripping and trigger selection reconstructed as $B^- \to K^+ \tau^- \mu^-$ (left) and $B^+ \to K^+ \tau^+ \mu^+$ (right). The different mass distributions are : visible in black, minimally corrected in red and DecayTreeFitter in blue. Natural units, where $c = 1$, are used.

in the plot, is that the mass distributions for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ data have sizeable differences. The number of candidates in the $B^+ \to K^+ \tau^+ \mu^-$ decay mode is approximately the double of that of the $B^+ \to K^+ \tau^- \mu^+$ decay mode. The fact that these two distributions are significantly different is expected since the content of the data samples is vastly dominated by background decays. Since the final state is different between the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, different types of processes contribute to the two data samples. In most cases, the backgrounds are B-meson decays partially reconstructed due to the presence of additional particles which are not used in the signal reconstruction. Often a sub-set of the final state particles are misidentified. The background content in the $B^+ \to K^+ \tau^- \mu^+$ sample is dominated by $B^+ \to \overline{D}^0 (\to K^+ \pi^- \pi^+ \pi^-) \mu^+ \nu_\mu$ decays while *e.g.* the $\overline{B}^0 \to D^0 (\to K^- \pi^+ \pi^- \pi^+) \pi^+ \mu^- \overline{\nu}_\mu$ decay mode is a dominant source of background for the $B^+ \to K^+ \tau^+ \mu^-$ sample. The decay topology of these two background decays is illustrated in Fig. 4.6. The various steps of the selection procedure used to reduce the proportion of these various background decay modes and their effect on the signal decay modes are explained in Secs. 4.3, 4.4 and 4.5. The efficiency of the full selection for the signal and contribution from the dominant background decays are presented in Sec. 4.6.
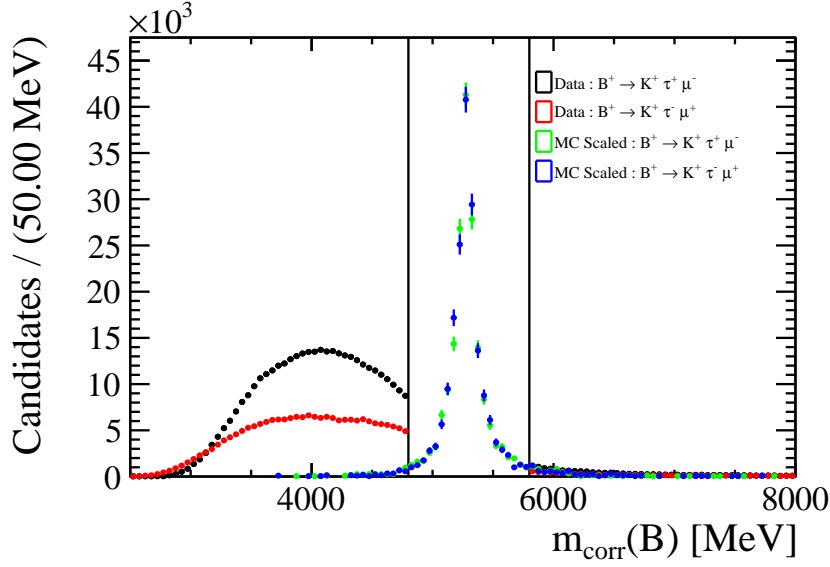
Figure 4.5 – Distribution of $m_{\mathrm{corr}}(B^+)$ for 2016 samples after applying the stripping and trigger requirements. The distributions in black and red are for real data where the candidates are reconstructed respectively as $B^+ \to K^+\tau^+\mu^-$ and $B^+ \to K^+\tau^-\mu^+$. The distributions in green and blue are for signal simulated samples, where $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$, for the $B^+ \to K^+\tau^+\mu^-$ and $B^+ \to K^+\tau^-\mu^+$ decay modes, respectively. The simulated distributions have been re-scaled arbitrarily to ease the comparison with the real data distributions. The vertical lines indicate the blinding window where the candidates in the real data have been removed. Natural units are used.
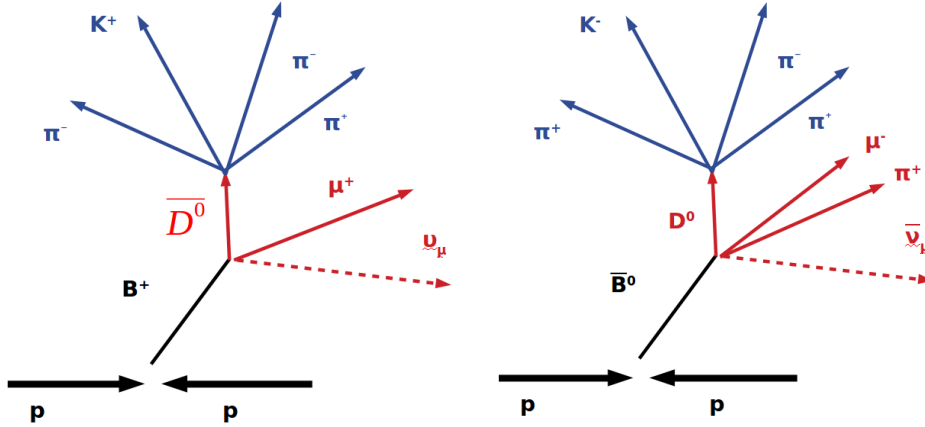
Figure 4.6 – Sketch of the generic topology of decay modes expected to give a dominant contribution to the background in the $B^+ \to K^+\tau^-\mu^+$ (left) and $B^+ \to K^+\tau^+\mu^-$ (right) data samples. This is illustrated by the $B^+ \to \overline{D}^0(\to K^+\pi^-\pi^+\pi^-)\mu^+\nu_\mu$ decay for the $B^+ \to K^+\tau^-\mu^+$ sample and the $\overline{B}^0 \to D^0(\to K^-\pi^+\pi^-\pi^+)\pi^+\mu^-\overline{\nu}_\mu$ decay for the $B^+ \to K^+\tau^+\mu^-$ sample.

4.3 Signal preselection

4.3.1 Stripping selection

During the Run 1 and Run 2 periods of the LHC, the LHCb detector has acquired a large amount of data corresponding to $9\,\mathrm{fb}^{-1}$ of pp collisions. As explained previously, it is essential to collect as much data as possible since it helps probe higher and higher energy scales in the search of new physics as well as improves the precision of the measurements in general. However, most of the data collected is of no interest in the search for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes. More precisely, the branching fractions for these two decay modes are expected to be very small according to NP models and the decays are forbidden in the SM. As a result, these decays are expected to be produced at a small rate, at best around one decay for every 10^5 B^+ meson decays, such that the vast majority of the data acquired is composed of background.

A sizeable proportion of these background decays can be suppressed by exploiting the kinematic and topological properties of the signal decays under study. Sets of loose requirements are designed in order to perform a first selection of the data by removing candidates whose properties are significantly different from what is expected for signal. These sets of requirements are called "stripping lines". Each analysis performed at LHCb needs to either use an already existing stripping line suitable for the study of the signal decays under consideration or define their own stripping line.

In this search, the `StrippingB2XTauMu_K_3pi_looseLine` line is used when reconstructing the $B^+ \to K^+\tau^\pm\mu^\mp$ decay candidates while the `StrippingB2XTau_DD0_Line` line is used when reconstructing the $B^+ \to \overline{D}^0 D_s^+$ candidates used as normalisation. The requirements imposed by these lines are given in Tabs. 4.1 and 4.2.

The various requirements can be split into various categories depending on the type of variable on which they are applied. It is important to use different types of variable in order to exploit properties useful in rejecting specific types of backgrounds. The most fundamental requirements are imposed on kinematic variables such as the (transverse) momentum $p_{(T)}$ of a given particle. By requesting a sufficiently high value for these quantities the amount of combinatorial background, *i.e.* random combinations of tracks, is suppressed. The reason why this is effective is because there are a many low-energy particles, mainly pions, produced during the hadronisation phase in proton-proton collisions. This results in a plethora of low energy pions, kaons and muons which can end up being used in the reconstruction of the signal candidate. Similarly, requirements are imposed on the invariant mass of intermediary unstable particles. Requesting a narrow window around the known mass of an intermediary particle suppresses the amount of background decays where this particle is not present. In the case of the $B^+ \to K^+\tau^\pm\mu^\mp$ decay modes, the requirements on the invariant masses of the tauon

Table 4.1 – Requirements applied by the `StrippingB2XTauMu_K_3pi_looseLine` stripping line used in the reconstruction of the $B^+ \to K^+ \tau^\pm \mu^\mp$ and $B^+ \to K^+ \tau^\pm \mu^\pm$ candidates.

π	$p_T > 250\,\mathrm{MeV}/c$
	$p > 2\,\mathrm{GeV}/c$
	$\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 4$
	$\chi^2_{\mathrm{IP}} > 16$
	`GhostProb` < 0.4
	`ProbNN`$_\pi > 0.55$
K	$p_T > 800\,\mathrm{MeV}/c$
	$p > 3\,\mathrm{GeV}/c$
	$\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 3$
	$\chi^2_{\mathrm{IP}} > 36$
	`GhostProb` < 0.5
	$\mathrm{DLL}_K > 5$
μ	$\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 3$
	$\chi^2_{\mathrm{IP}} > 36$
	`GhostProb` < 0.5
	$\mathrm{DLL}_\mu > 2$ and `isMuon`
$K\mu$	$p_T > 1\,\mathrm{GeV}/c$
	$p_T(K) + p_T(\mu) > 2\,\mathrm{GeV}/c$
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf} < 9$
τ	$500 < m < 2000\,\mathrm{MeV}/c^2$
	$p_T > 1\,\mathrm{GeV}/c$
	$p > 5\,\mathrm{GeV}/c$
	$\exists\, p_T(\pi) > 800\,\mathrm{MeV}/c$
	$\chi^2_{\mathrm{vtx}} < 16$
	$\chi^2_{\mathrm{IP}} > 36$
	$\chi^2_{\mathrm{FD}} > 16$
	`DIRA` > 0.99
	$0.1 < \rho = \sqrt{\Delta x^2_{PV} + \Delta y^2_{PV}} < 7\,\mathrm{mm}$
	$\Delta z_{PV} > 5\,\mathrm{mm}$
	`maxDOCA`$(\pi_i, \pi_j) < 0.2\,\mathrm{mm}$
B	$2 < m < 10\,\mathrm{GeV}/c^2$
	$p_T > 3\,\mathrm{GeV}/c$
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf} < 15$
	$\chi^2_{\mathrm{FD}} > 400$
	`DIRA` > 0.99

Table 4.2 – Requirements applied by the `StrippingB2XTau_DD0_Line` stripping line used in the reconstruction of the $B^+ \to \overline{D}^0 D_s^+$ candidates.

π and K	$p_{\mathrm{T}} > 250\,\mathrm{MeV}/c$
	$p > 2\,\mathrm{GeV}/c$
	$\chi^2_{\mathrm{trk}}/\mathrm{ndf} < 4$
	$\chi^2_{\mathrm{IP}} > 16$
	`GhostProb` < 0.4
π	`ProbNN`$_\pi > 0.55$
K	$\mathrm{DLL}_K > -5$
D^0 and D_s^+	$1800 < m < 2030\,\mathrm{MeV}/c^2$
	$p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$
	$\exists\, p_{\mathrm{T}}(\pi/K) > 800\,\mathrm{MeV}/c$
	$\chi^2_{\mathrm{vtx}} < 16$
	$\chi^2_{\mathrm{FD}} > 16$
	`DIRA` > 0.99
	$0.1 < \rho = \sqrt{\Delta x_{PV}^2 + \Delta y_{PV}^2} < 7\,\mathrm{mm}$
	$\Delta z_{PV} > 5\,\mathrm{mm}$
	$\mathrm{max}\texttt{DOCA} < 0.2\,\mathrm{mm}$
B	$5 < m < 7\,\mathrm{GeV}/c^2$
	$0 < m_{\mathrm{corr}} < 10\,\mathrm{GeV}/c^2$
	$p_{\mathrm{T}} > 2\,\mathrm{GeV}/c$
	$\exists\, p_{\mathrm{T}}(D^0/D_s^+) > 4\,\mathrm{GeV}/c$
	$\exists\, p_{\mathrm{T}}(\pi/K) > 2\,\mathrm{GeV}/c$
	$\sum_{\pi,K} p_{\mathrm{T}} > 7\,\mathrm{GeV}/c$
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf} < 90$
	$\chi^2_{\mathrm{FD}} > 225$
	$FD < 90\,\mathrm{mm}$
	`DIRA` > 0.99
	$\mathrm{max}_{D^0,D_s^+}(\chi^2_{\mathrm{IP}}) > 150$
	$\mathrm{min}_{D^0,D_s^+}(\chi^2_{\mathrm{IP}}) > 16$
	$\mathrm{max}\left(\mathrm{min}_{[\pi,K]_{D^0}}(\chi^2_{\mathrm{IP}}), \mathrm{min}_{[K,K,\pi]_{D_s^+}}(\chi^2_{\mathrm{IP}})\right) > 20$

and the B^+ meson, *i.e.* $500 < m(\tau^-) < 2000\,\text{MeV}/c^2$ and $2 < m(B^+) < 10\,\text{GeV}/c^2$, cannot be too stringent as these decays are only partially reconstructed and as such their reconstructed visible mass has a poor resolution, as can be seen in Fig. 4.2 and 4.3.

Requirements imposed on particle identification variables are used to reduce backgrounds associated with the presence of misidentified particles. The description of the various particle identification variables used in this analysis was given in Sec. 2.2.3. In the search for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes a sizeable part of the background coming from real b-hadron decays, also known as physics background, is due to decay modes where a sub-set of the final state particles are misidentified. As a result, it is primordial to require each particle in the final state to be well identified. The requirements from the stripping selection alone are arguably loose in order to keep a high PID selection efficiency for the signal. An additional requirement imposed on a combination of various PID variables is developed to further reject misidentified backgrounds while keeping a high efficiency on signal. This procedure is described in detail in Sec. 4.5.

Another category of variables used in the stripping selection is based on the decay topology. The idea is to make use of the information associated with the various vertices involved in the signal decay chain. This includes the fact that the particles in the final state come from the decay of intermediary particles and were not directly produced at the primary vertex. This information is encoded in the impact parameter (IP) which is defined as the minimal distance between a track and the primary vertex. Naturally, particles produced directly at the primary vertex, source of combinatorial background, have a low IP value while particles coming from the decay of intermediary states show a higher value. Instead of imposing requirements directly on the impact parameter itself, the requirements are imposed on χ^2_{IP} which is defined as the increase in the χ^2 of the primary vertex fit when including the track under consideration in the construction of the vertex. This variable behaves to first order as the impact parameter's significance, *i.e.* $\left(IP/\sigma(IP)\right)^2$ where $\sigma(IP)$ is the uncertainty on the IP. Another property of decays with the presence of unstable intermediary states with a sizeable lifetime, usually decaying through the weak interaction, is that their decay vertices show sizeable displacements with respect to their production vertices. In order to use this information lower bound requirements are imposed on the flight distance (FD), *i.e.* the distance between the production and decay vertices, of unstable particles as well as on χ^2_{FD} which is defined analogously to χ^2_{IP}. Additionally, it is required that the daughter particles coming from these unstable states form a good quality vertex. This is ensured by requesting a good vertex fit χ^2 meaning that upper bounds are imposed on χ^2_{vtx} or $\chi^2_{\text{vtx}}/\text{ndf}$. Similarly, upper bounds are imposed on the distance of closest approach (DOCA) between pair of tracks coming from a given vertex. This is the smallest distance between two tracks, which should be small if they come from the same vertex.

Other variables used in the stripping lines include `DIRA` which is defined as the cosine of the angle between the reconstructed momentum of a particle and its flight direction vector determined using its production and decay vertices. This value is expected to be very close to one for particles which are properly reconstructed. Some geometric properties are exploited as well, mainly by requesting the end vertex of intermediary states to be located, in the $x - y$ plane, within a ring centred around the primary vertex with given inner and outer radii. A lower bound on Δz_{PV} the distance, along the z-axis between the decay vertex of these unstable states and the primary vertex (PV) is also imposed.

4.3.2 Trigger selection

Additionally to the requirements imposed at the stripping level, a set of requirements based on trigger decisions is imposed. As explained in Sec. 2.2.4, various trigger lines which apply requirements based on the output of the hardware or software triggers are designed and used to select candidates with properties similar to what is expected for the signal decay under consideration. The trigger decisions are split into three categories which are TOS, TIS and TOB depending on whether the decision is positive based on particles belonging exclusively to the signal candidate decay, the rest of the underlying event, or both simultaneously. In this search, the decisions are always required to be TOS, both for the signal and normalisation decay modes.

At the hardware level, it is requested to have a positive decision from `L0MuonDecision` applied on the muon of the $B^+ \to K^+ \tau^- \mu^+$ or $B^+ \to K^+ \tau^+ \mu^-$ decay, while a positive decision from `L0HadronDecision` applied on the B^+ candidate of the $B^+ \to \overline{D}^0 D_s^+$ decay is requested for the normalisation channel. These two decisions increase the likelihood that the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay candidates contain a real muon and the $B^+ \to \overline{D}^0 D_s^+$ candidates contain a real B^+ meson. For Run 1 data, a positive decision from `Hlt1TrackAllL0Decision` is requested at the software level for the $B^+ \to \overline{D}^0 D_s^+$ decay candidates, while a positive decision from either `Hlt1TrackAllL0Decision`, `Hlt1TrackMuonDecision` or both is requested for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay candidates at the HLT1 level. For Run 2 data the `Hlt1TrackAllL0Decision` decision is replaced by `Hlt1TrackMVADecision`. These requirements essentially improve the rejection of candidates made of low-quality tracks. At the HLT2 level, the same requirements are imposed for the $B^+ \to K^+ \tau^\pm \mu^\mp$ and $B^+ \to \overline{D}^0 D_s^+$ decay candidates. For Run 1 data, a positive decision is requested from at least one line amongst `Hlt2TopoNBodyBBDTDecision`, where $N = 2$, 3 or 4 while for Run 2 data a positive decision is requested from at least one line amongst `Hlt2TopoNBodyDecision`, where $N = 2$, 3 or 4. These are topological requirements based on the presence of either 2, 3 or 4 particles in the final state. The trigger selection efficiency on signal candidates is mainly driven by the requirements at the L0 level, which result in efficiencies around 60% and 40% for the signal and normalisation decay modes, while

the combination of the software trigger decisions results in efficiencies at the 90% level, both for the signal and normalisation decay modes.

4.3.3   Additional requirements

While the application of stripping and trigger requirements is useful as a first step in selecting the data, the requirements which are imposed at this stage are still reasonably loose. This stems from the fact that the signal efficiency needs to be kept high since these requirements cannot be modified after the data has been recorded. The selection is then further tailored to reduce the specific backgrounds present in the data when reconstructing the signal candidates by applying offline requirements. The selection is firstly refined by applying a set of requirements aimed at reducing specific sources of physics background. Later on the selection is further optimised through the use of multivariate analysis techniques as well as by using a combination of PID variables. These procedures are described in Secs. 4.4 and 4.5.

While the selection has been kept identical until now for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, the additional requirements imposed on these decay modes need to differ due to the different background contributions. The first requirements which differ depending on the $K\mu$ charge combination are the following : $m(K^+ \pi^- \pi^+ \pi^-) > 2 \, \mathrm{GeV}/c^2$ is imposed on the $B^+ \to K^+ \tau^- \mu^+$ decay mode and $m(K^+ \mu^-) > 1 \, \mathrm{GeV}/c^2$ is imposed on the $B^+ \to K^+ \tau^+ \mu^-$ decay mode. In the first case, the requirement significantly reduces the proportion of background coming from charm meson decays. In particular, all candidates where the reconstructed kaon and pions actually come from a single $\overline{D}^0 \to K^+ \pi^- \pi^+ \pi^-$ decay are rejected. This source of background is quite sizeable as can be seen in Fig. 4.7, where there is a significant excess in the reconstructed $B^+ \to K^+ \tau^- \mu^+$ data of background candidates with a fully reconstructed $\overline{D}^0 \to K^+ \pi^- \pi^+ \pi^-$ decay. The second requirement reduces backgrounds from partially reconstructed semi-leptonic decays as well as some misidentified backgrounds such as $K^{*0}(892) \to K^+ \pi^-$ where the pion is misidentified as a muon or $\rho(770) \to \pi^+ \pi^-$ where one of the pions is misidentified as a kaon and the other as a muon.

Mass window requirements are imposed on the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes based on invariant mass windows for the B^+ and τ^- candidates, where the requirement is imposed on the minimally corrected mass of the tauon and not its visible mass (see Sec. 4.2). A good fit quality in the DTF reconstruction procedure is requested as well as a good DIRA and a sizeable displacement, in the positive z-direction, of the decay vertex of the τ^- candidate with respect to the $K\mu$ vertex. For the $B^+ \to \overline{D}^0 D_s^+$ decay mode, mass window requirements are imposed on the B^+, the $\overline{D}^0$ and the D_s^+ candidates. Additionally, a good quality vertex as well as a significant reconstructed lifetime are required for the B^+ candidate and sizeable displacements are required for the charm meson candidates. The full list of requirements imposed on both signal and normalisation channels is given in Tab. 4.3.
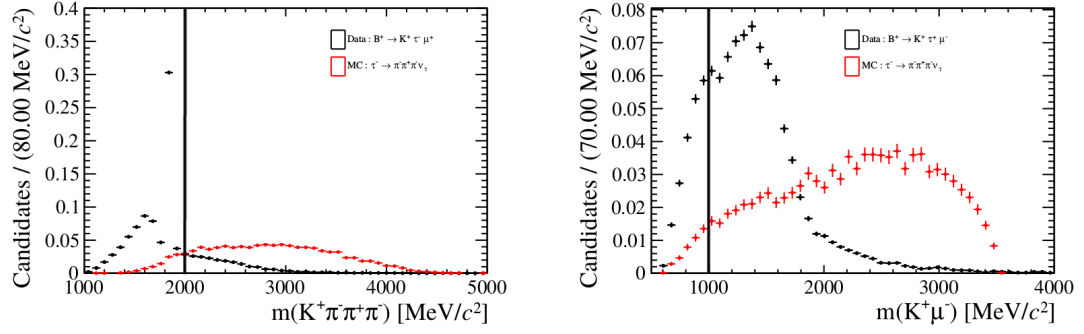
Figure 4.7 – The plot on the left shows the $m(K^+\pi^-\pi^+\pi^-)$ distribution for a sub-sample of 2018 data in black and simulated candidates in red, where candidates are reconstructed as $B^+ \to K^+\tau^-\mu^+$. The decay of the tauon in the simulation sample is given by $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$. Similarly, the plot on the right shows the $m(K^+\mu^-)$ distributions for candidates reconstructed as $B^+ \to K^+\tau^+\mu^-$. The vertical lines indicate the position of the invariant mass requirements imposed on these variables. The data has the stripping and trigger selections applied.

Table 4.3 – List of additional requirements, with respect to the stripping and trigger selections, imposed on the $B^+ \to K^+\tau^\pm\mu^\mp$ and $B^+ \to \overline{D}^0 D_s^+$ decay candidates.

$B^+ \to K^+\tau^\pm\mu^\mp$	$B^+ \to \overline{D}^0 D_s^+$
$4.5 < m_{\text{corr}}(B^+) < 8.0 \,\text{GeV}/c^2$	$5230 < m(B^+) < 5430 \,\text{MeV}/c^2$
Good DTF fit quality	$\chi^2_{\text{vtx}}(B^+) < 16$
$SV_z(\tau) - PV_z(K\mu) > 2\sigma$	$\chi^2_{\text{FD}}(D_s^+) > 4$
$1 < m_{\text{corr}}(\tau^-) < 2.5 \,\text{GeV}/c^2$	$\chi^2_{\text{FD}}(\overline{D}^0) > 4$
$\log_{10}(1 - \text{DIRA}) < -4$	$1930 < m(D_s^+) < 2000 \,\text{MeV}/c^2$
$m(K^+\pi^-\pi^+\pi^-) > 2 \,\text{GeV}/c^2$ ($B^+ \to K^+\tau^-\mu^+$ only)	$1820 < m(\overline{D}^0) < 1910 \,\text{MeV}/c^2$
$m(K^+\mu^-) > 1 \,\text{GeV}/c^2$ ($B^+ \to K^+\tau^+\mu^-$ only)	$\tau(B^+) > 10^{-3} \,\text{ns}$

In order to further suppress specific sources of physics background, various kaon, pion and muon invariant mass combinations are investigated for signal simulation and data. These distributions are studied after full selection, including the requirements imposed on the outputs of multivariate analysis techniques and the PID distributions discussed in Secs. 4.4 and 4.5. The various invariant mass distributions are used to investigate the presence of background that would survive the selection. These resonances are then vetoed. In some cases, an alternative mass hypothesis is assigned to one of the reconstructed tracks in order to investigate the presence of misidentified background decays. The relevant distributions are shown in Figs. 4.8 and 4.9. Based on these studies, an additional set of requirements imposed as the final stage of the selection procedure is determined. These requirements are given in Tab. 4.4. The aim of these requirements is to reject background decays involving the transitions $D^+ \to \pi^+ \pi^- \pi^+$, $D_s^+ \to \pi^+ \pi^- \pi^+$ as well as $D^+ \to K^- \pi^+ \pi^+$ where the kaon has been misidentified as a pion, for both the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ candidates. Additionally, for the $B^+ \to K^+ \tau^- \mu^+$ candidates background decays involving either a $D^0 \to K^- \pi^+$ transition or a $D^0 \to K^- \pi^+ \pi^- \pi^+$ where the kaon has been misidentified as a muon are further rejected. In the case of the $B^+ \to K^+ \tau^+ \mu^-$ candidates, the background decays involving the $D^0 \to K^- \pi^+$ transition where the pion has been misidentified as a muon are also rejected by these requirements. The selection efficiency associated with these requirements is given in Sec. 5.4 in Tab. 4.12.

Table 4.4 – Additional requirements imposed for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ candidates on the left and right, respectively. The squared brackets indicate a change in mass hypothesis, *e.g.* in the second line the kaon mass is assigned to the pion candidate.

$B^+ \to K^+ \tau^- \mu^+$	$B^+ \to K^+ \tau^+ \mu^-$
$m(\pi^- \pi^+ \pi^-) < 1.6\,\mathrm{GeV}/c^2$	$m(\pi^+ \pi^- \pi^+) < 1.6\,\mathrm{GeV}/c^2$
$m(\pi^+ [\to K^+] \pi^- \pi^-) \notin [1840, 1900]\,\mathrm{MeV}/c^2$	$m(\pi^- [\to K^-] \pi^+ \pi^+) \notin [1840, 1900]\,\mathrm{MeV}/c^2$
$m(K^+ \pi_1^-) \notin [1835, 1895]\,\mathrm{MeV}/c^2$	$m(K^+ \mu^- [\to \pi^-]) \notin [1835, 1895]\,\mathrm{MeV}/c^2$
$m(K^+ \pi_3^-) \notin [1835, 1895]\,\mathrm{MeV}/c^2$	
$m(\mu^+ [\to K^+] \pi^- \pi^+ \pi^-) \notin [1835, 1895]\,\mathrm{MeV}/c^2$	
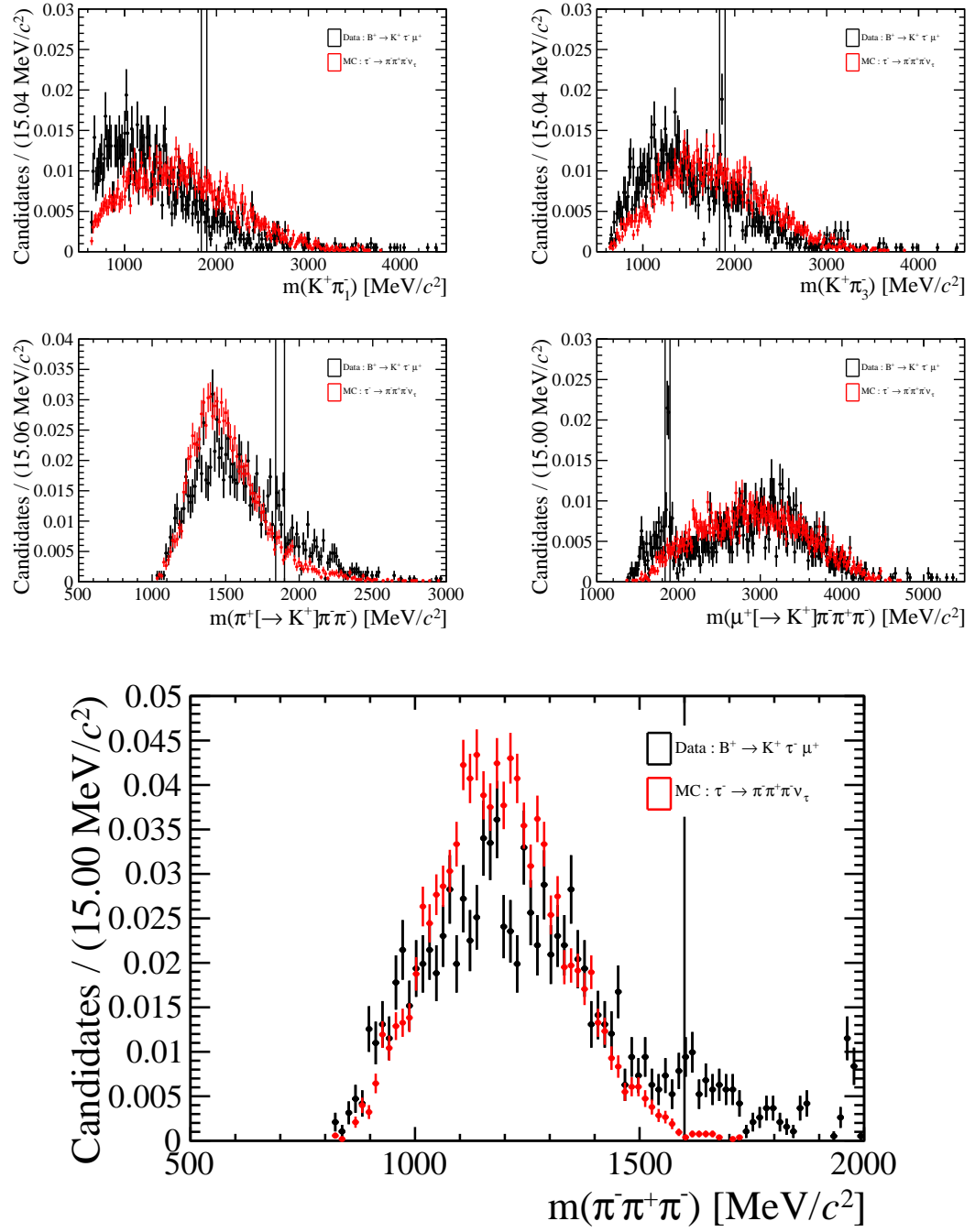
Figure 4.8 – Invariant mass combinations for the combined Run 2 data after full selection for candidates reconstructed as $B^+ \rightarrow K^+\tau^-\mu^+$. The distributions in red are for the $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ signal simulation and in black for data. The vertical lines indicate the values of the vetoes. The square brackets in the axis label indicate a change in mass hypothesis for the track under consideration.

Figure 4.9 – Invariant mass combinations for the combined Run 2 data after full selection for candidates reconstructed as $B^+ \to K^+ \tau^+ \mu^-$. The distributions in red are for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ signal simulation and in black for data. The vertical lines indicate the values of the vetoes. The square brackets in the axis label indicate a change in mass hypothesis for the track under consideration.
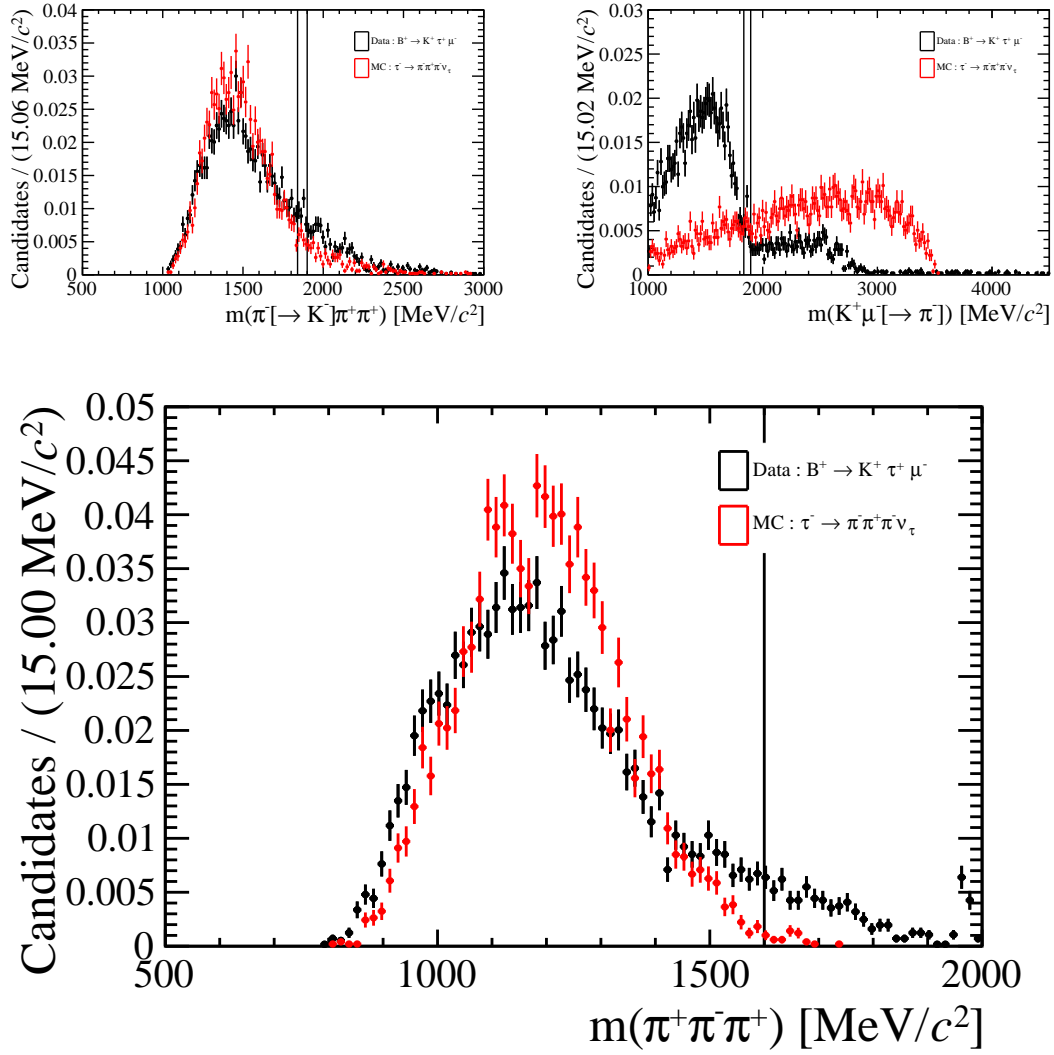
4.4 Multivariate analysis based selection

The next step in the selection procedure is based on the use of multivariate analysis (MVA) techniques. These MVA techniques are only used for the $B^+ \to K^+\tau^\pm\mu^\mp$ decay modes and not the $B^+ \to \overline{D}^0 D_s^+$ decay mode. The need to use this kind of techniques stems from the fact that after applying the preselection, as described in Sec. 4.3, there are no more trivial requirements which would significantly reduce background while keeping a high signal efficiency. Boosted Decision Trees (BDT) using xgboost as boosting algorithm [82] are used as MVA technique. The implementation of the BDT is performed using the scikit-learn [83] framework. Instead of using a single BDT with the purpose of generically reducing background, the strategy is to use two successive BDTs where the first is tailored to suppress combinatorial background while the second is designed to reduce physics background. Since the features of the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes differ significantly, especially in terms of the content of the physics background, the sets of BDTs are trained separately for both decay modes. Furthermore, different sets of BDTs are trained for Run 1 and Run 2 datasets since the data acquisition conditions are different.

When training a BDT, three important factors determine the resulting background suppression performance. These are the choice of signal and background training samples, the choice of input variables and the tuning of the hyper-parameters. The training samples need to approximate the properties of the signal and background as closely as possible in order to optimise their separation. The size of these samples is also important since the description of the variables used as input will be more accurate with higher statistics. Furthermore, if the statistics of the signal and background training samples are vastly different, an issue can arise where the training of the BDT is biased towards the sample with the highest statistics. Concerning the input variables, it is important that they show a sizeable discrimination between the signal and background distributions. However, adding variables which are extremely discriminating by themselves is ill-advised as the BDT will impose requirements almost exclusively on this particular variable and will not use efficiently the information encoded in the other distributions. In that case, it is better to apply a preselection requirement on that very discriminating variable. This is the reason for the application of the additional requirements described in the previous section. Finally, the choice of the value of the hyper-parameters will dictate the internal behaviour of the BDT which naturally will directly impact on its performance.

The samples provided for training are different in the case of the BDT aimed at reducing combinatorial background, referred as the "combinatorial BDT", and the BDT aimed at reducing physics background, referred as the "physics BDT". In the former case, the simulation sample with $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ is used to describe the signal while the high-mass sideband in data $m_{corr}(B^+) > 5.8\,\text{GeV}/c^2$ is used to describe the background. The reason for using the high mass

sideband is that no B hadron decay is expected to result in a reconstructed $m_{\text{corr}}(B^+)$ greater than $5.8\,\text{GeV}/c^2$. As such, only random combination of tracks, *i.e.* combinatorial background, is expected to populate the high mass sideband. In order to increase the available statistics, all years within a Run period are combined to make the training samples. For the physics BDT, signal simulation with $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ is used as the signal training sample, while the low mass sideband in data $4.5 < \text{m}_{\text{corr}}(B^+) < 4.8\,\text{GeV}/c^2$ is used as background training sample. This choice is motivated by the fact that the physics background is mostly located in the lower mass sideband. However, this region of data also contains combinatorial background. Since the aim of this BDT is to suppress physics background, it is undesirable to have these candidates whose features are significantly different from those of the physics background. In order to partially remedy this issue, a loose requirement on the output of the combinatorial BDT is imposed to the datasets used to build both the signal and background training samples. By doing that, the proportion of physics to combinatorial background in the low mass sideband region in data is increased. The requirement needs however to remain relatively loose in order not to reduce drastically the available statistics.

The input variables used for the training of the combinatorial BDT are the following : the DTF fit χ^2, the minimum and maximum of $m(\pi^+\pi^-)$, the kaon χ^2_{IP} and the cone p_T asymmetry of the kaon and the tauon, where the cone has an opening angle of $1\,\text{rad}$ and is defined as $\frac{p_T^s - p_T^{cone}}{p_T^s + p_T^{cone}}$ with p_T^s the transverse momentum of the signal candidate and p_T^{cone} the transverse component of the total vectorial momentum of all other tracks in the cone. The choice of these variables is based on physical arguments, since these variables are expected to be discriminant against combinatorial background. To ensure a good understanding of the BDT performance, it is also important that the variables used in the BDT are well modeled in simulation since simulated samples are provided as signal training samples. Additional variables were considered in the training but did not provide higher performance and as such were rejected. Since most variables used as input to the combinatorial BDT are also discriminant against physics background, they are also provided as input to the physics BDT, with the exception of the minimum of $m(\pi^+\pi^-)$, which provides no additional discriminating power. Aside from these variables, there are other variables which are mainly useful at discriminating against the physics background. The variables added are χ^2_{vtx} variables for various combinations of final state particles : $K^+\pi^-\pi^+\pi^-$, $K^+\pi^-\pi^+$, $\mu^+\pi^-\pi^+$, $\mu^+\pi^-\pi^+\pi^-$. The reason these combinations are useful is because of the presence of different types of charm background decays, mainly, which produce good quality vertices unlike the signal. Final state particles of these decays can also be misidentified, explaining why combinations involving the muon track are also useful.

