Measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction

Veronica Sølund Kirsebom

EPFL

To those who make me laugh. . .

Acknowledgements

The PhD has been five challenging, interesting and informative years full of experiences that I would not be without. I would therefore like to thank Michel De Cian, Olivier Schneider, Aurelio Bay, Tatsuya Nakada, Frederic Blanc and Guido Hafaeli for accepting me as a PhD student in this lab. My thanks extend to Mirco Dorigo who introduced me, when I was a summer student at CERN, to the exciting physics of the LHCb experiment and who motivated me to apply for a PhD position at EPFL. I am deeply grateful to Michel for competently supervising me from the start to the end of my PhD project and for always showing interest in my work. I would also like to thank my thesis director Olivier for always providing me with detailed and useful feedback and for giving me the freedom to organize my research. In addition to this, I enjoyed being a teaching assistant in courses taught by Oliver and Fred - both passionate and inspiring teachers.

I would like to thank all my colleagues during the years for making this lab such a nice place to work. In particular, I would like to thank my first office mates Violaine and Pavol, who made me feel at home and who made me laugh so much. I really appreciate all those random conversations that we had. Thanks should also go to the other PhD students: Brice, Vincenzo, Olivier, Axel, Guido, Matthieu, Vlado, Lino, Guillaume and Sebastian, who provided me with the best start of the PhD, and all those PhD students who joined after me: Surapat, Carina, Marie, Sara, Serhii, Anna, Sonia, Paul, Jennifer, Maria, Federico, Raphaël, Dimitrios and Anni - continuing the great atmosphere of the lab. In addition to this, I would like to express my gratitude to the secretaries Esther Hofmann and Corinne Craman for always being extremely helpful and for organising nice and memorable social events. A special thanks to Corinne, who helped me getting my dream apartment in Lausanne by communicating in French with the Régie.

A big thanks to all my old friends for holding on to me as I moved to Switzerland. Thanks for nice reunions in Denmark, and thanks to those who visited me in Lausanne. A special thanks goes to my friend Naya with whom I have shared so many good times full of interesting and useful conversations. I am also extremely grateful to Pavol, who (believe it or not) kept me sane during the most stressful times of the thesis, and who introduced me to the impressive Slovak diaspora in Switzerland with great people like Eva.

Finally, to my family, mum, dad and brothers. I am deeply grateful for your encouragements, patience and understanding. Thank you for always being there for me - I feel very lucky to have you.

Lausanne, May 24, 2023   V. S. K.

Abstract

The tension between different determinations of the CKM matrix element $|V_{ub}|$ using measurements of either exclusive or inclusive semileptonic decays represents a long-standing puzzle in flavour physics. To resolve this, measurements of new exclusive channels as well as more precise measurements are needed. This thesis presents the first measurement of the exclusive semileptonic decay $B^+ \to \rho^0 \mu^+ \nu_\mu$ at the LHCb experiment, using data collected during 2018. Profiting from the large number of B mesons produced in pp collisions at the LHC, a precise measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction as a function of the dilepton invariant mass squared is performed relative to the decay $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ with $\bar{D}^0 \to \pi^+ \pi^-$. This measurement will be used to obtain a determination of the $|V_{ub}|$ matrix element from the $B^+ \to \rho^0 \mu^+ \nu_\mu$ channel. All steps leading to the final measurement are documented and the current precision of the result is given, with the numerical values of the result still kept blind to avoid accidental biases before the final publication. The measurement presented in this thesis does not only pave the way for obtaining $|V_{ub}|$, but also for measuring the full differential branching fraction that expresses the rich helicity structure of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay and provides more variables to test the Standard Model and to probe new physics.

Résumé

La tension entre les différentes déterminations de l'élément $|V_{ub}|$ de la matrice CKM à partir des mesures de désintégrations semileptoniques exclusives ou inclusives représente une énigme de longue date en physique des saveurs. Pour résoudre ce problème, des mesures de nouveaux canaux exclusifs ainsi que des mesures plus précises sont nécessaires. Cette thèse présente la première mesure de la désintégration semileptonique exclusive $B^+ \to \rho^0 \mu^+ \nu_\mu$ à l'expérience LHCb, en utilisant les données recueillies au cours de l'année 2018. Profitant du grand nombre de mésons B produits dans les collisions pp au LHC, une mesure précise de la fraction d'embranchement différentielle de la désintégration $B^+ \to \rho^0 \mu^+ \nu_\mu$ en fonction du carré de la masse invariante des deux leptons est réalisée relativement au canal de désintégration $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ avec $\bar{D}^0 \to \pi^+ \pi^-$. Cette mesure sera utilisée pour obtenir une détermination de l'élement de matrice $|V_{ub}|$ à partir du canal du signal. Toutes les étapes menant à la mesure finale sont documentées et la précision actuelle du résultat est donnée, les valeurs numériques du résultat restant cachées pour éviter les biais accidentels avant la publication finale.

La mesure présentée dans cette thèse n'ouvre pas seulement la voie à l'obtention de $|V_{ub}|$, mais aussi à la mesure de la fraction d'embranchement différentielle complète qui exprime la riche structure d'hélicité de la désintégration du signal et fournit davantage de variables pour tester le Modèle Standard et sonder la nouvelle physique.

Contents

Contents

Contents

Introduction

While the philosophical idea of matter being made up by small invisible and indivisible *atoms* can be traced back to ancient times, the scientific development, eventually leading to the creation of the Standard Model of Particle Physics (SM), did not materialize before more than two thousand years later.

Starting in the early 19th century, based on experimental observations, Dalton [1] suggested that matter was made of unbreakable atoms, however, already by the beginning of the 20th century, Thomson [2] and Rutherford [3] had demonstrated that Dalton's atom was, in fact, made of *electrons* with negative electric charges orbiting a much heavier positively charged nucleus concentrated at the center of the atom. While the electron was the first elementary particle to be discovered, the nucleus turned out to be a bound state of two types of nucleons - the positive *proton* and the neutral *neutron*. During the second half of the 20th century, where the mathematical framework of the SM was developed along with advances in experimental particle physics, particle colliders powerful enough to probe the structure of the strongly bound proton revealed that it was made up by elementary particles called *quarks*.

The SM has not only proved powerful in explaining experimental observations, but also in predicting the existence of new particles and interaction properties that have subsequently been confirmed experimentally. Most recently, the success of the SM has been illustrated by the discovery of the last unobserved SM particle, the *Higgs boson*, by the ATLAS [4], and CMS [5] experiments in 2012. Today the SM describes twelve elementary particles, six *leptons* (e, μ, τ, ν_e, ν_μ, ν_τ) and six quarks (d, u, s, c, b, t), as well as their properties and interactions with the *electromagnetic*, *weak* and *strong* forces. However, despite the success of the SM, certain phenomena such as *dark matter*, *dark energy* and *neutrino masses* cannot be explained. For this reason physicists are trying to uncover the underlying new physics by challenging the predictions of the SM.

Persistent anomalies observed in studies of lepton and quark properties, also known as *flavour physics*, have received a growing interest. Among these is the long standing tension between different determinations of the $b \to u$ quark transition strength represented by the CKM matrix element V_{ub}. A discrepancy of $\sim 3\sigma$ is observed between measurements based on inclusive and exclusive semileptonic decays. To understand whether this arises from unexplored theoretical or experimental problems, or from a new physics phenomenon,

more precise measurements are needed.

Thus, in this thesis a measurement of the exclusive semileptonic $B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ decay is performed with data collected by CERN's heavy flavour experiment LHCb. Its differential branching fraction is measured as a function of q^2 relative to the normalisation channel $B^+ \to \bar{D}^0(\to \pi^+\pi^-)\mu^+\nu_\mu$. This measurement paves the way for obtaining a precise measurement of the $|V_{ub}|$ matrix element at LHCb, and ultimately, for measuring the full differential branching fraction that provides multiple variables to test the Standard Model and to probe new physics.

This thesis is organised as follows. Chapter 1 provides an introduction to the SM and the field of flavour physics. The motivation and strategy of the presented analysis is given in Chapter 2. Chapter 3 describes the LHC and the LHCb detector at CERN, which has collected the data used to measure the $B^+ \to \rho^0\mu^+\nu_\mu$ decay. In Chapter 4, the signal selection and the control studies are explained. The neutrino reconstruction and computation of q^2 is found in Chapter 5. The fit of the normalisation channel is described in Chapter 6, while the fit of the signal channel in bins of q^2 is performed in Chapter 7. The final result of the $B^+ \to \rho^0\mu^+\nu_\mu$ differential branching fraction is reported in Chapter 8. Finally, a summary and outlook are given in Chapter 9.

1 Flavour physics in the Standard Model

The following sections contain a brief introduction to the Standard Model of Particle Physics followed by a review of selected topics within the field of flavour physics relevant to the study presented in this thesis.

1.1   The Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) [6] describes the fundamental particles and their interactions, except gravity [7]:

- **The strong interaction**, responsible for holding together the nuclei.

- **The electromagnetic interaction**, responsible for electricity, light, magnetism and keeping electrons in orbits around the nucleus, *i.e.* holding together the atoms.

- **The weak interaction**, responsible for the radioactive decay of nuclei as well as the energy production inside the sun via nuclear fusion.

The mathematical formulation of the SM is a quantum field theory that incorporates both quantum mechanics and special relativity [8]. The fundamental particles and their interactions are represented by fields defined in space and time, and their dynamics are described by the gauge invariant SM Lagrangian [9], which conserves the local gauge symmetry:

$$\mathrm{SU(3)}_C \times \mathrm{SU(2)}_\mathrm{L} \times \mathrm{U(1)}_Y \ . \tag{1.1}$$

Here $\mathrm{SU(3)}_C$ is the symmetry group of the strong interaction, while the weak and electromagnetic interactions can be unified into the electroweak interaction (EW) [10] with the symmetry group $\mathrm{SU(2)}_\mathrm{L} \times \mathrm{U(1)}_Y$, which will be explained later in this chapter.
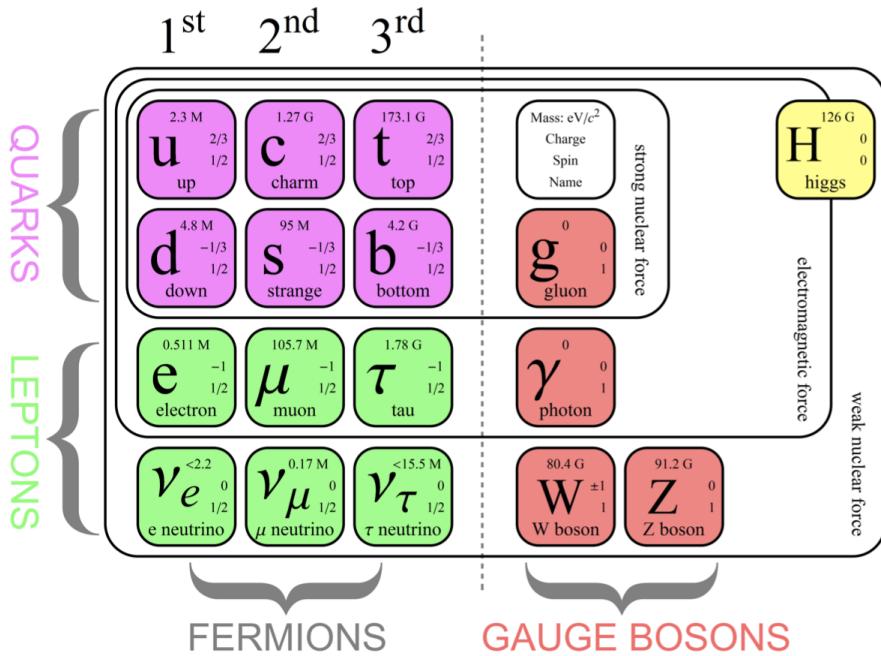
Figure 1.1 – Fundamental particles and their interactions unified in the SM [14].

1.1.1 Particles and their interactions

There are twelve fundamental particles of matter defined by their mass and quantum numbers that determine their interactions. For each particle, there is a corresponding antiparticle with the same mass, but opposite quantum numbers. They are all spin-1/2 particles, also known as fermions [11, 12] and, based on their interactions, they are categorised as either quarks or leptons. They are further divided into three generations, each being a heavier copy of the previous one. While both quarks and leptons interact via the electromagnetic and weak force, except the neutral leptons (neutrinos) that only interact weakly, only quarks can interact via the strong force, which is the defining difference between the two classes of fermions.