The choice of the BDT hyper-parameters dictates the internal behaviour of the algorithm, which impacts its performance. The value of these parameters cannot be a priori determined, *e.g.* from physical arguments. As a result, their values are determined by performing a scan within a reasonable range and assessing the corresponding BDT performance. The area under
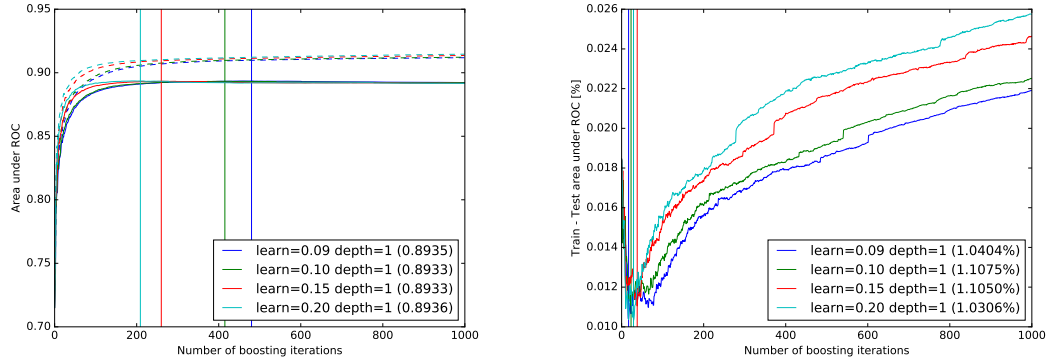
Figure 4.10 – The plot on the left shows the area under the ROC curve as a function of the number of decision trees in the BDT. The different colours correspond to various values of the learning rate parameter. The continuous curves are for the test set while the dashed curves are for the training set. The plot on the right shows the relative difference between the area under the ROC curve computed for the training and test sets. The smaller the difference the smaller the over-training of the BDT. The optimal value for the learning rate parameter minimises the relative difference between the training and test sets.

the receiver operating characteristic (ROC) curve is used as performance metric. The integral of the ROC curves is computed both for the training and an independent testing sample. If these values disagree significantly it means that the BDT is affected by over-training. The choice of the values of the hyper-parameters is then a balancing act between the maximisation of the performance and the minimisation of the over-training. The area under the ROC curve for both training and test sets for a given set of hyper-parameters values is shown in Fig. 4.10. The performance of the BDT keeps increasing with the number of decision trees in the case of the training set while it reaches a maximal value and then stabilises, or even decreases slowly, in the case of the test set. This reflects the situation in which the performance of a BDT applied on the same sample on which it was trained will naturally increase almost indefinitely with the complexity the BDT. However, the same is not true when applying the BDT to a statistically independent sample, such as the test sample. The reason for this behaviour is that by increasing excessively the complexity of the BDT, the latter will focus on the details of the training samples, *i.e.* mainly optimise for the observed statistical fluctuations, rather than on the general features of the training sample. As a result, it was decided to use 200 decision trees in the BDT.

The performances of the BDT are very similar for Run 1 and Run 2 data, therefore only the plots for Run 2 data are shown below. The area under the ROC curve is approximately 90% as can bee seen in Fig. 4.11. The multiple curves are due to the fact that a cross-validation with 5 folds is performed. Each of the curves in a plot corresponds to a given fold. All the folds

show equivalent performances, which indicates an absence of over-training. This absence of over-training can also be observed in Fig. 4.12 where the BDT output for the training and testing samples are seen to be compatible. Additionally, the feature importance for each input variable used in the BDT is shown in Fig. 4.13. The correlation matrices for the input variables are shown in Fig. 4.14.

A potential issue which can arise when using a BDT as part of the selection is that after imposing requirements on its output the resulting $m_{\text{corr}}(B^+)$ distribution for the background is reshaped into a peaking structure which could be misinterpreted as being due to signal. The $m_{\text{corr}}(B^+)$ distribution after various BDT requirements is shown for the data sidebands in Figs. 4.15–4.18. Since the mass distribution shapes are similar for all the BDT requirements it proves that the B^+ corrected mass distribution is not strongly distorted by imposing requirements on the BDT outputs. As mentioned previously, it is also primordial that the distribution of the variables used as input to the BDT are well reproduced in simulation. This is verified by comparing equivalent distributions in simulation and data using the $B^+ \to \overline{D}^0 D_s^+$ decay mode. In order for the comparison to be meaningful, the data needs to be as pure as possible, *i.e.* contain almost exclusively signal decays. For this reason, the comparisons are performed after full selection and are shown in Figs. 4.19 and 4.20. The distributions show good agreement between data and simulation.

Finally, a criteria needs to chosen in order to determine which requirements to apply on the output of the BDT. Since this analysis is a search for a rare process, there is a non-negligible possibility for the final signal yield to be non-significantly different from zero. If that were to be the case, the next step would be to compute an upper limit on the signal branching fractions. Thus, the requirements are determined using an optimisation procedure which consists in minimizing the expected upper limit on the signal branching fractions obtained from pseudo-experiments, as explained in Sec. 5.3. This procedure is used to determine simultaneously the optimal requirements on the output of the BDT as well as the particle identification requirements discussed in Sec. 4.5. From this procedure the requirements imposed on the BDT outputs are $\text{BDT}_{\text{comb}} > 0.8$ and $\text{BDT}_{\text{phys}} > 0.8$ for the $B^+ \to K^+ \tau^- \mu^+$ decay mode and $\text{BDT}_{\text{comb}} > 0.75$ and $\text{BDT}_{\text{phys}} > 0.7$ for the $B^+ \to K^+ \tau^+ \mu^-$ decay mode, respectively.
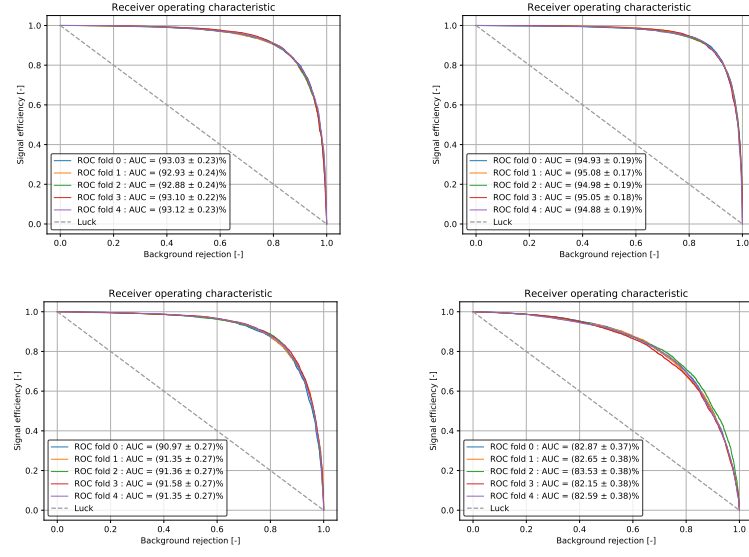
Figure 4.11 – ROC curves for the Run 2 sample for the $B^+ \rightarrow K^+\tau^-\mu^+$ (left) and $B^+ \rightarrow K^+\tau^+\mu^-$ (right) decay modes. The top and bottom rows show the performance associated to the combinatorial and physics BDT, respectively.
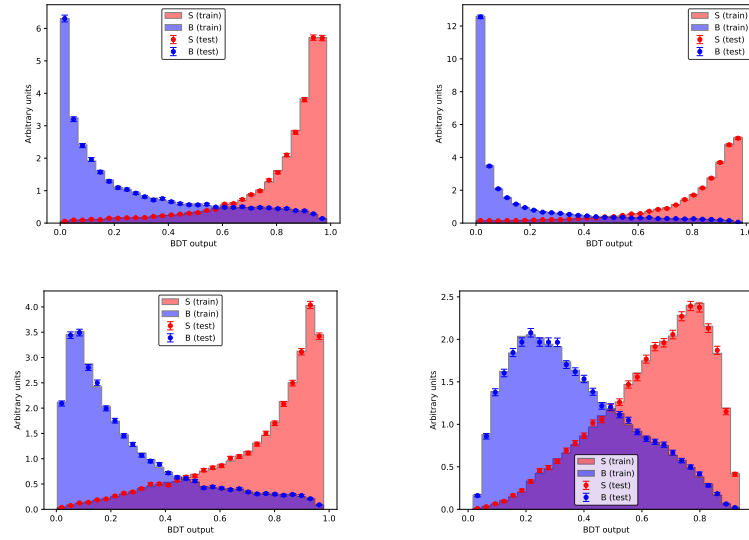
Figure 4.12 – BDT output for the signal in red and the background in blue for the Run 2 sample for the $B^+ \rightarrow K^+\tau^-\mu^+$ (left) and $B^+ \rightarrow K^+\tau^+\mu^-$ (right) decay modes. The top and bottom rows show the output for the combinatorial and physics BDT, respectively. The filled histograms correspond to the training sets while the data points correspond to the testing sets.
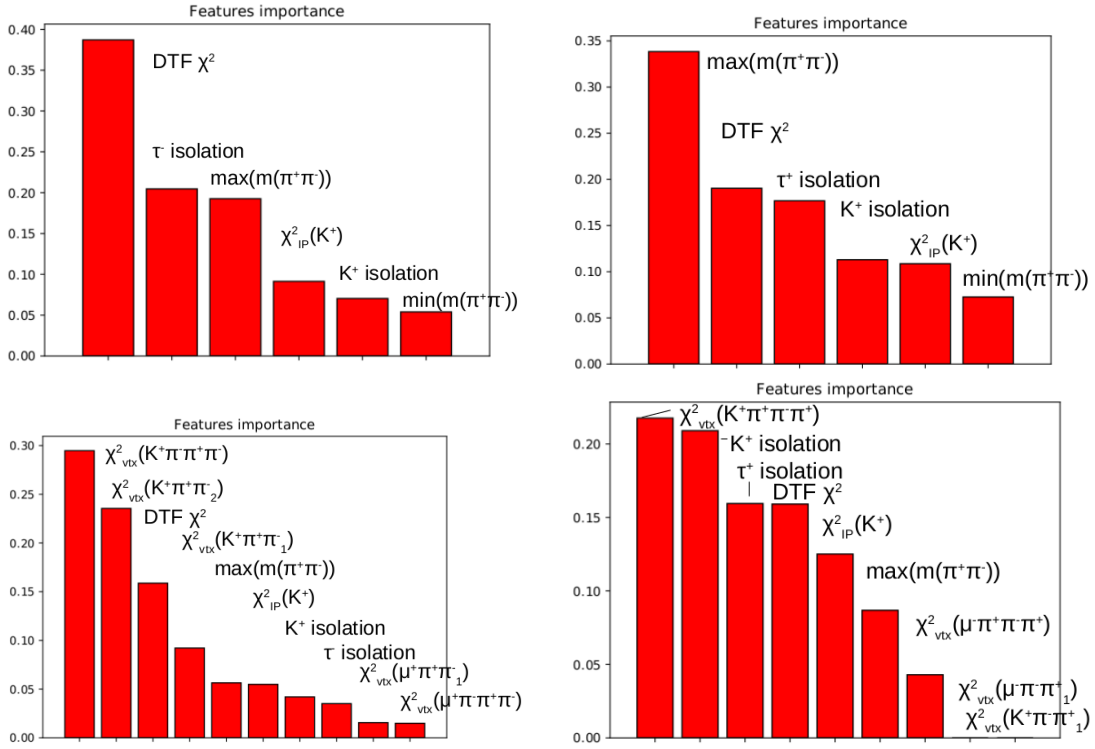
Figure 4.13 – BDT input variable importance for the Run 2 sample for the $B^+ \to K^+ \tau^- \mu^+$ (left) and $B^+ \to K^+ \tau^+ \mu^-$ (right) decay modes. The top and bottom rows show the variable importance for the combinatorial and physics BDT, respectively.
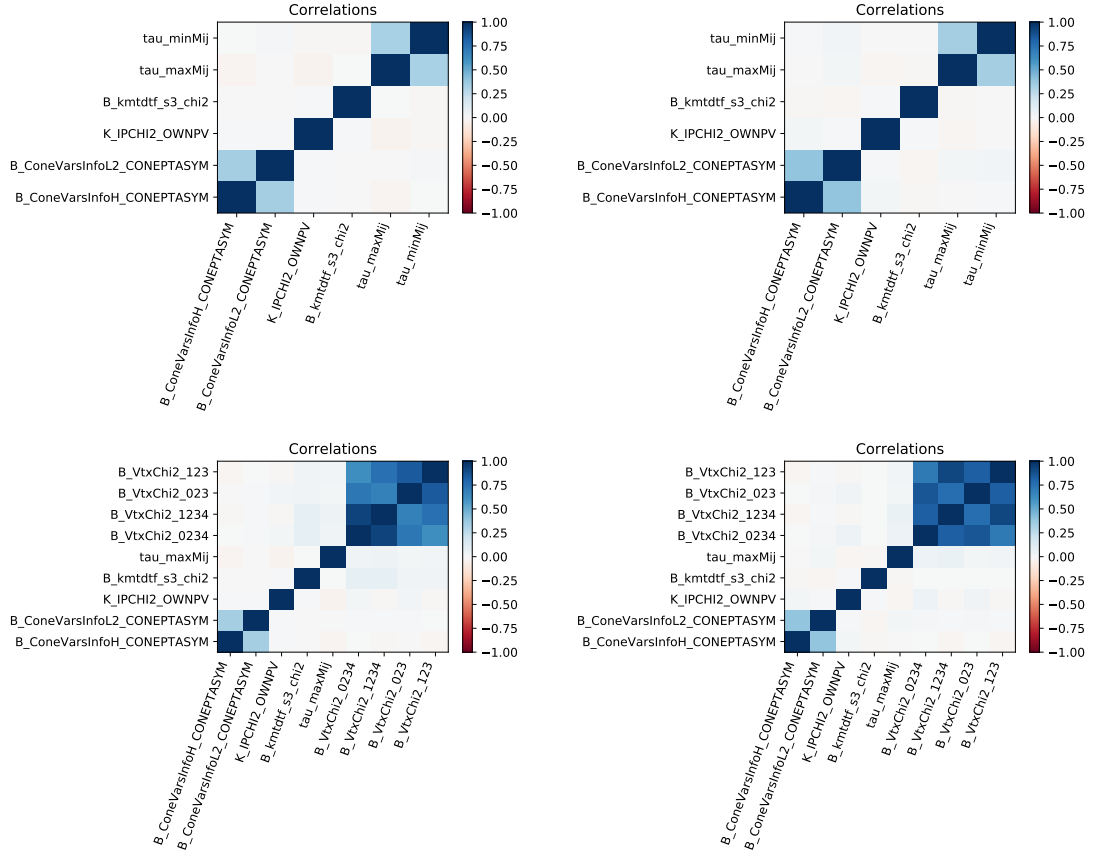
Figure 4.14 – Correlation matrices of the BDT input variables for the Run 2 signal training sample for the $B^+ \to K^+\tau^-\mu^+$ (left) and $B^+ \to K^+\tau^+\mu^-$ (right) decay modes. The variables of the combinatorial BDT and physics BDT are shown in the top and bottom rows, respectively.
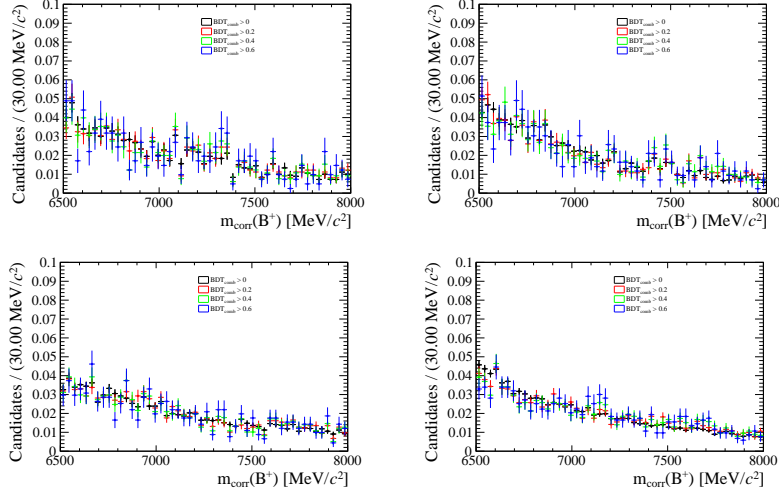
Figure 4.15 – Normalised $m_{corr}(B^+)$ distribution for various BDT_{comb} requirements for the Run 1 (top) and Run 2 (bottom) samples for the $B^+ \to K^+ \tau^- \mu^+$ (left) and $B^+ \to K^+ \tau^+ \mu^-$ (right) decay modes. The distributions are showed for the high mass sideband in data with $m_{corr}(B^+) > 6.5$ where only combinatorial background is expected. It can be observed that the mass distribution keeps the same shape for all requirements on the output of the BDT.
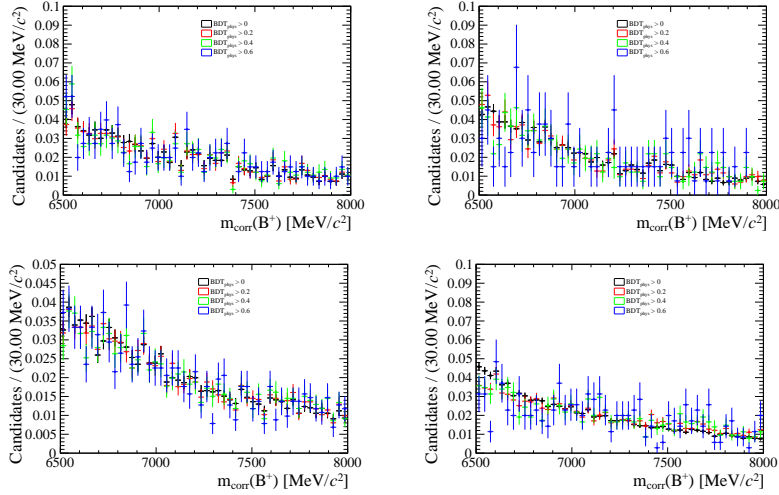
Figure 4.16 – Normalised $m_{corr}(B^+)$ distribution for various BDT_{phys} requirements for the Run 1 (top) and Run 2 (bottom) samples for the $B^+ \to K^+ \tau^- \mu^+$ (left) and $B^+ \to K^+ \tau^+ \mu^-$ (right) decay modes. The distributions are showed for the high mass sideband in data with $m_{corr}(B^+) > 6.5$ where only combinatorial background is expected. It can be observed that the mass distribution keeps the same shape for all requirements on the output of the BDT.
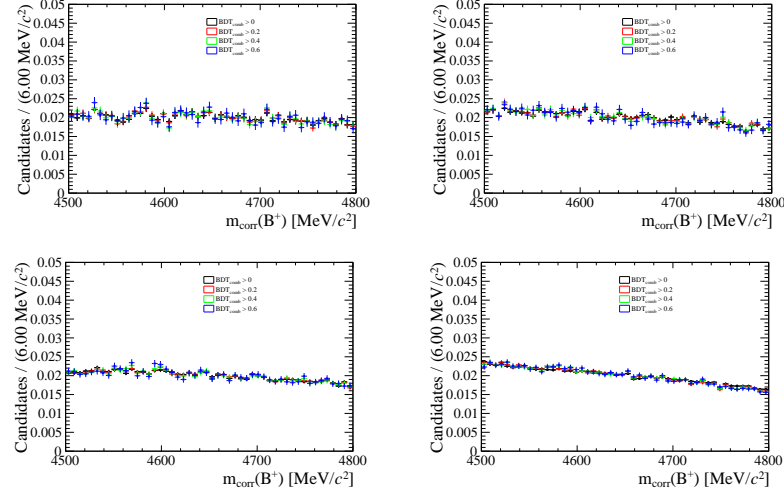
Figure 4.17 – Normalised $m_{corr}(B^+)$ distribution for various BDT_{comb} requirements for the Run 1 (top) and Run 2 (bottom) samples for the $B^+ \rightarrow K^+ \tau^- \mu^+$ (left) and $B^+ \rightarrow K^+ \tau^+ \mu^-$ (right) decay modes. The distributions are showed for the low mass sideband in data with $4.5 < m_{corr}(B^+) < 4.8$. It can be observed that the mass distribution keeps the same shape for all requirements on the output of the BDT.

Figure 4.18 – Normalised $m_{corr}(B^+)$ distribution for various BDT_{phys} requirements for the Run 1 (top) and Run 2 (bottom) samples for the $B^+ \rightarrow K^+ \tau^- \mu^+$ (left) and $B^+ \rightarrow K^+ \tau^+ \mu^-$ (right) decay modes. The distributions are showed for the high mass sideband in data with $4.5 < m_{corr}(B^+) < 4.8$. It can be observed that the mass distribution keeps the same shape for all requirements on the output of the BDT.
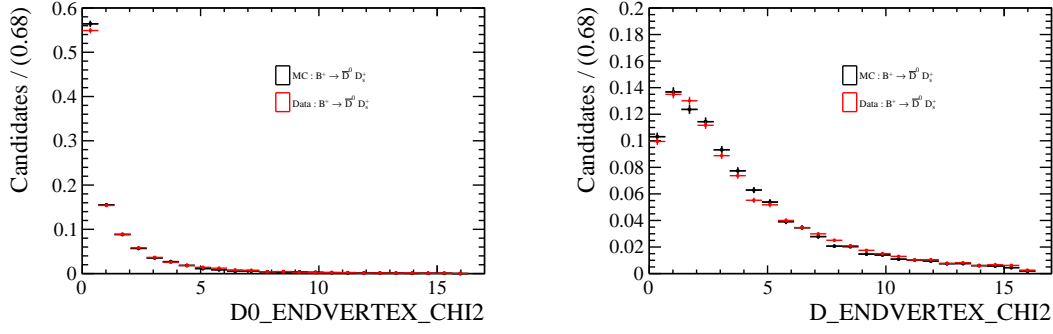
Figure 4.19 – Decay vertex χ^2 distributions for Run 2 data in black and simulation in red for the $B^+ \to \overline{D}^0 D_s^+$ decay mode after full selection. The left and right plots show the distributions for the $\overline{D}^0$ and D_s^+ candidates, respectively.
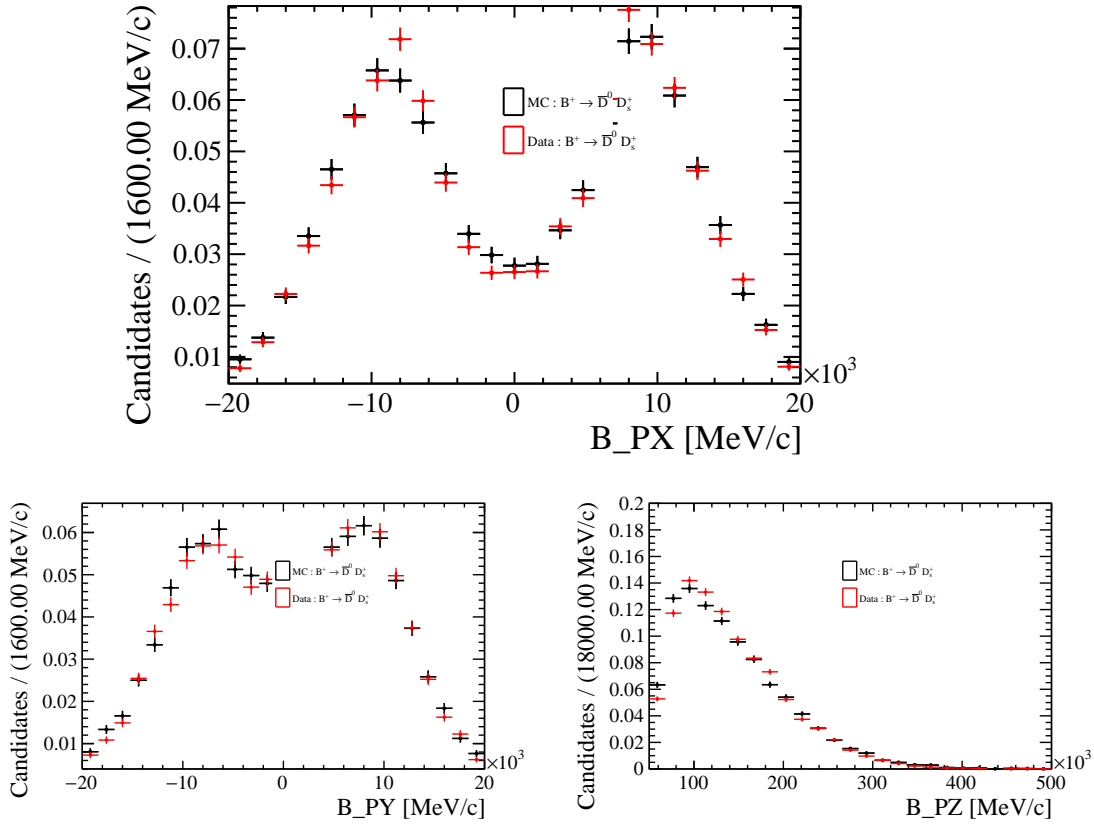
Figure 4.20 – B^+ candidate's momentum component distributions for Run 2 data in black and simulation in red for the $B^+ \to \overline{D}^0 D_s^+$ decay mode after full selection. The plot on the top shows the p_x distribution while the bottom left and right plots show the p_y and p_z components, respectively.

4.5   Particle identification

As previously discussed, the use of particle identification variables is essential in suppressing background decays where a subset of the final state particles are misidentified. However, the efficiency associated with requirements imposed on these variables cannot be reliably measured using standard simulation because the PID variable distributions in simulation do not reproduce well the distributions in data. Nonetheless, it is possible to correct the PID distributions in simulation such that they reproduce accurately the features observed in data using a data-driven technique. This can be done using calibration data where an abundant and well-known decay mode, *e.g. $J/\psi \to \mu^+ \mu^-$*, is reconstructed and the PID distributions associated with its daughters are studied. For example, in the case of the $J/\psi \to \mu^+ \mu^-$, this would correspond to the PID distributions applied to a muon candidate. The PID distributions depend on the kinematics and occupancy of the event, *i.e.* it depends on the values of p and p_T of the track under consideration and the number n_{tracks} of reconstructed tracks in the event. In order to keep only the signal contribution in the calibration data, the *sPlot* technique [84] is used. Then for each candidate in the $B^+ \to K^+ \tau^\pm \mu^\mp$ simulation, the value of a given PID variable is sampled from the calibration data samples according to the corresponding location in the ($p, p_T, n_{\text{tracks}}$) phase-space in order to obtain a new corrected distribution. The correlations between various tracks for a given type of PID variable are conserved in the correction procedure since the correction is based on the kinematics of the tracks, which preserve their correlations. In practice, the correction is implemented using the PIDGen package [85].

In order to perform this correction, it is necessary to provide a description of the signal distributions of p, p_T and $n_{\text{tracks}}$. Simulation can be used to describe both kinematic distributions but it is however unable to describe accurately the n_{tracks} distribution. Thus, it is necessary to correct the n_{tracks} distribution in simulation such that corresponds to that in the real data. To do so, the n_{tracks} distribution is compared between data and simulation for the $B^+ \to \overline{D}^0 D_s^+$ decay mode after full selection. The correction of the distribution for the simulated sample is performed by multiplying the value of n_{tracks} of each candidate by a constant scale factor w. The value of that scale factor is determined by minimising the discrepancy between the n_{tracks} distribution for data and corrected simulation. The discrepancy between both distributions is measured with the following metric :

$$\chi^2 = \sum_{i=1}^{N_{\text{bins}}} \frac{(y_i - y_{\text{sim},i})^2}{\sigma_i^2 + \sigma_{\text{sim},i}^2}, \tag{4.2}$$

where $N_{\text{bins}} = 60$ is the total number of bins in the histogram of n_{tracks}, y_i, σ_i and $y_{\text{sim},i}$, $\sigma_{\text{sim},i}$ are the yields and corresponding uncertainties in bin i for data and corrected simulation, respectively. The comparison between the simulation and data n_{tracks} distribution for the $B^+ \to \overline{D}^0 D_s^+$ decay mode before and after correction is shown in Fig. 4.21, while the minimisation of

the χ^2 of Eq.(4.2) as a function of the scale factor w is shown in Fig. 4.22. The n_{tracks} distribution in simulation is therefore corrected by using a scale factor of $w = 1.11$. The same scale factor is used to correct the n_{tracks} distribution for simulated samples of the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes since their uncorrected n_{tracks} distributions are similar to that of the $B^+ \to \overline{D}^0 D_s^+$ decay mode as shown in Fig. 4.23. While the particular value $w = 1.11$ was determined by using 2016 data samples, it has been checked that it is also suitable to correct datasets from all other years.
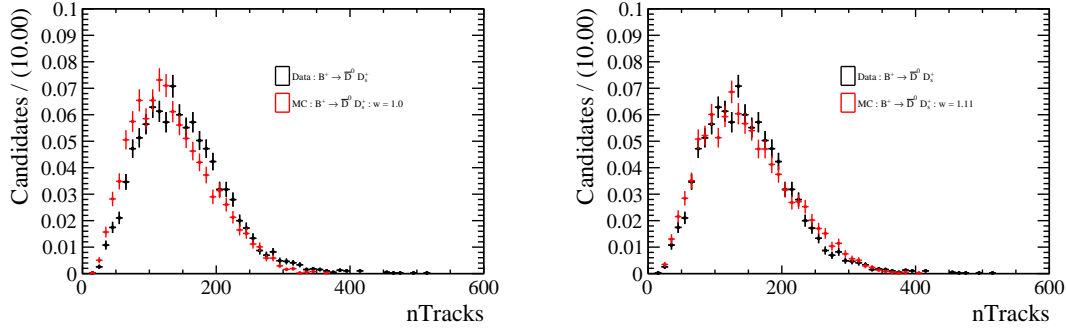
Figure 4.21 – Comparison between the n_{tracks} distribution in data (black) and simulation (red) for the $B^+ \to \overline{D}^0 D_s^+$ decay mode using 2016 data. The plot on the left shows an uncorrected distribution for the simulation while the plot on the right applies a scale factor of $w = 1.11$.
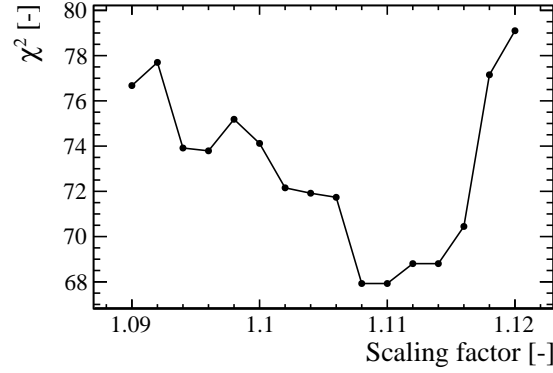
Figure 4.22 – Variation of the χ^2 variable of Eq. (4.2) as a function of the scale factor w for 2016 data.
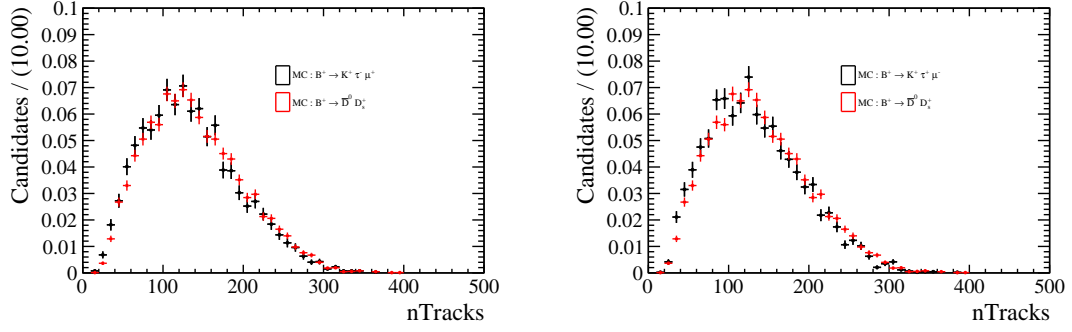
Figure 4.23 – Comparison between the n_tracks distributions of simulated $B^+ \to \overline{D}^0 D_s^+$ and $B^+ \to K^+\tau^-\mu^+$ (left) or $B^+ \to K^+\tau^+\mu^-$ (right) decays using 2016 data. The distributions for the signal and normalisation channels are in black and red, respectively.

The simulated samples have been processed such as to apply the PID requirements present in the stripping lines, described in Tabs. 4.1 and 4.2, on the corrected PID distributions. On top of the PID requirements imposed by the stripping line, the requirement $\texttt{ProbNN}_K(K) > 0.55$ is imposed on each kaon track in the $B^+ \to \overline{D}^0(\to K^+\pi^-)D_s^+(\to K^+K^-\pi^+)$ decay mode. The PID distributions for corrected simulation and data, after full selection, have been compared for the $B^+ \to \overline{D}^0 D_s^+$ decay mode and good agreement is observed. For the $B^+ \to K^+\tau^\pm\mu^\mp$ decay modes, the PID selection is further optimised with respect to the requirements imposed in the stripping selection through the use of a combined variable defined as

$$\mathscr{P}_3 = \texttt{ProbNN}_K(K) \times \prod_{i=1}^{3} \texttt{ProbNN}_\pi(\pi_i). \tag{4.3}$$

The requirement imposed on $\mathscr{P}_3$ is determined in conjunction with the requirements imposed on the BDT outputs, as described in Sec. 5.3. The resulting PID requirements are summarised in Tab. 4.5. The selection efficiencies associated with the PID requirements are given in Sec. 4.7.

Table 4.5 – PID requirements imposed on the $B^+ \to \overline{D}^0 D_s^+$, $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ candidates on the left, middle and right columns, respectively.

$B^+ \to \overline{D}^0 D_s^+$	$B^+ \to K^+\tau^-\mu^+$	$B^+ \to K^+\tau^+\mu^-$
$\texttt{ProbNN}_\pi(\pi) > 0.55 \quad \forall\pi$	$\texttt{ProbNN}_\pi(\pi) > 0.55 \quad \forall\pi$	$\texttt{ProbNN}_\pi(\pi) > 0.55 \quad \forall\pi$
$\texttt{ProbNN}_K(K) > 0.55 \quad \forall K$	$DLL_\mu(\mu) > 2$	$DLL_\mu(\mu) > 2$
$DLL_K(K) > -5 \quad \forall K$	$DLL_K(K) > 5$	$DLL_K(K) > 5$
	$\mathscr{P}_3 > 0.5$	$\mathscr{P}_3 > 0.6$

4.6 Expected background contributions

4.6.1 Expected yields for dominant non-resonant background decay modes

In order to have a better understanding of the data composition, various potential background decay modes are simulated and their expected yield in data after full selection is computed. The dominant sources of background are partially reconstructed decays, possibly with misidentifed particles in the final state. Therefore, their B^+ corrected mass distribution is not expected to peak in the signal window. A list of decays expected to be the dominant sources of background in the $B^+ \to K^+ \tau^+ \mu^-$ data sample is determined and for each of these decay modes 1M events after generator level requirements have been simulated. Each sample contains a mixture of decay modes leading to the same reconstructed final state, including potential pion or kaon misidentification. The relative proportion of the decay modes follows the expected ratios from their branching fractions taken from Ref. [33]. For each sample, the expected background yield after full selection, including the BDT+PID requirements described in Sec. 5.3, for the 2016 data sample in the mass range $4.5 < m_{\mathrm{corr}}(B^+) < 8.0 \,\mathrm{GeV}/c^2$ is computed. The background yields are computed using the fitted yield of the $B^+ \to \overline{D}^0 D_s^+$ decay mode in the 2016 data sample. For a given B^+ or B^0 background decay mode with a combined branching fraction $\mathscr{B}_{bkg}$ and selection efficiency ϵ_{bkg}, its expected yield in the 2016 dataset is given by

$$\frac{\mathscr{B}_{bkg}\epsilon_{bkg}N_{2016}^{norm}}{\epsilon_{norm}\mathscr{B}(B^+ \to \overline{D}^0 D_s^+)\mathscr{B}(\overline{D}^0 \to K^+\pi^-)\mathscr{B}(D_s^+ \to K^+K^-\pi^+)}, \qquad (4.4)$$

where N_{2016}^{norm} and ϵ_{norm} are respectively the $B^+ \to \overline{D}^0 D_s^+$ yield and selection efficiency for the 2016 dataset. If the head of the background decay is neither a B^+ nor a B^0, this number needs to be multiplied by the appropriate ratio of fragmentation fractions. The fitted $B^+ \to \overline{D}^0 D_s^+$ yield as well as the associated full selection efficiency are given in Tab. 4.6. The expected yields for the various background decays are given in Tab. 4.7 while their B^+ corrected mass distributions after stripping selection are shown in Fig. 4.24. As expected, no peaking structure is observed in these mass distributions. The low mass sideband in data is populated by a multitude of partially reconstructed background decay modes with kinematics and topologies similar to those given by the decays in Tab. 4.7. As a result, the overall $m_{\mathrm{corr}}(B^+)$ distribution in the data is composed of a sum a smoothly decreasing distributions. This observation motivates the use of an effective distribution in order to describe the overall contribution of the physics background content in the final mass fit. The mathematical description of that probability density function is given in Sec. 5.1. While the simulated background decays are expected to mainly contribute to the $B^+ \to K^+ \tau^+ \mu^-$ decay mode, a similar conclusion can be made about the background content of the $B^+ \to K^+ \tau^- \mu^+$ sample. Indeed, with the exception of the dominant source of background involving the transition $\overline{D}^0 \to K^+\pi^-\pi^+\pi^-$, which is removed through the application of the requirement $m(K^+\pi^-\pi^+\pi^-) > 2\,\mathrm{GeV}/c^2$, the sources of physics

background expected in that sample share similar properties to those of the decays studied for the $B^+ \to K^+\tau^+\mu^-$ decay mode. In particular, their corrected mass distributions are also smoothly decaying and entail the use of an effective shape as well.

Table 4.6 – The first line gives the full selection efficiency, including systematic uncertainties which are described in Sec. 5.4, for the 2016 $B^+ \to \overline{D}^0 D_s^+$ decay mode. The second line gives the fitted yields in data, where the uncertainty includes the systematic uncertainties. The following lines give the branching fractions associated to the various sub-decays. The last line gives the combination of the branching fractions, efficiency and yield used to estimate the expected yields in data of various background decays.

Variable	Value
ϵ_{norm}	$\left(4.98 \pm 0.05(stat) \pm 0.21(syst)\right) \times 10^{-4}$
N_{2016}	3780.16 ± 64.14
$\mathscr{B}(B^+ \to \overline{D}^0 D_s^+)$	$(9.0 \pm 0.9) \times 10^{-3}$
$\mathscr{B}(\overline{D}^0 \to K^+\pi^-)$	$(3.950 \pm 0.031) \times 10^{-2}$
$\mathscr{B}(D_s^+ \to K^+K^-\pi^+)$	$(5.39 \pm 0.15) \times 10^{-2}$
$\dfrac{N_{2016}}{\epsilon_{norm}\mathscr{B}(B^+\to\overline{D}^0 D_s^+)\mathscr{B}(\overline{D}^0\to K^+\pi^-)\mathscr{B}(D_s^+\to K^+K^-\pi^+)}$	$(3.96 \pm 0.45) \times 10^{11}$

Table 4.7 – The first column specifies the background decay mode, where h denotes either a pion or a kaon, while the second gives the associated selection efficiency when reconstructing the decays as $B^+ \to K^+\tau^+\mu^-$. The efficiencies are computed using uncorrected simulation. The third and fourth columns give respectively the combined branching fraction and the expected yield, in the mass region $4.5 < m_{\text{corr}}(B^+) < 8\,\text{GeV}/c^2$, in the 2016 data sample. These are approximate estimations since the simulated background decays contain a mixture of various decay modes. For the $B_s^0 \to K^+ D^0(\to 4h)D_s^-(\to \mu^-\overline{\nu}_\mu)$ and $\overline{B}_s^0 \to K^+ D^0(\to 4h)\pi^-$ decay modes, the value $\frac{f_s}{(f_u+f_d)} = 0.122 \pm 0.006$, taken from Ref.[86], is used to compute the expected yield while for the $\Lambda_b^0 \to \Lambda_c^+(\to 2hX)\pi^+\pi^-\mu^-\overline{\nu}_\mu$ decay the value $\frac{f_{\Lambda_b^0}}{(f_u+f_d)} = 0.259 \pm 0.018$ is used. In the event where there are no candidates left after the full selection, the $+1\sigma$ value is taken as the BDT efficiency in order to estimate an upper bound on the full selection efficiency.

Decay	Total efficiency	Total $\mathscr{B}$	Expected yield
$\overline{B}^0 \to \pi^+ D^0(\to 4h)\mu^-\overline{\nu}_\mu$	$(1.62 \pm 1.62) \times 10^{-7}$	$(3.37 \pm 0.42) \times 10^{-4}$	$(2.16 \pm 2.19) \times 10^1$
$B_s^0 \to K^+ D^0(\to 4h)D_s^-(\to \mu^-\overline{\nu}_\mu)$	$< 1.65 \times 10^{-7}$	$(3.62 \pm 1.07) \times 10^{-4}$	< 5.77
$\overline{B}_s^0 \to K^+ D^0(\to 4h)\pi^-$	$(3.68 \pm 2.60) \times 10^{-7}$	$(8.56 \pm 1.08) \times 10^{-5}$	3.04 ± 2.22
$B^0 \to K^+ \overline{D}^0(\to 4h)\pi^-$	$< 1.51 \times 10^{-7}$	$(7.24 \pm 1.40) \times 10^{-6}$	< 0.43
$\Lambda_b^0 \to \Lambda_c^+(\to 2hX)\pi^+\pi^-\mu^-\overline{\nu}_\mu$	$< 1.04 \times 10^{-7}$	$(3.52 \pm 1.96) \times 10^{-3}$	< 75.09
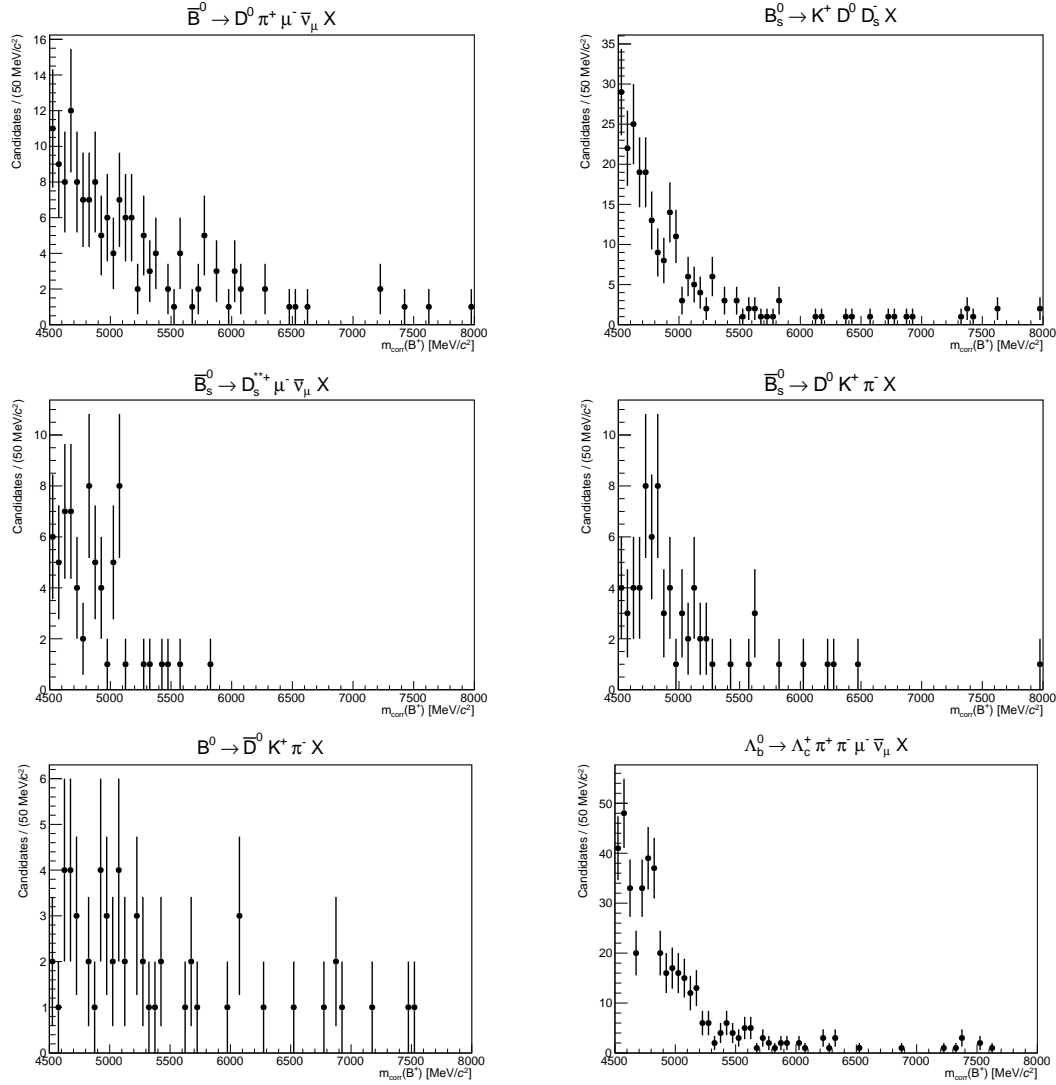
Figure 4.24 – Distribution of $m_{\mathrm{corr}}(B^+)$ for the inclusive simulated backgrounds described in Tab. 4.7. The decay modes from the top left to the bottom right are the following: $\overline{B}^0 \to D^0\pi^+\mu^-\overline{\nu}_\mu$, $B_s^0 \to K^+D^0D_s^-$, $\overline{B}_s^0 \to D_s^{**+}\mu^-\overline{\nu}_\mu$, $\overline{B}_s^0 \to D^0K^+\pi^-$, $B^0 \to \overline{D}^0K^+\pi^-$ and $\Lambda_b^0 \to \Lambda_c^+\pi^+\pi^-\mu^-\overline{\nu}_\mu$. All these distributions are shown after reconstruction plus stripping selection except for the $\overline{B}_s^0 \to D_s^{**+}\mu^-\overline{\nu}_\mu$ decay mode which is shown after the full preselection and for which only 100k events after the generator level requirements have been simulated.

4.6.2 Expected yields for resonant backgrounds

While the dominant sources of background are partially reconstructed and as such produce non-resonant-like mass shapes, there are a few decay modes which present topologies and particle content similar to that of the signal and are fully reconstructed. The production rate of these modes is significantly suppressed relative to the partially reconstructed decays discussed previously. However, since the decay modes are fully reconstructed their associated $m_{\mathrm{corr}}(B^+)$ distribution can produce a peaking structure, possibly at the signal location.

For the $B^+ \to K^+\tau^-\mu^+$ data sample, the $B^+ \to D^-\big(\to \pi^+\pi^-\pi^-\big)K^+\pi^+$ as well as the $B^+ \to D^-\big(\to K^+\pi^-\pi^-\big)K^+\pi^+$ decays where the pion from the B^+ decay is misidentified as a muon and the daughters of the D^- are reconstructed as the τ^- candidate are the most dangerous backgrounds a priori. Similarly, the $B^+ \to D^+(\to \pi^-\pi^+\pi^+)K^+\pi^-$ and $B^+ \to D^+(\to K^-\pi^+\pi^+)K^+\pi^-$ decays are expected to contribute to the $B^+ \to K^+\tau^+\mu^-$ data sample. Besides these decay modes, the $B^+ \to D_s^-\big(\to \tau^-\big(\to \pi^-\pi^+\pi^-(\pi^0)\nu_\tau\big)\overline{\nu}_\tau\big)K^+\pi^+$ decays could be dangerous as well since they involve a real tauon candidate and thus could create a peaking structure in the $m_{\mathrm{corr}}(B^+)$ distribution. These last two decay modes are expected to mainly contribute to the $B^+ \to K^+\tau^-\mu^+$ data sample but could in principle contribute to the $B^+ \to K^+\tau^+\mu^-$ decay mode as well, albeit with suppression due to topological differences with respect to the signal decay. In order to study the potential impact of these various decay modes, simulated samples of 1M events after generator requirements have been produced for each of them. Their $m_{\mathrm{corr}}(B^+)$ distribution after stripping requirements are shown in Fig. 4.25. Additionally, their expected yield in the 2016 data sample is estimated using the $B^+ \to \overline{D}^0 D_s^+$ yields. The resulting numbers are conservative estimates as the signal efficiency is assumed for the requirements on the BDT outputs. These yields are given in Tab. 4.8. Since the expected yield in data is completely negligible for each of the background decay modes, these decay channels are ultimately harmless and do not invalidate the use of the inclusive model to describe the physics background content in data.
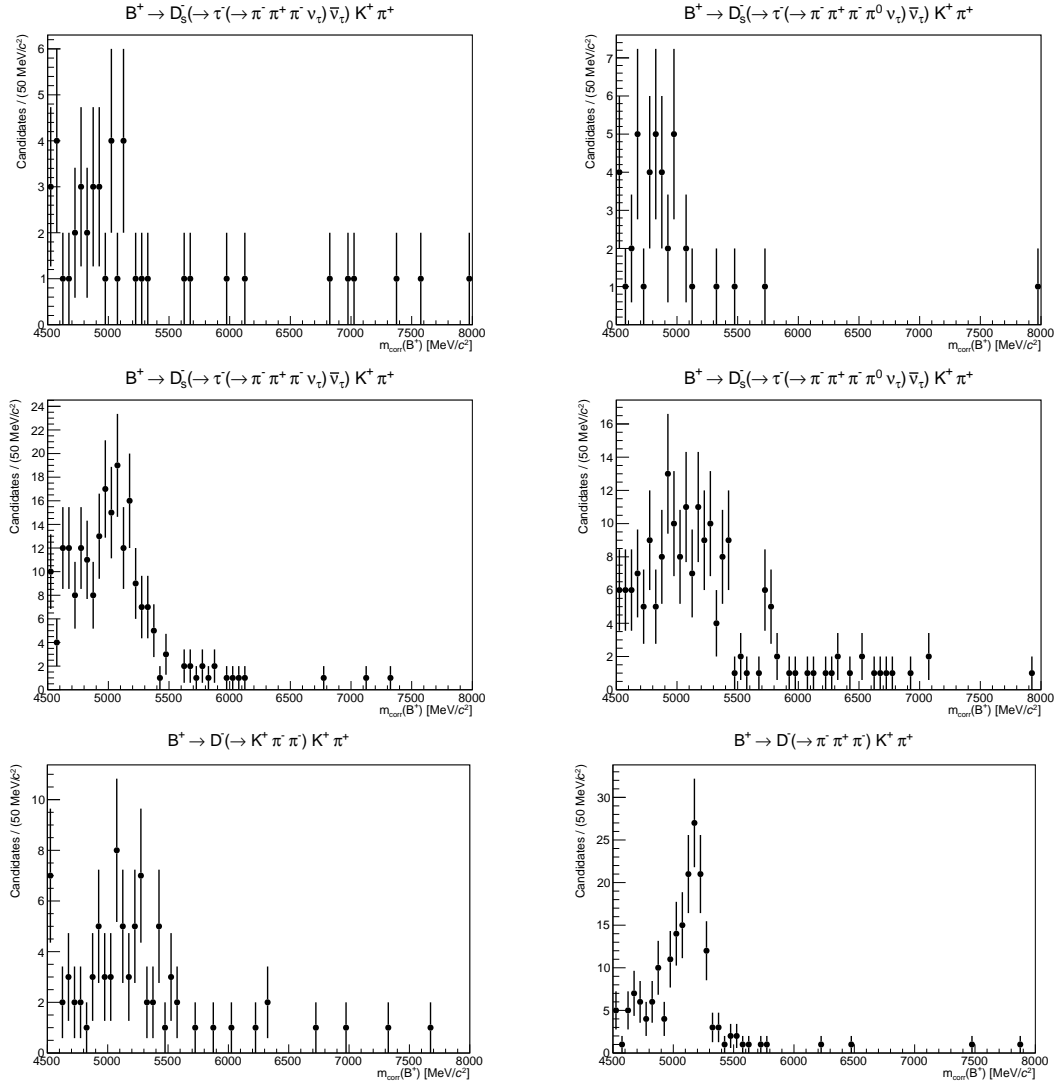
Figure 4.25 – Distribution of $m_{\mathrm{corr}}(B^+)$ for the resonant simulated backgrounds after reconstruction plus stripping selection. The first line shows the $B^+ \rightarrow D_s^- \big(\rightarrow \tau^- \big(\rightarrow \pi^- \pi^+ \pi^- \nu_\tau \big) \overline{\nu}_\tau \big) K^+ \pi^+$ and $B^+ \rightarrow D_s^- \big(\rightarrow \tau^- \big(\rightarrow \pi^- \pi^+ \pi^- \pi^0 \nu_\tau \big) \overline{\nu}_\tau \big) K^+ \pi^+$ decay modes reconstructed as $B^+ \rightarrow K^+ \tau^+ \mu^-$. The second line shows these same decay modes but reconstructed as $B^+ \rightarrow K^+ \tau^- \mu^+$. The last line shows the $B^+ \rightarrow D^- \big(\rightarrow K^+ \pi^- \pi^- \big) K^+ \pi^+$ and $B^+ \rightarrow D^- \big(\rightarrow \pi^+ \pi^- \pi^- \big) K^+ \pi^+$ decay modes reconstructed as $B^+ \rightarrow K^+ \tau^- \mu^+$ which are identical respectively to the $m_{\mathrm{corr}}(B^+)$ distributions of the $B^+ \rightarrow D^+ (\rightarrow K^- \pi^+ \pi^+) K^+ \pi^-$ and $B^+ \rightarrow D^+ (\rightarrow \pi^- \pi^+ \pi^+) K^+ \pi^-$ decay modes reconstructed as $B^+ \rightarrow K^+ \tau^+ \mu^-$, since the B^+ decay follows a phase-space model.

Table 4.8 – The first column specifies the background decay mode while the second gives the associated selection efficiency when reconstructing the decays as $B^+ \to K^+ \tau^- \mu^+$ (top) and $B^+ \to K^+ \tau^+ \mu^-$ (bottom). The efficiencies are computed using uncorrected simulation and the signal $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ decay mode efficiency is assumed for the requirements on the BDT, except for the $B^+ \to D_s^- (\to \tau^- (\to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau) \overline{\nu}_\tau) K^+ \pi^+$ decay mode, where the $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ efficiency is assumed instead. The third and fourth columns give respectively the combined branching fraction and the expected yield, in the mass region $4.5 < m_{\mathrm{corr}}(B^+) < 8\,\mathrm{GeV}/c^2$, in the 2016 data sample.

Decay	Total efficiency	Total $\mathscr{B}$	Expected yield
$B^+ \to D^- (\to K^+ \pi^- \pi^-) K^+ \pi^+$	$(3.00 \pm 1.50) \times 10^{-7}$	$(7.22 \pm 0.48) \times 10^{-6}$	$(8.55 \pm 4.42) \times 10^{-1}$
$B^+ \to D^- (\to \pi^+ \pi^- \pi^-) K^+ \pi^+$	$(3.21 \pm 0.48) \times 10^{-6}$	$(2.52 \pm 0.21) \times 10^{-7}$	$(3.20 \pm 0.66) \times 10^{-1}$
$B^+ \to D_s^- (\to \tau^- (\to \pi^- \pi^+ \pi^- \nu_\tau) \overline{\nu}_\tau) K^+ \pi^+$	$(1.35 \pm 0.22) \times 10^{-6}$	$(9.18 \pm 1.19) \times 10^{-7}$	$(4.91 \pm 1.17) \times 10^{-1}$
$B^+ \to D_s^- (\to \tau^- (\to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau) \overline{\nu}_\tau) K^+ \pi^+$	$(3.93 \pm 0.80) \times 10^{-7}$	$(4.56 \pm 0.59) \times 10^{-7}$	$(7.09 \pm 1.88) \times 10^{-2}$
$B^+ \to D^+ (\to K^- \pi^+ \pi^+) K^+ \pi^-$	$(5.96 \pm 1.99) \times 10^{-7}$	$(5.25 \pm 1.04) \times 10^{-7}$	$(1.24 \pm 0.50) \times 10^{-1}$
$B^+ \to D^+ (\to \pi^- \pi^+ \pi^+) K^+ \pi^-$	$(2.95 \pm 0.43) \times 10^{-6}$	$(1.83 \pm 0.37) \times 10^{-8}$	$(2.14 \pm 0.59) \times 10^{-2}$
$B^+ \to D_s^- (\to \tau^- (\to \pi^- \pi^+ \pi^- \nu_\tau) \overline{\nu}_\tau) K^+ \pi^+$	$(3.03 \pm 3.04) \times 10^{-8}$	$(9.18 \pm 1.19) \times 10^{-7}$	$(1.10 \pm 1.12) \times 10^{-2}$
$B^+ \to D_s^- (\to \tau^- (\to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau) \overline{\nu}_\tau) K^+ \pi^+$	$< 2.62 \times 10^{-8}$	$(4.56 \pm 0.59) \times 10^{-7}$	$< 4.71 \times 10^{-3}$

4.6.3 Expected yields for additional backgrounds

The most important sources of background have been discussed previously. A few other cross-checks are performed in order to make sure that the background description is under control. The decay $B^+ \to K^+ \psi(2S)\big(\to \tau^+(\to \pi^+\pi^-\pi^+\overline{\nu}_\tau)\tau^-(\to \mu^-\overline{\nu}_\mu\nu_\tau)\big)$ is considered as a potential background source. As usual, 1M events after generator requirements have been simulated and the expected yield of this background in the 2016 data sample is computed. The corresponding $m_{\mathrm{corr}}(B^+)$ distribution is also shown after the reconstruction plus stripping selection for candidates reconstructed as $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ in Fig. 4.26. Following the same procedure as previously, a negligible yield is expected for both the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ data samples. Other decay channels have been considered as well but all gave negligible contributions.
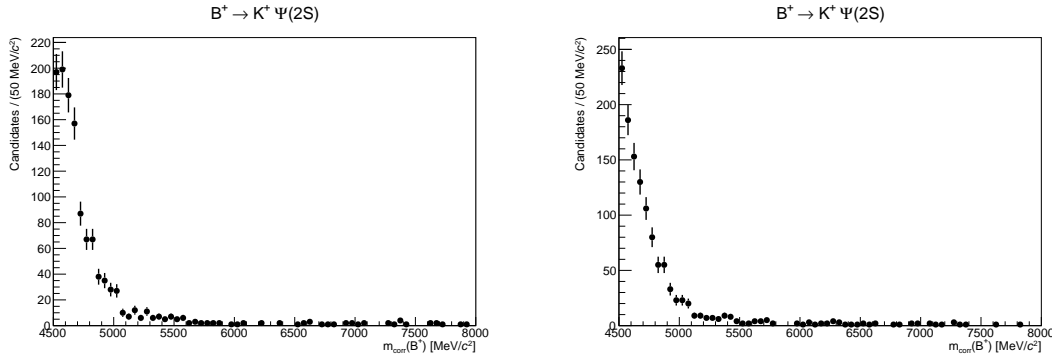
Figure 4.26 – Distribution of $m_{\mathrm{corr}}(B^+)$ after reconstruction plus stripping selection for the sum of the $B^+ \to K^+\psi(2S)\big(\to \tau^+(\to \pi^+\pi^-\pi^+\overline{\nu}_\tau)\tau^-(\to \mu^-\overline{\nu}_\mu\nu_\tau)\big)$ and $B^+ \to K^+\psi(2S)\big(\to \tau^-(\to \pi^-\pi^+\pi^-\nu_\tau)\tau^+(\to \mu^+\nu_\mu\overline{\nu}_\tau)\big)$ background decays reconstructed as $B^+ \to K^+\tau^-\mu^+$ on the left and $B^+ \to K^+\tau^+\mu^-$ on the right.

4.6.4 Simulation of expected combination of inclusive backgrounds in data

In order to further validate the use of an inclusive model in the description of the overall physics background content, an additional simulated sample is produced. This sample models the inclusive decay of b hadrons to at least five charged tracks. The resulting mass distribution for this sample is assumed to describe well the distribution of the physics background content in data. The sample is generated such as to contain 40M events after the generator level requirements. It is then reconstructed either as $B^+ \to K^+\tau^-\mu^+$ or $B^+ \to K^+\tau^+\mu^-$ and is fitted with the inclusive model described in Sec. 5.1 after applying the preselection requirements. The resulting fits are shown in Fig. 4.27 where the model is seen to describe adequately the data in the full mass range. While the statistics of the simulated sample are rather low after applying the preselection requirements, it still shows that within these limited statistics there

are no indication of peaking backgrounds in the signal region and that the fit model is able to describe the data successfully.
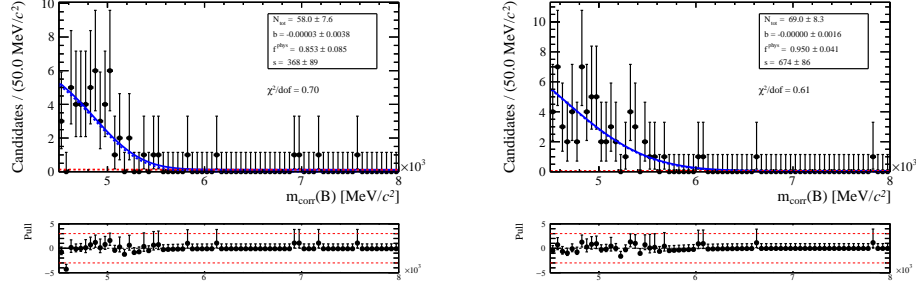
Figure 4.27 – Fit of the simulated sample of inclusive decays of b-hadrons to at least 5 charged tracks. The data is simulated with 2016 data acquisition conditions and has the preselection applied. The candidates are reconstructed as $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ in the left and right plots, respectively. The total pdf is drawn in solid blue while the individual components describing the physics and combinatorial backgrounds are drawn in dashed blue and red, respectively.

4.7 Selection efficiencies

In this section, the various selection efficiency terms are given. The full selection efficiency is expressed as follows

$$\epsilon_s = \epsilon_{gen} \times \epsilon_{reco+strip} \times \epsilon_{presel} \times \epsilon_{BDT} \times \epsilon_{PID} \times \epsilon_{trigger} \times \epsilon_{res}, \tag{4.5}$$

where ϵ_s, ϵ_{gen}, $\epsilon_{reco+strip}$, ϵ_{presel}, ϵ_{BDT}, ϵ_{PID}, $\epsilon_{trigger}$ and ϵ_{res} are the efficiencies associated with the requirements of the full selection for the signal channel, the generator level, the reconstruction and stripping without PID requirement, the additional preselection, the BDT, the PID, the trigger and the additional mass requirements aimed at removing resonant sub-decays, respectively. For the $B^+ \to \overline{D}^0 D_s^+$ decay mode, the structure is identical except for the fact that neither the ϵ_{BDT} term is present, since there is no BDT selection applied on that channel, nor the ϵ_{res} term. The trigger efficiency itself is further split into the contributions due to the requirements applied on the hardware and software stages, *i.e.* $\epsilon_{trigger} = \epsilon_{L0} \times \epsilon_{HLT}$.

The selection efficiencies are measured by using simulation with the sole exception being ε_{L0} which is measured using data-driven techniques, which are explained in detail in Sec. 5.4. The efficiencies associated with the generator level requirements are given in Tab. 4.9 for all decay modes. The values of $\varepsilon_{reco+strip}$, ε_{presel} and ε_{res} are given in Tabs. 4.10, 4.11 and 4.12 for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes while the former two terms are given in Tab. 4.13 for the $B^+ \to \overline{D}^0 D_s^+$ decay mode. The efficiency associated with the BDT requirements is given in Tab. 4.14 for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes. The values of the efficiencies associated with the PID requirements are given for all three decay modes in Tab. 4.15. Both trigger related efficiencies are given in Tab. 4.16 for the $B^+ \to \overline{D}^0 D_s^+$ decay mode while the values of ε_{L0} and ε_{HLT} are given respectively in Tabs. 4.17 and 4.18 for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes. Additionally, the combined full selection efficiency is given in Tab. 4.19 for all decay modes.

Table 4.9 – The second and third columns give the efficiencies ε_{gen} associated with the generator level requirements for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes with $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$, respectively. The last column gives the efficiency for the $B^+ \to \overline{D}^0 D_s^+$ decay mode.

Year	$\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$	$\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	$B^+ \to \overline{D}^0 D_s^+$
2011	$(6.17 \pm 0.01) \times 10^{-2}$	$(5.49 \pm 0.01) \times 10^{-2}$	$(13.31 \pm 0.05) \times 10^{-2}$
2012	$(6.45 \pm 0.01) \times 10^{-2}$	$(5.72 \pm 0.01) \times 10^{-2}$	$(13.57 \pm 0.04) \times 10^{-2}$
2015	$(7.36 \pm 0.02) \times 10^{-2}$	$(6.60 \pm 0.02) \times 10^{-2}$	$(14.52 \pm 0.03) \times 10^{-2}$
2016	$(7.41 \pm 0.02) \times 10^{-2}$	$(6.57 \pm 0.02) \times 10^{-2}$	$(14.53 \pm 0.03) \times 10^{-2}$
2017	$(7.34 \pm 0.02) \times 10^{-2}$	$(6.63 \pm 0.02) \times 10^{-2}$	$(14.55 \pm 0.04) \times 10^{-2}$
2018	$(7.35 \pm 0.02) \times 10^{-2}$	$(6.60 \pm 0.02) \times 10^{-2}$	$(14.52 \pm 0.04) \times 10^{-2}$

Table 4.10 – Efficiencies $\varepsilon_{reco+strip}$ associated with the reconstruction and the requirements of the `StrippingB2XTauMu_K_3pi_looseLine` stripping line, excluding particle identification requirements, for the $B^+ \rightarrow K^+\tau^-\mu^+$ (left) and $B^+ \rightarrow K^+\tau^+\mu^-$ (right) decay modes. The top and bottom parts of the table show the efficiencies for the $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \rightarrow \pi^-\pi^+\pi^-\pi^0\nu_\tau$ sub-decay modes.

Year	$B^+ \rightarrow K^+\tau^-\mu^+$	$B^+ \rightarrow K^+\tau^+\mu^-$
2011	$(5.56 \pm 0.05) \times 10^{-2}$	$(5.62 \pm 0.05) \times 10^{-2}$
2012	$(5.19 \pm 0.03) \times 10^{-2}$	$(5.21 \pm 0.03) \times 10^{-2}$
2015	$(5.42 \pm 0.06) \times 10^{-2}$	$(5.52 \pm 0.06) \times 10^{-2}$
2016	$(5.70 \pm 0.03) \times 10^{-2}$	$(5.80 \pm 0.03) \times 10^{-2}$
2017	$(4.45 \pm 0.03) \times 10^{-2}$	$(4.45 \pm 0.03) \times 10^{-2}$
2018	$(4.47 \pm 0.03) \times 10^{-2}$	$(4.47 \pm 0.03) \times 10^{-2}$
2011	$(5.54 \pm 0.04) \times 10^{-2}$	$(5.73 \pm 0.05) \times 10^{-2}$
2012	$(5.08 \pm 0.03) \times 10^{-2}$	$(5.26 \pm 0.03) \times 10^{-2}$
2015	$(5.59 \pm 0.06) \times 10^{-2}$	$(5.48 \pm 0.06) \times 10^{-2}$
2016	$(5.70 \pm 0.03) \times 10^{-2}$	$(5.99 \pm 0.03) \times 10^{-2}$
2017	$(4.46 \pm 0.03) \times 10^{-2}$	$(4.60 \pm 0.03) \times 10^{-2}$
2018	$(4.42 \pm 0.03) \times 10^{-2}$	$(4.50 \pm 0.03) \times 10^{-2}$

Table 4.11 – Efficiencies ε_{presel} associated with the preselection requirements for the $B^+ \rightarrow K^+\tau^-\mu^+$ (left) and $B^+ \rightarrow K^+\tau^+\mu^-$ (right) decay modes. The top and bottom parts of the table show the efficiencies for the $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \rightarrow \pi^-\pi^+\pi^-\pi^0\nu_\tau$ sub-decay modes.

Year	$B^+ \rightarrow K^+\tau^-\mu^+$	$B^+ \rightarrow K^+\tau^+\mu^-$
2011	$(4.30 \pm 0.04) \times 10^{-1}$	$(4.46 \pm 0.04) \times 10^{-1}$
2012	$(4.12 \pm 0.03) \times 10^{-1}$	$(4.40 \pm 0.03) \times 10^{-1}$
2015	$(4.02 \pm 0.05) \times 10^{-1}$	$(4.27 \pm 0.05) \times 10^{-1}$
2016	$(4.07 \pm 0.03) \times 10^{-1}$	$(4.17 \pm 0.03) \times 10^{-1}$
2017	$(4.18 \pm 0.03) \times 10^{-1}$	$(4.36 \pm 0.03) \times 10^{-1}$
2018	$(4.17 \pm 0.03) \times 10^{-1}$	$(4.45 \pm 0.03) \times 10^{-1}$
2011	$(3.49 \pm 0.04) \times 10^{-1}$	$(3.89 \pm 0.04) \times 10^{-1}$
2012	$(3.38 \pm 0.03) \times 10^{-1}$	$(3.76 \pm 0.03) \times 10^{-1}$
2015	$(3.29 \pm 0.05) \times 10^{-1}$	$(3.64 \pm 0.05) \times 10^{-1}$
2016	$(3.32 \pm 0.03) \times 10^{-1}$	$(3.65 \pm 0.03) \times 10^{-1}$
2017	$(3.36 \pm 0.03) \times 10^{-1}$	$(3.74 \pm 0.03) \times 10^{-1}$
2018	$(3.40 \pm 0.03) \times 10^{-1}$	$(3.81 \pm 0.03) \times 10^{-1}$

Table 4.12 – Efficiencies ε_{res} associated with the additional invariant mass requirements aimed at removing resonant sub-decays for the $B^+ \to K^+ \tau^- \mu^+$ (left) and $B^+ \to K^+ \tau^+ \mu^-$ (right) decay modes. The top and bottom parts of the table show the efficiencies for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes.

Year	$B^+ \to K^+ \tau^- \mu^+$	$B^+ \to K^+ \tau^+ \mu^-$
2011	$(8.89 \pm 0.12) \times 10^{-1}$	$(9.62 \pm 0.08) \times 10^{-1}$
2012	$(8.97 \pm 0.09) \times 10^{-1}$	$(9.51 \pm 0.07) \times 10^{-1}$
2015	$(9.10 \pm 0.16) \times 10^{-1}$	$(9.60 \pm 0.11) \times 10^{-1}$
2016	$(8.86 \pm 0.08) \times 10^{-1}$	$(9.48 \pm 0.06) \times 10^{-1}$
2017	$(8.95 \pm 0.07) \times 10^{-1}$	$(9.48 \pm 0.05) \times 10^{-1}$
2018	$(8.86 \pm 0.08) \times 10^{-1}$	$(9.56 \pm 0.05) \times 10^{-1}$
2011	$(8.93 \pm 0.16) \times 10^{-1}$	$(9.57 \pm 0.12) \times 10^{-1}$
2012	$(8.97 \pm 0.13) \times 10^{-1}$	$(9.67 \pm 0.08) \times 10^{-1}$
2015	$(8.61 \pm 0.25) \times 10^{-1}$	$(9.56 \pm 0.17) \times 10^{-1}$
2016	$(8.94 \pm 0.11) \times 10^{-1}$	$(9.59 \pm 0.08) \times 10^{-1}$
2017	$(8.80 \pm 0.12) \times 10^{-1}$	$(9.57 \pm 0.07) \times 10^{-1}$
2018	$(8.80 \pm 0.12) \times 10^{-1}$	$(9.59 \pm 0.08) \times 10^{-1}$

Table 4.13 – The second column gives the efficiencies associated with the reconstruction and the requirements of the `StrippingB2XTau_DD0_Line` stripping line, excluding particle identification requirements, for the $B^+ \to \bar{D}^0 D_s^+$ decay mode, while the last column gives the preselection efficiencies. These numbers correspond to $\epsilon_{reco+strip}$ and ϵ_{presel} respectively.

Year	$\epsilon_{reco+strip}^n$	ϵ_{presel}^n
2011	$(2.89 \pm 0.02) \times 10^{-2}$	$(5.56 \pm 0.04) \times 10^{-1}$
2012	$(2.69 \pm 0.02) \times 10^{-2}$	$(5.62 \pm 0.03) \times 10^{-1}$
2015	$(3.02 \pm 0.03) \times 10^{-2}$	$(5.50 \pm 0.05) \times 10^{-1}$
2016	$(3.14 \pm 0.02) \times 10^{-2}$	$(5.39 \pm 0.03) \times 10^{-1}$
2017	$(3.16 \pm 0.02) \times 10^{-2}$	$(5.41 \pm 0.03) \times 10^{-1}$
2018	$(3.15 \pm 0.02) \times 10^{-2}$	$(5.39 \pm 0.03) \times 10^{-1}$

Table 4.14 – The second and third columns give the efficiency ε_{BDT} associated with the optimal requirements on the output of the BDT (determined in Sec. 5.3), *i.e.* $\mathrm{BDT_{comb}} > 0.8$, $\mathrm{BDT_{phys}} > 0.8$ and $\mathrm{BDT_{comb}} > 0.75$, $\mathrm{BDT_{phys}} > 0.7$ for the $B^+ \rightarrow K^+\tau^-\mu^+$ and $B^+ \rightarrow K^+\tau^+\mu^-$ decay modes respectively, for the $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \rightarrow \pi^-\pi^+\pi^-\pi^0\nu_\tau$ sub-decay modes at the top and bottom.

Year	$B^+ \rightarrow K^+\tau^-\mu^+$	$B^+ \rightarrow K^+\tau^+\mu^-$
2011	$(4.84 \pm 0.06) \times 10^{-1}$	$(4.72 \pm 0.06) \times 10^{-1}$
2012	$(4.55 \pm 0.05) \times 10^{-1}$	$(4.41 \pm 0.05) \times 10^{-1}$
2015	$(4.44 \pm 0.09) \times 10^{-1}$	$(4.14 \pm 0.08) \times 10^{-1}$
2016	$(4.55 \pm 0.05) \times 10^{-1}$	$(4.01 \pm 0.04) \times 10^{-1}$
2017	$(4.65 \pm 0.05) \times 10^{-1}$	$(4.27 \pm 0.05) \times 10^{-1}$
2018	$(4.72 \pm 0.05) \times 10^{-1}$	$(4.29 \pm 0.05) \times 10^{-1}$
2011	$(3.02 \pm 0.06) \times 10^{-1}$	$(2.54 \pm 0.06) \times 10^{-1}$
2012	$(2.71 \pm 0.05) \times 10^{-1}$	$(2.36 \pm 0.04) \times 10^{-1}$
2015	$(2.67 \pm 0.08) \times 10^{-1}$	$(2.11 \pm 0.07) \times 10^{-1}$
2016	$(2.53 \pm 0.04) \times 10^{-1}$	$(2.03 \pm 0.04) \times 10^{-1}$
2017	$(2.69 \pm 0.05) \times 10^{-1}$	$(2.19 \pm 0.04) \times 10^{-1}$
2018	$(2.69 \pm 0.05) \times 10^{-1}$	$(2.22 \pm 0.04) \times 10^{-1}$

Table 4.15 – The second, third and fourth columns give the efficiencies ϵ_{PID} associated with the PID requirements for the $B^+ \rightarrow K^+\tau^-\mu^+$, $B^+ \rightarrow K^+\tau^+\mu^-$ and $B^+ \rightarrow \bar{D}^0 D_s^+$ decay modes, respectively. The top and bottom halves of the table refer respectively to the $\tau^- \rightarrow \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \rightarrow \pi^-\pi^+\pi^-\pi^0\nu_\tau$ sub-decay modes.

Year	$B^+ \rightarrow K^+\tau^-\mu^+$	$B^+ \rightarrow K^+\tau^+\mu^-$	$B^+ \rightarrow \bar{D}^0 D_s^+$
2011	$(4.57 \pm 0.09) \times 10^{-1}$	$(3.87 \pm 0.09) \times 10^{-1}$	$(3.15 \pm 0.05) \times 10^{-1}$
2012	$(4.48 \pm 0.07) \times 10^{-1}$	$(3.79 \pm 0.07) \times 10^{-1}$	$(3.10 \pm 0.04) \times 10^{-1}$
2015	$(5.77 \pm 0.13) \times 10^{-1}$	$(5.92 \pm 0.13) \times 10^{-1}$	$(5.65 \pm 0.07) \times 10^{-1}$
2016	$(5.54 \pm 0.07) \times 10^{-1}$	$(5.41 \pm 0.07) \times 10^{-1}$	$(5.71 \pm 0.04) \times 10^{-1}$
2017	$(6.45 \pm 0.07) \times 10^{-1}$	$(6.25 \pm 0.08) \times 10^{-1}$	$(5.87 \pm 0.04) \times 10^{-1}$
2018	$(6.42 \pm 0.07) \times 10^{-1}$	$(6.06 \pm 0.07) \times 10^{-1}$	$(5.85 \pm 0.04) \times 10^{-1}$
2011	$(4.65 \pm 0.13) \times 10^{-1}$	$(3.75 \pm 0.12) \times 10^{-1}$	-
2012	$(4.51 \pm 0.10) \times 10^{-1}$	$(3.88 \pm 0.10) \times 10^{-1}$	-
2015	$(6.02 \pm 0.18) \times 10^{-1}$	$(5.56 \pm 0.20) \times 10^{-1}$	-
2016	$(5.51 \pm 0.10) \times 10^{-1}$	$(5.44 \pm 0.11) \times 10^{-1}$	-
2017	$(6.51 \pm 0.11) \times 10^{-1}$	$(6.42 \pm 0.11) \times 10^{-1}$	-
2018	$(6.55 \pm 0.11) \times 10^{-1}$	$(6.24 \pm 0.11) \times 10^{-1}$	-

Table 4.16 – Efficiencies associated with the trigger requirements for the $B^+ \to \overline{D}^0 D_s^+$ decay mode. The second and third columns give ϵ_{L0} computed using the TISTOS method, explained in Sec. 5.4, and ϵ_{HLT}, respectively. The efficiency ϵ_{HLT} is computed with respect to the L0 requirements.

Year	$\epsilon_{L0}^n (TISTOS)$	ϵ_{HLT}^n
2011	$(3.65 \pm 0.35) \times 10^{-1}$	$(8.89 \pm 0.10) \times 10^{-1}$
2012	$(3.66 \pm 0.25) \times 10^{-1}$	$(9.18 \pm 0.07) \times 10^{-1}$
2015	$(4.25 \pm 0.47) \times 10^{-1}$	$(9.38 \pm 0.07) \times 10^{-1}$
2016	$(3.68 \pm 0.15) \times 10^{-1}$	$(9.61 \pm 0.03) \times 10^{-1}$
2017	$(3.93 \pm 0.14) \times 10^{-1}$	$(9.64 \pm 0.03) \times 10^{-1}$
2018	$(3.59 \pm 0.13) \times 10^{-1}$	$(9.58 \pm 0.03) \times 10^{-1}$

Table 4.17 – The second and third columns give the efficiencies ϵ_{L0} associated with the L0 trigger requirements, which have been determined using a data-driven method detailed in Sec. 5.4, for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively. The top and bottom halves of the table refer respectively to the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes.