Each interaction is mediated by one or more spin-1 particles, also known as gauge bosons [9]. In this way the electromagnetic force is mediated by the photon (γ), the strong force is mediated by the gluon (g) and the weak force is mediated by the three weak bosons W^+, W^- and Z^0. Finally, there is the Higgs boson (H^0) with spin 0 [13]. It is associated with the Higgs field that gives rise to particle masses. All SM particles and their interactions are summarised in Fig. 1.1.

Quarks

Quarks come in six different flavours. There are three up-type quarks with an electric charge of $+2/3e$ (e is the elementary charge of $1.602176634 \times 10^{-19}$ C [15]) corresponding to the quark flavours up (u), charm (c) and top (t). Left are three down-type quarks with

an electric charge of $-1/3e$ corresponding to the quark flavours down (d), strange (s) and bottom (b). These can be arranged into three generations:

$$1^{\text{st}} \ : \begin{pmatrix} u \\ d \end{pmatrix} \ , \ 2^{\text{nd}} \ : \begin{pmatrix} c \\ s \end{pmatrix} \ \text{and} \ 3^{\text{rd}} \ : \begin{pmatrix} t \\ b \end{pmatrix} . \tag{1.2}$$

The second and third generations are heavier copies of the first generation. Quark masses differ by several orders of magnitude between different generations, *e.g.* the u quark belonging to the first generation has a mass around $2.2 \, \text{MeV}/c^2$ [15], while the t quark belonging to the third generation is almost one hundred thousand times heavier with a mass of about $173 \, \text{GeV}/c^2$, making it the heaviest particle of the SM.

Leptons

Leptons come in three different flavours and are either electrically charged or neutral. Charged leptons carry $-1e$ and are known as the electron (e^-), muon (μ^-) and tau (τ^-). Neutral leptons, also known as neutrinos, come in the same three flavours corresponding to ν_e, ν_μ and ν_τ. Like the quarks, leptons are grouped into three generations:

$$1^{\text{st}} \ : \begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \ , \ 2^{\text{nd}} \ : \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \ \text{and} \ 3^{\text{rd}} \ : \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} . \tag{1.3}$$

The second and third generations of leptons are heavier copies of the first generation, however, this is only known to be true for the charged leptons. Neutrinos were for a long time assumed to be massless, until the discovery of neutrino oscillation [16–18]. This proved that at least two neutrino flavours had to be massive, however, these masses are tiny, at least six order of magnitudes below that of any other SM fermions, and they have not yet been precisely determined.

Matter

Most of the matter that we encounter on earth is made of particles belonging to the first generation of quarks and leptons, *i.e.* the u and d quarks and the electron. Together they form atoms where electrons are orbiting a nucleus consisting of protons and neutrons made up by *uud* and *udd* quarks, respectively.

Quarks and charged leptons from the second and third generations are unstable. They can be produced in high-energy collisions, after which they decay via the weak interaction into particles of the first generation.

Table 1.1 – Interaction strengths and ranges at $Mc^2 = 1\,\mathrm{GeV}$ [7].

Interaction	Gravitation	Weak	Electromagnetic	Strong
Relative strength	10^{-40}	10^{-5}	10^{-2}	1
Range (m)	∞	10^{-18}	∞	10^{-15}

Interactions

The strengths of the fundamental interactions relative to the strong interaction along with their interaction ranges are given in Table 1.1 for the energy scale of $Mc^2 = 1\,\mathrm{GeV}$. At this energy, the strengths of the four interactions are at completely different scales, *e.g.* the weak interaction is about one hundred thousand times weaker than the strong interaction. The very different nature of the strong, electromagnetic and weak interaction will be discussed in the following sections.

Strong interaction

The strong interaction is described by quantum chromodynamics (QCD), which is based on the $\mathrm{SU(3)_C}$ symmetry. The C refers to the colour charge associated with the strong interaction. The strong force is mediated by electrically neutral and massless gluons. There are eight independent gluons corresponding to different linear combinations of colour and anticolour states [7]. The possible colour charges are red (r), green (g) and blue (b) and their corresponding anticolours. Since quarks and gluons are colour-charged, quarks change colour by the emission of gluons and gluons self-interact.

Due to a special property of the strong force called confinement, quarks are not observed isolated, but only in bound states, except for the t quark, which decays too fast to form a bound state. Bound states of quarks, also known as hadrons, must be colour-neutral. The most common hadrons are baryons (antibaryons) made of three quarks (antiquarks) with three different colours rgb ($\bar{r}\bar{g}\bar{b}$), and mesons made of a quark and antiquark with a colour combination like $r\bar{r}$, $g\bar{g}$ or $b\bar{b}$. Other combinations of two quarks and two antiquarks (tetraquarks) and four quarks and one antiquark (pentaquarks) are also possible.

Another special property of the strong force is asymptotic freedom [19, 20], where the value of the strong coupling α_s becomes very high at low energies, while it decreases asymptotically with increasing energy. At high energies, it is therefore possible to make predictions for QCD processes using perturbation theory, however, at low energies where $\alpha_s \sim 1$, these methods can no longer be used, and instead other methods, such as lattice QCD have to be applied.

Electromagnetic interaction

The electromagnetic interaction is described by quantum electrodynamics (QED), first formulated in 1927 by Paul Dirac [21]. It is based on the $U(1)_{em}$ symmetry, which conserves electric charge. The force mediator is the massless, electrically neutral photon (γ). All particles with electric charge interact with the photon. Examples are the Coulomb repulsion between two electrons corresponding to the exchange of a photon, or the annihilation of an electron and its antiparticle (positron) into a photon. Since photons do not carry electric charge, they cannot self-interact like the strong force mediators, the gluons.

The weak interaction

The weak interaction is mediated by the three heavy bosons $W^\pm$ and Z^0 of mass $80.4\,\text{GeV}/c^2$ and $91.2\,\text{GeV}/c^2$ [15]. The two W bosons carry electric charge of $\pm 1e$, while the Z^0 boson is neutral. In interactions where the exchange of energy between weakly interacting particles is small compared to the mass of the mediator, the weak interaction becomes effectively independent on the momentum transfer. The weak coupling strength is then given by the Fermi constant [7, 15],

$$G_F = \frac{\sqrt{2}g_w^2}{8M_W^2} = 1.1663787(6) \times 10^{-5}\,\text{GeV}^{-2} \, , \tag{1.4}$$

where g_w is the weak coupling constant and M_W is the W mass. The weak interaction is therefore weak at low energies due to its heavy mediators.

All leptons and quarks can interact via the weak force. The coupling between leptons and weak bosons is the same regardless of the lepton flavour. This concept is known as lepton universality [7]. However, for quarks the coupling depends on their flavour, which will be explained in Sec. 1.2.1.

The weak interactions mediated by the $W^\pm$ boson are referred to as charge current (CC) interactions. In these interactions, the exchange of a W boson changes the electric charge of the lepton or quark by $\pm 1e$. For leptons this can only happen within the same generation. For instance, an electron can convert into an electron neutrino by emitting a W^- boson, but it cannot convert into a muon or tau neutrino, *i.e.* the lepton flavour must be conserved. For quarks the interaction with W bosons changes their flavour between up- and down-type quarks, both within and across generations. For instance, a b quark of electric charge $-1/3e$ can emit a W^- boson and decay into a c quark of electric charge $+2/3e$, which is lighter than the b quark. Interactions involving Z^0 bosons are referred to as neutral current interactions (NC). These interactions do not change the flavour or charge of leptons and quarks, instead, they mediate the exchange of momentum between particles.

In contrast to the other SM interactions, the weak interaction discriminates particles based on their chirality [22], which is an inherent property of particles. Particles with positive chirality are said to be left-handed (LH), while those with negative chirality are said to

be right-handed (RH). In the limit of zero mass, the chirality is the same as the helicity, which is the projection of the spin vector on the particle momentum, is the same. In this case, particles with momentum and spin pointing in the same (opposite) directions are RH (LH). Experiments show that the $W^\pm$ bosons only interact with LH particles and RH antiparticles, resulting in a maximal violation of parity (see the paragraph "Discrete symmetries"). For this reason, the weak interaction is expressed as a vector (V) minus axial-vector (A) interaction, also known as the chiral V–A interaction. The Z^0 boson interacts with both chiralities, however, the coupling to LH particles is stronger than to RH particles with the relative strength given by the Weinberg angle introduced in the next section. Although, RH neutrinos could in principle interact with the Z^0 boson, this has never been observed, and consequently RH neutrinos (if they exist) are unaffected by the weak force.

Electroweak symmetry breaking and the Higgs mechanism

Although the electromagnetic and weak interaction appear to be very different, they can be unified into one single electroweak (EW) interaction based on the $SU(2)_L \times U(1)_Y$ symmetry, as proposed by Glashow, Salam and Weinberg in the 1960's [23–25].

The weak couplings of the W (g_W) and Z^0 (g_Z) bosons are related to the electromagnetic coupling ($g_e = e\sqrt{4\pi/\hbar c} = \sqrt{4\pi\alpha}$) via the Weinberg angle $\sin^2\theta_W \approx 0.23$ [22]:

$$g_e = g_W \sin(\theta_W)\,, \tag{1.5}$$

$$g_e = g_Z \sin(\theta_W) \cos(\theta_W)\ . \tag{1.6}$$

The $SU(2)_L$ symmetry conserves weak isospin, T_3, and gives rise to the three gauge bosons W_1, W_2 and W_3. As indicated by the subscript L, only LH fermions can interact with the W bosons. The $U(1)_Y$ symmetry conserves weak hypercharge, Y_W, and gives rise to the B boson. At this point the SM faces two important problems. The gauge symmetry prevents mass terms for the gauge bosons, and the $SU(2)_L \times U(1)_Y$ symmetry forces also fermions to be massless. However, experimentally it is known that the weak bosons and fermions are massive.

In order to resolve this, a scalar field that exists throughout space, also known as the Higgs field [13], is introduced. Below a certain critical temperature [26], it acquires a non-zero vacuum expectation value of about 246 GeV [15]. This breaks the gauge symmetry and triggers the Higgs mechanism giving rise to two charged massive W^+ and W^- bosons. The W_3 and B bosons mix and form two neutral gauge bosons, the massless photon (γ) and the massive Z^0 boson. The number that must be conserved in EW interactions is the total electric charge given by a linear combination of Y_W and T_3 corresponding to $Q = T_3 + Y_W/2$. The quantum numbers Q, Y_W, and T_3 are listed for fermions in Table 1.2 and for bosons in Table 1.3. Moreover, the Higgs field interacts with the fermions via the Yukawa couplings, which gives rise to quark and charged lepton masses, but not neutrino

Table 1.2 – Fermions and their EW quantum numbers. The sign of the quantum numbers for the corresponding antiparticles are reversed [22].

LH fermions	Q	Y_W	T_3	RH fermions	Q	Y_W	T_3
ν_e, ν_μ, ν_τ	0	-1	$+1/2$	ν_e, ν_μ, ν_τ	0	0	0
e^-, μ^-, τ^-	-1	-1	$-1/2$	e^-, μ^-, τ^-	-1	-2	0
u, c, t	$+2/3$	$+1/3$	$+1/2$	u, c, t	$+2/3$	$+4/3$	0
d, s, b	$-1/3$	$+1/3$	$-1/2$	d, s, b	$-1/3$	$-2/3$	0

Table 1.3 – Bosons and their EW quantum numbers [22].

Boson	Q	Y_W	T_3
$W^\pm$	± 1	± 1	0
Z^0	0	0	0
γ	0	0	0
g	0	0	0
H^0	0	$+1$	$-1/2$

masses[1]. The Higgs field introduces a new SM boson, also known as the Higgs boson, which is a scalar boson (spin 0) with no electric charge.

Discrete symmetries

According to the *CPT* theorem [27] all interactions are predicted to be invariant under the successive operations of charge conjugation (C), parity transformation (P) and time reversal (T), where:

- C transforms a particle into its antiparticle.

- P inverts the three axes of the coordinate system in space such that $(x, y, z) \rightarrow (-x, -y, -z)$. For fermions this results in a change of chirality.

- T changes the sign of the time coordinate.

A consequence of the *CPT* theorem is that particles and their antiparticles have the same mass and lifetime, as well as the same, but opposite, electric charges and magnetic moments. So far, experimental observations have been consistent with *CPT* invariance.

While the electromagnetic and strong interactions are invariant under each of the operations C, P and T, the weak interaction violates both P and C [28,29]. In fact, the weak interaction violates P maximally, which can be illustrated by considering that 100% of the neutrinos are LH. By inverting the spatial coordinates, the neutrino momentum is reversed, while its

---

[1]If the Higgs mechanism was responsible for neutrino masses, we would observe RH neutrinos.

spin is unchanged, resulting in a RH neutrino, which has never been observed. In the same way, violation of C in the weak interaction can be illustrated by applying the C operator to the LH neutrino resulting in a LH antineutrino, which has also never been observed.