Year	$B^+ \to K^+ \tau^- \mu^+$	$B^+ \to K^+ \tau^+ \mu^-$
2011	$(5.99 \pm 0.13) \times 10^{-1}$	$(6.19 \pm 0.14) \times 10^{-1}$
2012	$(5.63 \pm 0.11) \times 10^{-1}$	$(5.69 \pm 0.11) \times 10^{-1}$
2015	$(3.46 \pm 0.16) \times 10^{-1}$	$(3.79 \pm 0.16) \times 10^{-1}$
2016	$(5.59 \pm 0.09) \times 10^{-1}$	$(5.54 \pm 0.10) \times 10^{-1}$
2017	$(6.15 \pm 0.09) \times 10^{-1}$	$(6.44 \pm 0.09) \times 10^{-1}$
2018	$(5.78 \pm 0.09) \times 10^{-1}$	$(5.65 \pm 0.10) \times 10^{-1}$
2011	$(5.75 \pm 0.18) \times 10^{-1}$	$(5.92 \pm 0.21) \times 10^{-1}$
2012	$(5.70 \pm 0.15) \times 10^{-1}$	$(5.69 \pm 0.16) \times 10^{-1}$
2015	$(4.04 \pm 0.23) \times 10^{-1}$	$(3.75 \pm 0.26) \times 10^{-1}$
2016	$(5.65 \pm 0.14) \times 10^{-1}$	$(5.64 \pm 0.14) \times 10^{-1}$
2017	$(5.90 \pm 0.13) \times 10^{-1}$	$(6.50 \pm 0.14) \times 10^{-1}$
2018	$(5.60 \pm 0.14) \times 10^{-1}$	$(5.85 \pm 0.14) \times 10^{-1}$

Table 4.18 – Efficiencies ε_{HLT} associated with the software trigger requirements for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes in the second and third columns, respectively. The top and bottom parts of the table correspond to the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$ sub-decay modes.

Year	ϵ^s_{HLT} $B^+ \to K^+\tau^-\mu^+$	ϵ^s_{HLT} $B^+ \to K^+\tau^+\mu^-$
2011	$(8.94 \pm 0.11) \times 10^{-1}$	$(8.71 \pm 0.12) \times 10^{-1}$
2012	$(9.31 \pm 0.07) \times 10^{-1}$	$(9.25 \pm 0.08) \times 10^{-1}$
2015	$(9.81 \pm 0.08) \times 10^{-1}$	$(9.78 \pm 0.08) \times 10^{-1}$
2016	$(9.79 \pm 0.04) \times 10^{-1}$	$(9.83 \pm 0.03) \times 10^{-1}$
2017	$(9.75 \pm 0.04) \times 10^{-1}$	$(9.76 \pm 0.04) \times 10^{-1}$
2018	$(9.77 \pm 0.04) \times 10^{-1}$	$(9.87 \pm 0.03) \times 10^{-1}$
2011	$(8.63 \pm 0.17) \times 10^{-1}$	$(8.91 \pm 0.17) \times 10^{-1}$
2012	$(9.17 \pm 0.11) \times 10^{-1}$	$(9.30 \pm 0.11) \times 10^{-1}$
2015	$(9.89 \pm 0.07) \times 10^{-1}$	$(9.79 \pm 0.12) \times 10^{-1}$
2016	$(9.83 \pm 0.05) \times 10^{-1}$	$(9.84 \pm 0.05) \times 10^{-1}$
2017	$(9.64 \pm 0.07) \times 10^{-1}$	$(9.84 \pm 0.04) \times 10^{-1}$
2018	$(9.80 \pm 0.05) \times 10^{-1}$	$(9.83 \pm 0.05) \times 10^{-1}$

Table 4.19 – Signal efficiencies associated with the full selection for the $B^+ \to K^+\tau^-\mu^+$, $B^+ \to K^+\tau^+\mu^-$ and $B^+ \to \overline{D}^0 D_s^+$ decay modes on the left, middle and right columns, respectively. The top and bottom halves of the table show the efficiencies for the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$ decay modes.

Year	$B^+ \to K^+\tau^-\mu^+$	$B^+ \to K^+\tau^+\mu^-$	$B^+ \to \overline{D}^0 D_s^+$
2011	$(1.55 \pm 0.06) \times 10^{-4}$	$(1.47 \pm 0.06) \times 10^{-4}$	$(2.19 \pm 0.05) \times 10^{-4}$
2012	$(1.33 \pm 0.04) \times 10^{-4}$	$(1.24 \pm 0.04) \times 10^{-4}$	$(2.14 \pm 0.04) \times 10^{-4}$
2015	$(1.27 \pm 0.08) \times 10^{-4}$	$(1.51 \pm 0.08) \times 10^{-4}$	$(5.43 \pm 0.11) \times 10^{-4}$
2016	$(2.10 \pm 0.06) \times 10^{-4}$	$(2.01 \pm 0.05) \times 10^{-4}$	$(4.98 \pm 0.05) \times 10^{-4}$
2017	$(2.20 \pm 0.06) \times 10^{-4}$	$(2.27 \pm 0.06) \times 10^{-4}$	$(5.52 \pm 0.06) \times 10^{-4}$
2018	$(2.08 \pm 0.06) \times 10^{-4}$	$(2.03 \pm 0.05) \times 10^{-4}$	$(4.97 \pm 0.05) \times 10^{-4}$
2011	$(6.59 \pm 0.37) \times 10^{-5}$	$(5.90 \pm 0.35) \times 10^{-5}$	-
2012	$(5.64 \pm 0.25) \times 10^{-5}$	$(5.31 \pm 0.24) \times 10^{-5}$	-
2015	$(6.71 \pm 0.54) \times 10^{-5}$	$(5.43 \pm 0.48) \times 10^{-5}$	-
2016	$(8.58 \pm 0.33) \times 10^{-5}$	$(8.43 \pm 0.33) \times 10^{-5}$	-
2017	$(8.70 \pm 0.34) \times 10^{-5}$	$(9.81 \pm 0.36) \times 10^{-5}$	-
2018	$(8.46 \pm 0.33) \times 10^{-5}$	$(8.66 \pm 0.34) \times 10^{-5}$	-

5 Mass fit and results

In this chapter are presented the various models used to fit the invariant mass of the $B^+ \to \overline{D}^0 D_s^+$ candidates and the corrected B^+ mass of the $B^+ \to K^+ \tau^\pm \mu^\mp$ candidates. The limit setting procedure is explained and the requirements imposed on the BDT outputs and the combined PID variable $\mathscr{P}_3$, based on the expected limits on the signal branching fractions, are determined. Finally, the data is fitted and the values of $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$ and $\mathscr{B}(B^+ \to K^+ \tau^+ \mu^-)$ are measured.

5.1 Mass fit model

In order to obtain the expected limits on $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$ and $\mathscr{B}(B^+ \to K^+ \tau^+ \mu^-)$ necessary for the determination of the requirements on the BDT outputs as well as on $\mathscr{P}_3$, the mass fit models for the various samples need to be determined first. As a general rule, the shape parameters of a given probability density function (pdf) are determined by fitting combined samples which merge all years within a Run. The samples for the individual years are then fitted and share the same shape parameters within a Run.

5.1.1 $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes

In the case of the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes, there are two sources of signal. These are the $B^+ \to K^+ \tau^\pm \mu^\mp$ decays that proceed either through the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ or $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes. Naturally, the $m_{\text{corr}}(B^+)$ distributions associated with these two sub-processes will differ and as such the pdf describing their corrected mass distributions have to be determined separately. In order to determine these pdf, an extended maximum likelihood fit to the $m_{\text{corr}}(B^+)$ distribution is performed for the corresponding simulated samples. The functional form used in the fit is identical for both the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay modes and is given by the sum of two Gaussian functions $\mathscr{G}$ and a Crystal Ball (CB)

function $\mathscr{CB}$ with a tail on the upper mass sideband, all three sharing the same mean parameter. For the CB function, the constraint $n = 2$ is imposed in order to stabilise the fit. This parameter dictates the order of the power law tail and is highly correlated with α which indicates the location of the transition between the Gaussian core and the power law tail. The other parameters are left free in the fit. The CB function is defined as follows

$$\mathscr{CB}(m; \mu, \sigma, \alpha, n) = N \begin{cases} \exp\left(-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2\right), & \text{if } \frac{m-\mu}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{\alpha^2}{2}\right)\left(\frac{n}{|\alpha|} - |\alpha| - \frac{m-\mu}{\sigma}\right)^{-n}, & \text{otherwise} \end{cases}, \quad (5.1)$$

with μ and σ being the mean and standard deviation of the Gaussian core of the distribution while the meaning of α and n have been explained above. The parameter N is a normalisation factor whose expression is

$$N = \frac{1}{\sigma\left(\frac{n}{|\alpha|(n-1)}\exp\left(-\frac{\alpha^2}{2}\right) + \sqrt{\frac{\pi}{2}}\left(1 + \text{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right)\right)}, \quad (5.2)$$

with $\text{erf}(x)$ the error function. The functional form of the signal is then given by :

$$f_{sig} = N_{sig} \cdot \mathscr{G}(m_{\text{corr}}; m, s) + N_{sig2} \cdot \mathscr{G}(m_{\text{corr}}; m, s_2) + N_{sig3} \cdot \mathscr{CB}(m_{\text{corr}}; m, s_3, \alpha_3, n_3 = 2), \quad (5.3)$$

where N_{sig}, N_{sig2} and N_{sig3} are the yields associated with the three functions, m is the mean parameter shared between all functions, s_1, s_2 and s_3 are the widths, α_3 is the location of the transition between the Gaussian and power law tail and finally n_3 is the order of the power law tail. The fits to the Run 1 and Run 2 simulated samples are shown respectively in Figs. 5.1 and 5.2 for the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$ decay modes. The fits to the individual years are well described by the common pdf.

Concerning the description of the background content of the $B^+ \to K^+\tau^\pm\mu^\mp$ data samples, the strategy is to use an inclusive model to describe the combination of all sources of physics background, as explained previously, and a component to describe the combinatorial background. The resulting pdf is the sum of an Argus function $\mathscr{A}$ convolved with a Gaussian function in order to describe the physics background and a decreasing exponential to describe the combinatorial background. The Argus function is defined as follows

$$\mathscr{A}(m; m_0, c_0) = Nm\sqrt{1 - \left(\frac{m}{m_0}\right)^2}\exp\left(c_0\left(1 - \left(\frac{m}{m_0}\right)^2\right)\right), \quad (5.4)$$

where m_0 and c_0 are respectively the endpoint and slope parameters while N is a normalisation factor. The endpoint parameter of the Argus function is fixed to $5\,\text{GeV}/c^2$ and $4.75\,\text{GeV}/c^2$ for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes, respectively. The different values account for the fact that the composition of the physics background is different for the $B^+ \to K^+\tau^-\mu^+$

Figure 5.1 – Mass fit of the Run 1 simulated samples after applying the optimal 3D require-ments for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes in the left and right column, respectively. The top and bottom lines show the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes, respectively. The blue solid line is the total pdf while the dashed lines are the individual components of the fit.

and $B^+ \to K^+ \tau^+ \mu^-$ samples and that the mean value of the associated $m_{\text{corr}}(B^+)$ distribution is higher for $B^+ \to K^+ \tau^- \mu^+$ candidates than for $B^+ \to K^+ \tau^+ \mu^-$ candidates. These values are fixed in order to stabilise the fit and the effect of this choice will be accounted for by imposing an associated systematic uncertainty. Additionally, the slope parameter of the Argus function is fixed to -1 in order to increase the stability when fitting low statistics samples. Since the slope parameter is highly correlated with the width of the Gaussian function, this choice has virtually no impact on the final result. The resulting functional form of the model is given by

$$f_{bkg} = N_{bkg}\Big(f_{phys} \cdot \mathscr{A}\big(m_{\text{corr}}; m_0, c_0 = -1\big) \otimes \mathscr{G}\big(m_{\text{corr}}; m = 0, s\big) + (1 - f_{phys})e^{bm_{\text{corr}}}\Big), \quad (5.5)$$

where N_{bkg} is the total background yield, f_{phys} is the fraction of the background due to the physics background component, m_0 (fixed to $5\,\text{GeV}/c^2$ for $B^+ \to K^+ \tau^- \mu^+$ and $4.75\,\text{GeV}/c^2$ for $B^+ \to K^+ \tau^+ \mu^-$) and c_0 are the endpoint and slope of the Argus function, m and s are the mean and width of the Gaussian function and finally b is the slope of the exponential function. Unlike for the other models, the shape parameters are not determined through fits to combined data samples but instead a simultaneous fit is performed on all years. The shape parameters are shared for all years within a Run while the background yields float separately for each year. The fraction of physics background f_{phys} is shared for all years and a systematic uncertainty will account for this choice. Furthermore, the width of the Gaussian function is fixed to $400\,\text{MeV}/c^2$
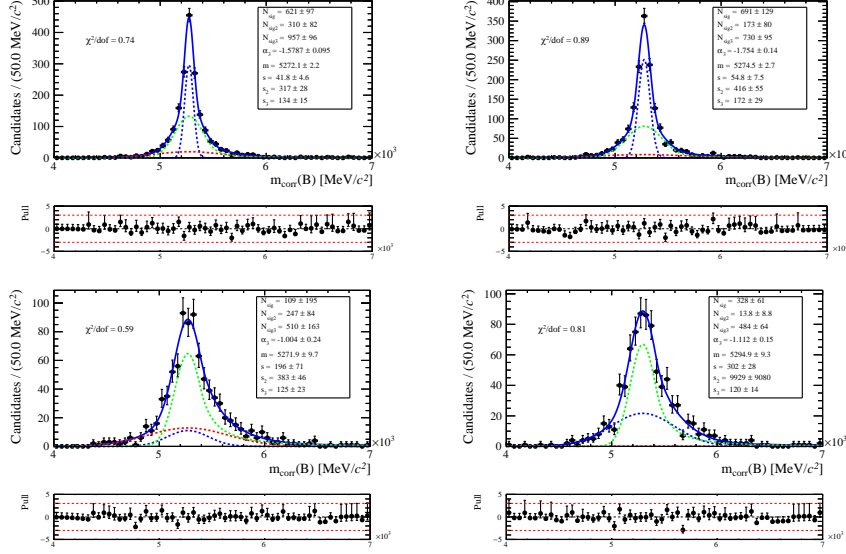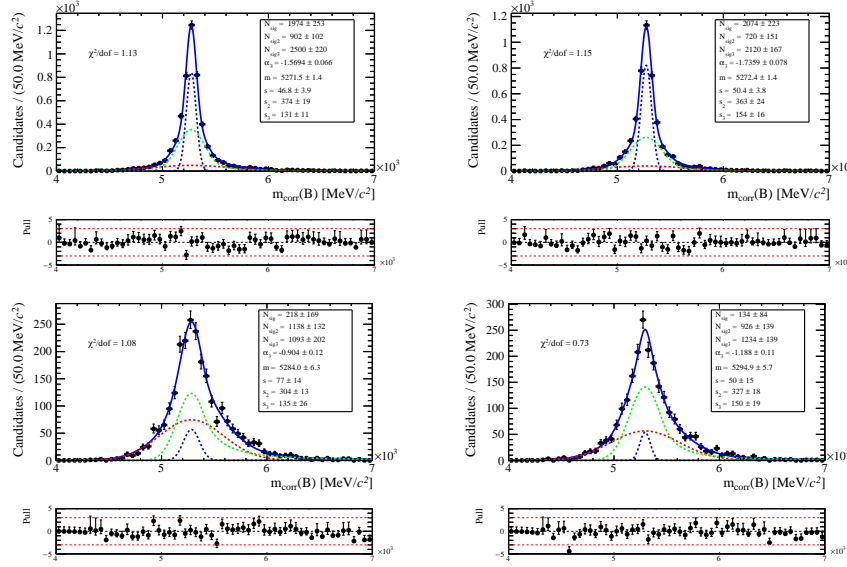
Figure 5.2 – Mass fit of the Run 2 simulated samples after applying the optimal 3D requirements for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes in the left and right column, respectively. The top and bottom lines show the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes, respectively. The blue solid line is the total pdf while the dashed lines are the individual components of the fit.

when fitting the blind data. This is done in order to improve the stability of the fits when fitting only the data mass sidebands during the optimisation process of the 3D requirements. However, when performing fits over the full mass range, *e.g.* in order to determine the expected limits on $\mathcal{B}(B^+ \to K^+ \tau^- \mu^+)$ and $\mathcal{B}(B^+ \to K^+ \tau^+ \mu^-)$, the width parameter is also left free in the fit.

While there is no component in the fit that describes an individual background decay mode, the use of this inclusive model has been validated by fitting the inclusive simulation of b-hadron decays to at least five charged tracks, as explained in Sec. 4.6. The fits of this simulated sample show that this pdf is able to describe the expected background distributions in the full mass range. However, since the statistics of this simulated sample is limited, an additional cross-check is performed in the form of a fit to the real data, after full selection, of the full $m_{\mathrm{corr}}(B^+)$ mass range for candidates in the $m_{\mathrm{corr}}(\tau^-) > 2.5 \, \mathrm{GeV}/c^2$ region. The resulting fits are shown in the in Figs. 5.3 and 5.4. The $B^+ \to K^+ \tau^- \mu^+$ sample displays low statistics and as such only the exponential part of the pdf is used in the fit while the full extent of the background pdf is used for the $B^+ \to K^+ \tau^+ \mu^-$ candidates. Nonetheless, the background pdf is shown to describe well the data for both decay modes, within the limited statistics. The fits to the real blind data after full selection are shown in Figs. 5.5 and 5.6.

Figure 5.3 – Simultaneous Run 1 and Run 2 background mass fit of the $B^+ \to K^+\tau^-\mu^+$ data samples after full selection in the region $m_{\text{corr}}(\tau^-) > 2.5$. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively.

Figure 5.4 – Simultaneous Run 1 and Run 2 background mass fit of the $B^+ \to K^+ \tau^+ \mu^-$ data samples after full selection in the region $m_{\mathrm{corr}}(\tau^-) > 2.5$. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively.

Figure 5.5 – Simultaneous Run 1 and Run 2 background mass fit of the $B^+ \to K^+\tau^-\mu^+$ blinded data samples after full selection. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively.

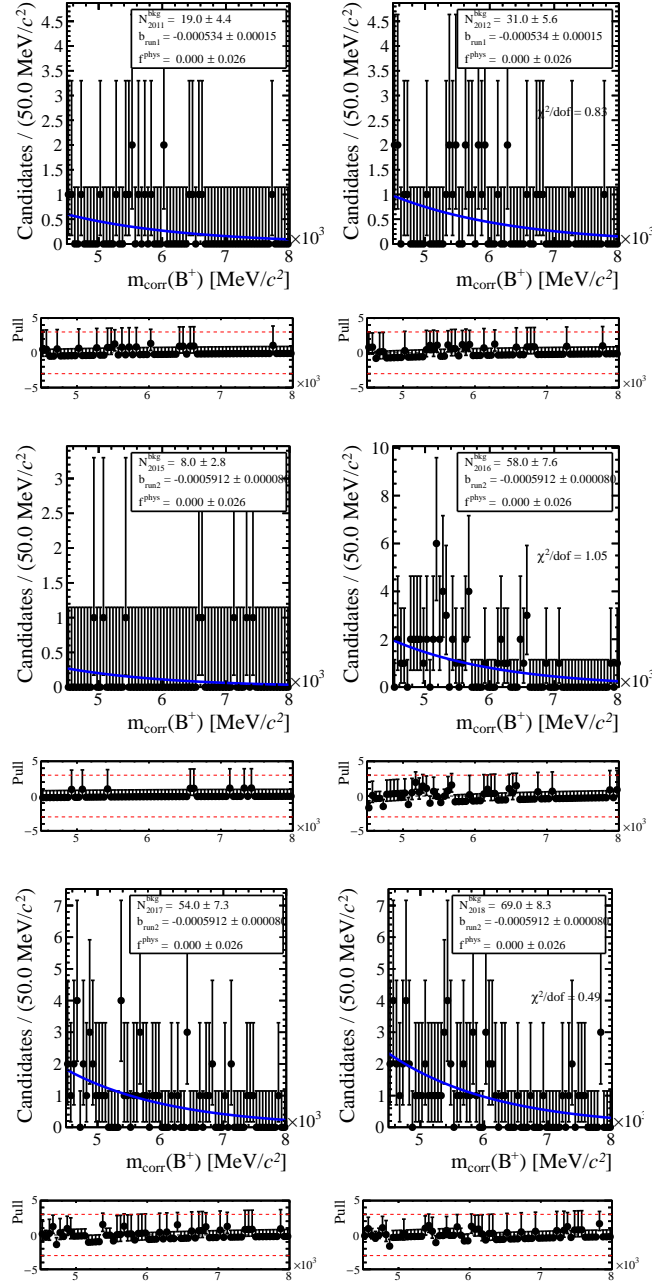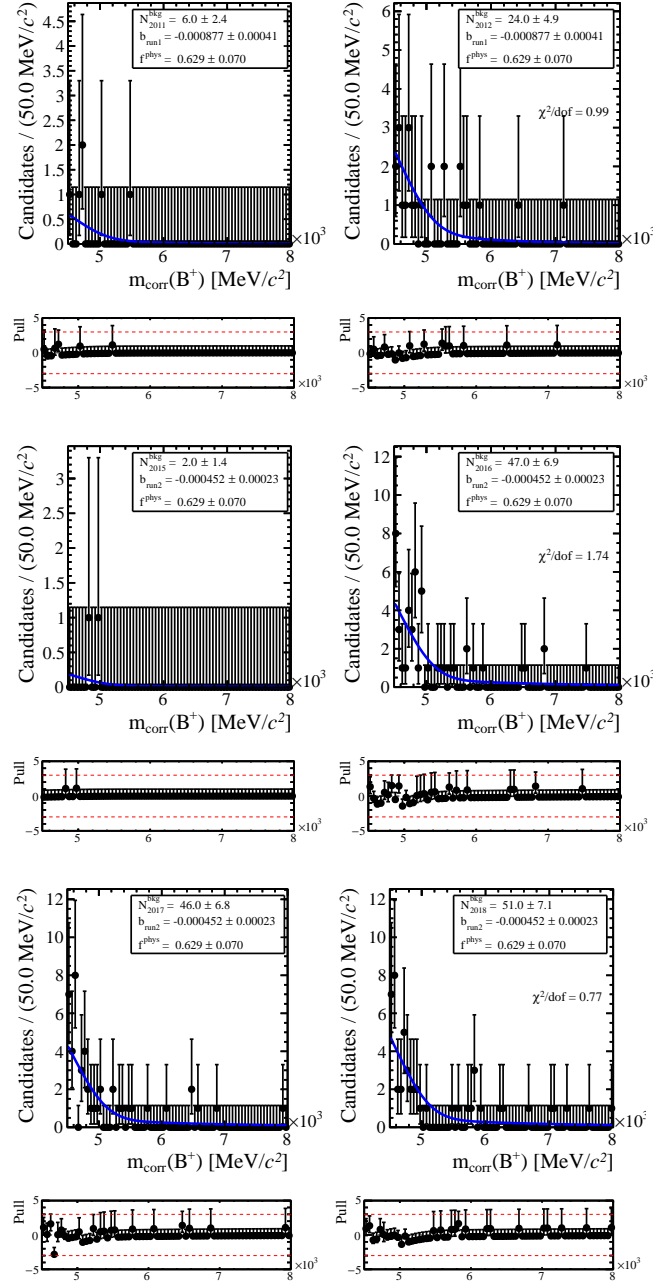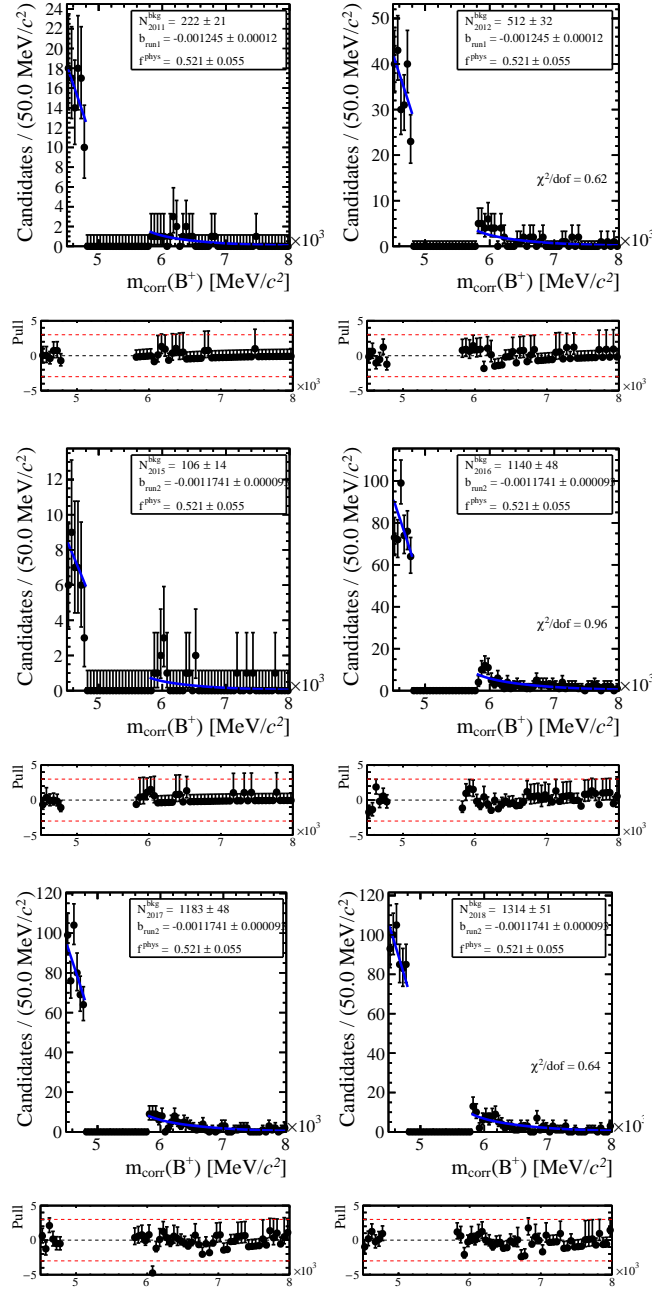Figure 5.6 – Simultaneous Run 1 and Run 2 background mass fit of the $B^+ \to K^+ \tau^+ \mu^-$ blinded data samples after full selection. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively.
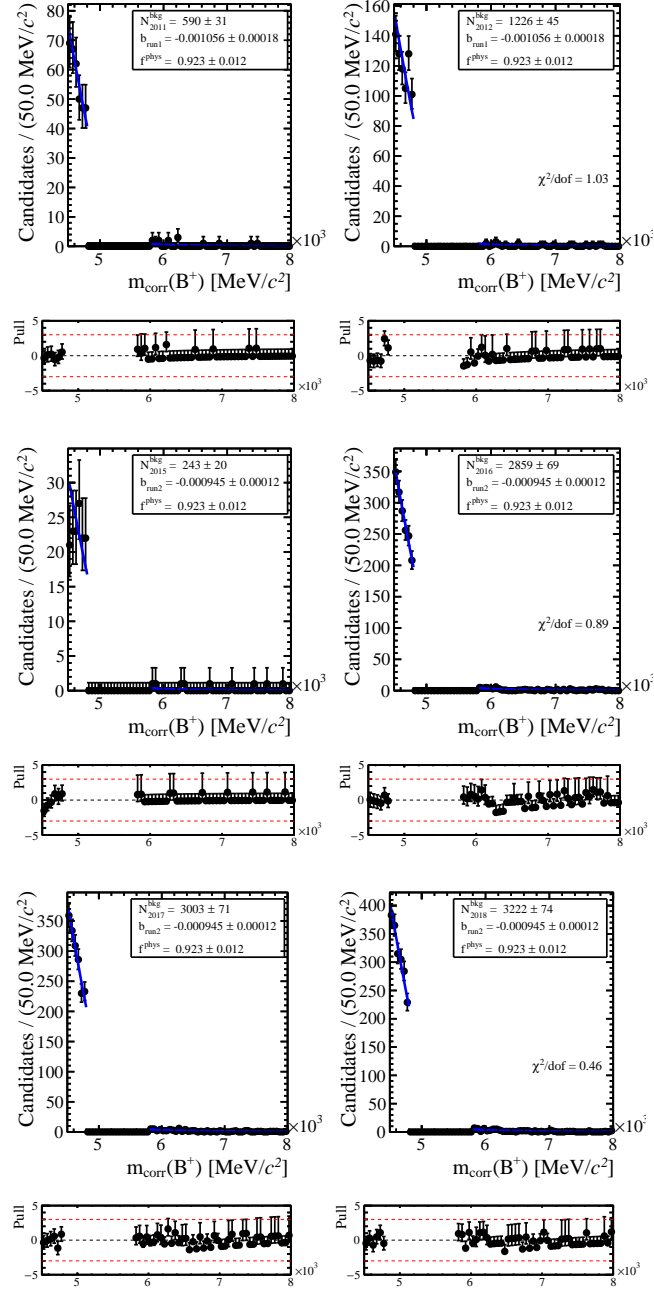
5.1.2 $B^+ \rightarrow \bar{D}^0 D_s^+$ decay mode

In order to determine the signal mass shape of the $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)D_s^+(\rightarrow K^+K^-\pi^+)$ candidates, fits to the corresponding signal simulated samples are performed. The model used in the fit is given by the sum of two CB function with power law tails on opposite sides. The mean parameter is shared between both CB functions and the order of the power law is fixed to 2 for both pdf. The resulting functional form is the following:

$$f_{sig} = N_{sig}\Big(f \cdot \mathscr{CB}\big(m(B^+); m, s, \alpha, n = 2\big) + (1 - f) \cdot \mathscr{CB}\big(m(B^+); m, s_{\text{ratio}} \cdot s, \alpha_2, n_2 = 2\big)\Big), \quad (5.6)$$

where N_{sig} is the total signal yield, f is the fraction of the signal yield associated with the CB function with the low-mass sideband tail, m and s are the mean and width parameters, s_{ratio} is the ratio between the widths associated with the high-mass and low-mass CB functions and finally, n and n_2 are the power of the power law tails.

When fitting the real data, besides describing the signal component it is necessary to describe the various background components as well. After full selection, the contribution due to the combinatorial background is small and is described simply by a constant. In terms of physics background, there is a sizeable contribution due to the single-charm background decay $B^+ \rightarrow \bar{D}^0 K^+K^-\pi^+$. A specific difficulty due to the presence of this background decay is that its associated invariant mass distribution is very similar, both in terms of shape and location, to the one associated with the signal decay under consideration. As a result, it is not possible to simply add another peaking pdf to describe these decays and leave it free in the fit as the fit would be unable to distinguish between these background candidates and the signal candidates. As such, it is necessary to determine a way to fully constrain its associated pdf. In order to determine the shape of the pdf associated with the $B^+ \rightarrow \bar{D}^0 K^+K^-\pi^+$ candidates, a fit is performed to the B^+ invariant mass associated with candidates which are outside the $1850 < m(K^+K^-\pi^+) < 1890 \,\text{MeV}/c^2$ and $1930 < m(K^+K^-\pi^+) < 2000 \,\text{MeV}/c^2$ mass regions. This choice excludes contributions due to $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)D^+(\rightarrow K^+K^-\pi^+)$ and $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)D_s^+(\rightarrow K^+K^-\pi^+)$ candidates. The resulting $m(B^+)$ distribution is fitted with the sum of a Gaussian function describing the $B^+ \rightarrow \bar{D}^0 K^+K^-\pi^+$ candidates and a decaying exponential describing the combinatorial background. The fit as well as the $m(K^+K^-\pi^+)$ distribution are shown in Fig. 5.7. In the fit to the data, the contribution due to the $B^+ \rightarrow \bar{D}^0 K^+K^-\pi^+$ background decays is described by a Gaussian function whose width is fixed to the value found by fitting the candidates in the $m(K^+K^-\pi^+)$ sidebands while its mean is shared with the signal's. In order to constrain the yield associated with this background, it is assumed that its associated $m(K^+K^-\pi^+)$ distribution is described by a constant polynomial which can then be used to extrapolate the number of candidates in the $1930 < m(K^+K^-\pi^+) < 2000 \,\text{MeV}/c^2$ signal region. As a result, when fitting the $m(B^+)$ distribution the pdf associated with this background is fully constrained except for the mean parameter, which is shared with the
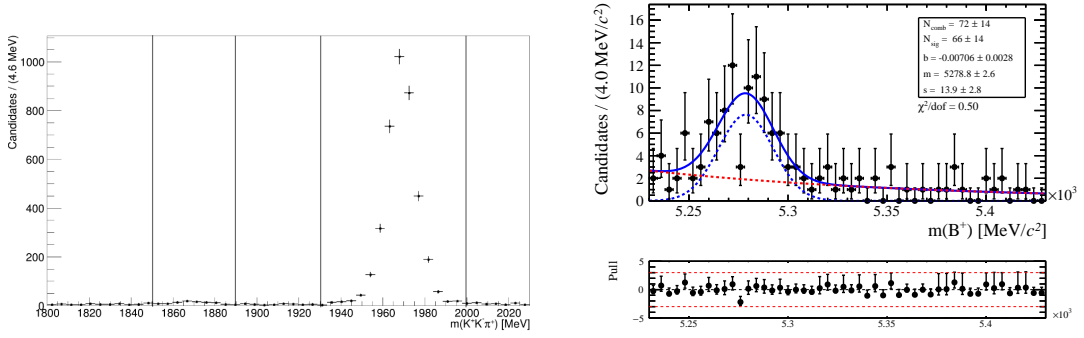
Figure 5.7 – The left plot show the $m(K^+K^-\pi^+)$ distribution for 2016 data. The vertical lines indicate the two mass windows which are rejected when studying the $B^+ \to \overline{D}^0 K^+K^-\pi^+$ background. The right plot shows the fit to the B^+ mass distribution for events outside the D^+ and D_s^+ mass windows. The total pdf is the sum of a Gaussian function and a falling exponential drawn in dashed blue and red, respectively. The full pdf is drawn in solid blue.

signal's. A similar procedure is used to describe the $B^+ \to K^+\pi^- D_s^+$ and $B^+ \to K^+\pi^- K^+K^-\pi^+$ background decays. However, no sizeable contribution from these decay modes is observed, as can be seen in Fig. 5.8. Thus no extra component is added in the fit to the data.

The resulting model used to fit the $B^+ \to \overline{D}^0 D_s^+$ data is the sum of the signal pdf determined by fitting the signal simulation samples, the constrained function to describe the contribution of the $B^+ \to \overline{D}^0 K^+K^-\pi^+$ background decay and a constant to describe the combinatorial background. In the fit the parameters f, s_{ratio}, α and α_2 are fixed to the values found in the fit to the signal simulation samples. The resulting functional form is thus given by :

$$
\begin{aligned}
f_{data} = N_{sig}\Big(& f^{MC} \cdot \mathscr{CB}\big(m(B^+); m, s, \alpha^{MC}, n = 2\big) \\
& + (1 - f^{MC}) \cdot \mathscr{CB}\big(m(B^+); m, s_{\text{ratio}}^{MC} \cdot s, \alpha_2^{MC}, n_2 = 2\big)\Big) \\
& + N_{comb} \cdot \mathscr{H}\big(m(B^+) - 5.23\,\text{GeV/}c^2\big) \cdot \mathscr{H}\big(5.43\,\text{GeV/}c^2 - m(B^+)\big) \\
& + N_{D^0 KK\pi} \cdot \mathscr{G}\big(m(B^+); m, s_{D^0 KK\pi}\big),
\end{aligned}
\tag{5.7}
$$

where N_{sig}, N_{comb} and $N_{D^0 KK\pi}$ are respectively the signal, combinatorial and $B^+ \to \overline{D}^0 K^+K^-\pi^+$ yields, $\mathscr{H}(x)$ is the Heaviside step function, the MC superscript denotes the fact that the parameter is fixed to the value found in the simulation and the $D^0 KK\pi$ subscript denotes that the parameter is fixed to the value found in the $K^+K^-\pi^+$ sideband fit, as explained previously. In general, the momentum scale used in data is slightly inaccurate. For this reason, a correction to that momentum scale in data is applied. The intrinsic uncertainty of this momentum scale is $3 \times 10^{-4}\big(m(B^+) - m(D^0) - m(D_s^+)\big)$ which results in an uncertainty of $0.4\,\text{MeV/}c^2$ on the mean of the signal [87]. As a result, the measured mean value of the $B^+ \to \overline{D}^0 D_s^+$ signal candidates'
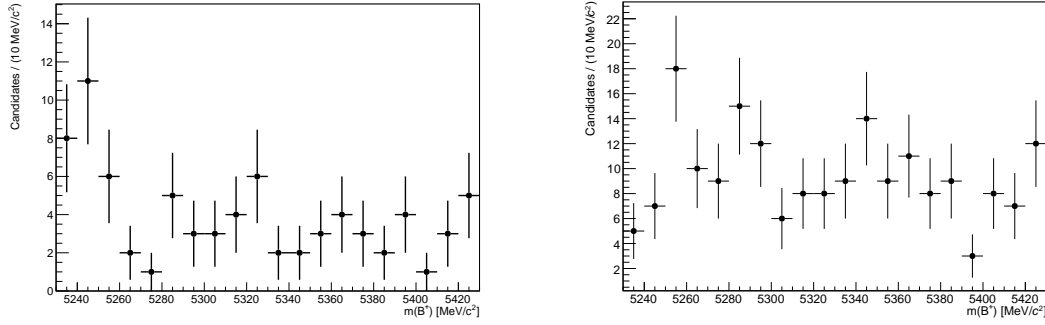
Figure 5.8 – The left plot shows the B^+ mass distribution for candidates outside the D^0 mass window for 2016 data. There is no significant contribution of $B^+ \rightarrow K^+\pi^- D_s^+$ background decays. The right plot shows the B^+ mass distribution for candidates where the value of $m(K^+K^-\pi^+)$ is outside the D^+ and D_s^+ mass windows, as defined in Figure 5.7, and the value of $m(K^+\pi^-)$ is outside the D^0 mass window. The plot is made using the combined Run 2 data. Once again, there is no significant contribution of charmless $B^+ \rightarrow K^+\pi^- K^+K^-\pi^+$ background decays.

distribution is consistent within uncertainties with the world-average value of of the mass of the B^+ meson which is $5279.34 \pm 0.12\,\mathrm{MeV}/c^2$ [33]. The mass fits of the signal simulation and real data samples are shown in Fig. 5.9 where it can be noticed that the resolution for the simulation and data agree.
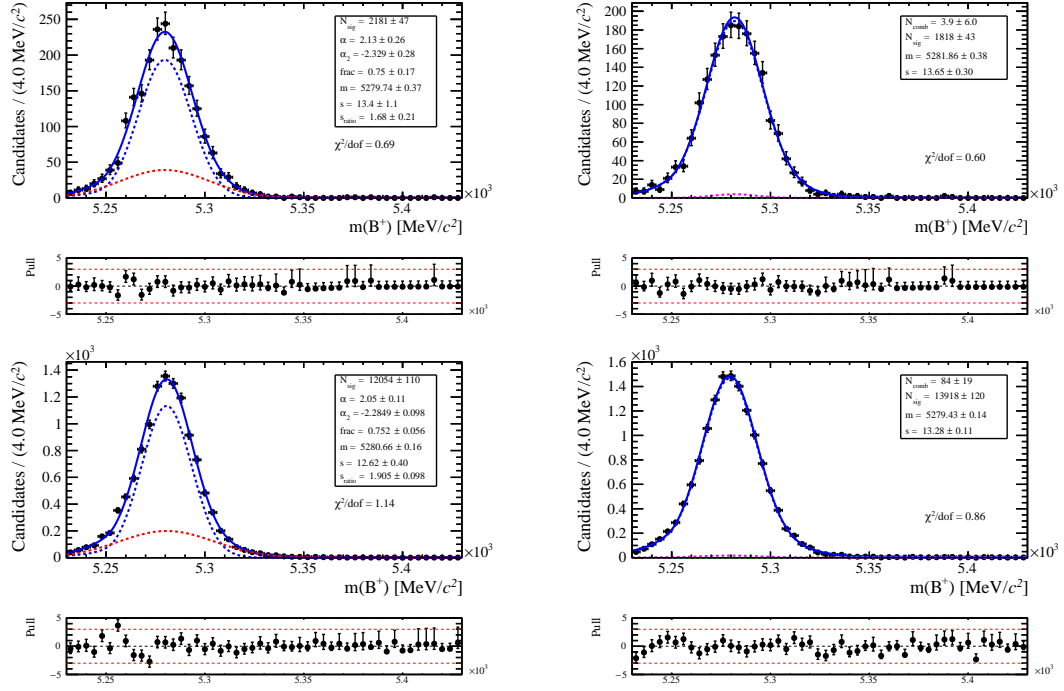
Figure 5.9 – Mass fit of the $B^+ \to \overline{D}^0 D_s^+$ decay mode for the Run 1 samples in the top row and the Run 2 samples in the bottom row. The left plots show signal simulation while the right plots show the fit to the data. The solid blue curve is the total pdf while the dashed curves are the individual components of the fit. In the case of the simulated samples, the dashed curves represent the two CB functions used to construct the total signal pdf while in the data samples the blue, green and purple components describe the total signal pdf, the combinatorial background contribution and the $B^+ \to \overline{D}^0 K^+ K^- \pi^+$ background component, respectively.