For some time the CP symmetry was believed to be conserved in all interactions, however, in 1964 an experiment by Christenson *et al.* [30] showed that CP is in fact violated in weak decays of K^0_L mesons. Since then, CP violation has been observed in different weak interactions between quarks [31–33], but not in weak interactions between leptons, nor in strong and electromagnetic interactions. It is therefore only the weak interaction that gives rise to an asymmetry between particles and their antiparticles in the SM.

1.1.2 Successes and shortcomings

The SM was formulated in the 1960's and 1970's and it has since then proven successful at describing and predicting phenomena related to elementary particles and their interactions. For instance, the SM prediction of the electron magnetic dipole moment agrees with measurements within 1 part per 100 billion [34]. Moreover, the SM has predicted the existence of several particles that have later been confirmed experimentally, most recently, the Higgs boson discovered in 2012 [4, 5].

Despite the success of the SM, there are important shortcomings, including:

- **Gravity**, described by Einstein's theory of general relativity [35] cannot be explained within the SM framework [35].

- **Matter-antimatter asymmetry**. There is almost no antimatter in the universe compared to matter [36]. Consequently, matter and antimatter must have been treated differently after the Big Bang. Although, the SM allows for such an asymmetry via CP violation, the observed CP violation is not large enough to account for the abundance of matter in the universe.

- **Neutrino masses**. Neutrino oscillations have been observed meaning that at least two neutrinos must be massive [16], however, in the SM neutrinos are massless. One possible solution comes from the seesaw mechanism [37] where heavy ($\sim 10^{14}$ GeV) RH neutrinos are added to the SM resulting in masses of ~ 1 eV for LH neutrinos.

- **Generations and masses**. There are three known generations of leptons and quarks, however, this is not predicted nor explained by the SM. Furthermore, the particle masses differ by several orders of magnitude between generations, which is also not explained by the SM (but also not forbidden).

- **Dark matter and energy**. Several astrophysical observations [38] confirm the presence of so-called dark matter which has, so far, only been measured through its gravitational interactions. In fact, dark matter accounts for 85% of all matter in the universe [39]. The accelerating expansion of the universe caused by dark energy [40],

estimated to make up 68% of the energy in the universe, is also not described by the SM [39].

Today, much of the effort in particle physics goes into understanding phenomena like these. It is clear that the SM is not a complete theory, so particle physicists are constantly testing it, looking for discrepancies between predictions and observations, which could point to new physics (NP).

1.2 Flavour physics and hints of new physics

The field of flavour physics studies the nature of the weak interaction in flavour-changing quark transitions. Precise measurements of fundamental SM parameters probe NP that can explain, for instance, the matter-antimatter asymmetry or dark matter. In the following sections, flavour physics, in particular the CKM matrix and its matrix elements $|V_{ub}|$ and $|V_{cb}|$, and possible hints of NP will be reviewed.

1.2.1 Quark mixing and the CKM matrix

As quarks acquire masses through the Yukawa couplings between the Higgs field and fermionic fields, their mass eigenstates become different from their weak interaction eigenstates. This allow quarks to change flavour between up- and down-type quarks within and across generations in CC interactions. The weak and mass eigenstate bases are related by the complex unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix [41, 42]. By convention, the weak eigenstates of the up-type quarks (u, c, t) are the same as their mass eigenstates, such that the CKM matrix transforms the mass eigenstates of the down-type quarks (d, s, b) into their weak interaction eigenstates (d', s', b'):

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} . \tag{1.7}$$

The elements of the CKM matrix, V_{ij}, represent the couplings between up-type (i) and down-type (j) quark flavours. The probability of a transition between quark flavours j and i is proportional to $|V_{ij}|^2$. The magnitudes and phases of the CKM matrix elements are not predicted by the SM and must therefore be determined experimentally. It turns out that elements on the diagonal are close to one, while off-diagonal elements are small, but non-zero. Consequently, the quark mixing within the same generation is favoured over those across generations, which are said to be CKM suppressed.

The CKM matrix can be parameterised by three mixing angles (θ_{12}, θ_{13}, θ_{23}) and one

CP-violating phase (δ), which is the only known source of CP violation in the quark sector:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}c_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}. \tag{1.8}$$

Here $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$. By definition $s_{ij}, c_{ij} \geq 0$, and experimentally, it is observed that $s_{13} \ll s_{23} \ll s_{12} \ll 1$ and $\delta \sim \mathcal{O}(1)$ [15].

Since the CKM matrix elements are fundamental SM parameters used to predict the transition probability of flavour-changing processes and the amount of CP violation, their precise determination is important.

1.2.2   Measuring the CKM matrix elements

The absolute value of the CKM elements V_{ij} are determined from flavour-changing $q_j \to q_i$ quark transitions. However, as quarks do not exist as free particles, these transitions have to be studied inside the bound state of hadrons. In general, the heaviest quark of the hadron decays into a lighter quark, while the other quarks are unchanged and therefore called spectator quarks.

Common matter composed of protons and neutrons can be used to study $d \to u$ quark transitions, however, the unstable hadrons containing quarks from the second and third generation must be produced in particle collisions. This can happen when energetic cosmic rays hit the atmosphere or in experiments where leptons or hadrons are made to collide at high energies. Some of the hadrons created in particles collisions are:

Pions : $\pi^+(u\bar{d})$, $\pi^0(u\bar{u}, d\bar{d})$, $\pi^-(d\bar{u})$ *etc.*

Kaons : $K^+(\bar{s}u)$, $K^0(\bar{s}d)$, $K^-(s\bar{u})$ *etc.*

D mesons : $D^0(c\bar{u})$, $D^+(c\bar{d})$, $D_s^+(c\bar{s})$ *etc.*

B mesons : $B^+(\bar{b}u)$, $B_s^0(\bar{b}s)$, $B_c^+(\bar{b}c)$ *etc.*

There are three different types of weak decays, based on the absence or presence of hadrons and leptons in the final state [43]:

- **Leptonic decays**, where the hadron decays into a final state consisting of only leptons, *e.g.* the decay $B^+ \to \mu^+\nu_\mu$ illustrated in Fig. 1.2, where V_{ub} determines the strength by which the two quarks annihilate.

- **Semileptonic decays**, where the hadron decays into a final state consisting of both hadrons and leptons, *e.g.* the beta-decay $n \to p\,e^-\bar{\nu}_e$ illustrated in Fig. 1.2, where V_{ud} determines the strength of the transition.
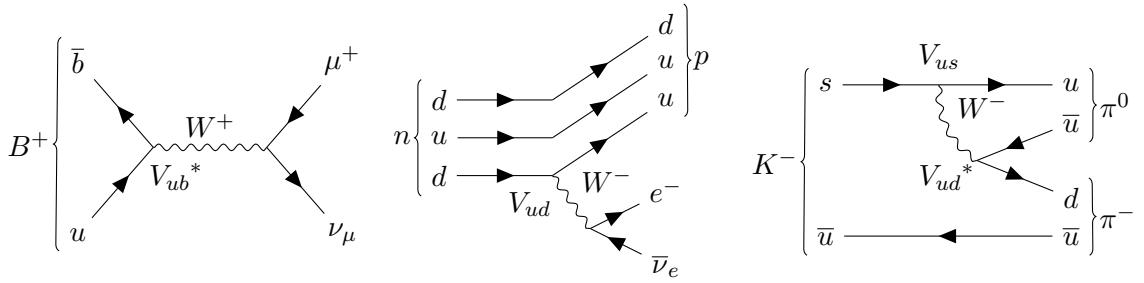
Figure 1.2 – Feynman diagram of decays used to determine CKM matrix elements, *i.e.* example of a leptonic decay (left), a semileptonic decay (middle) and a hadronic decay (right).

- **Non-leptonic decays**, where the hadron decays into a final state consisting of only hadrons, *e.g.* the decay $K^- \to \pi^0 \pi^-$ illustrated in Fig. 1.2, where V_{us} and V_{ud} determine the strength of the transition.

The theoretical predictions of these decays pose the same challenge, namely that quarks are not free particles, but confined inside hadrons via the exchange of soft gluons. At these energies, the strong interaction cannot be calculated perturbatively, and instead numerical methods, *e.g.* the light-cone sum rules (LCSR) [44], which will be introduced in Chapter 2, or lattice QCD (LQCD) [45], where the theory of QCD is formulated on a lattice consisting of points in space and time. Quark fields are defined at lattice sites, while the gluon fields are defined on the links. In order for the approximation to approach the QCD continuum, the spacing between lattice sites have to be minimised, which causes the computational cost of the numerical simulations to drastically increase. It is therefore necessary to find a compromise between lattice spacing and computational cost.

In general, leptonic and semileptonic decays are theoretically simpler than non-leptonic decays, because the effect of the strong interaction can be isolated [46]. In both cases, the decay amplitude can be written as the product of a well understood leptonic current and a more complicated hadronic current. For leptonic decays, the effects of the strong interaction are described by a decay constant, while for semileptonic decays, they are described by form factors (FFs). Experimentally, semileptonic modes involving heavy quarks (*c* or *b*) are easier to study than leptonic modes with a neutrino in the final state which are helicity suppressed.

In order to precisely determine the CKM matrix elements a global fit of all the different CKM measurements over the years is performed. It is assumed that the CKM matrix is unitary and that there are three generations of quarks. This results in the CKM matrix where the magnitudes of the elements are given by [15]:

$$V_{CKM} = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361^{+0.00011}_{-0.00009} \\ 0.22636 \pm 0.00048 & 0.973220 \pm 0.00011 & 0.04053^{+0.00083}_{-0.00061} \\ 0.00854^{+0.00023}_{-0.00016} & 0.03978^{+0.00082}_{-0.00060} & 0.999172^{+0.000024}_{-0.000035} \end{pmatrix} . \tag{1.9}$$
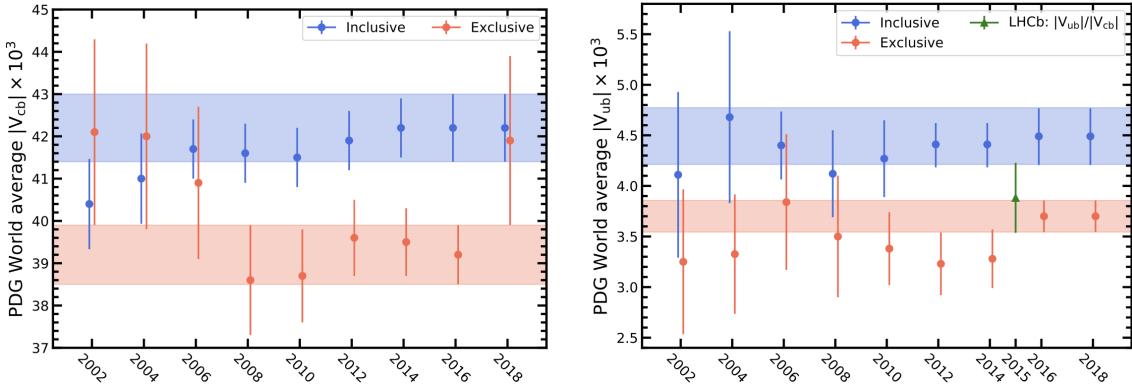
Figure 1.3 – PDG averaged $|V_{cb}|$ and $|V_{ub}|$ values from inclusive and exclusive semileptonic decays over the last twenty years [47].

The diagonal elements are close to one meaning that quarks belonging to the same generation have the strongest coupling, while quarks from the third and first generation have the smallest coupling.

1.2.3 Semileptonic determinations of $|V_{cb}|$ and $|V_{ub}|$

One of the big puzzles in flavour physics is the long-standing tension between inclusive and exclusive semileptonic determinations of the two CKM matrix elements $|V_{cb}|$ and $|V_{ub}|$ [47].

In exclusive semileptonic decays, one measures the decay of a b hadron into a final state containing a specific c hadron (D^*, D) or u hadron (π, ω, ρ), while in inclusive semileptonic decays one measures the sum over all possible hadronic states containing a c or u quark [48]. In both cases, the CKM matrix element is determined from the measured decay rate combined with theoretical calculations of the hadronic effects.

The experimental and theoretical methods associated with inclusive and exclusive decays are largely independent, and can therefore provide a cross check for $|V_{cb}|$ and $|V_{ub}|$ determinations. However, over the past decade the inclusive and exclusive determinations of both matrix elements have continued to disagree as illustrated in Fig. 1.3. Currently, the tension for $|V_{cb}|$ (left) is around 3σ, without taking the latest exclusive determination [49] into account, and around 3.5σ for $|V_{ub}|$ (right).

Inclusive approach

The inclusive determinations of $|V_{cb}|$ and $|V_{ub}|$ are mainly coming from B factories [50] where e^- and e^+ are made to collide at a centre-of-mass energy of 10.58 GeV corresponding to the $\Upsilon(4S)$ resonance, which almost exclusively decays into either $B^0\bar{B}^0$ or B^+B^- pairs. This results in a production cross-section of $b\bar{b}$ of approximately 1.1 nb [48].

The theoretical formulation of the inclusive $B \to X_q\ell\nu_\ell$ decay rate with $q = c, u$ and $\ell = e, \mu$

is based on the Operator Product Expansion resulting in the Heavy Quark Expansion (HQE) [15]. The total rate is expressed as the product of the CKM matrix element $|V_{qb}|$ and a series in $1/m_b$ and $1/m_q$ with perturbative and non-perturbative parameters determined from the measured moments of the lepton energy and hadronic invariant mass spectra. For $|V_{cb}|$ the total rate is given by [15]

$$
\begin{aligned}
\Gamma = {} & \frac{|V_{cb}|^2 G_F m_b^5(\mu)}{192\pi^3} \eta_{ew} \\
& \times [z_0^{(0)}(r) + \frac{\alpha_s(\mu)}{\pi} z_0^{(1)}(r) + (\frac{\alpha_s(\mu)}{\pi})^2 z_0^{(2)}(r) + ... \\
& + \frac{\mu_\pi^2}{m_b^2}(z_2^{(0)}(r) + \frac{\alpha_s(\mu)}{\pi} z_2^{(1)}(r) + ...) \\
& + \frac{\mu_G^2}{m_b^2}(y_2^{(0)}(r) + \frac{\alpha_s(\mu)}{\pi} y_2^{(1)}(r) + ...) \\
& + \frac{\rho_D^3}{m_b^3}(z_3^{(0)}(r) + \frac{\alpha_s(\mu)}{\pi} z_3^{(1)}(r) + ...) \\
& + \frac{\rho_{LS}^3}{m_b^3}(y_3^{(0)}(r) + \frac{\alpha_s(\mu)}{\pi} y_3^{(1)}(r) + ...) + ...] \,,
\end{aligned}
\tag{1.10}
$$

where μ is the energy scale, m_b is the b-quark mass, η_{ew} is the leading electroweak corrections to the four-fermion operator, $r = m_c/m_b$ where m_c is the c-quark mass, y_i and z_i are perturbatively calculable coefficients at different orders of the HQE. The parameters μ_π, μ_G, ρ_D and μ_{LS} are the non-perturbative inputs to the HQE.

The experimental inclusive determination of $|V_{cb}|$ is precise with a relative uncertainty of about 2% [15] where the main source of uncertainty is coming from the ignorance of higher-order perturbative and non-perturbative corrections.

The inclusive decay rate of $B \to X_u \ell \nu_\ell$ can also be formulated in terms of the HQE, however, it is difficult to measure due to the large background from CKM favoured $B \to X_c \ell \nu_\ell$ processes. This limits the regions of phase-space where the partial decay rate can actually be measured, and consequently, HQE cannot be used in the same way as for $b \to c$. Instead, non-perturbative distribution functions called shape functions are introduced. Their shapes are not well known, but they can be determined from fits to parameters obtained from $B \to X_c \ell \nu_\ell$ measurements [15]. This results in a less precise inclusive determination of $|V_{ub}|$ where the relative uncertainty is about 7% with similar contributions from theory and experiment.

Exclusive approach

The exclusive $|V_{cb}|$ and $|V_{ub}|$ determinations are provided by the B-factory experiments [50] and by the LHCb experiment [51] at the Large Hadron Collider (LHC) [52] where protons are made to collide at a center-of-mass energy $(\sqrt{s})$ up to 14 TeV. The production cross-section for $b\bar{b}$ pairs in the LHCb detector volume is around $144\,\mu\text{b}$ [53] at $\sqrt{s} = 13$ TeV,

which is one hundred thousand times higher than at the B factories. Furthermore, heavier b hadrons such as B_s, B_c and Λ_b can be produced at the LHC, however, the hadronic environment makes the background rejection harder than at the B factories.

Theoretically, the exclusive decay rate is expressed as the product of the CKM matrix element squared and a form factor that depends on the initial (X) and final (Y) state hadron. The form factors are functions of the momentum transfer, $q^2 = (p_X - p_Y)^2 = (p_\ell + p_{\bar{\nu}_\ell})^2$, and their shapes need to be parameterised. One of the most general parameterisations comes from Boyd, Grinstein and Lebed (BGL) [54], where the form factors are expressed as a series in powers of q^2 (explained in more detail in Chapter 2). The non-perturbative parameters can be calculated using LQCD or light-cone sum rules (LCSR) [55, 56]. In general, LQCD provides the most precise predictions at high q^2, while the opposite is true for LCSR.

LQCD calculations involving heavy quarks like the b and c quark are difficult, since the lattice spacing a must be small to satisfy the requirement of $m_q \leq 1/a$ for $m_b \approx 4.5\,\text{GeV}/c^2$ and $m_c \approx 1.5\,\text{GeV}/c^2$ [15]. This is to avoid large discretization errors, which are proportional to powers of am_q. For heavy quarks, it is therefore necessary to either reduce the lattice spacing resulting in an increased computational cost, or apply more sophisticated methods.

The most precise exclusive determinations of $|V_{cb}|$ are based on the semileptonic decays $B \to \bar{D}^* \ell \bar{\nu}_\ell$ and $B \to \bar{D} \ell \bar{\nu}_\ell$. The relative uncertainty on $|V_{cb}|$ is around 2%–3% [15] with similar contributions from experiment and theory. The most precise determination of $|V_{ub}|$ comes from the exclusive decay $B^+ \to \pi \ell \nu_\ell$ where a combined fit to data and theory as a function of q^2 results in a relative uncertainty of about 4% [15]. Other determinations have been obtained from $B \to \rho \ell \bar{\nu}_\ell$, $B \to \omega \ell \bar{\nu}_\ell$ and $\Lambda_b \to p \mu \bar{\nu}_\mu$.

Determinations of $|V_{cb}|$ and $|V_{ub}|$

The exclusive and inclusive determinations of $|V_{cb}|$ are found to be [15]:

$$|V_{cb}| = (42.2 \pm 0.8) \times 10^{-3} \text{ (inclusive)} \,,$$

$$|V_{cb}| = (39.5 \pm 0.9) \times 10^{-3} \text{ (exclusive)} \,.$$

The exclusive and inclusive determinations $|V_{ub}|$ are:

$$|V_{ub}| = (4.25 \pm 0.12^{+0.15}_{-0.14} \pm 0.23) \times 10^{-3} \text{ (inclusive)} \,,$$

$$|V_{ub}| = (3.70 \pm 0.10 \pm 0.12) \times 10^{-3} \text{ (exclusive)} \,.$$

The combined exclusive determinations of $|V_{ub}|$ and $|V_{cb}|$ are illustrated in Fig. 1.4 and compared to the corresponding combined inclusive determination.
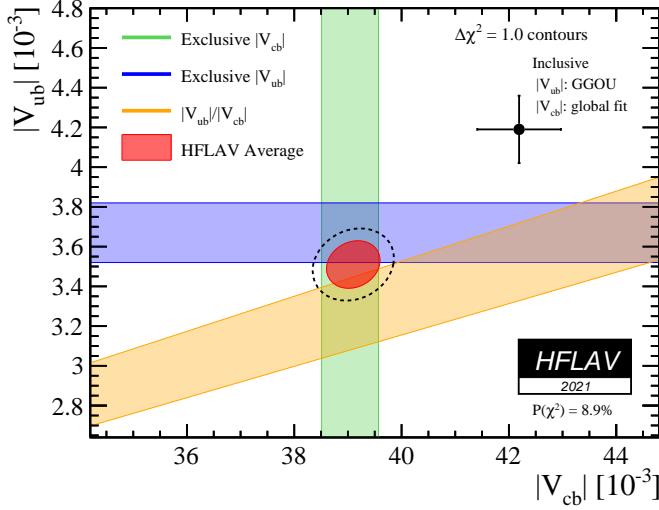
Figure 1.4 – Combined exclusive $|V_{ub}|$ and $|V_{cb}|$ determinations compared to the combined inclusive determination [57].

For both $|V_{ub}|$ and $|V_{cb}|$ the exclusive and inclusive determinations are not in agreement [2]. It is possible that the tension in $|V_{ub}|$ could be due to not yet well understood experimental or theoretical issues, however, NP physics has also been proposed as a possible explanation [61].

1.2.4 New physics explanation for $|V_{ub}|$ tension

The $|V_{ub}|$ puzzle does not only consist of a tension between exclusive and inclusive semileptonic determinations, but also between different exclusive decay channels. This is illustrated in Fig. 1.5. Determinations from $B \to \rho \ell \bar{\nu}_\ell$ and $B \to \omega \ell \bar{\nu}_\ell$ decays are significantly smaller than the one from $B \to \pi \ell \bar{\nu}_\ell$. Moreover, the two determinations from $B \to \rho \ell \bar{\nu}_\ell$ by Belle [62] and BaBar [63] are also not in agreement.

One of the explanations for the tension involves NP in the form of a right-handed weak current [61]. According to the SM, the weak interaction mediated by W bosons only interacts with left-handed fermions, however, by allowing for a small right-handed admixture the tension could be eased. In this case, the effective Lagrangian with one new parameter ϵ_R would read:

$$\mathcal{L}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{ub}^L (\bar{u}\gamma_\mu P_L b + \epsilon_R \bar{u}\gamma_\mu P_R b)(\bar{\nu}\gamma^\mu P_L l) + h.c. \, , \tag{1.11}$$

where $P_{L,R} = (1 \pm \gamma_5)/2$ are the chiral projectors, and as ϵ_R is allowed to be non-zero the measured CKM matrix element is denoted $|V_{ub}^L|$ [66]. The SM Lagrangian is recovered when $\epsilon_R \to 0$.

---

[2]According to recent analyses [58–60] the tension between $|V_{cb}|$ determinations can be eased by using a different method to calculate the $B \to \bar{D}^* \ell \nu_\ell$ form factors, however, the V_{cb} puzzle is still not resolved and more studies are needed.
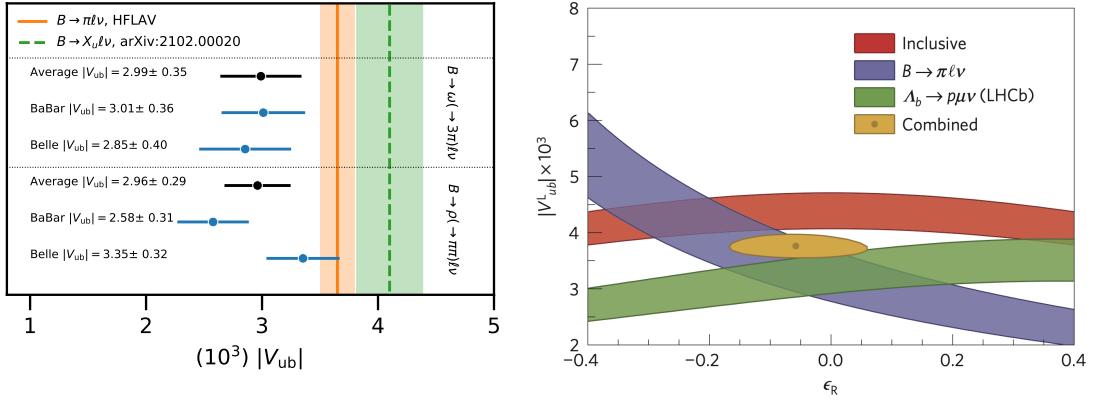
Figure 1.5 – Extracted $|V_{ub}|$ values from $B \to \rho \ell \nu_\ell$, $B \to \omega \ell \nu_\ell$, $B \to \pi \ell \nu_\ell$ and inclusive semileptonic measurements (left) [64], and experimental constraints on $|V_{ub}^L|$ and ϵ_R (right) [65].

The experimental constraints on $|V_{ub}^L|$ and ϵ_R from the inclusive $B \to X_u \ell \nu_\ell$, exclusive $B \to \pi \ell \nu_\ell$ and exclusive $\Lambda_b \to p \mu \nu$ measurements are illustrated in Fig. 1.5 [65]. These measurements allow for a small non-zero right-handed coupling, however, the uncertainties are still too large to make any precise conclusion, and more measurements are therefore needed to understand the $|V_{ub}|$ puzzle.