5.2 Maximum likelihood fit and limit setting

The final measurement of $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$ and $\mathscr{B}(B^+ \to K^+ \tau^+ \mu^-)$ is performed through the use of an unbinned extended maximum likelihood fit to the $m_{\text{corr}}(B^+)$ distribution. In the scenario where the measured values are compatible with zero, upper limits will be set using the CL_{s} method [88]. For this reason, the 3D requirements on the BDT outputs and the combined PID variable $\mathscr{P}_3$ are optimised based on the expected limits on the signal branching fractions. For a given year the signal branching fraction, *e.g.* $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$, is linked to the fitted $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to \overline{D}^0 D_s^+$ yields as follows

$$
\begin{aligned}
N_s = {} & \mathscr{B}(B^+ \to K^+ \tau^- \mu^+) \\
& \times \frac{\mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau)\varepsilon_{3\pi} + \mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau)\varepsilon_{3\pi\pi^0}}{\mathscr{B}(B^+ \to \overline{D}^0 D_s^+)\mathscr{B}(\overline{D}^0 \to K^+ \pi^-)\mathscr{B}(D_s^+ \to K^+ K^- \pi^+)} \\
& \times \frac{N_n}{\varepsilon_n},
\end{aligned}
\tag{5.8}
$$

where N_s and N_n are the signal and normalisation yields respectively, $\varepsilon_{3\pi}$, $\varepsilon_{3\pi\pi^0}$ and ε_n are the full selection efficiencies for the $B^+ \to K^+ \tau^- \mu^+$ decay where $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ and for the $B^+ \to \overline{D}^0 D_s^+$ decay mode, respectively. In the fit, the model describing the signal is given by

$$
\mathscr{M}_s = N_s\big(\mathscr{F}_{3\pi} f_{3\pi}(m_{\text{corr}}; \hat{\boldsymbol{\theta}}_{3\pi}) + (1 - \mathscr{F}_{3\pi}) f_{3\pi\pi^0}(m_{\text{corr}}; \hat{\boldsymbol{\theta}}_{3\pi\pi^0})\big),
\tag{5.9}
$$

where $f_{3\pi}$ and $f_{3\pi\pi^0}$ are respectively the pdf used to fit the signal $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes, $\hat{\boldsymbol{\theta}}_{3\pi}$ and $\hat{\boldsymbol{\theta}}_{3\pi\pi^0}$ are the maximum likelihood estimators of the shape parameters of these pdfs obtained in the fit to the simulation samples. Finally, $\mathscr{F}_{3\pi}$ is the fraction of the total signal yield given from the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ component,

$$
\mathscr{F}_{3\pi} = \frac{\mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau)\varepsilon_{3\pi}}{\mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau)\varepsilon_{3\pi} + \mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau)\varepsilon_{3\pi\pi^0}}.
\tag{5.10}
$$

The background model is given by

$$
\mathscr{M}_b = N_b f_b(m_{\text{corr}}; \boldsymbol{\theta}_b),
\tag{5.11}
$$

where N_b is the background yield, f_b is the background pdf and $\boldsymbol{\theta}_b$ are the shape parameters associated with f_b. The complete model used in the fit is then naturally given by

$$
\mathscr{M}_{s+b} = \mathscr{M}_s + \mathscr{M}_b.
\tag{5.12}
$$

The three pdfs $f_{3\pi}$, $f_{3\pi\pi^0}$ and f_b are described in Section 5.1.

When fitting the data, the free parameters are $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$, N_b and the background shape parameters $\boldsymbol{\theta}_b$. From the point of view of the limit setting procedure, the parameter of interest is $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$, while all the other parameters are considered as nuisances. In order to determine an expected limit on $\mathscr{B}(B^+ \to K^+ \tau^- \mu^+)$, a pseudo-dataset is generated according to $\mathscr{M}_b(m_{corr}; \hat{\boldsymbol{\theta}}_b)$, where $\hat{\boldsymbol{\theta}}_b$ are the maximum likelihood estimators for the shape parameters associated with f_b, obtained in the fit to the blind Run 1 and Run 2 datasets. The number of expected events over the full mass range is extrapolated from the background yield obtained by fitting the sidebands of the blind data samples. The number of events generated in the pseudo-dataset then follows a Poisson distribution whose mean is this expected number of events over the full mass range. The pseudo-dataset is then fitted with $\mathscr{M}_{s+b}$ and $\mathscr{M}_b$ in order to obtain the one-sided profile likelihood ratio test statistic for the signal+background and background-only hypotheses. The profile likelihood ratio is obtained by using the asymptotic approximation [89]. The ratio of these two profile likelihood ratios provide the value of CL_s from which the upper limit on the signal branching fraction can be computed for the desired confidence level (CL).

The extension of the fit model to the full dataset when performing the simultaneous fit to all years is straightforward. The total signal yield can be simply expressed as the sum of the individual signal yields for every year, which can be generalised from the previous equations as

$$
\begin{aligned}
N_{tot,s} &= \sum_i N_{i,s} \\
&= \mathscr{B}(B^+ \to K^+ \tau^- \mu^+) \\
&\times \sum_i \frac{\mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau)\epsilon_i^{3\pi} + \mathscr{B}(\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau)\epsilon_i^{3\pi\pi^0}}{\mathscr{B}(B^+ \to \overline{D}^0 D_s^+)\mathscr{B}(\overline{D}^0 \to K^+ \pi^-)\mathscr{B}(D_s^+ \to K^+ K^- \pi^+)} \frac{N_{i,n}}{\epsilon_{i,n}},
\end{aligned}
\tag{5.13}
$$

where $N_{tot,s}$ is the total signal yield, the index i corresponds to a given year and runs in the sum over all years, *i.e.* 2011, 2012, 2015, 2016, 2017 and 2018. The multiple branching fractions needed in the computation are given in Tab. 5.1 while the various selection efficiencies have been discussed in Sec. 4.7.

5.3 BDT and PID requirement optimisation

The selection requirements imposed on the outputs of the two BDT and the combined PID variable, *i.e.* BDT_{comb}, BDT_{phys} and $\mathscr{P}_3$, are optimised under the hypothesis of an absence of significant signal. The optimisation is performed simultaneously on the three variables. For a given triplet of values $(x; y; z) \in [0, 1] \times [0, 1] \times [0, 1]$, the requirements $\text{BDT}_{\text{comb}} > x$, $\text{BDT}_{\text{phys}} > y$ and $\mathscr{P}_3 > z$ are imposed and the corresponding expected limit at 90% CL is computed using

Table 5.1 – Branching fractions used in the obtention of the limits on $\mathscr{B}(B^+ \to K^+\tau^-\mu^+)$ and $\mathscr{B}(B^+ \to K^+\tau^+\mu^-)$. The values are taken from Ref. [33].

Decay	Branching fraction
$\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$	$(9.31 \pm 0.05)\%$
$\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$	$(4.62 \pm 0.05)\%$
$B^+ \to \overline{D}^0 D_s^+$	$(9.0 \pm 0.9) \times 10^{-3}$
$\overline{D}^0 \to K^+\pi^-$	$(3.950 \pm 0.031)\%$
$D_s^+ \to K^+K^-\pi^+$	$(5.39 \pm 0.15)\%$

the procedure detailed in Sec. 5.2. This procedure is repeated for several sets of requirement values and the optimal set of requirements is the one providing the lowest expected limit. The optimal sets of requirements after optimising on the full Run 1 + Run 2 data are $(0.80, 0.80, 0.50)$ and $(0.75, 0.70, 0.60)$ for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes, respectively. The expected upper limits corresponding to these requirements are

$$\mathscr{B}(B^+ \to K^+\tau^-\mu^+) < 1.8(2.2) \times 10^{-6} \text{ at } 90(5)\% \text{ CL (expected)},$$
$$\mathscr{B}(B^+ \to K^+\tau^+\mu^-) < 2.3(2.8) \times 10^{-6} \text{ at } 90(5)\% \text{ CL (expected)}.$$

The CL_s scans from which these upper limits are derived are shown in Fig. 5.10. The optimisation is performed in three dimensions, however for illustration purposes the three-dimensional scans are projected to two-dimensional scans on two of the three variables corresponding to the optimal requirement on the third variable. These two-dimensional projections are shown in Fig. 5.11. The mass fits corresponding the optimal requirements are shown in Figs. 5.1, 5.2 for simulated signal data and 5.5, 5.6 for the blinded data sets. As explained in Sec. 5.2, these fits to the blind data are then used to produce pseudo-datasets from which the expected limit can be computed. The specific pseudo-datasets used in the computation of the limit for the optimal requirements are shown in Figs. 5.12 and 5.13. Naturally, once that the effect of the systematic uncertainties, discussed in Sec. 5.4, are included in the fit the expected limits will deteriorate slightly.
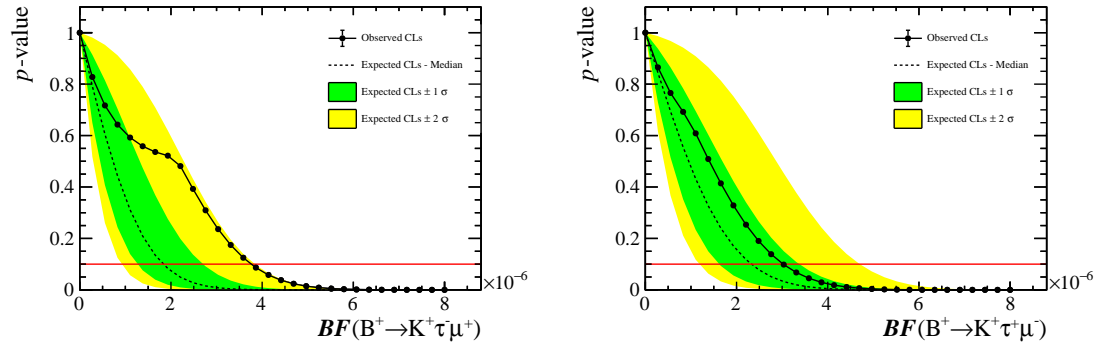
Figure 5.10 – CL$_s$ scans associated with the best 3D requirements for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes on the left and right plots, respectively. The red line indicates the 90% CL.
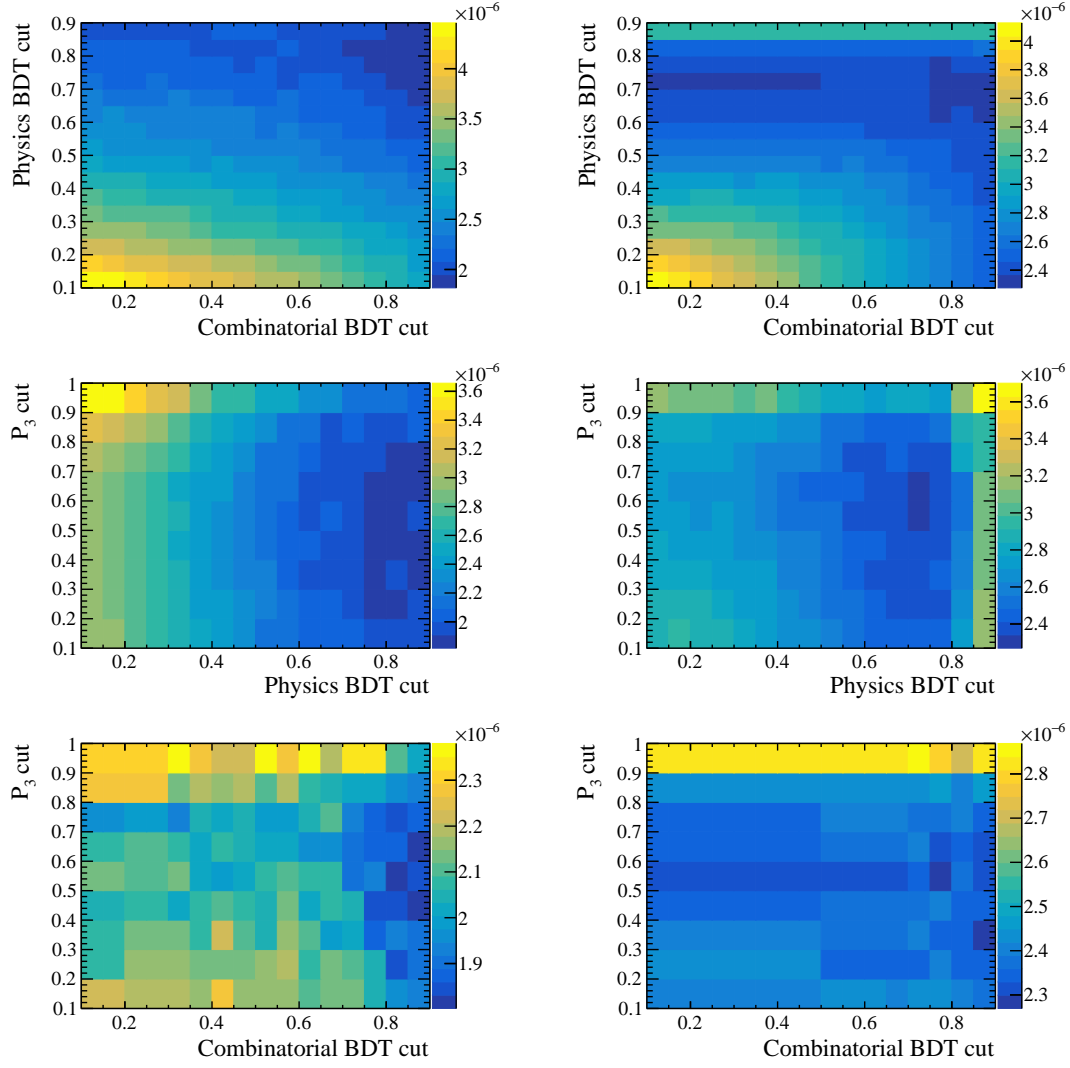
Figure 5.11 – The top, middle and bottom rows show respectively the expected limit as a function of the requirements imposed on $(\mathrm{BDT}_{\mathrm{comb}}; \mathrm{BDT}_{\mathrm{phys}})$, $(\mathrm{BDT}_{\mathrm{phys}}; \mathscr{P}_3)$ and $(\mathrm{BDT}_{\mathrm{comb}}; \mathscr{P}_3)$ in the x-axis and y-axis. The plots on the left and right columns are for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively. The colour scale indicates the value of the expected limit at 90% CL.

Figure 5.12 – Pseudo-datasets used to compute the expected limit on $\mathscr{B}(B^+ \to K^+\tau^-\mu^+)$ after applying the optimal 3D requirements. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total pdf. The dashed blue and red lines are the background and signal components, respectively.
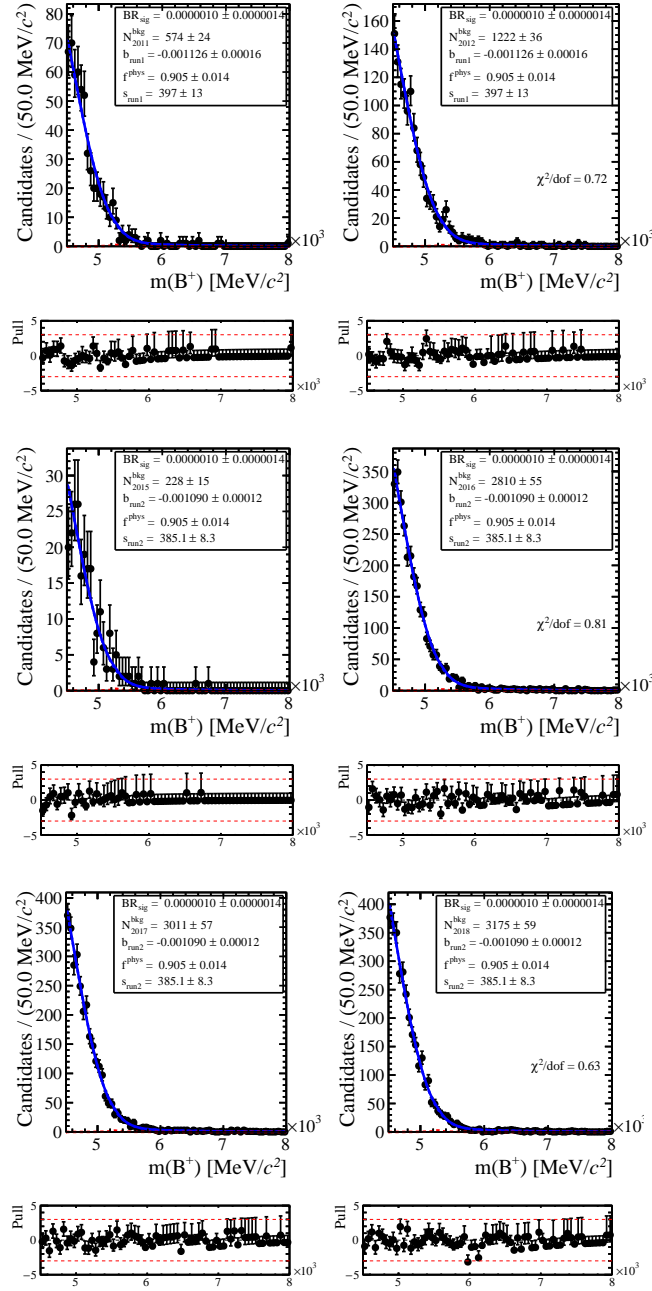
Figure 5.13 – Pseudo-datasets used to compute the expected limit on $\mathscr{B}(B^+ \to K^+ \tau^+ \mu^-)$ after applying the optimal 3D requirements. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total pdf. The dashed blue and red lines are the background and signal components, respectively.

5.4   Systematic uncertainties and checks

As is the case for any experimental measurement, various sources of systematic uncertainty affect the final result. In particular, in the search for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes the sources of systematic uncertainty can be split into two main categories, depending on whether they impact the measurement of selection efficiencies or yields. The methods used to evaluate the uncertainties differ depending on the category the source of uncertainty belongs to. In the following are presented various sources of systematic uncertainty impacting the final result and strategies to evaluate them.

5.4.1   Efficiencies

In order to treat the systematic uncertainties in a consistent way, the efficiencies associated with similar types of variables are grouped together. Additionally, the ordering of the selection steps can have an influence when evaluating systematic uncertainties. As previously mentioned, the full selection efficiency is expressed as follows

$$\epsilon_s = \epsilon_{gen} \times \epsilon_{reco+strip} \times \epsilon_{presel} \times \epsilon_{BDT} \times \epsilon_{PID} \times \epsilon_{trigger} \times \epsilon_{res}, \tag{5.14}$$

where ϵ_s, ϵ_{gen}, $\epsilon_{reco+strip}$, ϵ_{presel}, ϵ_{BDT}, ϵ_{PID}, $\epsilon_{trigger}$ and ϵ_{res} are the efficiencies associated with the requirements of the full selection for the signal channel, the generator level, the reconstruction and stripping without PID requirement, the additional preselection, the BDT, the PID, the trigger and the additional mass requirements aimed at removing resonant sub-decays, respectively. For the $B^+ \to \overline{D}^0 D_s^+$ decay mode, the structure is identical except for the fact that neither the ϵ_{BDT} term is present, since there is no BDT selection applied on that channel, nor the ϵ_{res} term. The trigger efficiency itself is further split into the contributions due to the requirements applied on the hardware and software stages, *i.e.* $\epsilon_{trigger} = \epsilon_{L0} \times \epsilon_{HLT}$. Performing this separation is useful in handling the different sources of systematic uncertainty affecting the trigger selection.
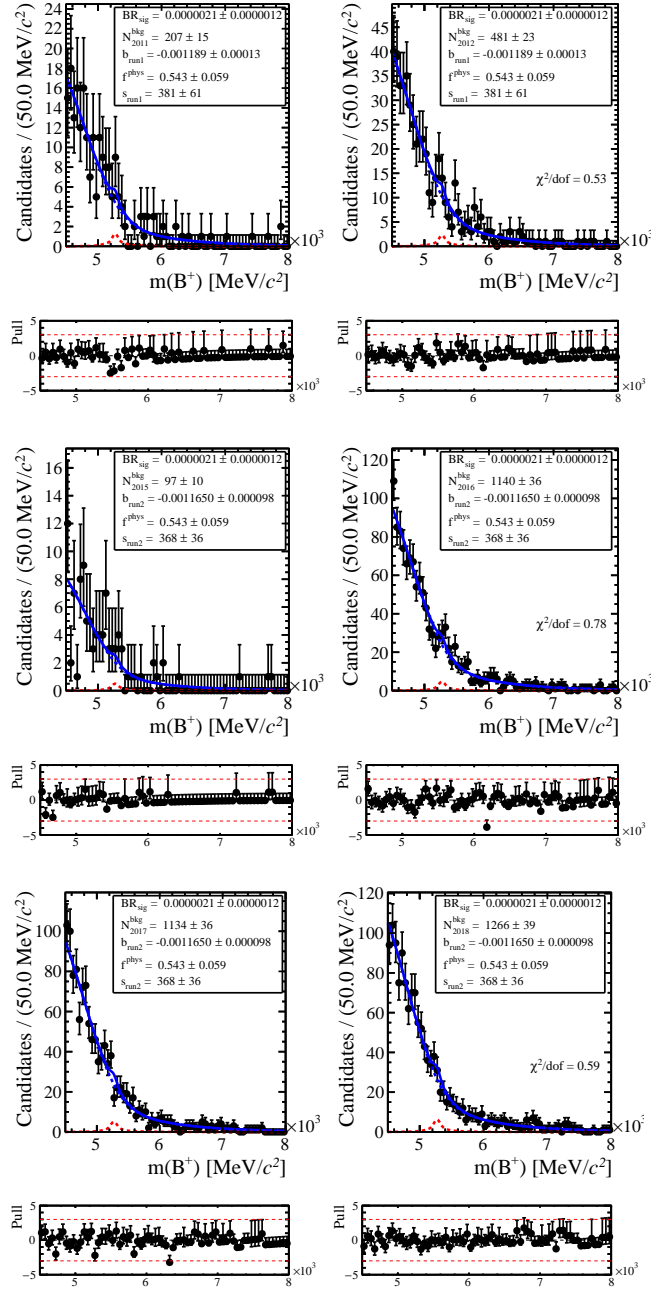
As a general rule, the systematic uncertainty assigned to a given efficiency is given by

$$\sigma_{sys} = \sqrt{(\epsilon_{nom} - \epsilon_{alt})^2 + |\sigma_{nom}^2 - \sigma_{alt}^2|}, \tag{5.15}$$

where ϵ_{nom} and σ_{nom} are the efficiency and its uncertainty computed using the nominal analysis procedure while ϵ_{alt} and σ_{alt} are the alternative efficiency and its uncertainty computed using an alternative method. In principle, it could be possible to simply assign the difference between the central values of the nominal and alternative efficiencies as the systematic uncertainty. However, the majority of the systematic uncertainties are evaluated by using pseudo-datasets where the uncertainty on ε_{alt} is correlated with ε_{nom}. Using Eq. (5.15) allows

to account for these correlations.

The selection efficiencies are nominally measured by using simulation. If the simulation describes accurately a given distribution then the selection efficiency associated with a requirement imposed on that variable will be correctly measured and as such no systematic uncertainty needs to be assigned on that particular efficiency. Another case scenario where there is no need to assign a systematic uncertainty is if there are discrepancies between simulation and data but they are expected to cancel when taking the ratio of signal and normalisation efficiencies, as is the case when measuring the signal branching fractions. The simulation can be reliably used to measure the $\varepsilon_{gen}, \varepsilon_{reco+strip}, \varepsilon_{presel}$ and ε_{res} efficiencies and no systematic uncertainties are assigned to these terms.

BDT

In order for the efficiency associated with a requirement imposed on the output of a BDT to be accurate, it is necessary that all the variables used as input to the BDT are well described in simulation. However, the 2017 and 2018 data samples are affected by a known issue where the error estimates for the VELO sub-detector measurements are different in data and simulation, which has a small impact on the χ^2_{IP} distributions. Since the $\chi^2_{\mathrm{IP}}(K^+)$ is used as input to both BDT, the efficiency measurement using the simulation is inaccurate. Thus a systematic uncertainty needs to be assigned to the BDT efficiencies for the 2017 and 2018 data samples. In order to determine the systematic uncertainty, it is necessary to evaluate the efficiency using an alternative method. This is done by applying a correction procedure to the nominal $\chi^2_{\mathrm{IP}}(K^+)$ distribution and evaluating the BDT using this corrected distribution instead of the nominal one. The alternative efficiency is then measured for the requirements imposed on these new BDT output distributions and the systematic uncertainty is assigned according to Eq. (5.15).

The procedure used to correct the $\chi^2_{\mathrm{IP}}(K^+)$ distribution is analogous to the one used to correct the n_{tracks} distribution explained in Sec. 4.5. The corrected $\chi^2_{\mathrm{IP}}(K^+)$ distribution is obtained by multiplying the $\chi^2_{\mathrm{IP}}(K^+)$ value of each candidate by a constant scale factor. This scale factor is determined by comparing the $\chi^2_{\mathrm{IP}}(K^+)$ distribution, for the kaon candidate from the $\overline{D}^0$ decay, between data and simulation using the $B^+ \to \overline{D}^0 D_s^+$ decay mode and minimising a χ^2 function analogous to the one described in Eq. (4.2). The variation of this χ^2 function for various values of the scale factor is shown in Fig. 5.14. The systematic uncertainties associated with the requirements imposed on the BDT outputs, *i.e.* $\mathrm{BDT_{comb}} > 0.8$, $\mathrm{BDT_{phys}} > 0.8$ and $\mathrm{BDT_{comb}} > 0.75$, $\mathrm{BDT_{phys}} > 0.7$ for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes respectively, are given in Tab. 5.2.

Concerning the other variables used as input to the BDT, the $B^+ \to \overline{D}^0 D_s^+$ decay mode was
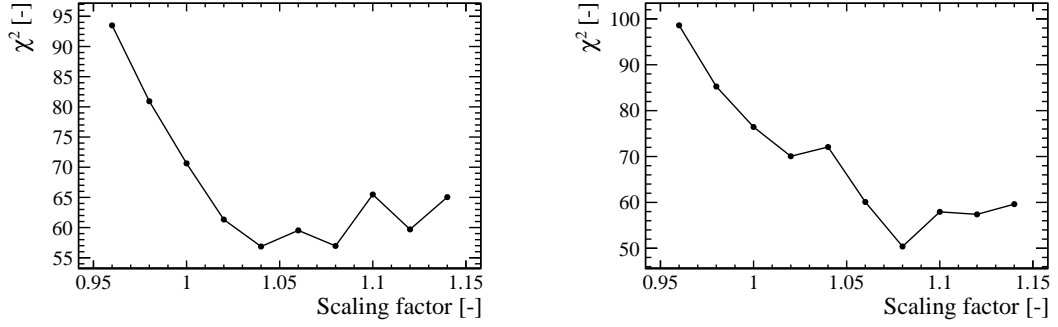
Figure 5.14 – The left and right plots show the χ^2 variable, used to correct the $\chi^2_{\text{IP}}(K^+)$ distribution, as a function of the scale factor w for 2017 and 2018 data, respectively.

Table 5.2 – The second and third columns give the absolute systematic uncertainty associated with the optimal requirements on the output of the BDT, *i.e.* $\text{BDT}_{\text{comb}} > 0.8$, $\text{BDT}_{\text{phys}} > 0.8$ and $\text{BDT}_{\text{comb}} > 0.75$, $\text{BDT}_{\text{phys}} > 0.7$ for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes respectively, for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes at the top and bottom.

Year	$B^+ \to K^+ \tau^- \mu^+$	$B^+ \to K^+ \tau^+ \mu^-$
2017	0.03×10^{-1}	0.03×10^{-1}
2018	0.06×10^{-1}	0.06×10^{-1}
2017	0.03×10^{-1}	0.02×10^{-1}
2018	0.04×10^{-1}	0.05×10^{-1}

used to check for potential discrepancies between data and simulation for analogous variables, as explained in Sec. 4.4. Overall, the simulation describes well the data for all tested variables. The comparisons between data and simulation, using Run 2 data, for these various variables are shown in Figs. 4.19 and 4.20.

PID

The nominal efficiencies associated with the PID requirements are computed using the corrected PID distributions. However, there are multiple sources of uncertainty in the PIDGen correction procedure whose effects naturally need to be reflected on the efficiencies. The three main sources of systematic uncertainty in the PIDGen correction procedure are due to the parametrisation of weighted PID control samples based on the kernel density estimator, the limited statistics of the control samples and finally the use of the *sPlot* technique [84] in order to isolate the signal component in the calibration data. For the latter source, an absolute systematic uncertainty of 0.2% on the PID efficiency is assigned. In order to asses the systematic uncertainty due to the parametrisation of the control samples, new corrected PID distributions are determined by increasing the kernel width by 50% in the correction procedure. These alternative PID distributions are then used to compute the efficiencies associated with the PID requirements and an associated systematic uncertainty σ_{PID}^{KDE} is calculated based on Eq. (5.15). The associated systematic uncertainties are given in Tab. 5.3 for the $B^+ \rightarrow \overline{D}^0 D_s^+$, $B^+ \rightarrow K^+ \tau^- \mu^+$ and $B^+ \rightarrow K^+ \tau^+ \mu^-$ decay modes. Concerning the systematic uncertainty due to the limited statistics of the calibration samples, five alternative bootstrapped calibration samples are used to produce five alternative corrected PID distributions from which alternative efficiencies $\epsilon_{PID}(stat_0) \ldots \epsilon_{PID}(stat_4)$ can be computed. For each of these alternative efficiencies an associated systematic uncertainty is computed and the largest systematic uncertainty among them is selected as the uncertainty σ_{PID}^{stat} associated with the limited statistics and is given in Tab. 5.4 for the $B^+ \rightarrow \overline{D}^0 D_s^+$, $B^+ \rightarrow K^+ \tau^- \mu^+$ and $B^+ \rightarrow K^+ \tau^+ \mu^-$ decay modes. The total PID systematic uncertainty is then taken as the sum in quadrature of the systematic uncertainties associated with the three different sources, *i.e.* $\sigma_{PID}(syst) = \sigma_{PID}^{KDE} \oplus \sigma_{PID}^{stat} \oplus 0.2\%$ and is given respectively for the $B^+ \rightarrow \overline{D}^0 D_s^+$, $B^+ \rightarrow K^+ \tau^- \mu^+$ and $B^+ \rightarrow K^+ \tau^+ \mu^-$ decay modes in Tabs. 5.5, 5.6 and 5.7.

Table 5.3 – Systematic uncertainty σ_{PID}^{KDE} associated with the variation of the kernel's width used in the kernel density estimation procedure to correct the PID distributions. The values are shown in the second, third and fourth columns for the $B^+ \to \overline{D}^0 D_s^+$, $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively. The top and bottom parts of the table refer respectively to the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay modes.

Year	$B^+ \to \overline{D}^0 D_s^+$	$B^+ \to K^+ \tau^- \mu^+$	$B^+ \to K^+ \tau^+ \mu^-$
2011	0.07×10^{-1}	0.03×10^{-1}	0.14×10^{-1}
2012	0.11×10^{-1}	0.01×10^{-1}	0.13×10^{-1}
2015	0.04×10^{-1}	0.29×10^{-1}	0.32×10^{-1}
2016	0.02×10^{-1}	0.05×10^{-1}	0.02×10^{-1}
2017	0.03×10^{-1}	0.18×10^{-1}	0.10×10^{-1}
2018	0.01×10^{-1}	0.00×10^{-1}	0.14×10^{-1}
2011	-	0.01×10^{-1}	0.20×10^{-1}
2012	-	0.15×10^{-1}	0.06×10^{-1}
2015	-	0.04×10^{-1}	0.12×10^{-1}
2016	-	0.16×10^{-1}	0.05×10^{-1}
2017	-	0.09×10^{-1}	0.23×10^{-1}
2018	-	0.07×10^{-1}	0.03×10^{-1}

Table 5.4 – Systematic uncertainty σ_{PID}^{stat} associated with the bootstrapped samples used to quantify the effect of the limited statistics of the calibration samples on the procedure to correct the PID distributions. The values are shown in the second, third and fourth columns for the $B^+ \to \overline{D}^0 D_s^+$, $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively. The top and bottom parts of the table refer respectively to the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ decay modes.

Year	$B^+ \to \overline{D}^0 D_s^+$	$B^+ \to K^+ \tau^- \mu^+$	$B^+ \to K^+ \tau^+ \mu^-$
2011	0.09×10^{-1}	0.16×10^{-1}	0.14×10^{-1}
2012	0.05×10^{-1}	0.05×10^{-1}	0.11×10^{-1}
2015	0.14×10^{-1}	0.24×10^{-1}	0.45×10^{-1}
2016	0.06×10^{-1}	0.18×10^{-1}	0.10×10^{-1}
2017	0.04×10^{-1}	0.33×10^{-1}	0.14×10^{-1}
2018	0.08×10^{-1}	0.07×10^{-1}	0.16×10^{-1}
2011	-	0.26×10^{-1}	0.15×10^{-1}
2012	-	0.15×10^{-1}	0.31×10^{-1}
2015	-	0.16×10^{-1}	0.40×10^{-1}
2016	-	0.29×10^{-1}	0.29×10^{-1}
2017	-	0.12×10^{-1}	0.22×10^{-1}
2018	-	0.25×10^{-1}	0.35×10^{-1}

Table 5.5 – Combined systematic uncertainty $\sigma_{PID}(syst)$ associated with the PID requirements for the $B^+ \to \bar{D}^0 D_s^+$ decay mode, expressed absolutely and relatively in the third and fourth columns respectively. The second column gives the efficiency ϵ_{PID}.

Year	ϵ_{PID}^n	$\sigma_{PID}^n(syst)$	$\sigma_{PID}^n(syst)$ relative
2011	$(3.15 \pm 0.05) \times 10^{-1}$	0.12×10^{-1}	3.67%
2012	$(3.10 \pm 0.04) \times 10^{-1}$	0.12×10^{-1}	3.93%
2015	$(5.65 \pm 0.07) \times 10^{-1}$	0.15×10^{-1}	2.65%
2016	$(5.71 \pm 0.04) \times 10^{-1}$	0.06×10^{-1}	1.11%
2017	$(5.87 \pm 0.04) \times 10^{-1}$	0.05×10^{-1}	0.89%
2018	$(5.85 \pm 0.04) \times 10^{-1}$	0.08×10^{-1}	1.45%

Table 5.6 – Combined systematic uncertainty $\sigma_{PID}(syst)$ associated with the PID requirements for the $B^+ \to K^+ \tau^- \mu^+$ decay mode, for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ at the top and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay mode at the bottom, expressed absolutely and relatively in the third and fourth columns respectively. The second column gives the efficiency ϵ_{PID}.

Year	ϵ_{PID}^s	$\sigma_{PID}^s(syst)$	$\sigma_{PID}^s(syst)$ relative
2011	$(4.57 \pm 0.09) \times 10^{-1}$	0.16×10^{-1}	3.48%
2012	$(4.48 \pm 0.07) \times 10^{-1}$	0.05×10^{-1}	1.16%
2015	$(5.77 \pm 0.13) \times 10^{-1}$	0.38×10^{-1}	6.59%
2016	$(5.54 \pm 0.07) \times 10^{-1}$	0.19×10^{-1}	3.35%
2017	$(6.45 \pm 0.07) \times 10^{-1}$	0.38×10^{-1}	5.82%
2018	$(6.42 \pm 0.07) \times 10^{-1}$	0.07×10^{-1}	1.10%
2011	$(4.65 \pm 0.13) \times 10^{-1}$	0.26×10^{-1}	5.56%
2012	$(4.51 \pm 0.10) \times 10^{-1}$	0.21×10^{-1}	4.70%
2015	$(6.02 \pm 0.18) \times 10^{-1}$	0.17×10^{-1}	2.76%
2016	$(5.51 \pm 0.10) \times 10^{-1}$	0.33×10^{-1}	6.02%
2017	$(6.51 \pm 0.11) \times 10^{-1}$	0.15×10^{-1}	2.33%
2018	$(6.55 \pm 0.11) \times 10^{-1}$	0.26×10^{-1}	3.98%

Table 5.7 – Combined systematic uncertainty $\sigma_{PID}(syst)$ associated with the PID requirements for the $B^+ \to K^+ \tau^+ \mu^-$ decay mode, for the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ at the top and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay mode at the bottom, expressed absolutely and relatively in the third and fourth columns respectively. The second column gives the efficiency ϵ_{PID}.

Year	ϵ_{PID}^s	$\sigma_{PID}^s(syst)$	$\sigma_{PID}^s(syst)$ relative
2011	$(3.87 \pm 0.09) \times 10^{-1}$	0.20×10^{-1}	5.11%
2012	$(3.79 \pm 0.07) \times 10^{-1}$	0.17×10^{-1}	4.54%
2015	$(5.92 \pm 0.13) \times 10^{-1}$	0.55×10^{-1}	9.28%
2016	$(5.41 \pm 0.07) \times 10^{-1}$	0.10×10^{-1}	1.91%
2017	$(6.25 \pm 0.08) \times 10^{-1}$	0.17×10^{-1}	2.75%
2018	$(6.06 \pm 0.07) \times 10^{-1}$	0.21×10^{-1}	3.49%
2011	$(3.88 \pm 0.10) \times 10^{-1}$	0.25×10^{-1}	6.44%
2012	$(5.56 \pm 0.20) \times 10^{-1}$	0.31×10^{-1}	5.58%
2015	$(5.56 \pm 0.20) \times 10^{-1}$	0.42×10^{-1}	7.58%
2016	$(5.44 \pm 0.11) \times 10^{-1}$	0.29×10^{-1}	5.33%
2017	$(6.42 \pm 0.11) \times 10^{-1}$	0.32×10^{-1}	4.94%
2018	$(6.24 \pm 0.11) \times 10^{-1}$	0.35×10^{-1}	5.61%

In the case of the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes there is an additional complication. In general discrepancies between data and simulation are expected to cancel at first order when taking the ratio of signal to normalisation mode efficiencies, as is the case when measuring the signal branching fractions. However, the requirement imposed on $\mathscr{P}_3$ affects only the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes and not the $B^+ \to \overline{D}^0 D_s^+$ decay mode. Thus, a similar $\mathscr{P}_3'$ is constructed for the $B^+ \to \overline{D}^0 D_s^+$ decay mode in order to study potential discrepancies between data and simulation. This $\mathscr{P}_3'$ variable is defined as $\mathtt{ProbNN}_K(K) \times \mathtt{ProbNN}_K(K_1) \times \mathtt{ProbNN}_K(K_2) \times \mathtt{ProbNN}_\pi(\pi_D)$, where K is the kaon from the $\overline{D}^0$ decay while K_1, K_2 and π_D are the daughters of the D_s^+ meson. The comparison between data and simulation after full selection for every year is shown in Fig. 5.15. In order to account for the discrepancies, efficiencies associated with requirements equivalent to those applied for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes, *i.e.* $\mathscr{P}_3 > 0.5$ and $\mathscr{P}_3 > 0.6$ for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes respectively, are measured. Then the relative uncertainty given by $|(\varepsilon_{data} - \varepsilon_{sim})/\varepsilon_{sim}|$, where ε_{data} and ε_{sim} are respectively the efficiencies measured on data and simulation, is computed and is added in quadrature to the relative PID systematic uncertainty due to all other sources described previously. These numbers are given in Tab.5.8. Naturally, the additional source of systematic uncertainty impacts only the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes.