2 Motivation and strategy for measuring the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay

There exists a long-standing tension between determinations of $|V_{ub}|$ in inclusive and exclusive semileptonic decays, as discussed in Sections 1.2.3 and 1.2.4. This tension could be due to experimental or theoretical issues, however, it could also be a hint of NP, such as the existence of a right-handed weak current as discussed in Section 1.2.4. In any case, understanding this issue is important, since $|V_{ub}|$ is a fundamental parameter of the SM, and a precise determination of it is required in various tests of the SM.

The decay process $B^+ \to \rho^0 \mu^+ \nu_\mu$, which is illustrated in Fig. 2.1, provides an exclusive semileptonic determination of $|V_{ub}|$. Moreover, the differential decay rate of $B^+ \to \rho^0 \mu^+ \nu_\mu$ with the vector meson ρ^0 almost exclusively decaying into two charged pions, has a rich angular structure providing more observables besides $|V_{ub}|$ that can be used to probe NP.

The $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay has already been measured at the B factories, *e.g.* BaBar and Belle, however, the precision of these measurements have been limited by a small signal yield. At LHCb the signal yield is much larger due to the high $b\bar{b}$ production cross-section. This is expected to permit a more precise determination of $|V_{ub}|$ based on the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay as well as a more precise measurement of the $B^+ \to \rho^0$ form factors.

In the following sections, the theoretical framework behind the $B^+ \to \rho^0 \mu^+ \nu_\mu$ measurement is explained. Previous measurements are presented and the prospects of an LHCb measurement is reviewed. Finally, the strategy of this analysis is laid out.
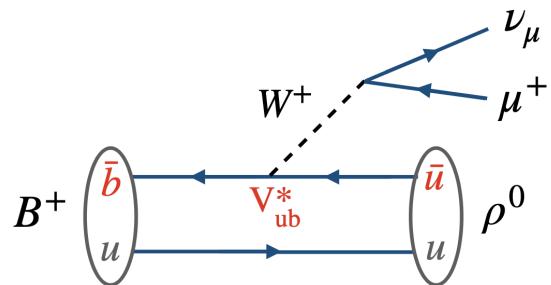
Figure 2.1 – Feynman diagram of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay.

2.1 Differential decay rate and form factors

The decay rate for the process $B \to f_1 f_2..$, where f_i are final state particles, can be expressed as [62]:

$$d\Gamma(B \to f_1 f_2..) = \frac{1}{2m_B}|M(B \to f_1 f_2..)|^2 d\Omega \ , \tag{2.1}$$

where m_B is the B meson mass, M is the decay matrix element and Ω is the total phase space:

$$d\Omega = (2\pi)^4 \delta^{(4)}(p_B - \sum_i p_i) \prod_i \frac{d^3 \vec{p_i}}{(2\pi)^3 2E_i} \ . \tag{2.2}$$

Here p_B is the B-meson four-momentum and $p_i = (E_i/c, \vec{p_i})$ are the four-vectors of f_i.

In semileptonic B decays the momentum transfer between the B meson and its decay products is limited by the b-quark mass of $\sim 4\,\text{GeV}/c^2$ [15], which is much smaller than the W boson mass of $\sim 80\,\text{GeV}/c^2$. It is therefore possible to describe the interaction with an effective theory where the heavy W boson is integrated out giving rise to the effective weak Lagrangian [64]:

$$\mathcal{L}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{ub}(\bar{u}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu P_L \nu_\ell) + h.c. \ , \tag{2.3}$$

where G_F is the Fermi constant, $P_{L,R} = (1 \pm \gamma_5)/2$ are the chiral projectors, γ_μ are Dirac matrices and $h.c.$ is the charged-conjugated term.

As mentioned earlier, the matrix element of semileptonic decays can be factorised into a well understood leptonic current, L^μ, and a hadronic current, H_μ, which is difficult to calculate due to the confinement of quarks inside the mesons. The matrix element of the semileptonic $B \to X_u \ell \bar{\nu}$ decay is given by [62]:

$$M(B \to X_u \ell \bar{\nu}) = \frac{G_F}{\sqrt{2}} V_{ub} L^\mu H_\mu \ , \tag{2.4}$$

where the leptonic current is given by $L^\mu = \bar{\ell}\gamma^\mu(1 - \gamma^5)\nu_\ell$, and the hadronic current for $X_u = \rho^0$ can be expressed as $H_\mu = \langle \rho^0(P_{\rho^0}, \varepsilon)|\bar{u}\gamma^\mu(1 - \gamma_5)b|B(P_B)\rangle$, where ε is the ρ^0 polarisation vector and P_{ρ^0} and P_B are the ρ^0 and B meson four-momenta, respectively.

The polarization vector leads to a hadronic current that depends on three independent form factors $A_1(q^2)$, $A_2(q^2)$ and $V(q^2)$ (under the assumption that the charged lepton mass can be neglected), where q^2 is the di-lepton invariant mass squared. The hadronic

current can then be expressed as [62]:

$$
\begin{aligned}
\langle \rho^0(P_{\rho^0}, \varepsilon) | V_\mu - A_\mu | B(P_B) \rangle &= -i\varepsilon_\mu^*(m_B + m_{\rho^0})A_1(q^2) \\
&\quad + i(p_B + p_{\rho^0})_\mu (\varepsilon^* \cdot p_B) \frac{A_2(q^2)}{m_B + m_{\rho^0}} \\
&\quad + \epsilon_{\mu\nu\rho\sigma} \varepsilon^{*\nu} p_B^\rho p_{\rho^0}^\sigma \frac{2V(q^2)}{m_B + m_{\rho^0}} \ .
\end{aligned}
\tag{2.5}
$$

It is written in terms of the V–A interaction, *i.e.* $V_\mu = \bar{u}\gamma_\mu b$ and $A_\mu = \bar{u}\gamma_\mu\gamma_5 b$, and $\epsilon_{\mu\nu\rho\sigma}$ is the Levi Civita tensor.

The form factors $A_1(q^2)$, $A_2(q^2)$ and $V(q^2)$ cannot be predicted using perturbative methods, and since the ρ^0 meson is produced with a high momentum due to the large mass difference between the B and ρ^0 meson, current LQCD calculations cannot provide precise predictions for the form factors. A better method is the light-cone sum rules (LCSR) [44], which provides predictions for the region $q^2 \lesssim \mathcal{O}(m_B\Lambda_{\mathrm{QCD}}) \sim 14\,\mathrm{GeV}^2/c^4$. Several parameterisation have been developed to extend the predictions of the form factors to the full q^2 region [44, 54, 67–69]. One of these is the Bourrely-Caprini-Lellouch (BCL) parameterisation [69] that expresses the form factors as a power expansion in the variable [44]

$$
z(q^2) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}} \ ,
\tag{2.6}
$$

where the pair-creation threshold is $t_+ = (m_B + m_{\rho^0})^2$ and the z-origin is determined by $t_0 = (m_B + m_{\rho^0})(\sqrt{m_B} - \sqrt{m_{\rho^0}})^2$. By rearranging the power expansion in z as a simplified series expansion around $q^2 = 0$, the form factors can be expressed with the Bharucha-Straub-Zwicky (BSZ) [44] parameterisation as

$$
F_i(q^2) = P_i(q^2) \sum_k \alpha_k^i (z(q^2) - z(0))^k \ ,
\tag{2.7}
$$

where $P_i(q^2) = (1 - q^2/m_{R,i}^2)^{-1}$ is the pole of the first resonance of mass $m_{R,i}$ in the spectrum. Experimentally, the three parameters $\alpha_{0,1,2}$ are found to be sufficient for fitting each of the form factors [44]. With this parameterisation, the form factor at $q^2 = 0$ corresponds to one of the fit parameters, *i.e.* $F_i(0) = \alpha_0^i$.

The form factors in semileptonic decays, where the B meson decays into a vector meson, are often expressed in the standard helicity basis with $H_\pm$ and H_0. They relate to the A_1, A_2 and V form factors in the following way [64]:

$$
\begin{aligned}
H_\pm(q^2) &= \frac{2m_B|P_{\rho^0}|V(q^2)}{m_B + m_{\rho^0}} \pm (m_B + m_{\rho^0})A_1(q^2) \, , \\
H_0(q^2) &= 8m_B m_{\rho^0} A_2(q^2)/\sqrt{q^2} \, .
\end{aligned}
\tag{2.8}
$$
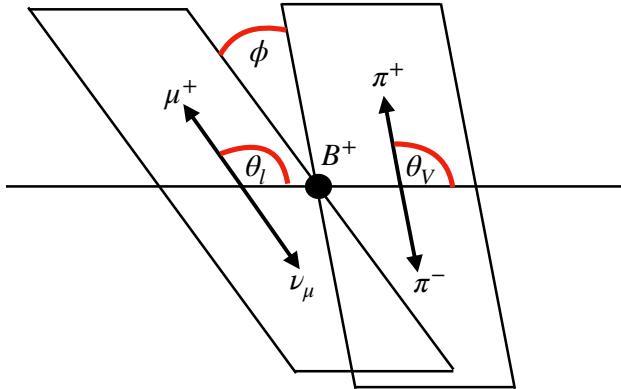
Figure 2.2 – Definition of the angular variables of the $B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ decay: the angles θ_l, θ_V and ϕ are measured in the rest frames of the ρ^0 meson, dilepton system and B^+ meson, respectively.

The differential decay rate in q^2 can then be expressed as

$$
\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{ub}|^2 c_V^2}{96\pi^3} |P_{\rho^0}| \frac{q^2}{m_B^2} (1 - \frac{1 - m_\ell^2}{q^2})^2
$$
$$
\times ((1 + \frac{m_\ell^2}{2q^2})(H_+^2(q^2) + (H_-^2(q^2) + (H_0^2(q^2)))) \,,
$$

(2.9)

where c_V is the vector-current coupling constant, which depends on the Weinberg angle [22]. By writing out the angular dependencies of $H_\pm$ and H_0 and grouping together the prefactors that depend on q^2 into functions denoted J_i, one arrives at Ref. [61].

$$
\frac{d\Gamma}{dq^2 d\cos\theta_V d\cos\theta_\ell d\phi} = \frac{G_F^2 |V_{ub}|^2 m_B^3}{2\pi^4} \times \bigg(J_{1s} \sin^2\theta_V + J_{1c} \cos^2\theta_V
$$
$$
+ (J_{2s} \sin^2\theta_V + J_{2c} \cos^2\theta_V) \cos 2\theta_\ell + J_3 \sin^2\theta_V \sin^2\theta_\ell \cos 2\phi
$$
$$
+ J_4 \sin 2\theta_V \sin 2\theta_\ell \cos\phi + J_5 \sin 2\theta_V \sin\theta_\ell \cos\phi
$$
$$
+ (J_{6s} \sin^2\theta_V + J_{6c} \cos^2\theta_V) \cos\theta_\ell + J_7 \sin 2\theta_V \sin\theta_\ell \sin\phi
$$
$$
+ J_8 \sin 2\theta_V \sin 2\theta_\ell \sin\phi + J_9 \sin^2\theta_V \sin^2\theta_\ell \sin 2\phi \bigg). \quad (2.10)
$$

This is the full differential decay rate of $B^+ \to \rho^0(\to \pi^+\pi^-)\ell^+\nu_\ell$ that in addition to q^2 depends on three angular variables θ_V, θ_ℓ and ϕ. Together, these four variables completely define the four-body final state. The fully differential decay rate expresses the rich helicity structure of the decay and provides more variables to test the SM and to probe NP than by measuring $|V_{ub}|$ alone. The three angles θ_V, θ_ℓ and ϕ describe the relative orientation of the final state particles as illustrated in Fig. 2.2. The twelve J_i coefficients are functions of q^2 that can be predicted within the Standard Model. They contain the non-perturbative hadronic effects expressed in terms of form factors, and can be sensitive to possible NP effects, *e.g.* a right-handed weak current as introduced in Eq. 1.11.

The ultimate goal would be to perform a fit of the differential decay rate in the three angular variables in bins of q^2 to obtain the J_i coefficients and compare with theoretical predictions. However, the full fit of the differential decay rate is expected to be complicated by the large amount of background, especially from $b \to c l \bar{\nu}$ and $b \to u l \bar{\nu}$ decays. Since these backgrounds might be correlated in the three angular variables, extracting information about NP is expected to be very challenging.

2.2 Previous measurements by BaBar and Belle

The decay $B^+ \to \rho^0 \ell^+ \nu_\ell$ where $\ell = e, \mu$ has been measured by the B factory experiments with the most precise measurements coming from BaBar [63] and Belle [62].