Table 5.8 – Systematic uncertainty $\sigma_{PID}^{\mathscr{P}_3}$ associated with the discrepancy between data and simulation for the variable $\mathscr{P}_3$, expressed relatively in the fourth column. The efficiencies associated with the PID requirements on the $\mathscr{P}_3'$ variable in the $B^+ \to \overline{D}^0 D_s^+$ samples, for data and corrected simulation, are given in the second and third columns respectively. The top and bottom halves of the table correspond to the numbers associated with the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively.

Year	ϵ_{data}	ϵ_{sim}	$\sigma_{PID}^{\mathscr{P}_3}$ relative
2011	$(6.92 \pm 0.19) \times 10^{-1}$	$(7.10 \pm 0.57) \times 10^{-1}$	2.50%
2012	$(6.40 \pm 0.14) \times 10^{-1}$	$(6.89 \pm 0.43) \times 10^{-1}$	7.20%
2015	$(9.38 \pm 0.09) \times 10^{-1}$	$(9.34 \pm 0.19) \times 10^{-1}$	0.41%
2016	$(9.32 \pm 0.04) \times 10^{-1}$	$(9.51 \pm 0.07) \times 10^{-1}$	1.94%
2017	$(9.37 \pm 0.04) \times 10^{-1}$	$(9.53 \pm 0.11) \times 10^{-1}$	1.75%
2018	$(9.21 \pm 0.04) \times 10^{-1}$	$(9.38 \pm 0.16) \times 10^{-1}$	1.83%
2011	$(4.99 \pm 0.20) \times 10^{-1}$	$(5.01 \pm 0.45) \times 10^{-1}$	0.29%
2012	$(4.59 \pm 0.14) \times 10^{-1}$	$(4.79 \pm 0.25) \times 10^{-1}$	4.22%
2015	$(8.27 \pm 0.14) \times 10^{-1}$	$(8.51 \pm 0.34) \times 10^{-1}$	2.83%
2016	$(8.39 \pm 0.06) \times 10^{-1}$	$(8.62 \pm 0.14) \times 10^{-1}$	2.65%
2017	$(8.36 \pm 0.06) \times 10^{-1}$	$(8.63 \pm 0.07) \times 10^{-1}$	3.10%
2018	$(8.08 \pm 0.06) \times 10^{-1}$	$(8.37 \pm 0.20) \times 10^{-1}$	3.51%

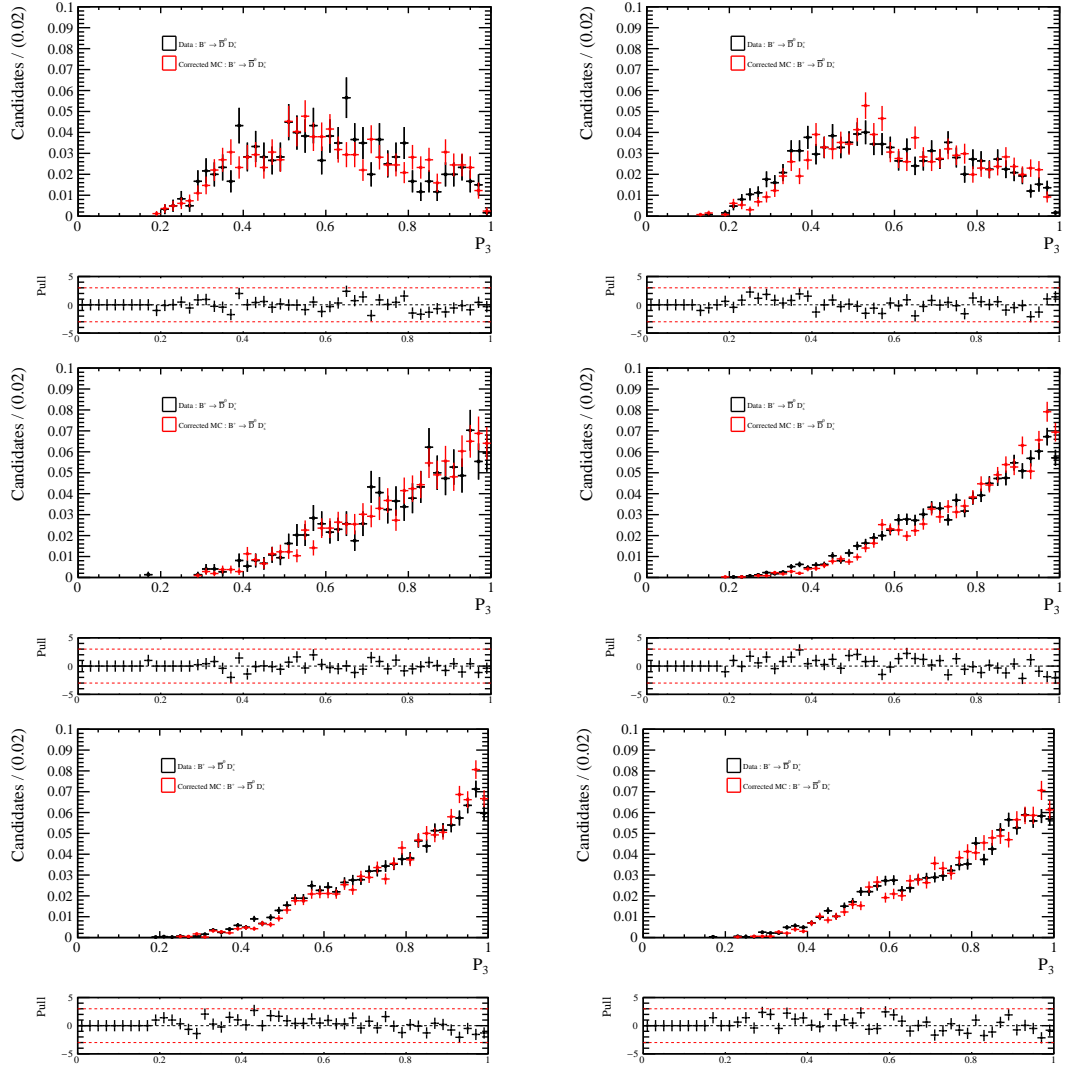Figure 5.15 – Comparison of the $\mathscr{P}_3'$ distributions between $B^+ \to \overline{D}^0 D_s^+$ data and corrected simulation after full selection in black and red, respectively. The plots from the top left to the bottom right show the 2011, 2012, 2015, 2016, 2017 and 2018 years. The distributions for simulated data apply a scale factor of $w = 1.11$ on the n_{tracks} distribution.

Trigger

The selection efficiency associated with the trigger requirements is split into two sources according to whether the requirements are imposed on the hardware (L0) or software (HLT) triggers *i.e.* $\varepsilon_{trigger} = \varepsilon_{L0} \times \varepsilon_{HLT}$, as mentioned previously. The effect of the software triggers is generally well described in simulation and as such no systematic uncertainty is assigned on ε_{HLT}. However, this is not the case for the hardware triggers where systematic uncertainties need to be assigned in order to account for data-simulation discrepancies. While data-driven methods are used to obtain corrected hardware trigger efficiencies for both the $B^+ \rightarrow K^+ \tau^\pm \mu^\mp$ and $B^+ \rightarrow \overline{D}^0 D_s^+$ decay modes, the specific methods differ depending on the decay mode.

In the case of the $B^+ \rightarrow \overline{D}^0 D_s^+$ decay mode, the so-called `TISTOS` method [90] is used to measure the efficiency associated with the requirement of a positive decision (TOS) from `L0HadronDecision` applied to the B^+ candidate. The data-driven efficiency is given by

$$\epsilon_{L0} = \frac{N_{TISTOS}}{N_{TIS}}, \tag{5.16}$$

where N_{TISTOS} is the number of signal candidates remaining in data after requesting simultaneously a positive TOS decision from `L0HadronDecision` and a positive TIS decision from `L0MuonDecision`, both applied on the B^+ candidate, while N_{TIS} is the number of signal candidates surviving the requirement of a positive TIS decision from `L0MuonDecision` applied on the B^+ candidate. Both N_{TIS} and N_{TISTOS} are measured by fitting the $B^+ \rightarrow \overline{D}^0 D_s^+$ data after full selection, excluding the nominal trigger requirements, and applying the aforementioned L0 requirements. The resulting `TISTOS` corrected efficiency is then considered as the nominal efficiency ε_{L0} and its uncertainty is considered as the systematic uncertainty on ε_{L0}. The nominal, corrected efficiencies and their associated systematic uncertainties are given in Tab. 5.9.

Table 5.9 – The systematic uncertainty σ_{L0}^n associated with the L0 trigger requirements for the $B^+ \rightarrow \overline{D}^0 D_s^+$ decay mode is given in the fourth column. It is given by the statistical uncertainty of the `TISTOS` efficiency, expressed relatively. The second and third columns give ϵ_{L0} computed using simulation and the `TISTOS` method, respectively.

Year	ϵ_{L0}^n	$\epsilon_{L0}^n(TISTOS)$	$\sigma_{L0}^n(syst)$ relative
2011	$(3.59 \pm 0.09) \times 10^{-1}$	$(3.65 \pm 0.35) \times 10^{-1}$	9.60%
2012	$(3.15 \pm 0.07) \times 10^{-1}$	$(3.66 \pm 0.25) \times 10^{-1}$	6.76%
2015	$(3.86 \pm 0.09) \times 10^{-1}$	$(4.25 \pm 0.47) \times 10^{-1}$	11.09%
2016	$(3.72 \pm 0.05) \times 10^{-1}$	$(3.68 \pm 0.15) \times 10^{-1}$	4.04%
2017	$(4.10 \pm 0.05) \times 10^{-1}$	$(3.93 \pm 0.14) \times 10^{-1}$	3.66%
2018	$(3.40 \pm 0.05) \times 10^{-1}$	$(3.59 \pm 0.13) \times 10^{-1}$	3.68%

Since the `TISTOS` procedure cannot be used to determine the hardware trigger efficiency for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes, a different method is used. The strategy is to correct the nominal L0 efficiency measured in simulation by applying weights determined on data. More precisely, the ratio between the L0 efficiencies measured in data and simulation when reconstructing $B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$ candidates is used as weight to correct the value of ε_{L0} associated with the requirement of a positive TOS decision of `L0MuonDecision` measured with $B^+ \to K^+ \tau^\pm \mu^\mp$ simulation. Since this efficiency is dependent on the transverse momentum of the muon candidate, the correction is performed in bins of muon p_T. The weights used in the correction procedure have been in measured in Ref. [17]. The resulting corrected efficiency and its uncertainty are then taken as the nominal L0 efficiency and the difference between the central values of the corrected and uncorrected efficiencies is assigned as systematic uncertainty. These efficiencies are given in Tabs. 5.10 and 5.11 for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively.

Table 5.10 – The fourth column gives the systematic uncertainty σ_{L0}^s associated with the L0 trigger requirements for the $B^+ \to K^+ \tau^- \mu^+$ decay mode, with the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes at the top and bottom, respectively. The nominal efficiency is given by ϵ_{L0}^s while the alternative efficiency is given by $\epsilon_{L0,corr}^s$.

Year	ϵ_{L0}^s	$\epsilon_{L0,corr}^s$	$\sigma_{L0}^s(syst)$
2011	$(6.10 \pm 0.13) \times 10^{-1}$	5.99×10^{-1}	0.11×10^{-1}
2012	$(5.68 \pm 0.11) \times 10^{-1}$	5.63×10^{-1}	0.05×10^{-1}
2015	$(3.62 \pm 0.16) \times 10^{-1}$	3.46×10^{-1}	0.15×10^{-1}
2016	$(5.68 \pm 0.09) \times 10^{-1}$	5.59×10^{-1}	0.09×10^{-1}
2017	$(6.21 \pm 0.09) \times 10^{-1}$	6.15×10^{-1}	0.06×10^{-1}
2018	$(5.88 \pm 0.09) \times 10^{-1}$	5.78×10^{-1}	0.09×10^{-1}
2011	$(5.87 \pm 0.18) \times 10^{-1}$	5.75×10^{-1}	0.12×10^{-1}
2012	$(5.74 \pm 0.15) \times 10^{-1}$	5.70×10^{-1}	0.04×10^{-1}
2015	$(4.17 \pm 0.23) \times 10^{-1}$	4.04×10^{-1}	0.13×10^{-1}
2016	$(5.74 \pm 0.14) \times 10^{-1}$	5.65×10^{-1}	0.09×10^{-1}
2017	$(5.95 \pm 0.13) \times 10^{-1}$	5.90×10^{-1}	0.05×10^{-1}
2018	$(5.69 \pm 0.14) \times 10^{-1}$	5.60×10^{-1}	0.08×10^{-1}

Table 5.11 – The fourth column gives the systematic uncertainty σ_{L0}^s associated with the L0 trigger requirements for the $B^+ \to K^+ \tau^+ \mu^-$ decay mode, with the $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes at the top and bottom, respectively. The nominal efficiency is given by ϵ_{L0}^s while the alternative efficiency is given by $\epsilon_{L0,corr}^s$.

Year	ϵ_{L0}^s	$\epsilon_{L0,corr}^s$	$\sigma_{L0}^s(syst)$
2011	$(6.30 \pm 0.14) \times 10^{-1}$	6.19×10^{-1}	0.11×10^{-1}
2012	$(5.74 \pm 0.11) \times 10^{-1}$	5.69×10^{-1}	0.05×10^{-1}
2015	$(3.94 \pm 0.16) \times 10^{-1}$	3.79×10^{-1}	0.14×10^{-1}
2016	$(5.64 \pm 0.10) \times 10^{-1}$	5.54×10^{-1}	0.10×10^{-1}
2017	$(6.50 \pm 0.09) \times 10^{-1}$	6.44×10^{-1}	0.06×10^{-1}
2018	$(5.74 \pm 0.10) \times 10^{-1}$	5.65×10^{-1}	0.09×10^{-1}
2011	$(6.02 \pm 0.21) \times 10^{-1}$	5.92×10^{-1}	0.11×10^{-1}
2012	$(5.74 \pm 0.16) \times 10^{-1}$	5.69×10^{-1}	0.05×10^{-1}
2015	$(3.89 \pm 0.26) \times 10^{-1}$	3.75×10^{-1}	0.14×10^{-1}
2016	$(5.72 \pm 0.14) \times 10^{-1}$	5.64×10^{-1}	0.08×10^{-1}
2017	$(6.58 \pm 0.14) \times 10^{-1}$	6.50×10^{-1}	0.08×10^{-1}
2018	$(5.93 \pm 0.14) \times 10^{-1}$	5.85×10^{-1}	0.08×10^{-1}

5.4.2 Fit model

When performing a fit to the data there are a number of models able to describe accurately the observed distribution. As such, there is a degree of arbitrariness involved in any choice of model. It is thus important to quantify the effect on the final result due to this intrinsic uncertainty by setting systematic uncertainties associated with the various models used in the analysis. In order to do so, the total signal yield is measured with the nominal and an alternative model and a systematic uncertainty is derived in a similar way as was done for the selection efficiencies, *i.e.*

$$\sigma_{sys} = \sqrt{(N_{nom} - N_{alt})^2 + |\sigma_{nom}^2 - \sigma_{alt}^2|}, \tag{5.17}$$

where N_{nom} and N_{alt} are the nominal and alternative signal yields and σ_{nom} and σ_{alt} are their respective uncertainties. Once again, not only does this equation take into account the bias between the nominal and alternative models but also potential differences in statistical uncertainties. These systematic uncertainties are typically evaluated on large pseudo-datasets where the fitted yields are distributed according to a Gaussian distribution whose mean and standard deviation are taken as estimators of the fitted yield and its uncertainty. In particular, this procedure is necessary in the case of the $B^+ \rightarrow K^+ \tau^\pm \mu^\mp$ decay modes since the data is blind and as such the real signal yield cannot be directly measured. However, for the final result after unblinding, the systematic uncertainties on the $B^+ \rightarrow K^+ \tau^\pm \mu^\mp$ signal yields are reevaluated using the data. This is discussed in Sec. 5.5.

Fit model for the $B^+ \rightarrow K^+ \tau^\pm \mu^\mp$ decay modes

The model used to fit the $B^+ \rightarrow K^+ \tau^\pm \mu^\mp$ data samples can be split into two parts describing either the signal or background contributions. Each of these is impacted by different sources of systematic uncertainty. For the background model, the dominant source of systematic uncertainty is due to the fact that the endpoint and slope parameters of the Argus function are fixed in the final fit. The reason why this is the dominant source is because the endpoint parameter has a great impact on the overall shape of the background model. However, this parameter is correlated with the width of the Gaussian function used in the convolution and as long as the Gaussian's width is free in the fit, the model will largely be able to properly describe the data. Nonetheless, the impact from fixing this parameter cannot be neglected. The strategy used to assess a related systematic uncertainty, is to create two alternative models where the values of the endpoint and slope parameters are either increased or decreased by a constant value with respect to their values in the nominal model. Both parameters are positively correlated and their variation is performed in phase. For the slope parameter, a 5% variation is performed. Concerning the endpoint parameter, a reasonable size for the variation needs to be determined first. In order to do so, one thousand pseudo-datasets are

generated according to the nominal background model determined by fitting the blind data. These datasets are subsequently fitted with an alternative model where the Gaussian width is free, as will be the case in the nominal final fit, as well as the endpoint parameter. The distribution of fitted value of the endpoint parameter is recorded, and the standard deviation of the resulting distribution is used as the constant value to be used in the construction of the aforementioned alternative models. These distributions for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes are shown in Fig. 5.16. As a result, the endpoint parameter for the two alternative background models is varied by $\pm 109\,\mathrm{MeV}/c^2$ and $\pm 52\,\mathrm{MeV}/c^2$ for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes respectively, while the slope parameter is varied by $\pm 5\%$ coherently with the endpoint. In order to determine a systematic uncertainty by using Eq. (5.17), it is first necessary to determine the nominal signal yield. Thus, the one thousand pseudo-datasets generated with the nominal background-only model are fitted with the nominal signal plus background model. The resulting distribution of the fitted signal yield is shown in Fig. 5.17 for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes. These distributions are fitted with a Gaussian pdf, and their associated mean and standard deviation provide the values of N_{nom} and σ_{nom} to be used in Eq. (5.17). These pseudo-datasets are then fitted with the nominal signal pdf plus the alternative background models. Similarly to the nominal model, the resulting distribution of the fitted signal yield, shown in Fig. 5.18, is fitted by a Gaussian function and its associated mean and standard deviation correspond to N_{alt} and σ_{alt} to be used in Eq. (5.17). For each decay mode the largest systematic uncertainty out of the two is assigned as the systematic uncertainty due to fixing these parameters in the fit. The nominal and alternative yields as well as the associated absolute systematic uncertainties on the signal yield are given in Tab. 5.12.
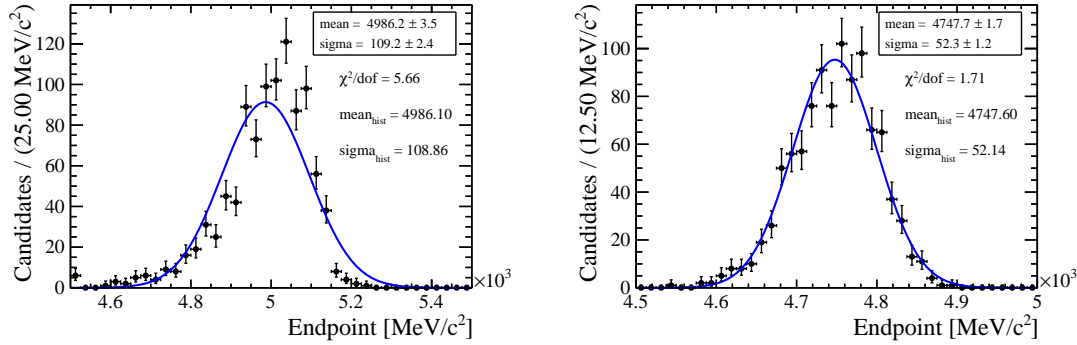

Figure 5.16 – Distribution of the value of the endpoint parameter obtained after fitting 1000 pseudo-datasets generated with the background-only model and fitted with an alternative background model where the endpoint parameter is free. The plots on the left and right correspond to the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively. The blue curve is a Gaussian fit to the distribution.
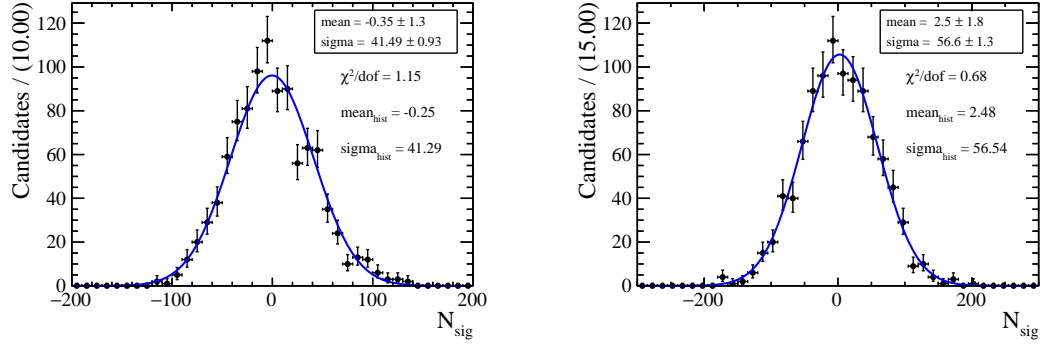
Figure 5.17 – Distribution for the $B^+ \rightarrow K^+\tau^-\mu^+$ (left) and $B^+ \rightarrow K^+\tau^+\mu^-$ (right) decay modes of the fitted total signal yield using the nominal signal plus background model for 1000 pseudo-datasets generated according to the nominal background-only model. The blue curve is a Gaussian fit to the distribution.

Table 5.12 – The first column gives the variation of the endpoint in the alternative model. The second and third columns give the total signal yield found by fitting 1000 background-only pseudo-datasets using the nominal and alternative models, respectively. The top and bottom halves of the table refer to the $B^+ \rightarrow K^+\tau^-\mu^+$ and $B^+ \rightarrow K^+\tau^+\mu^-$ decay modes, respectively.

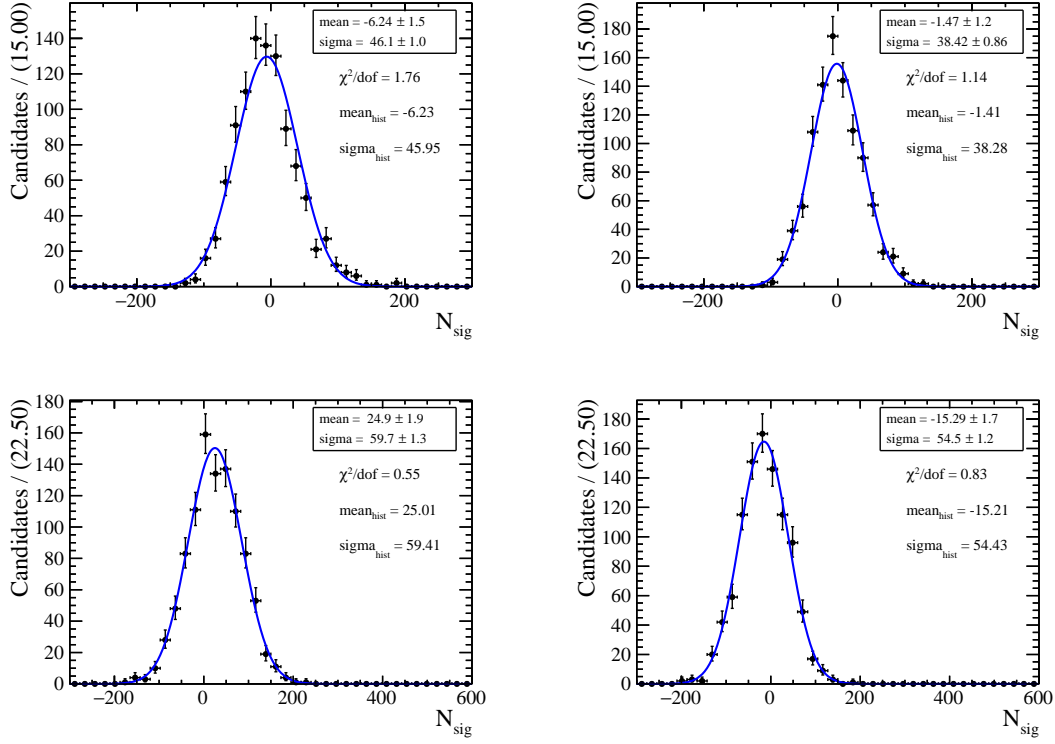| Variation	N_{nom}	N_{alt}	σ_{sys}
$+109$ MeV/c^2	-0.35 ± 41.49	-6.24 ± 46.11	20.98
-109 MeV/c^2	-0.35 ± 41.49	-1.47 ± 38.42	15.69
$+52$ MeV/c^2	2.52 ± 56.60	24.90 ± 59.71	29.37
-52 MeV/c^2	2.52 ± 56.60	-15.29 ± 54.50	23.45

Figure 5.18 – Distribution for the $B^+ \to K^+ \tau^- \mu^+$ (top) and $B^+ \to K^+ \tau^+ \mu^-$ (bottom) decay modes of the fitted total signal yield using the nominal signal pdf plus the alternative background models, for 1000 pseudo-datasets generated according to the nominal background-only model. The left and right columns show the distributions associated with the model where the endpoint parameter's value has been increased and decreased, respectively. The blue curve is a Gaussian fit to the distribution.
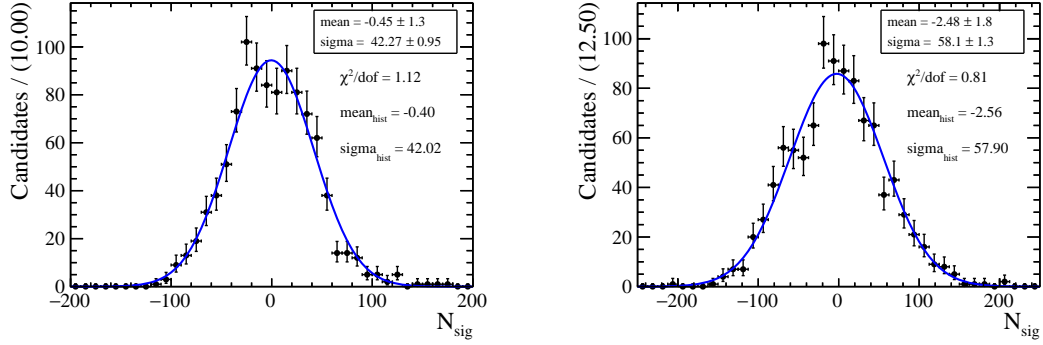
Figure 5.19 – Distribution for the $B^+ \rightarrow K^+ \tau^- \mu^+$ (left) and $B^+ \rightarrow K^+ \tau^+ \mu^-$ (right) decay modes of the fitted total signal yield using the nominal full fit model, for 1000 pseudo-datasets generated according to the background-only alternative pdf where f_{phys} is independent for each run. The blue curve is a Gaussian fit to the distribution.

A second source of systematic uncertainty linked to the background model is due to the fact that the fraction of physics background f_{phys} is shared amongst all years in the nominal model. In order to assign a systematic uncertainty, the blind data is fitted with an alternative background model where f_{phys} is only shared for years within a given Run. These fits are subsequently used to generate one thousand pseudo-datasets which are then fitted with the nominal full model. The resulting distribution of fitted signal yield, shown in Fig. 5.19, is fitted with a Gaussian function whose mean and standard deviation describe respectively N_{alt} and σ_{alt}. In this particular instance, the systematic uncertainty is not given by Eq. (5.17). Indeed, the usual procedure to assign a systematic uncertainty on the signal yield is to fit the same pseudo-datasets with two different models while in this case the same model is used to fit two different sets of pseudo-experiments. Thus, the systematic uncertainty assigned in this case is given by $(N_{nom} - N_{alt}) \oplus \sigma_{alt}/\sqrt{1000}$. The nominal and alternative yields as well as the corresponding systematic uncertainties are given in Tab. 5.13.

Table 5.13 – The first and second columns give the total signal yield found by fitting 1000 background-only pseudo-datasets using the nominal signal plus background model. In the first column, the pseudo-datasets were generated according to the nominal background pdf while in the second column, the data was generated according to an alternative background pdf where f_{phys} is independent for each run. The last column gives the associated systematic uncertainty. The top and bottom halves of the table refer to the $B^+ \rightarrow K^+ \tau^- \mu^+$ and $B^+ \rightarrow K^+ \tau^+ \mu^-$ decay modes, respectively.

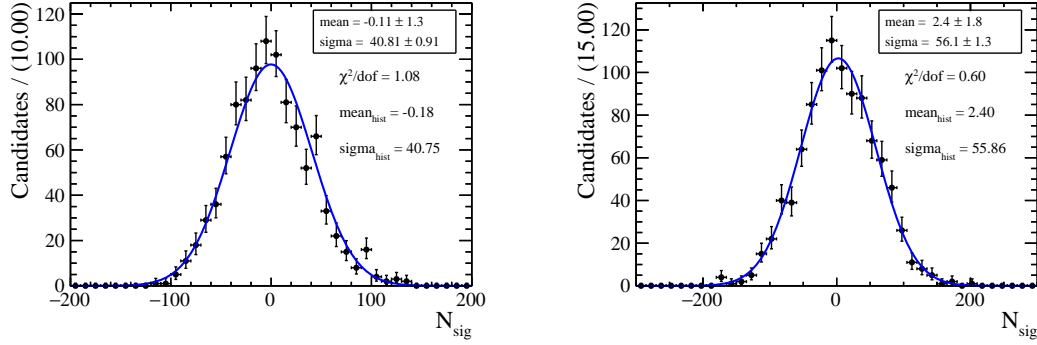| Decay mode	N_{nom}	N_{alt}	σ_{sys}
$B^+ \rightarrow K^+ \tau^- \mu^+$	-0.35 ± 41.49	-0.45 ± 42.27	1.34
$B^+ \rightarrow K^+ \tau^+ \mu^-$	2.52 ± 56.60	-2.48 ± 58.10	5.33

Figure 5.20 – Distribution for the $B^+ \to K^+\tau^-\mu^+$ (left) and $B^+ \to K^+\tau^+\mu^-$ (right) decay modes of the fitted total signal yield using the nominal background plus alternative signal models for 1000 pseudo-datasets generated according to the nominal background-only model.

The remaining sources of systematic uncertainty associated with the models used to fit the $B^+ \to K^+\tau^\pm\mu^\mp$ candidates are due to the signal contribution to the models. The first source is due to the functional form itself of the signal pdf. The nominal signal model is given by the sum of two Gaussian functions and a CB function with an upper-mass tail. As an alternative signal model, the CB function is replaced by a Gaussian function whose mean parameter is freely floating in the fit. The number of free parameters in the fit is identical for the nominal and alternative models. Similarly to the nominal signal model, the shape parameters of the alternative signal model are determined by fitting the signal simulated samples. The usual procedure is used to set the systematic uncertainty, *i.e.* one thousand pseudo-datasets generated according to the background-only model are fitted with the nominal background model plus the nominal or alternative signal model. The corresponding distributions of the fitted signal yield are fitted with a Gaussian function whose mean and standard deviation give the values to be used in Eq. (5.17) and are shown in Fig 5.20 for the alternative model. These yields and their associated systematic uncertainties are given in Tab. 5.14.

Table 5.14 – The first and second columns give the total signal yield found by fitting 1000 background-only pseudo-datasets using the nominal background model plus the nominal and alternative signal models, respectively. The last column gives the associated systematic uncertainty. The top and bottom halves of the table refer to the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes, respectively.

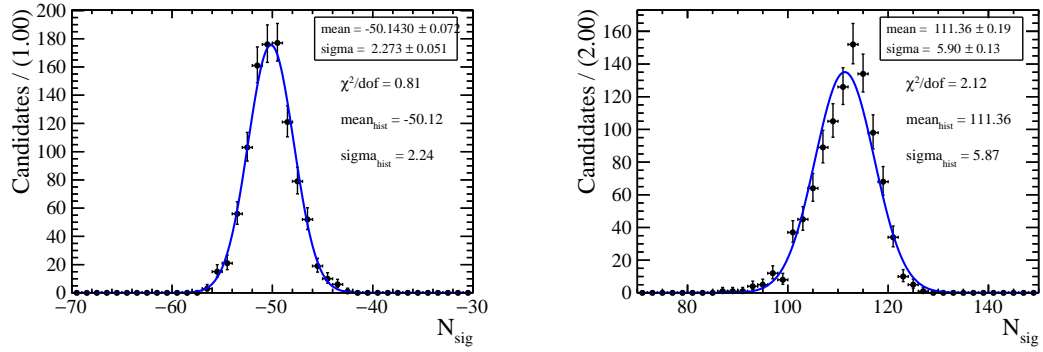| Decay mode	N_{nom}	N_{alt}	σ_{sys}
$B^+ \to K^+\tau^-\mu^+$	-0.35 ± 41.49	-0.11 ± 40.81	7.45
$B^+ \to K^+\tau^+\mu^-$	2.52 ± 56.60	2.43 ± 56.12	7.32

Figure 5.21 – Distribution for the $B^+ \to K^+ \tau^- \mu^+$ (left) and $B^+ \to K^+ \tau^+ \mu^-$ (right) decay modes of the fitted total signal yield, with the full model where the shape parameters of the signal pdf have been Gaussian fluctuated according to the fits to the signal simulation samples, for 1000 pseudo-datasets generated according to the nominal background-only model.

Another source of systematic uncertainty associated with the signal model is due to the fact that the shape parameters are fixed to the values found by fitting the signal simulation samples. The impact of this procedure is studied by fitting a single pseudo-dataset generated according to the background-only nominal model using the nominal background model plus an alternative signal pdf. The values of shape parameters of that signal pdf are sampled from a multivariate Gaussian function whose mean parameters are given by the estimators of the various shape parameters obtained in the nominal fit to the signal simulation sample while the off-diagonal elements of this Gaussian are given by the values of the covariance matrix obtained in the aforementioned fits. This same pseudo-dataset is fitted one thousand times. For each of these fits, the values of the shape parameters of the signal pdf are randomly sampled from the multivariate Gaussian function. The resulting distribution of fitted signal yields, shown in Fig. 5.21, is fitted with a Gaussian function and its associated standard deviation is assigned as systematic uncertainty.

Finally, the fit bias determined by fitting one thousand pseudo-datasets generated according to the background-only model and fitted with the nominal full model needs to be accounted for. As reminder, the resulting distribution of fitted yields is shown in Fig. 5.17. The fit bias itself is subtracted from the measured signal yield in the final fit and the quadratic sum between the bias and its uncertainty is added as an additional source of systematic uncertainty. The resulting systematic uncertainties for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes respectively are $\sigma_{sys} = 1.35$ and $\sigma_{sys} = 3.08$.

Fit model for the $B^+ \to \bar{D}^0 D_s^+$ decay mode

Some sources of systematic uncertainty are similar between the $B^+ \to K^+ \tau^{\pm} \mu^{\mp}$ and $B^+ \to \bar{D}^0 D_s^+$ fit models. For example, it is necessary to assess a systematic uncertainty due to the functional form of the fit model. In the case of the fit to the $B^+ \to \bar{D}^0 D_s^+$ candidates, the functional form of the signal pdf is nominally given by the sum of two CB functions with opposite side power law tails. As an alternative model, the sum of two Gaussian functions with shared mean parameter is considered. The data is fitted with the nominal and alternative full models and systematic uncertainties are assigned for each year according to Eq. (5.17). The fits to the combined Run 1 and Run 2 datasets using the alternative models are shown in Fig. 5.22 while the associated yields and systematic uncertainties are given in Tab. 5.15.

The model used to fit the data includes a component which describes the mass distribution of $B^+ \to \bar{D}^0 K^+ K^- \pi^+$ background candidates. In the nominal fit of the data, the yield associated with this component corresponds approximately to 1.5% of the total yield. Since there is some uncertainty in the procedure used to determine the model describing the mass distribution of these background candidates, a systematic uncertainty corresponding to 0.5% of the signal yield is assigned. This value corresponds to approximately one third of this background yield and reflects the degree of confidence in the mass shape determination procedure.

The last component in the model used to fit the $B^+ \to \bar{D}^0 D_s^+$ data samples describes the mass distribution due to the combinatorial background. Its nominal functional form is a uniform distribution. As an alternative model, it is replaced by a first degree polynomial. The data is fitted with this alternative model and the resulting yields are used to set a systematic uncertainty following the standard procedure. The fits to the Run 1 and Run 2 data samples using this alternative model are shown in Fig. 5.23. The corresponding yields and associated systematic uncertainties are given for every year in Tab. 5.16.
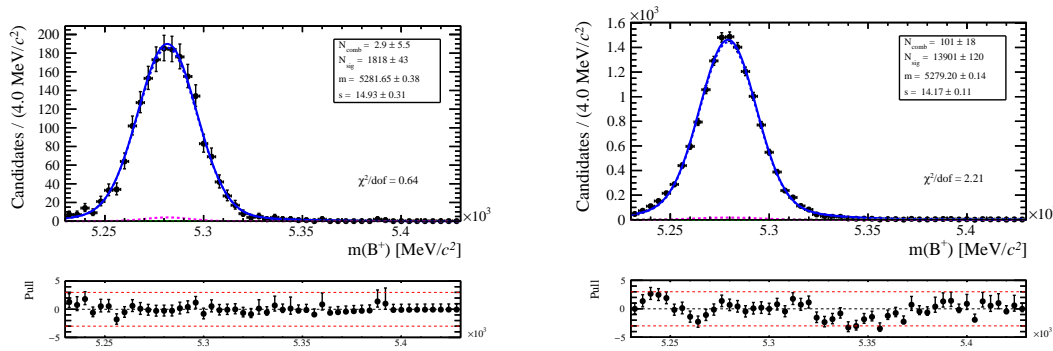

Figure 5.22 – Fit of the Run 1 (left) and Run 2 (right) data samples for the $B^+ \to \bar{D}^0 D_s^+$ decay mode using the alternative signal model.

Table 5.15 – The second and third columns give the signal yield, for the corresponding year, found by fitting the $B^+ \to \overline{D}^0 D_s^+$ data samples using the nominal and alternative signal models, respectively. The last column gives the associated systematic uncertainty.

Year	N_{nom}	N_{alt}	σ_{sys}
2011	590.52 ± 24.09	590.82 ± 24.10	0.83
2012	1227.09 ± 34.72	1227.72 ± 34.74	1.28
2015	726.44 ± 26.70	725.58 ± 26.67	1.56
2016	3780.16 ± 60.92	3775.69 ± 60.84	5.36
2017	4379.07 ± 65.56	4373.89 ± 65.49	6.08
2018	5031.97 ± 70.28	5026.01 ± 70.20	6.86

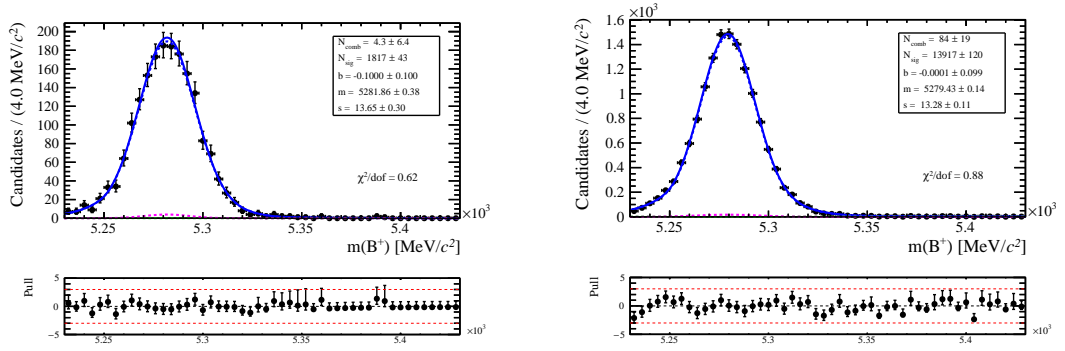
Figure 5.23 – Fit of the Run 1 (left) and Run 2 (right) data samples for the $B^+ \to \overline{D}^0 D_s^+$ decay mode using the alternative model describing the combinatorial background.