Different approaches are used to measure the $B^+ \to \rho^0 \ell^+ \nu_\ell$ decay in $e^+ e^- \to \Upsilon(4S) \to B\bar{B}$ events. In the tagged approach, the B meson is fully reconstructed despite having an unmeasured neutrino in its final state. This is done by fully reconstructing the other B meson, also known as the tag B. This approach results in clean signal samples with low systematic uncertainties, however, since the tag B is fully reconstructed the signal efficiency is decreased. A higher signal efficiency can be achieved with the untagged approach, where the missing energy and momentum of the event are used to reconstruct the neutrino, however, this results in less clean signal samples with larger systematic uncertainties.

2.2.1 BaBar measurement

The decay $B^+ \to \rho^0 \ell^+ \nu_\ell$ was measured by the BaBar experiment in 2010 [63]. In this untagged analysis, the signal yield is:

$$N_{\text{sig}} = 1403 \pm 143 \ . \tag{2.11}$$

It is extracted from a binned maximum likelihood fit of signal and backgrounds in three variables corresponding to the beam-energy substituted B mass (m_{ES}), the difference between the expected and reconstructed B energy (ΔE) and q^2, as illustrated in Fig. 2.3. Signal candidates are required to have a two-pion invariant mass of $650 < m(\pi^+ \pi^-) < 850 \, \text{MeV}/c^2$. The simulation of signal used in the fit is based on form factors from LCSR calculations by Ball and Zwicky [70]. The uncertainties on the form factors are assumed to be 10%–13% depending on the q^2 region.

Although the $B \to D^{(*)} \ell \bar{\nu}_\ell$ background is dominant prior to selection, it is the $B \to X_u \ell \bar{\nu}_\ell$ inclusive background that ends up causing the largest challenge. This is due to its similarity to signal resulting in a large correlation between the two components in the fit. The shape and yield of the $B \to X_u \ell \bar{\nu}_\ell$ decays are therefore fixed to predictions from simulations. However, these are associated with large uncertainties, partly due to the nonresonant $\pi^+ \pi^-$ component that is not well known, and consequently, this becomes the dominant systematic uncertainty on the partial and total branching fractions for $B^+ \to \rho^0 \ell^+ \nu_\ell$ in
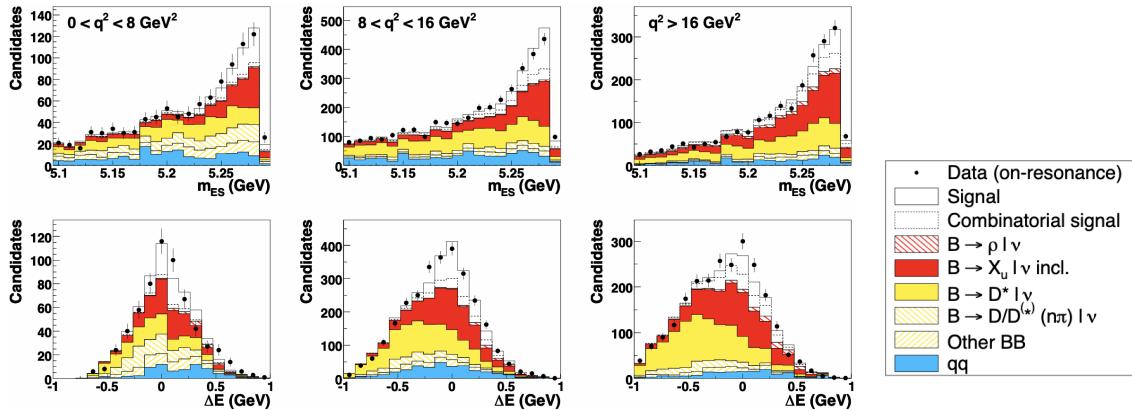
Figure 2.3 – m_{ES} and ΔE distributions for the $B^+ \to \rho^0 \ell^+ \nu_\ell$ signal and backgrounds in each q^2 bin after the BaBar fit [63].

the BaBar measurement. The total branching fraction is measured to be

$$\mathcal{B}(B^+ \to \rho^0 \ell^+ \nu_\ell) = (0.87 \pm 0.09(\text{stat}) \pm 0.15(\text{syst})) \times 10^{-4} \ , \tag{2.12}$$

where the first uncertainty is statistical (stat) and the second is systematic (syst). The differential decay rate is computed in three bins of q^2 with the limits $[0, 8, 16, 20.3]\,\text{GeV}^2/c^4$. A good agreement is found between data and the $B^+ \to \rho^0 \ell^+ \nu_\ell$ form-factor predictions from LCSR [70] and ISGW2 [1] quark models [71]. Moreover, a good agreement is found between the total branching fraction and partial branching fraction measurements. However, the uncertainty of the differential branching fraction is at the level of 15–30% depending on the q^2 bin. Due to the limited experimental precision, the three form factors are not extracted in this analysis. Finally, $|V_{ub}|$ is extracted by combining the measured partial branching fractions with integrals of the form-factor calculations based on LCSR [70] in each q^2 bin below $q^2 = 16\,\text{GeV}^2/c^4$, resulting in

$$|V_{ub}| = (2.75 \pm 0.24) \times 10^{-3} \ (q^2 < 16\,\text{GeV}^2/c^4) \ , \tag{2.13}$$

The uncertainty related to the form factors used for the calculation of $|V_{ub}|$ are not given in the references used for the result.

2.2.2   Belle measurement

Another measurement of the decay $B^+ \to \rho^0 \ell^+ \nu_\ell$ was performed by Belle in 2013 [62]. In this tagged analysis, the signal yield is:

$$N_{\text{sig}} = 621.7 \pm 35.0 \ . \tag{2.14}$$

---

[1]The Isgur-Scora-Grinstein-Wise model (ISGW2) is used to calculate form factors. It is a constituent quark model with relativistic corrections, and its predictions cover the full q^2 range, however, uncertainties of the predictions are difficult to quantify.
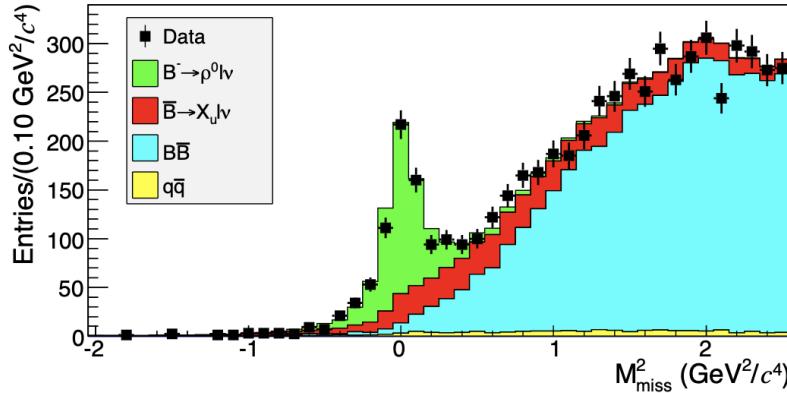
Figure 2.4 – Belle fit of M_{miss}^2 for the $B^+ \to \rho^0 \ell^+ \nu_\ell$ signal and backgrounds [62].

It is extracted from a binned maximum likelihood fit of the missing mass, *i.e.* $M_{\text{miss}}^2 = p_{\text{miss}}^2/c^2 = (p_{B_{\text{sig}}} - p_{\text{vis}})/c^2$ where $p_{B_{\text{sig}}}$ is the four-momentum of the reconstructed signal B meson inferred from tagging, and p_{vis} is the four-momentum of the visible final state particles of the signal. For correctly reconstructed semileptonic decays the signal peaks at $M_{\text{miss}}^2 = 0$, as illustrated in Fig. 2.4. Signal candidates are required to have a two-pion invariant mass of $477 < m(\pi^+\pi^-) < 1074\,\text{MeV}/c^2$. As in the BaBar measurement, the signal is modelled using LCSR form-factor predictions from Ref. [70]. The uncertainty related to the form-factor shapes is estimated as the difference in the total efficiencies resulting from the two different form-factor predictions from LCSR [70] and ISGW2 [71] calculations.

Also in this measurement, the $B \to D^{(*)}\ell\bar{\nu}_\ell$ background is dominant (represented by the $B\bar{B}$ component in Fig. 2.4), while $B \to X_u(\to \pi^+\pi^-)\ell\bar{\nu}_\ell$ with nonresonant final state pions constitutes the most difficult background, as it cannot be distinguished from signal in the M_{miss}^2 distribution. The nonresonant component is therefore estimated by performing a binned two-dimensional fit in M_{miss}^2 and $m(\pi^+\pi^-)$, where the nonresonant component is either floated or fixed to zero. In both cases, the signal yield is consistent with the result from the one-dimensional fit in M_{miss}^2 and the nonresonant component is found to be consistent with zero, thus this component is excluded from this analysis. Moreover, the $B^- \to f_2\ell\bar{\nu}_\ell$ background is found to be 3 times larger than the ISWG2 model prediction, which is not further investigated. The total branching ratio is measured to be

$$\mathcal{B}(B^+ \to \rho^0 \ell^+ \nu_\ell) = (1.83 \pm 0.10(\text{stat}) \pm 0.10(\text{syst})) \times 10^{-4} \ , \tag{2.15}$$

which is almost a factor two larger than the one measured by BaBar. The origin of this significant tension between the two results are not clear, however, it could be caused by the different treatments of the nonresonant contribution.

The relative uncertainty related to the $B \to X_u(\to \pi^+\pi^-)\ell\bar{\nu}_\ell$ background is estimated to be 2.4% of the total branching ratio, and the relative uncertainty related to the form-factor shapes is estimated to be 1.3%. To compute the partial branching fractions as a function

of q^2, the signal yield is extracted from a two-dimensional binned maximum likelihood fit in the two variables M^2_{miss} and $m(\pi^+\pi^-)$ in bins of q^2. There are 11 bins of q^2 between 0 and $22\,\text{GeV}^2/c^4$ with equal bin width of $2\,\text{GeV}^2/c^4$. A good agreement is found between the total and partial branching fraction measurements.

Finally, the measured partial branching fractions are combined with integrals of the form factor based on LCSR [70] in each q^2 bin and $|V_{ub}|$ is extracted as

$$|V_{ub}| = (3.56 \pm 0.11 \pm 0.09^{+0.54}_{-0.37}) \times 10^{-3} \ \ (q^2 < 16\,\text{GeV}^2/c^4) \ , \tag{2.16}$$

where the first uncertainty is statistical, the second is experimental systematic related to detector simulations and the third is theoretical related to form-factor uncertainties.

In a more recent tagged analysis [72] based on data collected by the Belle II experiment [73], the total branching fraction $\mathcal{B}(B^+ \to \rho^0 \ell \nu_\ell) = (1.77 \pm 0.23(\text{stat}) \pm 0.36(\text{syst})) \times 10^{-4}$ is determined with 205 ± 27 signal candidates. However, due to the limited signal statistics the partial branching fraction and $|V_{ub}|$ are not measured. The large systematic uncertainty of $\sim 20\%$ originates from the modeling of the nonresonant $B^+ \to (\pi^+\pi^-)\ell\nu_\ell$ contribution.

2.2.3   Combined Belle and BaBar measurement

In a paper by F. U. Bernlochner *et al.* from 2021 [64] the differential q^2 spectra from the aforementioned BaBar [63] and Belle [62] measurements are combined to determine $|V_{ub}|$ with higher precision and improved form-factor predictions [44].

In Fig. 2.5 the differential spectra from BaBar (blue) and Belle (orange and green) as well as the averaged differential spectrum (black) are illustrated (the green spectrum is the isospin transformed $B^0 \to \rho^- \ell^+ \nu_\ell$ measurement by Belle [62]).