Table 5.16 – The second and third columns give the signal yield, for the corresponding year, found by fitting the $B^+ \to \overline{D}^0 D_s^+$ data samples using the nominal and alternative models for the combinatorial background, respectively. The last column gives the associated systematic uncertainty.

Year	N_{nom}	N_{alt}	σ_{sys}
2011	590.52 ± 24.09	590.38 ± 24.08	0.55
2012	1227.09 ± 34.72	1226.80 ± 34.71	0.81
2015	726.44 ± 26.70	726.44 ± 26.70	0.09
2016	3780.16 ± 60.92	3780.15 ± 60.92	0.18
2017	4379.07 ± 65.56	4379.05 ± 65.56	0.20
2018	5031.97 ± 70.28	5031.94 ± 70.28	0.21

A systematic uncertainty has already been assigned on the signal model in order to reflect the uncertainty on the functional model. However, another source of systematic uncertainty affecting the signal description is due to fact that some shape parameters of the signal pdf are fixed when fitting the data. Indeed, when fitting the data the parameters f, s_{ratio}, α and α_2 are fixed to the values found in the fits to the simulation samples. In order to assign a systematic uncertainty, a procedure similar to the one used in the analogous case for the $B^+ \to K^+ \tau^\pm \mu^\mp$ decay modes is used. It consists in performing one thousand fits to the data where the values of these four parameters are randomly sampled from the multivariate Gaussian associated with the fits to the simulated samples. The resulting distributions of the fitted signal yields for each year, shown in Fig. 5.24, are then fitted with a Gaussian function and its standard deviation estimate is assigned as systematic uncertainty. In the case of the distributions associated with the 2011 and 2012 samples, there is a peculiar structure located at high fitted signal yields. This is due to the cases where the fitted yield associated with the combinatorial background, constrained to be positive, reaches zero. In these cases, since the yield associated with the $B^+ \to \overline{D}^0 K^+ K^- \pi^+$ background component is fully fixed, the signal yield reaches an upper bound. This is not an issue in the fits to samples within Run 2 since the contribution of combinatorial background decays is significantly different from zero. The systematic uncertainty derived from these distributions is given by the standard deviation of the Gaussian fits. Since these Gaussian fits describe accurately the observed 2011 and 2012 distributions before the signal yield upper bound, the estimate of the systematic uncertainty is sufficiently reliable.

Finally the statistical uncertainty from the nominal fit to the $B^+ \to \overline{D}^0 D_s^+$ data is also added as a systematic uncertainty.
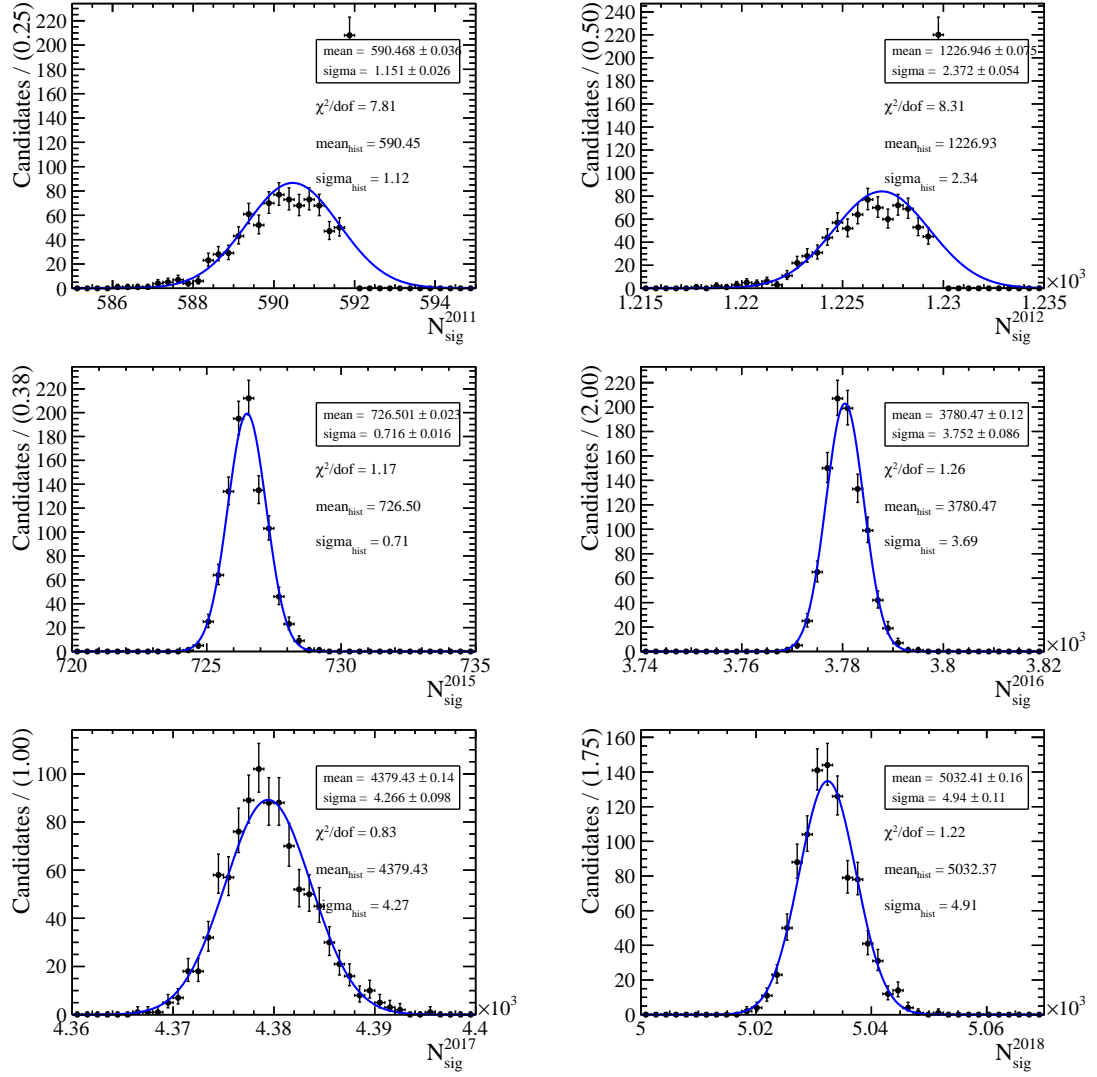
Figure 5.24 – Distributions of the measured signal yield when fitting 1000 times the $B^+ \to \overline{D}^0 D_s^+$ data. In each fit, the values of the fixed shape parameters in the signal model are randomly sampled from a multivariate Gaussian function calibrated on the fits to the simulated signal samples. The top, middle and bottom rows show the distributions for 2011 and 2012, 2015 and 2016, 2017 and 2018 data, respectively. The blue curve is a Gaussian fit.

5.4.3 Summary of the systematic uncertainties

A summary of the various systematic uncertainties considered in the analysis are given hereafter. The systematic uncertainties on the selection efficiencies associated with the $B^+ \to K^+\tau^-\mu^+$, $B^+ \to K^+\tau^+\mu^-$ and $B^+ \to \overline{D}^0 D_s^+$ decay modes are given in Tabs. 5.17, 5.18 and 5.19, respectively. The absolute systematic uncertainties on the $B^+ \to \overline{D}^0 D_s^+$ signal yields are given in Tab. 5.20 while the uncertainties on the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ signal yield are given in Tab. 5.21. These uncertainties on the $B^+ \to K^+\tau^\pm\mu^\mp$ decay modes yields are used before unblinding. The systematic uncertainties used in the final fit are determined on data and given in Tab. 5.25 in Sec. 5.5.

Table 5.17 – Relative systematic uncertainties in % on the selection efficiencies associated with the $B^+ \to K^+\tau^-\mu^+$ decay mode for $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ and $\tau^- \to \pi^-\pi^+\pi^-\pi^0\nu_\tau$ sub-decay modes at the top and bottom, respectively. The last column gives the total systematic uncertainty which is the sum in quadrature of the uncertainties of all the individual sources.

Year	BDT	PID	Trigger	MC stats	Total
2011	-	4.31	1.82	3.94	6.12
2012	-	7.31	0.84	3.10	7.98
2015	-	6.61	4.43	6.11	10.03
2016	-	3.89	1.53	2.65	4.95
2017	0.69	6.08	0.99	2.59	6.72
2018	1.35	2.16	1.57	2.65	4.00
2011	-	6.11	2.12	5.58	8.54
2012	-	8.61	0.71	4.46	9.73
2015	-	2.81	3.19	8.01	9.06
2016	-	6.34	1.58	3.89	7.60
2017	0.95	2.93	0.82	3.88	5.02
2018	1.38	4.39	1.49	3.93	6.23

Table 5.18 – Relative systematic uncertainties in % on the selection efficiencies associated with the $B^+ \to K^+ \tau^+ \mu^-$ decay mode for $\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$ and $\tau^- \to \pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ sub-decay modes at the top and bottom, respectively. The last column gives the total systematic uncertainty which is the sum in quadrature of the uncertainties of all the individual sources.

Year	BDT	PID	Trigger	MC stats	Total
2011	-	5.15	1.78	4.05	6.79
2012	-	6.22	0.89	3.21	7.05
2015	-	9.71	3.80	5.59	11.83
2016	-	3.29	1.73	2.70	4.59
2017	0.77	4.16	0.91	2.55	5.02
2018	1.51	4.96	1.61	2.69	6.06
2011	-	6.63	1.83	5.90	9.06
2012	-	9.12	0.91	4.60	10.25
2015	-	8.10	3.68	8.92	12.60
2016	-	5.96	1.47	3.92	7.29
2017	0.75	5.84	1.17	3.66	7.03
2018	2.35	6.62	1.42	3.88	8.15

Table 5.19 – Relative systematic uncertainties in % on the selection efficiencies associated with the $B^+ \to \overline{D}^0 D_s^+$ decay mode. The last column gives the total systematic uncertainty which is the sum in quadrature of the uncertainties of all the individual sources.

Year	PID	Trigger	MC stats	Total
2011	3.72	9.60	2.32	10.56
2012	3.98	6.76	1.71	8.03
2015	2.67	11.09	2.00	11.58
2016	1.16	4.04	1.08	4.34
2017	0.95	3.66	1.06	3.93
2018	1.49	3.68	1.07	4.11

Table 5.20 – Absolute systematic uncertainties on the signal yields of the $B^+ \rightarrow \overline{D}^0 D_s^+$ decay mode for each year. The second to the fifth columns show the systematic uncertainties on the parameters fixed from the fit to the simulation samples, the signal model, the description of the $B^+ \rightarrow \overline{D}^0 K^+ K^- \pi^+$ component and the model for the combinatorial background, respectively. The sixth column shows the statistical uncertainty on the data fit. The last column gives the total systematic uncertainty which is the sum in quadrature of the systematic uncertainties of all the individual sources.

Year	Fixed	Signal	$\overline{D}^0 K^+ K^- \pi^+$	Combinatorial	Stats	Total
2011	1.15	0.83	2.95	0.55	24.09	24.32
2012	2.37	1.28	6.14	0.81	34.72	35.37
2015	0.72	1.56	3.63	0.09	26.70	27.00
2016	3.75	5.36	18.90	0.18	60.92	64.12
2017	4.27	6.08	21.90	0.20	65.56	69.52
2018	4.94	6.86	25.16	0.21	70.28	75.13

Table 5.21 – Pre-unblinding absolute systematic uncertainties on $B^+ \rightarrow K^+ \tau^- \mu^+$ and $B^+ \rightarrow K^+ \tau^+ \mu^-$ signal yields. The last line gives the total systematic uncertainty which is the sum in quadrature of the systematic uncertainties of all the individual sources. The final systematic uncertainties determined on data after unblinding are given in Tab. 5.25 in Sec. 5.5.

Source	$B^+ \rightarrow K^+ \tau^- \mu^+$	$B^+ \rightarrow K^+ \tau^+ \mu^-$
Fit bias	1.35	3.08
Fixed background parameters	20.98	29.37
Shared f_{phys} parameter	1.34	5.33
Signal model	7.45	7.32
Fixed signal parameters	2.27	5.90
Total	22.46	31.45

5.4.4 Expected results including systematic effects

The values of the expected limits on $\mathcal{B}(B^+ \to K^+\tau^-\mu^+)$ and $\mathcal{B}(B^+ \to K^+\tau^+\mu^-)$ after full selection have been given in Sec. 5.3. However, these values did not take into account the effect of the systematic uncertainties. In order to take into account the effect of the systematic uncertainties determined previously, the likelihood function used in the fit to the data has to be modified with respect to the regular one. Assuming that there are N parameters with an associated systematic uncertainty, the likelihood function becomes

$$\mathcal{L}_c\big(m_{\mathrm{corr}}(B^+);\vec{\theta},\mu_1,\ldots,\mu_N\big) = \mathcal{L}\big(m_{\mathrm{corr}}(B^+);\vec{\theta}\big) \prod_{i=1}^{N} \mathcal{G}(\widetilde{\mu_i};\mu_i,\sigma_i), \tag{5.18}$$

where $\mathcal{L}_c$ and $\mathcal{L}$ are the modified and regular likelihood functions, $\vec{\theta}$ is the set of parameters of the regular model, μ_i and σ_i are the mean value and corresponding systematic uncertainty of the i-th parameter with an associated systematic uncertainty. In this specific case there are two parameters with an associated systematic uncertainty. Indeed, based on Eq. (5.13), the fitted signal yield in data can be expressed as

$$N_{sig} = C \times \mathcal{B}_{sig} + N_{sig}^{syst}, \tag{5.19}$$

where C is a normalisation factor including the various branching fractions involved in the $B^+ \to \overline{D}^0(\to K^+\pi^-)D_s^+(\to K^+K^-\pi^+)$ decay chain, the signal yield of the $B^+ \to \overline{D}^0 D_s^+$ decay mode as well as the selection efficiencies of both the $B^+ \to K^+\tau^\pm\mu^\mp$ and $B^+ \to \overline{D}^0 D_s^+$ decay modes. The fit bias on the signal yield is given by N_{sig}^{syst}. As a result, the values of C and N_{sig}^{syst} are left free in the final fit and the Gaussian functions associated with these parameters in the likelihood function are centred around the measured values of these parameters with a standard deviation parameter given by the systematic uncertainties. The measured values of the normalisation factor are $C = (39.7\pm4.3)\times10^6$ and $C = (39.3\pm4.2)\times10^6$ for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes, respectively. A small technicality is that the fit bias is directly subtracted to the fitted signal yield and the Gaussian function associated with N_{sig}^{syst} is centred around zero instead of the fit bias.

The usual pseudo-dataset generated according to the background-only model is fitted with the full model including the effect of the systematic uncertainties. The resulting $\mathrm{CL_s}$ scans are shown in Fig. 5.25. The corresponding expected limits at 90% and 95% confidence level are

$$\mathcal{B}(B^+ \to K^+\tau^-\mu^+) < 2.0(2.5) \times 10^{-6} \text{ at } 90(5)\% \text{ CL (expected)},$$
$$\mathcal{B}(B^+ \to K^+\tau^+\mu^-) < 2.7(3.3) \times 10^{-6} \text{ at } 90(5)\% \text{ CL (expected)}.$$

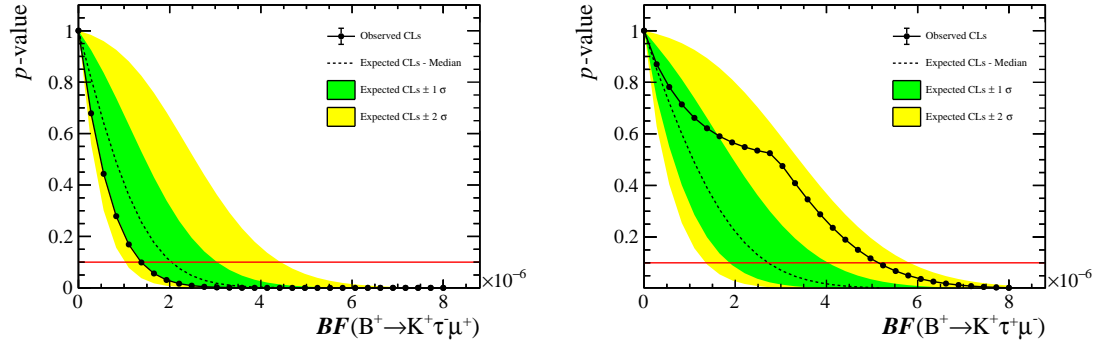These limits are slightly degraded compared to the non-constrained case, as expected.

Figure 5.25 – CL$_\text{s}$ scan scan using full Run 1 + 2 data for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes on the left and right plots, respectively. The red line indicates the 90% CL. These scans include the effect of the systematic uncertainties.

5.5 Results

The data is fitted with the models described previously including the effect of the systematic uncertainties determined on the pseudo-datasets. The results of the fits for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes are shown respectively in Figs. 5.26 and 5.27. In order to obtain the most accurate measurement of the systematic uncertainties as possible, the systematic uncertainties associated with the signal yield are redetermined using the real data. The various sources of systematic uncertainty and the alternative models studied are identical to what has been described in Sec. 5.4 but they are determined by measuring directly the total signal yield in the data. The total signal yields measured using the nominal fit model as well as the alternative models where the values of the Argus function endpoint and slope parameters have been varied are given in Tab. 5.22. The data is also fitted with the full model where the parameter f_{phys} is determined independently for Run 1 and Run 2. The fitted yields and the associated systematic uncertainties are given in Tab. 5.23. Additionally, the data is fitted with the full model where the alternative signal pdf is used. The resulting fitted yields and the associated systematic uncertainties are given in Tab. 5.24. Concerning the parameters fixed from the fits to the signal simulation samples, the systematic uncertainty is determined by fitting one thousand times the data with the full model where the fixed parameters in the signal pdf are sampled from a multivariate Gaussian determined in the aforementioned fits to the simulation samples. The distributions of the fitted signal yield are shown in Fig. 5.28 and the standard deviation of these distributions is assigned as systematic uncertainty. Finally, the values obtained in the study of pseudo-experiments are still used for the bias. The summary of the updated systematic uncertainties is given in Tab. 5.25. Using these systematic uncertainties the final fit is carried on. The results of the fits are shown in Figs. 5.29 and 5.30 for the $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$ decay modes, respectively. The measured values of the branching fractions are

$$\mathcal{B}(B^+ \to K^+ \tau^- \mu^+) = (-1.52 \pm 1.15) \times 10^{-6},$$
$$\mathcal{B}(B^+ \to K^+ \tau^+ \mu^-) = (-4.70 \pm 1.72) \times 10^{-6}.$$

Since the measured values are not significantly positive, upper limits are set on these two decay modes following the procedure described in Sec. 5.2. The CL_s scans are shown in Fig. 5.31 and the corresponding observed limits are

$$\mathcal{B}(B^+ \to K^+ \tau^- \mu^+) < 1.3(1.7) \times 10^{-6} \text{ at } 90(5)\% \text{ CL},$$
$$\mathcal{B}(B^+ \to K^+ \tau^+ \mu^-) < 1.2(1.5) \times 10^{-6} \text{ at } 90(5)\% \text{ CL}.$$
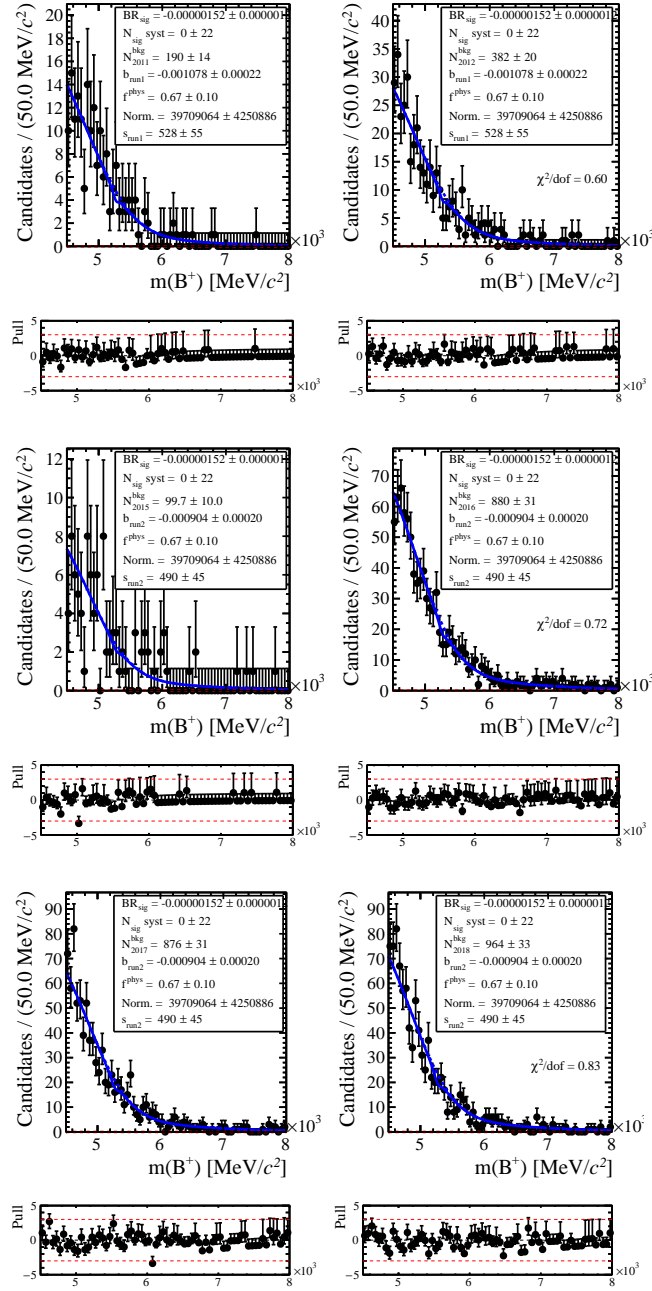
Figure 5.26 – Fit of the $B^+ \to K^+ \tau^- \mu^+$ data using the full Run 1 + 2 dataset including systematic uncertainties derived from pseudo-datasets. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.
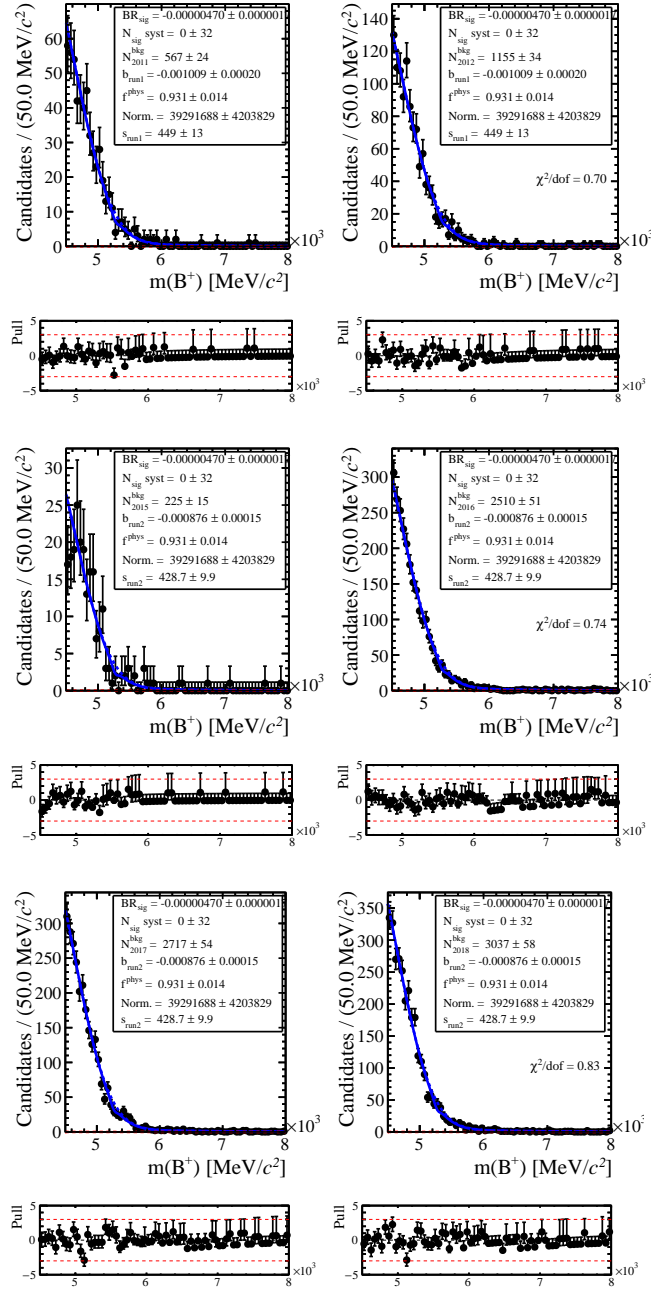
145

Figure 5.27 – Fit of the $B^+ \to K^+ \tau^+ \mu^-$ data using the full Run 1 + 2 dataset including systematic uncertainties derived from pseudo-datasets. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.
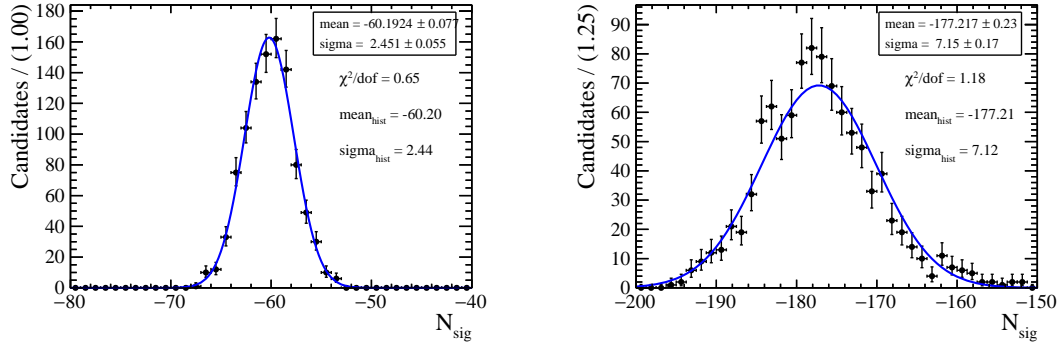
Figure 5.28 – Distribution of total signal yield obtained by fitting the data for the $B^+ \to K^+\tau^-\mu^+$ (left) and $B^+ \to K^+\tau^+\mu^-$ (right) decay modes. The fits use the full model where the shape parameters of the signal pdf have been Gaussian fluctuated according to the fits to the signal simulation samples. The procedure has been repeated 1000 times.

Table 5.22 – The first column gives the variation of the Argus' endpoint in the alternative model. The second and third columns give the total signal yield found by fitting the data using the nominal and alternative models, respectively. The top and bottom halves of the table refer respectively to the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes while the last column gives the associated systematic uncertainty.

Variation	N_{nom}	N_{alt}	σ_{sys}
$+109$ MeV/c^2	-60.73 ± 41.15	-71.27 ± 42.70	15.52
-109 MeV/c^2	-60.73 ± 41.15	-52.63 ± 39.13	15.11
$+52$ MeV/c^2	-182.21 ± 60.43	-163.25 ± 63.21	26.51
-52 MeV/c^2	-182.21 ± 60.43	-195.77 ± 59.12	18.44

Table 5.23 – The first and second columns give the total signal yield found by fitting the data using the nominal model and an alternative model where the fraction of physics background f_{phys} varies independently per run, respectively. The last column gives the associated systematic uncertainty. The top and bottom halves of the table refer to the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes, respectively.

N_{nom}	N_{alt}	σ_{sys}
-60.73 ± 41.15	-60.46 ± 40.15	9.03
-182.21 ± 60.43	-181.65 ± 61.09	8.99

Table 5.24 – The first and second columns give the total signal yield found by fitting the data using the nominal background model plus the nominal and alternative signal models, respectively. The last column gives the associated systematic uncertainty. The top and bottom halves of the table refer to the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes, respectively.

N_{nom}	N_{alt}	σ_{sys}
-60.73 ± 41.15	-59.86 ± 40.31	8.33
-182.21 ± 60.43	-177.65 ± 59.00	13.85

Table 5.25 – Systematic uncertainties on the total signal yields, in absolute value, of the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes. The last line gives the total systematic uncertainty which is the sum in quadrature of all the individual sources. The value associated with the fit bias is taken from the previous pseudo-experiment studies.

Source	$B^+ \to K^+\tau^-\mu^+$	$B^+ \to K^+\tau^+\mu^-$
Fit bias	1.35	3.08
Fixed background parameters	15.52	26.51
Shared f_{phys} parameter	9.03	8.99
Signal model	8.33	13.85
Fixed signal parameters	2.45	7.15
Total	19.99	32.19

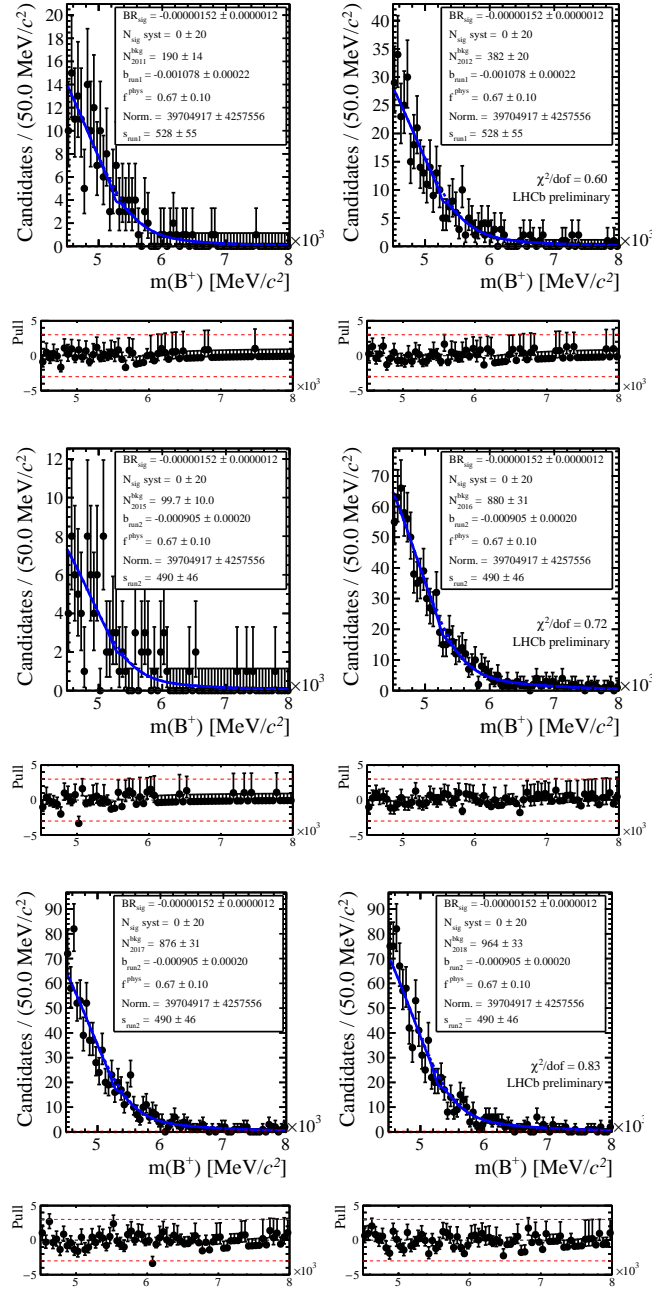Figure 5.29 – Fit of the $B^+ \rightarrow K^+\tau^-\mu^+$ data using the full Run 1 + 2 dataset including systematic uncertainties derived on the real data. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.
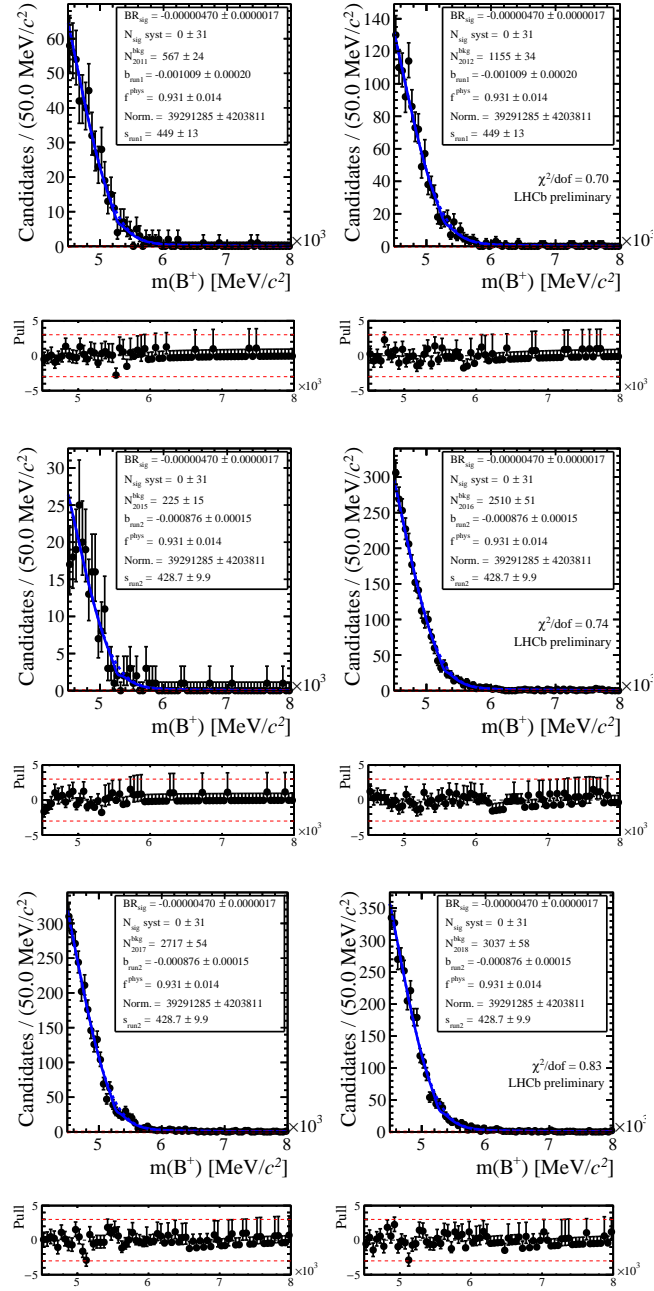
Figure 5.30 – Fit of the $B^+ \to K^+ \tau^+ \mu^-$ data using the full Run 1 + 2 dataset including systematic uncertainties derived on the real data. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.
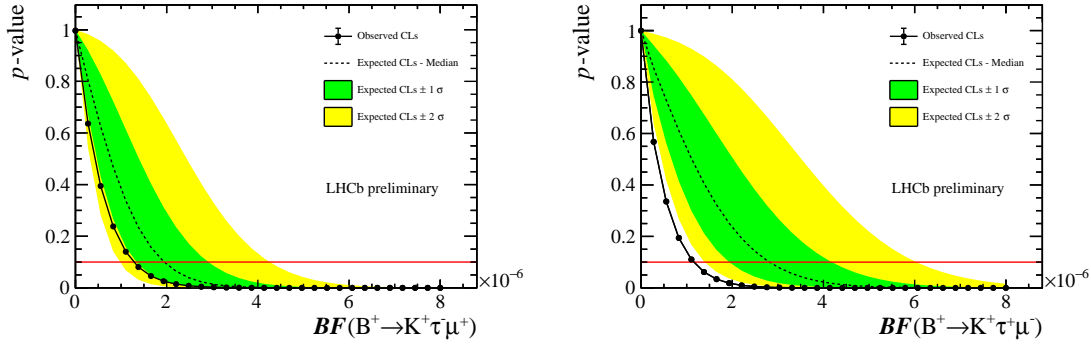
Figure 5.31 – CL$_s$ scan scan using full Run 1 + 2 data for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decay modes on the left and right plots, respectively. The red line indicates the 90% CL. These scans include the effect of the systematic uncertainties determined on the real data.

5.5.1 Further study of the $B^+ \to K^+\tau^+\mu^-$ decay mode

In the case of the $B^+ \to K^+\tau^+\mu^-$ decay mode, the measured branching fraction displays a sizeable tension with the null hypothesis corresponding to $\mathcal{B}(B^+ \to K^+\tau^+\mu^-) = 0$. If the measured value is assumed to be Gaussian distributed, the tension corresponds to 2.7σ where σ is the standard deviation. This corresponds to a one-sided p-value of 0.35%. The usual convention within the LHCb collaboration is to define the 3σ and 5σ values as the limits corresponding respectively to an "evidence" and a "discovery" in the case of a positive signal. These cases correspond respectively to a one-sided p-value of 0.13% and 2.87×10^{-7}. In this case since the measured branching fraction is negative it means that the fluctuation cannot be due to real signal and as such there clearly is no indication of the presence of real $B^+ \to K^+\tau^+\mu^-$ candidates in the data. Given that the tension of -2.7σ is relatively sizeable there are mainly two possible interpretations of the result. Either the tension is due to a statistical downwards fluctuation of the background, or the fit model used to describe the background content in data is not sufficiently accurate and as such overestimates the background yield in the signal region, which in turn is artificially compensated with a negative signal yield. In the following, various cross-checks are presented in order to shed light on which of the two hypotheses should be accepted.

The first study which is performed is to determine whether the hypothesis of a Gaussian distributed measurement holds true or not and whether the previously quoted p-values are correct. In order to do so, the background component parameters from the final fit, shown in Fig. 5.30, is used to generate 1000 pseudo-datasets which are then fitted using the full model, including the effect of the systematic uncertainties. The distribution of the fitted branching fraction for the 1000 pseudo-datasets is shown in Fig. 5.32 and is fitted with a Gaussian function. The p-value corresponding to the measured branching fraction in

the nominal fit is then determined in two ways, either by counting the number of pseudo-datasets where the fitted branching fraction is more extreme than the nominal result, or by integrating the fitted Gaussian up to the nominal result. From the 1000 pseudo-datasets 3 give a measured branching fraction more extreme than the nominal one, corresponding to a p-value of $\left(0.30^{+0.23}_{-0.13}\right)\%$ where the uncertainties are given by the 68% confidence interval. The p-value obtained by integrating the fitted Gaussian distribution is 0.14%. This indicates that the measured branching fraction does indeed follow a Gaussian distribution and as such the previously quoted p-values are valid.
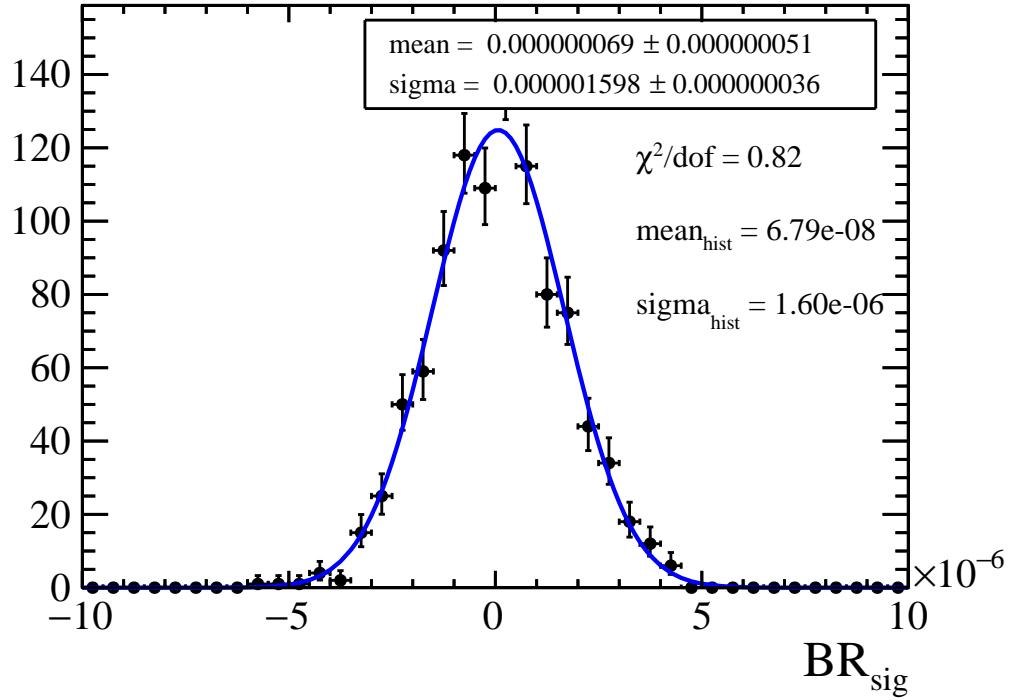

Figure 5.32 – Distribution of the fitted $\mathscr{B}(B^+ \rightarrow K^+ \tau^+ \mu^-)$ from 1000 pseudo-datasets generated according to the background contribution in the nominal fit to the data and fitted with the full model.

In order to test the validity of the background model, the data is fitted from three different starting points : $4.7\,\mathrm{GeV}/c^2$, $4.9\,\mathrm{GeV}/c^2$ and $5.0\,\mathrm{GeV}/c^2$. If the value of the fitted signal branching fraction varies significantly as a function of value of the fit starting point it would favour the hypothesis of a failure from the background model to appropriately describe the background content in the data. The fits are shown in the appendix in Figs. A.1, A.2 and A.3 while the values of the fitted branching fraction are summarised in Tab. 5.26. Instead of converging towards zero, the value of $\mathscr{B}(B^+ \rightarrow K^+ \tau^+ \mu^-)$ is stable for the different starting mass values considered

in the fit. To further investigate the statistical fluctuation that can be expected, a set of 1000 pseudo-experiments are performed by generating data according to the background model in the nominal fit to the data and fitting them from various starting points. As previously, the distributions of fitted branching fraction are fitted with a Gaussian function. These are shown in Figs. 5.33. Additionally, the distributions of the difference between the fitted branching fraction for the nominal and alternative starting points are also shown in Fig. 5.33 and fitted with a Gaussian function. The p-values associated with these distributions are computed as previously described and given in Tab. 5.27 and can be compared to the values in Tab. 5.26 found by assuming that the measurements are Gaussian distributed. Based on these observations, the measured branching fraction is statistically constant, independently of the starting point used in the fit and as such this supports the hypothesis that the measured value of $\mathcal{B}(B^+ \to K^+\tau^+\mu^-)$ is indeed due to a real downwards statistical fluctuation of the background.

Table 5.26 – The second line gives the measured $\mathcal{B}(B^+ \to K^+\tau^+\mu^-)$ when fitting the data with the nominal model including the systematic uncertainties determined on the real data for various values of the lowest $m_{\mathrm{corr}}(B^+)$ considered in the fit. The third line gives the associated p-value assuming a Gaussian measurement.

Starting point	4.5 GeV/c^2	4.7 GeV/c^2	4.9 GeV/c^2	5.0 GeV/c^2
$\mathcal{B}(B^+ \to K^+\tau^+\mu^-) \times 10^6$	-4.70 ± 1.72	-4.89 ± 1.88	-5.66 ± 1.84	-5.18 ± 1.81
p-value	0.32%	0.47%	0.11%	0.21%

Table 5.27 – The second line gives the p-value corresponding to the measured $\mathcal{B}(B^+ \to K^+\tau^+\mu^-)$ when fitting the data with the nominal model including the systematic uncertainties determined on the real data for various values of the lowest $m_{\mathrm{corr}}(B^+)$ considered in the fit by counting the number of pseudo-experiments yielding a value more extreme. The third line gives the associated p-value computed by integrating the Gaussian fit to the pseudo-experiments distribution shown in Fig. 5.33.

Starting point	4.5 GeV/c^2	4.7 GeV/c^2	4.9 GeV/c^2	5.0 GeV/c^2
p-value from counting	$0.30^{+0.23}_{-0.13}$%	$0.50^{+0.28}_{-0.18}$%	$0.40^{+0.25}_{-0.16}$%	$0.70^{+0.32}_{-0.22}$%
p-value from integration	0.14%	0.14%	0.04%	0.17%

The final test is to fit the data with the background-only model to see if it is able to describe accurately the data without the need for the signal component. The result of the fit is shown in the appendix in Fig. A.4 where it can be observed that the background model is indeed able to describe well the observed data. There is no indication from these fits that the background model shape is unable to describe reliably the B^+ corrected mass distribution.
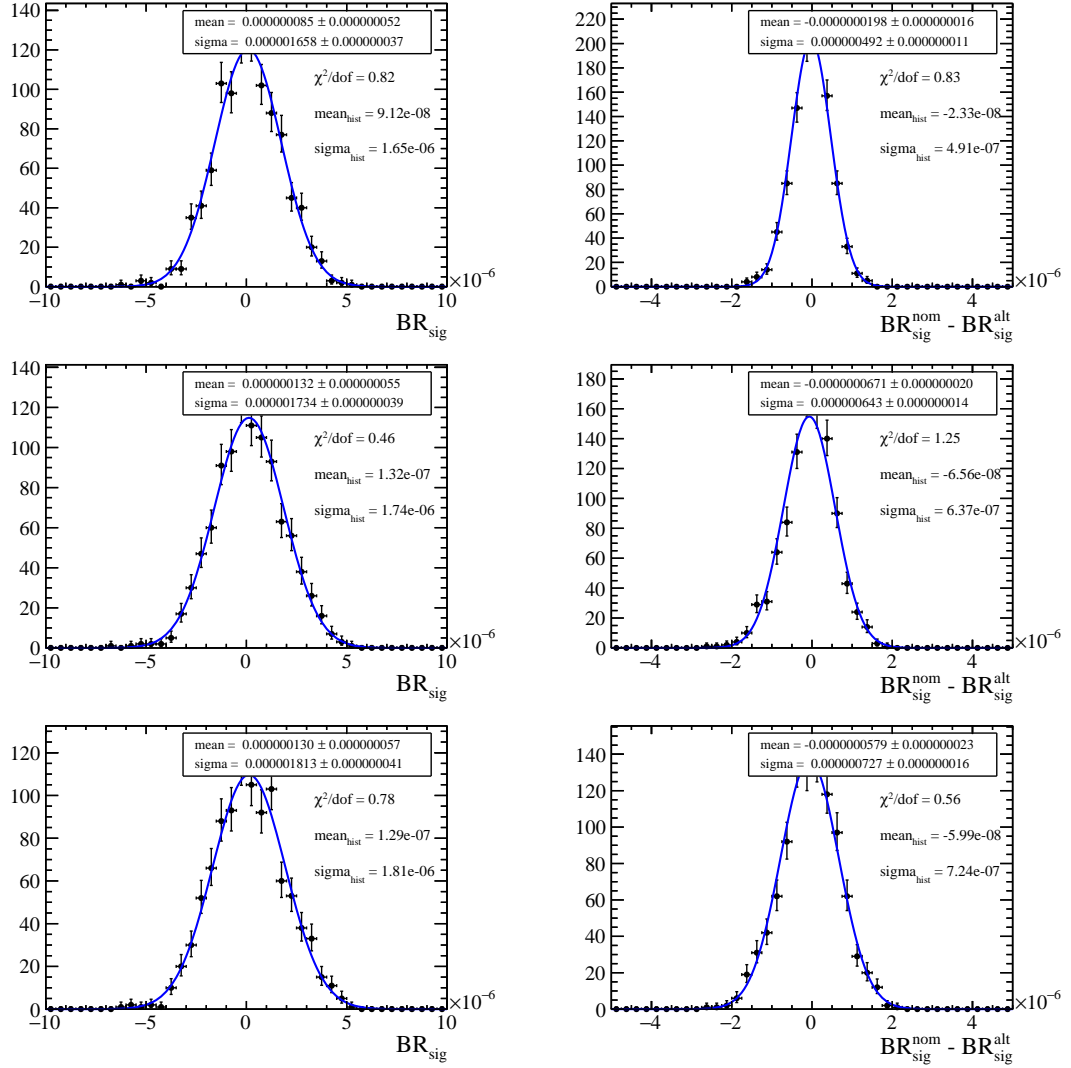
Figure 5.33 – Each line shows on the left the distribution of the fitted $\mathscr{B}(B^+ \to K^+\tau^+\mu^-)$ from 1000 pseudo-datasets generated according to the background contribution in the fit to the data where the starting point has been increased with respect to the nominal fit and fitted with the full model while the plot on the right shows the difference between the fitted branching fraction for a given pseudo-dataset associated with the nominal or alternative starting point. The first, second and third line correspond respectively to the cases where the alternative fit starts at 4.7, 4.9 and 5.0 GeV/c^2.

Based on these various tests, the tension observed in the measured value of $\mathscr{B}(B^+ \to K^+\tau^+\mu^-)$ is interpreted as being due to a statistical fluctuation of the background and therefore no action is taken to alter the measurement. The previously mentioned upper limits

$$\mathscr{B}(B^+ \to K^+\tau^-\mu^+) < 1.3(1.7) \times 10^{-6} \text{ at } 90(5)\% \text{ CL},$$
$$\mathscr{B}(B^+ \to K^+\tau^+\mu^-) < 1.2(1.5) \times 10^{-6} \text{ at } 90(5)\% \text{ CL},$$

are therefore the final results of the searches for the $B^+ \to K^+\tau^-\mu^+$ and $B^+ \to K^+\tau^+\mu^-$ decays.

6 Conclusion

The field of elementary particle physics is in a pivotal moment of its history. The current formulation of the standard model was finalised in the middle of the 1970s and the experimental discovery of the last remaining elementary particle postulated by the SM, the Higgs boson, was made in 2012. The SM has proven to be extremely successful in describing the properties and interactions of elementary particles and there has been a plethora of measurements which agree to a very high degree of accuracy with its predictions. Based on the known limitations of the SM, many extensions of the SM have been proposed to predict and describe phenomenons referred to as physics "beyond the standard model" (BSM). However, while there have been many measurements searching for BSM physics, there are currently no measurement, excluding neutrino oscillations, which can only be explained by BSM physics.

Currently the sub-field of high precision measurements of decays containing *b* or *c* quarks has been receiving a lot of attention. In particular, there has been a multitude of measurements whose combined results point towards a breaking of the hypothesis of lepton flavour universality. As a result, a significant effort is being made, both from the theoretical and experimental communities, to clarify the situation and possibly provide a consistent explanation of the observed anomalies. As part of this effort, there is a growing interest in searches for lepton flavour violating (LFV) decays.

This thesis presented the search for two LFV decay modes : $B^+ \to K^+ \tau^- \mu^+$ and $B^+ \to K^+ \tau^+ \mu^-$. The measurements have been performed by using data acquired by the LHCb detector at CERN, corresponding to $9.1 \, \text{fb}^{-1}$ of integrated luminosity in pp collisions at centre-of-mass energies of $\sqrt{s} = 7, 8$ and $13 \, \text{TeV}$. An efficient selection procedure of the data, which included the use of machine learning techniques, had to be devised in order to maximise the purity of the signal and an innovative mass reconstruction algorithm was used in order to manage the presence of neutral particles in the final state. No significant signal was found for any of the decay modes and therefore upper limits have been set on their branching fractions. The upper

limits at 90% confidence level are $\mathcal{B}(B^+ \to K^+\tau^-\mu^+) < 1.3 \times 10^{-6}$ and $\mathcal{B}(B^+ \to K^+\tau^+\mu^-) < 1.2 \times 10^{-6}$, which corresponds to more than an order of magnitude improvement compared to the current world best values. These values are at a level where they can be used to constrain some new physics models, which predict the existence of these decay modes, as can be seen for example in Fig. 1.5.

Lepton flavour violation searches in general, and the two decay modes studied in this thesis in particular, are important tools to further the knowledge of elementary particle physics. Concerning the LHCb experiment, the current and future upgrades are designed with the goal of collecting data corresponding to an integrated luminosity of $50\,\mathrm{fb}^{-1}$ by the end of 2030. With such an increase in statistics, measurements of LFV decay modes will allow to probe tiny branching fractions and therefore highly constrain new physics models. The methods developed in this thesis, in particular the modification of the "Decay Tree Fitter" algorithm could prove useful for future searches involving neutrinos and neutral pions and help in improving the sensitivity reach of these analyses.

A Appendix

Here are presented mass fit plots used in the validation of the measured result for the $B^+ \to K^+ \tau^+ \mu^-$ decay mode. The validity of the background model used when fitting the data is tested by fitting from three different starting points. The fits for the $4.7\,\text{GeV}/c^2$, $4.9\,\text{GeV}/c^2$ and $5.0\,\text{GeV}/c^2$ starting points are shown respectively in Figs. A.1, A.2 and A.3. The data is also fitted with the background-only model to ensure a good fit quality even without including the signal component. The fit is shown in Fig. A.4.
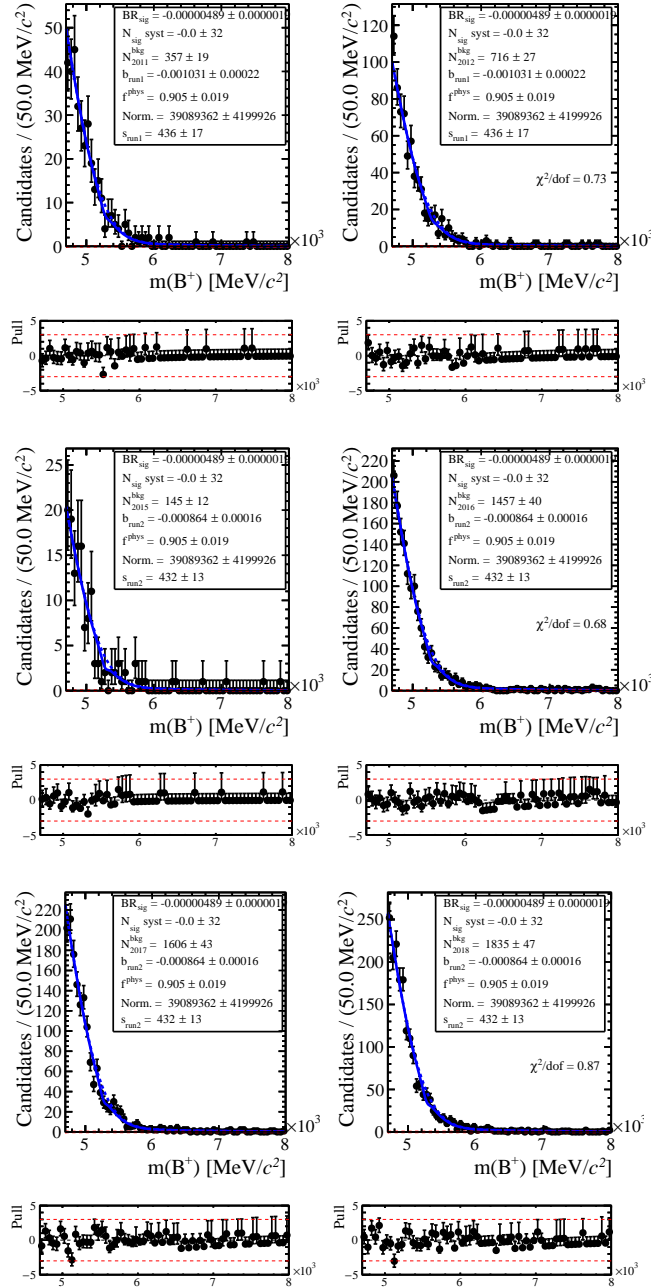
Figure A.1 – Fit of the $B^+ \to K^+ \tau^+ \mu^-$ data using the full Run 1 + 2 dataset including systematic uncertainties derived on the real data. Instead of fitting the data from 4.5 GeV/c^2 as usual it is instead fitted from 4.7 GeV/c^2. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.

Figure A.2 – Fit of the $B^+ \to K^+ \tau^+ \mu^-$ data using the full Run 1 + 2 dataset including systematic uncertainties derived on the real data. Instead of fitting the data from $4.5\,\mathrm{GeV}/c^2$ as usual it is instead fitted from $4.9\,\mathrm{GeV}/c^2$. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.
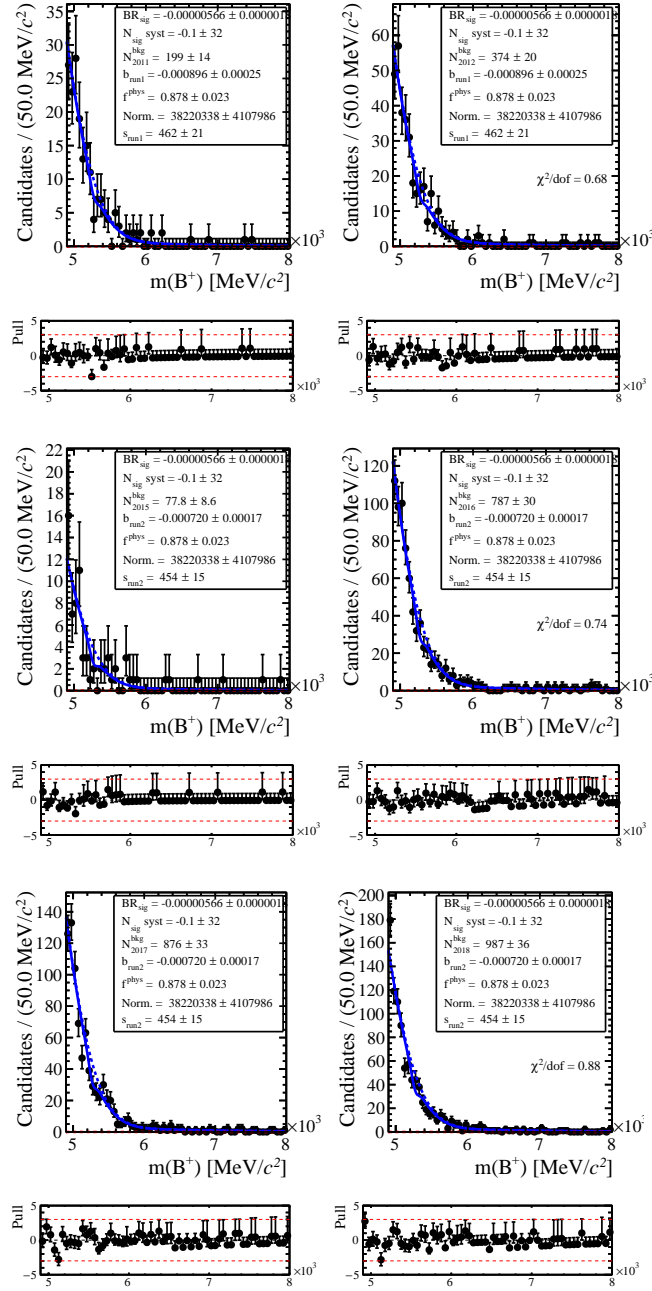
Figure A.3 – Fit of the $B^+ \to K^+ \tau^+ \mu^-$ data using the full Run 1 + 2 dataset including systematic uncertainties derived on the real data. Instead of fitting the data from 4.5 GeV/c^2 as usual it is instead fitted from 5.0 GeV/c^2. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model. The dashed blue and red lines are the background and signal components, respectively.
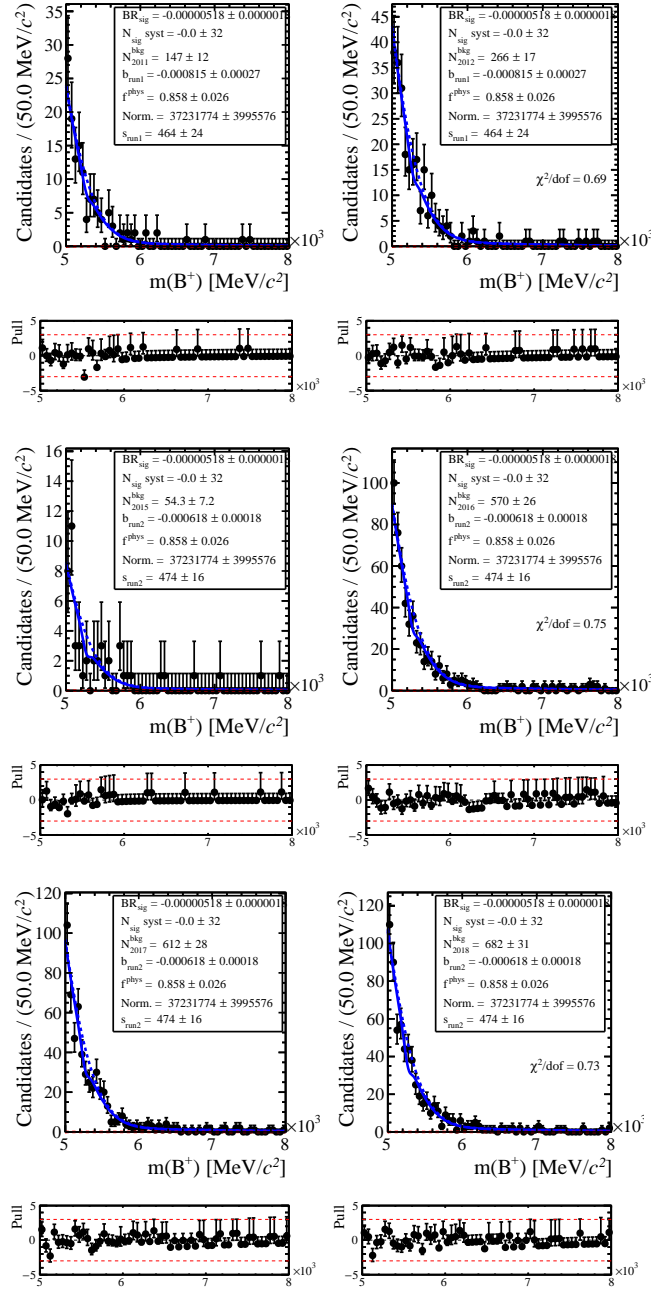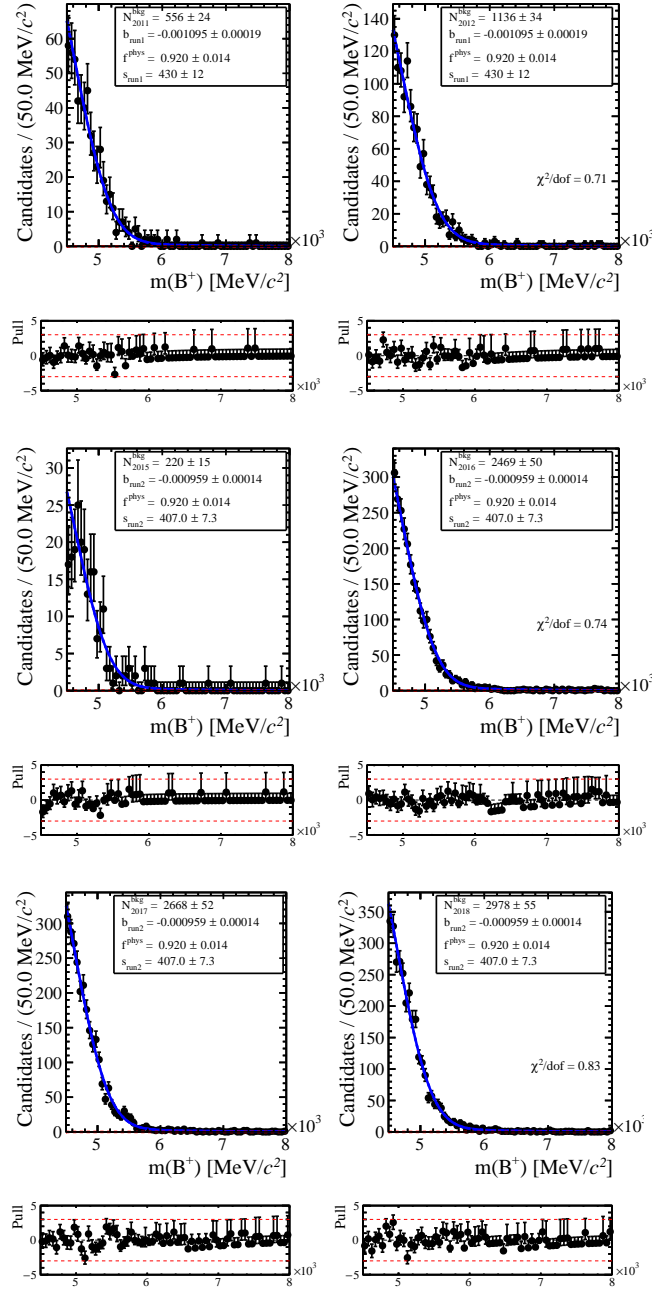
Figure A.4 – Fit of the $B^+ \to K^+ \tau^+ \mu^-$ data using only the background model. The top, middle and bottom lines show the 2011 and 2012, 2015 and 2016, 2017 and 2018 samples, respectively. The blue solid line is the total model.

Bibliography

[1] ATLAS, 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*, Phys. Lett. B **716** (2012) 1, `arXiv:1207.7214`.

[2] CMS, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, `arXiv:1207.7235`.

[3] M. Lubej, *Standard Model*, https://www.physik.uzh.ch/groups/serra/StandardModel.html.

[4] D. Hanneke, S. Hoogerheide, and G. Gabrielse, *Cavity Control of a Single-Electron Quantum Cyclotron: Measuring the Electron Magnetic Moment*, Phys. Rev. A **83** (2011) 052122, `arXiv:1009.4831`.

[5] A. Shomer, *A Pedagogical explanation for the non-renormalizability of gravity*, `arXiv:0709.3555`.

[6] P. Salucci, *The distribution of dark matter in galaxies*, Astron. Astrophys. Rev. **27** (2019), no. 1 2, `arXiv:1811.08843`.

[7] Supernova Search Team, A. G. Riess *et al.*, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, Astron. J. **116** (1998) 1009, `arXiv:astro-ph/9805201`.

[8] Supernova Cosmology Project, S. Perlmutter *et al.*, *Measurements of Ω and Λ from 42 high redshift supernovae*, Astrophys. J. **517** (1999) 565, `arXiv:astro-ph/9812133`.

[9] A. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Sov. Phys. Usp. **34** (1991), no. 5 392.

[10] A. Riotto, *Theories of baryogenesis*, in *ICTP Summer School in High-Energy Physics and Cosmology*, pp. 326–436, 7, 1998. `arXiv:hep-ph/9807454`.

[11] C. Baker *et al.*, *An Improved experimental limit on the electric dipole moment of the neutron*, Phys. Rev. Lett. **97** (2006) 131801, `arXiv:hep-ex/0602020`.

Bibliography

[12] J. M. Pendlebury *et al.*, *Revised experimental upper limit on the electric dipole moment of the neutron*, Phys. Rev. D **92** (2015), no. 9 092003, `arXiv:1509.04411`.

[13] G. F. Giudice, *Naturalness after LHC8*, PoS **EPS-HEP2013** (2013) 163, `arXiv:1307.7879`.

[14] B. Cleveland *et al.*, *Measurement of the solar electron neutrino flux with the Homestake chlorine detector*, Astrophys. J. **496** (1998) 505.

[15] MINOS, D. Michael *et al.*, *Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam*, Phys. Rev. Lett. **97** (2006) 191801, `arXiv:hep-ex/0607088`.

[16] Super-Kamiokande, K. Abe *et al.*, *Evidence for the Appearance of Atmospheric Tau Neutrinos in Super-Kamiokande*, Phys. Rev. Lett. **110** (2013), no. 18 181802, `arXiv:1206.0328`.

[17] LHCb, R. Aaij *et al.*, *Search for lepton-universality violation in $B^+ \to K^+ \ell^+ \ell^-$ decays*, Phys. Rev. Lett. **122** (2019), no. 19 191801, `arXiv:1903.09252`.

[18] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0} \ell^+ \ell^-$ decays*, JHEP **08** (2017) 055, `arXiv:1705.05802`.

[19] LHCb, 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*, Phys. Rev. D **97** (2018), no. 7 072013, `arXiv:1711.02505`.

[20] LHCb, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathscr{B}(\bar{B}^0 \to D^{*+} \tau^- \bar{\nu}_\tau)/\mathscr{B}(\bar{B}^0 \to D^{*+} \mu^- \bar{\nu}_\mu)$*, Phys. Rev. Lett. **115** (2015), no. 11 111803, `arXiv:1506.08614`, [Erratum: Phys.Rev.Lett. 115, 159901 (2015)].

[21] LHCb, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0} \mu^+ \mu^-$ decay using $3\,fb^{-1}$ of integrated luminosity*, JHEP **02** (2016) 104, `arXiv:1512.04442`.

[22] S. L. Glashow, D. Guadagnoli, and K. Lane, *Lepton flavor Violation in B decays?*, Phys. Rev. Lett. **114** (2015) 091801, `arXiv:1411.0565`.

[23] L. Galli, *Lepton Flavour violation in muon decays*, in *17th Conference on Flavor Physics and CP Violation*, 6, 2019. `arXiv:1906.10483`.

[24] LHCb, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, `arXiv:2103.11769`.

[25] D. Buttazzo, A. Greljo, G. Isidori, and D. Marzocca, *B-physics anomalies: a guide to combined explanations*, JHEP **11** (2017) 044, `arXiv:1706.07808`.

[26] G. Hiller, D. Loose, and K. Schönwald, *Leptoquark flavor patterns & B decay anomalies*, JHEP **12** (2016) 027, `arXiv:1609.08895`.

166

[27] A. Angelescu, D. Bečirević, D. Faroughy, and O. Sumensari, *Closing the window on single leptoquark solutions to the B-physics anomalies*, JHEP **10** (2018) 183, `arXiv:1808.08179`.

[28] M. Bordone, C. Cornella, J. Fuentes-Martín, and G. Isidori, *Low-energy signatures of the* PS^3 *model: from B-physics anomalies to LFV*, JHEP **10** (2018) 148, `arXiv:1805.09328`.

[29] C. Cornella, J. Fuentes-Martin, and G. Isidori, *Revisiting the vector leptoquark explanation of the B-physics anomalies*, JHEP **07** (2019) 168, `arXiv:1903.11517`.

[30] BaBar, J. Lees *et al.*, *A search for the decay modes* $B^{+-} \to h^{+-}\tau^{+-}l$, Phys. Rev. D **86** (2012) 012004, `arXiv:1204.2852`.

[31] LHCb, R. Aaij *et al.*, *Search for the lepton flavour violating decay* $B^+ \to K^+\mu^-\tau^+$ *using* B_{s2}^{*0} *decays*, JHEP **06** (2020) 129, `arXiv:2003.04352`.

[32] E. Fermi, *An attempt of a theory of beta radiation. 1.*, Z. Phys. **88** (1934) 161.

[33] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (2020) .

[34] A. Pich, *Effective field theory: Course*, in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pp. 949–1049, 6, 1998. `arXiv:hep-ph/9806303`.

[35] B. Grinstein, R. P. Springer, and M. B. Wise, *Effective Hamiltonian for Weak Radiative B Meson Decay*, Phys. Lett. B **202** (1988) 138.

[36] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, Rev. Mod. Phys. **68** (1996) 1125, `arXiv:hep-ph/9512380`.

[37] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[38] W. Altmannshofer and P. Stangl, *New Physics in Rare B Decays after Moriond 2021*, `arXiv:2103.13370`.

[39] HFLAV, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. **C81** (2021) 226, `arXiv:1909.12524`, updated results and plots available at `https://hflav.web.cern.ch/`.

[40] R. A. Diaz, R. Martinez, and C. E. Sandoval, *Improving bounds on flavor changing vertices in the two Higgs doublet model from B0 - anti-B0 mixing*, Eur. Phys. J. C **46** (2006) 403, `arXiv:hep-ph/0509194`.

[41] A. Ilakovac, *Lepton flavor violation in the standard model extended by heavy singlet Dirac neutrinos*, Phys. Rev. D **62** (2000) 036010, `arXiv:hep-ph/9910213`.

Bibliography

[42] J. C. Pati and A. Salam, *Lepton number as the fourth color*, Phys. Rev. **D10** (1974) 275.

[43] A. Crivellin *et al.*, *Lepton-flavour violating B decays in generic Z' models*, Phys. Rev. D **92** (2015), no. 5 054013, `arXiv:1504.07928`.

[44] D. Bečirević, S. Fajfer, N. Košnik, and O. Sumensari, *Leptoquark model to explain the B-physics anomalies, R_K and R_D*, Phys. Rev. D **94** (2016), no. 11 115021, `arXiv:1608.08501`.

[45] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[46] L. Evans and P. Bryant, *LHC machine*, Journal of Instrumentation **3** (2008) S08001.

[47] S. Myers, *The LEP Collider, from design to approval and commissioning*, John Adams' Lecture, CERN, Geneva, 1991. Delivered at CERN, 26 Nov 1990, doi: 10.5170/CERN-1991-008.

[48] J. Jowett, *Colliding Heavy Ions in the LHC*, .

[49] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[50] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[51] CMS, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[52] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, Dec, 2008.

[53] CERN, *Proceedings of the Workshop on Standard Model Physics (and more) at the LHC: CERN, Geneva, Switzerland 25 - 26 May 1999. CERN Workshop on Standard Model Physics (and more) at the LHC*, (Geneva), CERN, 2000. doi: 10.5170/CERN-2000-004.

[54] LHCb Collaboration, *Large hadron collider beauty experiment public results website*, http://lhcb-public.web.cern.ch/.

[55] LHCb collaboration, C. Elsässer, $\overline{b}b$ *production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html.

[56] F. Follin and D. Jacquet, *Implementation and experience with luminosity levelling with offset beam*, in *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, pp. 183–187, 2014. `arXiv:1410.3667`. doi: 10.5170/CERN-2014-004.183.

[57] LHCb Silicon Tracker Group, C. Abellan Beteta *et al.*, *Monitoring radiation damage in the LHCb Tracker Turicensis*, JINST **15** (2020), no. 08 P08016, `arXiv:1809.05063`.

[58] LHCb Collaboration, S. Amato *et al.*, *LHCb magnet: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2000.

[59] LHCb Outer Tracker Group, R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014), no. 01 P01002, `arXiv:1311.3893`.

[60] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, `arXiv:1211.6759`.

[61] C. Lippmann, *Particle identification*, Nucl. Instrum. Meth. A **666** (2012) 148, `arXiv:1101.3276`.

[62] LHCb collaboration, *Trigger schemes*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html.

[63] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[64] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.

[65] LHCb, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[66] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[67] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, `arXiv:hep-ph/0506026`.

[68] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[69] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[70] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81.

[71] *LHCb Trigger and Online Upgrade Technical Design Report*, Tech. Rep. CERN-LHCC-2014-016. LHCB-TDR-016, May, 2014.

[72] L. Collaboration, *LHCb PID Upgrade Technical Design Report*, Tech. Rep. CERN-LHCC-2013-022. LHCB-TDR-014, Nov, 2013.

Bibliography

[73] L. Collaboration, *LHCb VELO Upgrade Technical Design Report,* Tech. Rep. CERN-LHCC-2013-021. LHCB-TDR-013, Nov, 2013.

[74] L. Collaboration, *LHCb Tracker Upgrade Technical Design Report,* Tech. Rep. CERN-LHCC-2014-001. LHCB-TDR-015, Feb, 2014.

[75] LHCb SciFi Tracker Collaboration, P. Hopchev, *SciFi: A large Scintillating Fibre Tracker for LHCb. SciFi: A large Scintillating Fibre Tracker for LHCb,* Tech. Rep. arXiv:1710.08325, Oct, 2017. Presented at The Fifth Annual Conference on Large Hadron Collider Physics.

[76] O. G. Girard, *Development of the scintillating fibre tracker technology for the LHCb upgrade and the LHC beam profile monitoring system,* PhD thesis, Ecole Polytechnique, Lausanne, 2018, doi: 10.5075/epfl-thesis-8851.

[77] A. Kuonen, *Development and Characterisation of Silicon Photomultiplier Multichannel Arrays for the Readout of a Large Scale Scintillating Fibre Tracker,* PhD thesis, Ecole Polytechnique, Lausanne, 2018, doi: 10.5075/epfl-thesis-8842.

[78] S. Jadach, J. H. Kuhn, and Z. Was, *TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons,* Comput. Phys. Commun. **64** (1990) 275.

[79] BaBar, P. del Amo Sanchez *et al.*, *Dalitz plot analysis of $D_s^+ \to K^+ K^- \pi^+$,* Phys. Rev. D **83** (2011) 052001, `arXiv:1011.4190`.

[80] SLD, K. Abe *et al.*, *A measurement of R_b using a vertex mass tag,* Phys. Rev. Lett. **80** no. 4 .

[81] W. D. Hulsbergen, *Decay chain fitting with a kalman filter,* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **552** (2005) 566–575.

[82] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system,* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016. doi: 10.1145/2939672.2939785.

[83] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in Python,* Journal of Machine Learning Research **12** (2011) 2825.

[84] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions,* Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`.

[85] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2,* EPJ Tech. Instrum. **6** (2019), no. 1 1, `arXiv:1803.00824`.

[86] LHCb, R. Aaij *et al.*, *Measurement of b hadron fractions in 13 TeV pp collisions*, Phys. Rev. D **100** (2019), no. 3 031102, `arXiv:1902.06794`.

[87] LHCb, R. Aaij *et al.*, *Observation of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays*, Phys. Rev. D **87** (2013), no. 11 112012, `arXiv:1304.4530`, [Addendum: Phys.Rev.D 89, 019901 (2014)].

[88] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. G **28** (2002) 2693.

[89] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, `arXiv:1007.1727`, [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[90] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014.
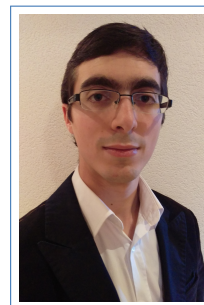
Lino Ferreira Lopes

Rue des Vergers 15
1994 Aproz
☏ *078 878 34 32*
✉ *linol17_prt@hotmail.com*

Education

2017–2021	**Ph.D. in Experimental Particle Physics**, *EPFL, Lausanne*, Laboratory of High Energy Physics (LPHE).
	Thesis title: *Search for lepton flavour violation in B meson decays at LHCb*
2015–2017	**Master of Science in Physics**, *EPFL, Lausanne*, LPHE.
	Thesis title: *Study of the $B_s^0 \to D_s^{(*)-}\mu^+\nu_\mu$ decays at LHCb*
	Advanced courses *e.g.* Quantum mechanics, Quantum field theory, Cosmology
2012–2015	**Bachelor of Science in Physics**, *EPFL*, Lausanne.

Work Experience

Ph.D. in particle physics

- Used statistical and machine learning tools to analyse large datasets acquired by the LHCb experiment at CERN in order to study hypothetical extensions of the current best description of the physical laws of the universe. I developed innovative data reconstruction techniques and modelled complex structures in order to maximise signal sensitivity and perform the current world's best measurement of the parameters of interest.
- Acted as the liaison between two different working groups and was responsible of setting up the properties of Monte Carlo simulation needed by my colleagues.
- Involved in the manufacture and quality assurance testing of a tracking detector using a novel scintillating fibre technology.
- Designed and supervised a data analysis project of a Master level student.
- Responsible of teaching and grading bachelor level students courses, both theoretical and experimental.
- Popularised particle physics at various events aimed at high school students as well as the general public.

Technical and Language Skills

Programming	Python, C/C++, Matlab
Data analysis	Scikit-learn, XGBoost, ROOT, Pandas
Other	LaTeX, Git, Microsoft Office, OpenOffice
Languages	French (native), Portuguese (native), English (fluent)

173