A χ^2 fit to the averaged differential spectrum combined with the newest LCSR predictions [44] and $|V_{ub}|$ is performed in the full q^2 range. The form factors are parameterised with the BSZ parameterisation [44], thus the fit generates new predictions for the BSZ expansion parameters beyond the $q^2 < 14\,\text{GeV}^2/c^4$, where the LCSR predictions are valid. The averaged total branching ratio is found to be:

$$\mathcal{B}(B^+ \to \rho^0 \ell^+ \nu_\ell) = (1.35 \pm 0.12) \times 10^{-4} \ , \tag{2.17}$$

where the uncertainty is the combined statistical and systematic uncertainty. This result is significantly lower than the PDG value of $\mathcal{B}(B^+ \to \rho^0 \ell^+ \nu_\ell) = 1.58 \pm 0.11$ [15]. The fit determination of $|V_{ub}|$ is:

$$|V_{ub}| = (2.96 \pm 0.29) \times 10^{-3} \ (\text{full } q^2 \text{ region}) \ , \tag{2.18}$$

where the uncertainty is the combined statistical and systematic uncertainty. The extracted $|V_{ub}|$ is consistently smaller than previous determinations from $B \to \pi \ell \bar{\nu}_\ell$. Due to the large
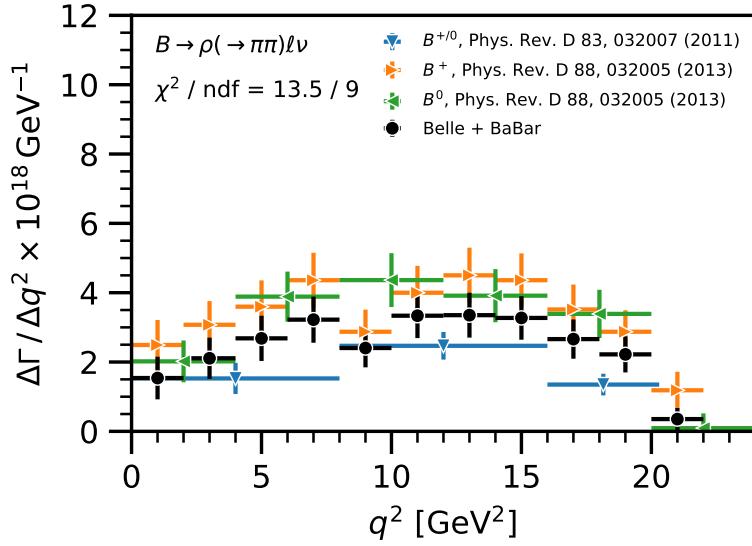
Figure 2.5 – Differential q^2 spectra for $B^+ \to \rho^0 \ell^+ \nu_\ell$ measured by BaBar (blue) and Belle (orange), for the isospin transformed decay $B^0 \to \rho^- \ell^+ \nu_\ell$ measured by Belle (green), together with their average (black) [64].

discrepancy in the measured $B^+ \to \rho^0 \ell^+ \nu_\ell$ spectra between BaBar and Belle visible in Fig. 2.5, $|V_{ub}|$ is also extracted for each experiment separately, which results in a slight tension between the two determinations as illustrated in the previous chapter in Fig. 1.5.

2.3 Prospects for a LHCb measurement

The existing $B^+ \to \rho^0 \ell^+ \nu_\ell$ measurements do not agree well, *i.e.* the measured branching fraction at Belle is about twice as high as the one measured at BaBar. New and more precise measurements are therefore needed to understand this tension.

Thanks to the high $b\bar{b}$ production cross-section at LHCb [53] a larger sample of $B^+ \to \rho^0 \ell^+ \nu_\ell$ decays can be obtained than at the B factories. In this analysis, the obtained signal yield is 25–60 times higher than the ones obtained by BaBar and Belle. However, the hadronic environment at LHCb poses several challenges. The collision of protons results in the production of hundreds of charged and neutral particles, which makes it harder to isolate signal from background. Moreover, the number of B mesons produced at LHCb cannot be precisely determined due to uncertainties in beam luminosity and $b\bar{b}$ production cross-section. For this reason, the branching fraction is measured relative to another decay channel, also called a normalisation channel. Furthermore, due to the composite nature of protons, the energies and momenta of the partons participating in the production of a B meson are not well constrained, and consequently, the momentum of the B meson decaying into a final state with neutrinos cannot be fully reconstructed. It is therefore necessary to apply certain corrections.
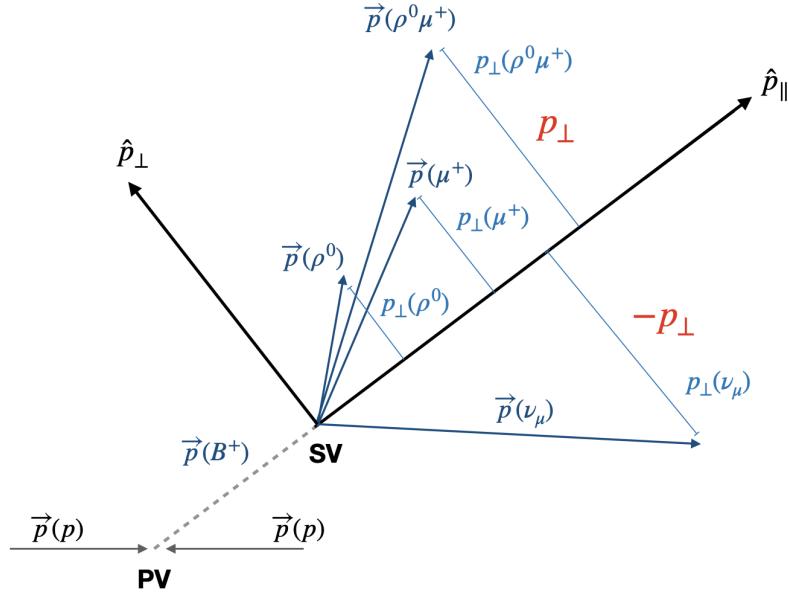
Figure 2.6 – The $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay seen in the laboratory frame in a coordinate system with one axis parallel to the B flight direction, $\hat{p}_\parallel$, and the other perpendicular to it, $\hat{p}_\perp$ leading to $\vec{p}_\perp(\mu^+) + \vec{p}_\perp(\rho^0) = -\vec{p}_\perp(\nu_\mu) \equiv \vec{p}_\perp$.

For the $B^+ \to \rho^0 \mu^+ \nu_\mu$ measurement at LHCb presented in this thesis, the direction of the B^+ momentum is constrained to the unit vector defined from the point where the B^+ meson was created (the primary vertex, PV) to the point where it decayed (the secondary vertex, SV), illustrated in Fig. 2.6. As the transverse momentum with respect to the B^+ flight direction of the final state must be zero, the transverse momentum of the neutrino is given by $\vec{p}_\perp(\nu_\mu) = -(\vec{p}_\perp(\mu^+) + \vec{p}_\perp(\rho^0)) \equiv -\vec{p}_\perp$. This information is used to apply a kinematic correction to the invariant mass of the visible final state particles, m_{vis}, resulting in a better discriminating variable called the corrected B mass [74] (derivation in Appendix A):

$$m_{\mathrm{corr}}(B^+) = \sqrt{m_{\mathrm{vis}}^2 + p_\perp^2} + p_\perp \; . \tag{2.19}$$

Using the kinematic constraints from before together with energy and momentum conservation, the neutrino momentum parallel to the B flight direction, $p_\parallel(\nu_\mu)$, can be determined up to a twofold ambiguity resulting from the quadratic equation (derivation in Sec. 5.1)

$$a \times p_\parallel(\nu_\mu)^2 + b \times p_\parallel(\nu_\mu) + c = 0 \; , \tag{2.20}$$

with the solution

$$p_\parallel(\nu_\mu)^\pm = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \; . \tag{2.21}$$

Using the speed of light $c = 1$, the parameters are

$$a = (m_{\text{vis}}^2 + p_\perp^2)\,, \tag{2.22}$$

$$b = -(m_B^2 - m_{\text{vis}}^2 - 2p_\perp^2)p_\parallel\,, \tag{2.23}$$

$$c = E^2 p_\perp^2 - \frac{1}{4}(m_B^2 - m_{\text{vis}}^2 - 2p_\perp^2)^2\,, \tag{2.24}$$

with E and $p_\parallel$ being the energy and parallel momentum of the visible final state with respect to the B^+ flight direction. Since the neutrino momentum is $\vec{p}(\nu_\mu) = \vec{p}_\perp(\nu_\mu) + \vec{p}_\parallel(\nu_\mu)^\pm$ the computation of q^2 results in two solutions where the correct one is not known a priori.

Despite the challenges associated with the hadronic environment, the significant increase in signal yield is expected to result in a more precise measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay at LHCb. In particular, this is expected to result in a more precise determination of $|V_{ub}|$ as well as the $B^+ \to \rho^0$ form factors compared to the previous determinations. This expectation is supported by the comparison performed in Sec. 8.4.2 between the precision of the measured $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential decay rate in this thesis and that performed by the Belle experiment.

2.4 Analysis strategy

The goal of this analysis is to measure the differential branching fraction of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay as a function of q^2 relative to the normalisation channel $B^+ \to \bar{D}^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ using data collected by the LHCb experiment during 2018. In future steps of the analysis, this measurement will be used to extract the CKM matrix element $|V_{ub}|$ as well as the $B^+ \to \rho$ form factors. The main steps of the analysis are outlined below.

- **Discriminating variables:** To extract the signal yield from data the variables $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$ are used.

- **Normalisation:** The decay $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ with $\bar{D}^0 \to \pi^+\pi^-$ is used as normalisation channel. This mode is chosen because it is a B-meson decay with the same final state as the signal. Consequently, many uncertainties cancel out when calculating the ratio between the differential branching fraction of the signal and the integrated branching fraction of the normalisation channel, *i.e.* $d\mathcal{B}(B^+ \to \rho^0\mu^+\nu_\mu)/dq^2 \times 1/\mathcal{B}(B^+ \to D^0\mu^+\nu_\mu)$. As the q^2 distribution of the normalisation channel does not fully cover the one of the signal, only the signal yield is extracted in bins of q^2.

- **MVA selection:** Two Multivariate Algorithms (MVAs) are used to select the signal. First, a charge isolation MVA is trained on a simulated background to recognise candidates with additional tracks. Next, the main MVA is trained with charge isolation variables based on the first MVA together with kinematic and geometric variables to separate the signal and background.

- **Control:** The decay $B^0 \to J/\psi \rho^0$ with $J/\psi \to \mu^+ \mu^-$ is used as a control channel, because it has the same topology and final state as the signal if the reconstruction of one of the muons is omitted. Most importantly, it is used to verify the signal simulation and selection.

- **Reconstructing q^2:** Due to the missing neutrino, q^2 cannot be fully reconstructed, and instead, a quadratic equation with two possible solutions is obtained. To select the correct q^2 solution, a multivariate regression method is used.

- **Template fits:** Fitting templates of $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+\pi^-)$ are extracted from simulations or data to model the signal and background processes. The two-dimensional fit is performed in bins of the reconstructed q^2, and the extracted signal yields are unfolded to get the true q^2 distribution. The yield of the normalisation channel is extracted from a one-dimensional template fit to $m_{\mathrm{corr}}(B^+)$.

- **Measuring the differential branching fraction:** The efficiency corrected yields of the signal channel in bins of q^2 and of the normalisation channel in the full q^2 region are combined with the external determinations of $\mathcal{B}(B^+ \to \bar{D}^0 \mu^+ \nu_\mu)$ and $\mathcal{B}(\bar{D}^0 \to \pi^+\pi^-)$ into a measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction as a function of q^2.

3 The LHCb experiment at the LHC

All data used to study the decay of $B^+ \to \rho^0 \mu^+ \nu_\mu$ in this thesis have been collected in proton-proton (*pp*) collisions at the Large Hadron Collider (LHC) [52] by the LHCb detector [51]. The experimental setup enabling this measurement will be explained in this chapter.

3.1 The Large Hadron Collider

The LHC is a 27 kilometre circular accelerator designed to collide high-intensity beams of protons at a centre-of-mass energy, $\sqrt{s}$, of up to 14 TeV, making it the largest and most powerful particle accelerator in the world. The increase in $\sqrt{s}$ compared to previous colliders allows physicists to search for new heavy particles, and in 2012, this lead to the discovery of the Higgs boson by the ATLAS [4] and CMS [5] experiments.

The LHC is located at the European Organisation for Nuclear Research (CERN) at the French-Swiss border in the Geneva region. It is placed approximately 100 meters underground in the tunnel originally built for the Large Electron-Positron Collider (LEP) [75] as illustrated in Fig. 3.1. Inside the LHC two beams of protons are circulating in opposite directions. They are separated in two beam pipes kept at ultrahigh vacuum. This is achieved by cryogenic pumping [77] where the beam pipes are set in contact with a bath of liquid helium at $1.9\,\mathrm{K}$ resulting in an efficient condensation of residual gas molecules on the walls of the vacuum chamber. At four locations along the LHC ring the two beams pass into one chamber where they can be made to collide. Time varying electric fields produced in sixteen different radio frequency (RF) [78] cavities are used to accelerate the protons to their nominal energy. Due to the way the protons are being accelerated, the beams are not continuous, but divided into packets of protons called bunches. Each beam is made up of 2556 bunches [79] with approximately 1.1×10^{11} protons per bunch. Magnetic fields of dipole magnets are used to guide the beams around the ring [80]. To keep the bending radius of the protons, ρ, constant, the magnetic field, B, has to increase with the proton
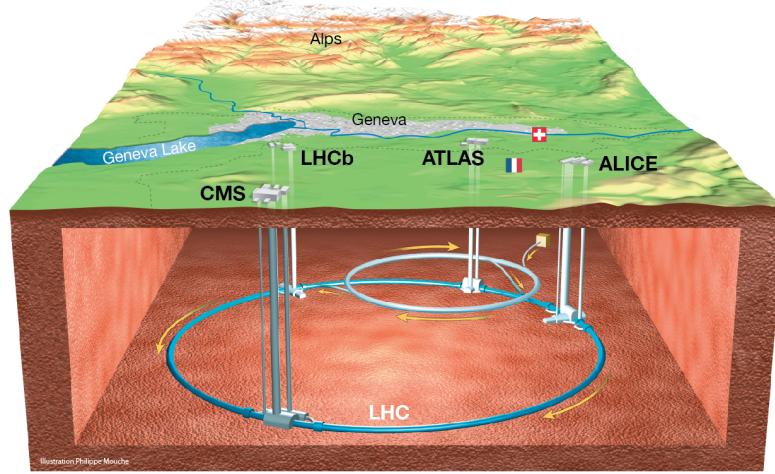
Figure 3.1 – The LHC situated at CERN in the former LEP tunnel approximately 100 meters beneath the French-Swiss border [76].

momenta, p, as expressed in Eq. 3.1 where q is the charge of the proton [81]:

$$B\rho = \frac{p}{q} \,. \tag{3.1}$$

At maximum energy, each proton has an energy of about 7 TeV requiring a magnetic field of 8 T, which can only be achieved with superconducting electromagnets. At the LHC, these are made of copper stabilised niobium-titanium (NbTi) Rutherford cables [82] and are operated in superfluid helium at a temperature of 1.9 K. Besides steering the beam with dipole magnets, magnetic multipoles are needed to focus the beam and counteract unwanted phenomena [83] such as gravitational effects, electromagnetic interactions among the protons, interactions with the pipe wall or effects of the magnetic fields of the experiments.

When the beams have reached their maximum energy, they are made to collide at four interaction points corresponding to the location of the experiments ATLAS [84], CMS [85], ALICE [86] and LHCb [51]. Each detector is distinct and run by a different collaboration of scientists. ATLAS and CMS are general purpose particle detectors, while the ALICE detector is specialised to study the physics of strongly interacting matter at extreme densities. The LHCb detector is optimised for studying heavy flavour physics, in particular, decays of b and c hadrons. Before protons are injected into the LHC ring, they are prepared and accelerated through different stages in the CERN accelerator complex [52] as illustrated in Fig. 3.2. To begin with, a metal cylinder (Duoplasmatron) [87] containing hydrogen gas surrounded by an electrical field that breaks down the gas into electrons and protons. The protons are then accelerated inside the linear accelerator LINAC 2 [88] before being injected into a series of circular accelerators that successively increase their energy. The first step after LINAC 2 is the Proton Synchrotron Booster (PSB) [89] where protons reach an energy of 1.4 GeV. This is followed by the Proton Synchrotron (PS) [90] where
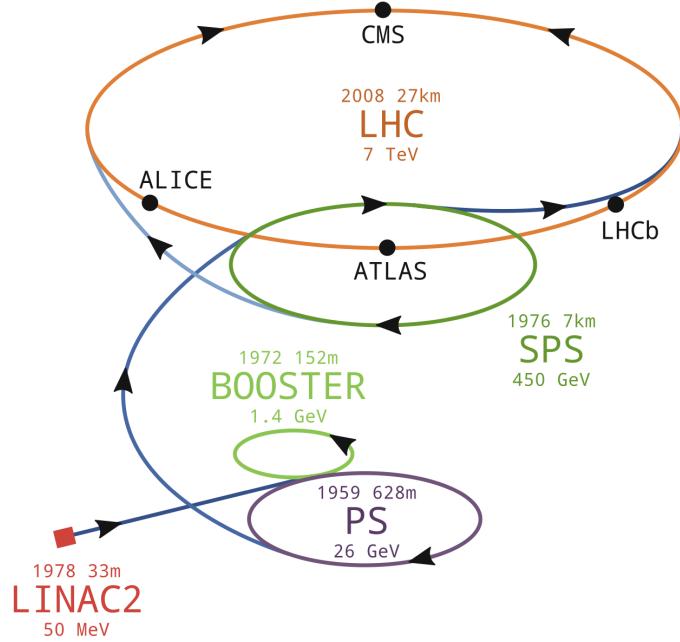
Figure 3.2 – The proton acceleration chain in the CERN Accelerator Complex during LHC Run 2. The maximum energy per beam, accelerator circumference and startup year are specified [92].

protons are accelerated in bunches separated by 25 ns [52]. After reaching 26 GeV, they are passed to the Super Proton Synchrotron (SPS) [91] where they are accelerated to 450 GeV, before, finally, being injected into the LHC ring in both the clockwise and counterclockwise directions.

The two key parameters to quantify the performance of a particle collider, such as the LHC, are the centre of mass energy, $\sqrt{s}$, and the luminosity, $\mathcal{L}$. The centre of mass energy defines the available energy in the collision. At the LHC, where identical particles, protons, of the same energy, E, are made to collide head-on, the available energy is simply given by $\sqrt{s} = 2E$. The luminosity, together with the particle production cross-section, σ, determines the rate of collisions corresponding to $dN_{\mathrm{coll}}/dt = \sigma\mathcal{L}$. In other words, the higher the luminosity, the more collisions. The luminosity depends on the number of protons, N_p, the frequency of bunch crossings, f_{bc}, and the transverse size of the bunch, σ_b, in the following way [93]:

$$\mathcal{L} = \frac{f_{bc}N_p^2}{4\pi\sigma_b^2} \,.$$
(3.2)

The transverse beam size is usually expressed in terms of the β-function [94], which is related to the transverse oscillation of the beam along the nominal trajectory (z axis). It is given by $\beta(z) = \pi\sigma_b(z)/\varepsilon$ where ε is the transversal emitance, which is a constant that depends on the initial preparation of the beam. To maximize the interaction rate, quadrupole magnets are used to focus the beam such that the magnitude of the β-function

at the interaction point, known as β^*, is as small as possible. In 2018 the LHC achieved a β^* of 25 cm [79] in ATLAS and CMS resulting in a peak luminosity of $2.1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which is twice as high as the LHC design luminosity and resulting in more than one billion inelastic proton-proton collisions every second [95]. The main limitation on β^* at the LHC comes from machine protection considerations [96]. This has to do with the evolution of the β-function around the interaction point that can be expressed as $\beta(z) = \beta^* + \frac{z^2}{\beta^*}$. This relation implies that the smaller β^* is, the faster the β-function, and thereby the beam size will rise as one goes away from the interaction point.

The first operational run of the LHC (Run 1) took place from the beginning of 2010 until the end of 2012 at $\sqrt{s} = 7\,\mathrm{TeV}$ for the first two years, and $8\,\mathrm{TeV}$ during the last year [97]. After a two-year shut down period where the accelerator and detectors where made ready to operate at $\sqrt{s} = 13\,\mathrm{TeV}$ [98], the second operational run of the LHC (Run 2) took place from 2015 until the end of 2018 [99]. The total integrated luminosity for ATLAS and CMS after Run 1 and Run 2 was $189.3\,\mathrm{fb}^{-1}$. The third operational period of the LHC (Run 3) started in July 2022 and will continue until 2026 with a centre-of-mass energy reaching $13.6\,\mathrm{TeV}$ [100].

3.2 The LHCb experiment

The main goal of the LHCb experiment [51] is to study matter-antimatter asymmetry and rare decays of b and c hadrons to test the SM and search for NP.

Since the beginning of its operation, the LHCb experiment has provided various important results. Among these are the first measurement of the CKM matrix element $|V_{ub}|$ from the baryonic decay channel $\Lambda_b^0 \to p\mu^-\bar{\nu}$ [65] as well as the first measurement, together with CMS, of the leptonic decay $B_s^0 \to \mu^+\mu^-$ [101, 102]. In addition, several high precision tests of lepton flavour universality have been performed by LHCb. This includes the branching fraction ratio measurements of $R(D^{(*)}) = \mathcal{B}(B \to D^{(*)}\tau\bar{\nu}_\tau)/\mathcal{B}(B \to D^{(*)}\mu\bar{\nu}_\mu)$ [103, 104] that agree with the SM within 2σ leading to a deviation of the world average with respect to the SM of 3.2σ [104, 105], which could hint to NP. Moreover, LHCb has measured the ratios $R(K^{(*)}) = \mathcal{B}(B \to K^{(*)}\mu\mu)/\mathcal{B}(B \to K^{(*)}ee)$ [106, 107]. However, while the previous measurements found a deviation between $R(K^{(*)})$ and the SM prediction of up to 3.1σ, the latest measurement with improved background descriptions and larger data samples finds $R(K^{(*)})$ to be in perfect agreement with the SM [108].

The geometry of the LHCb detector, illustrated in Fig. 3.4, has been optimised for studying b hadrons produced in pp collisions at the LHC. In these conditions, most b hadrons are produced in gluon fusion [109], as illustrated by the two Feynman diagrams in Fig. 3.3.

The momenta of the two incoming gluons are highly asymmetric and consequently the $b\bar{b}$ pair is emitted along the same direction of the high momentum gluon with a small angle with respect to the beam pipe. Subsequently, b hadrons are formed from the $b\bar{b}$ pairs in the nonperturbative process of fragmentation, and thus, the b hadrons are emitted in the same
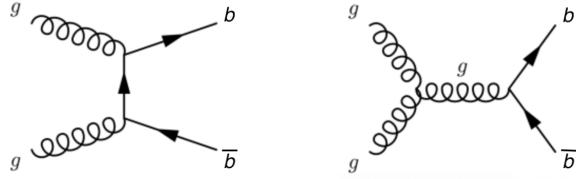
Figure 3.3 – Feynman diagrams illustrating gluon fusion processes at the LHC resulting in $b\bar{b}$ pairs.
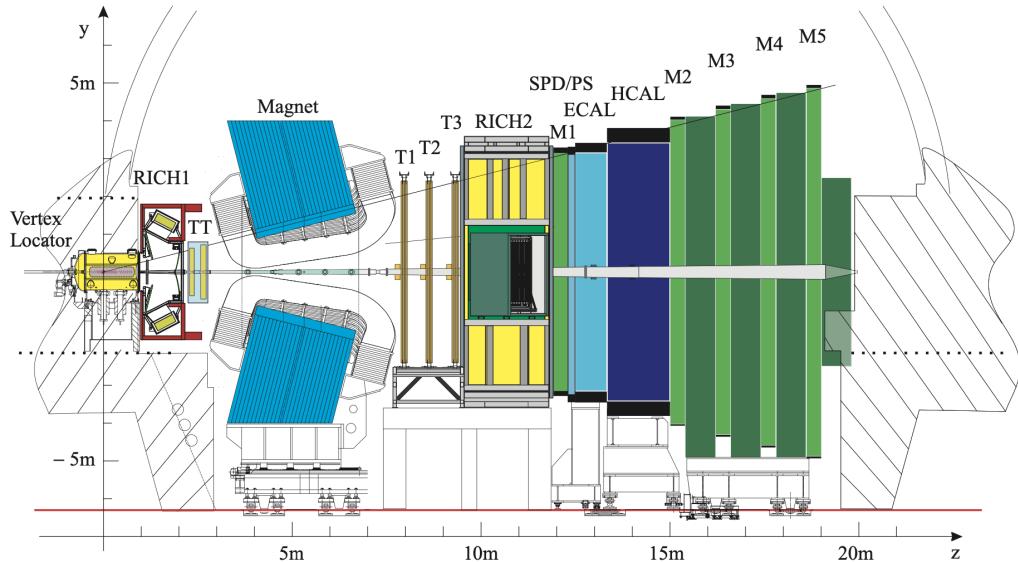
Figure 3.4 – Illustration of the LHCb detector and its subdetectors arranged along the beam pipe [51].

forward or backward cone. Due to space limitations, the LHCb detector only covers the forward region with an angular coverage of approximately 10–300 mrad in the horizontal bending plane of the LHCb magnet, and 10–250 mrad in the vertical non-bending plane. This ensures a high geometric efficiency for detecting all decay products from b hadrons emitted in the forward direction.

Although, the LHCb acceptance is optimised for detecting b hadrons, they are only present in 1% of all inelastic interactions, and consequently, a hard and efficient selection is required. This is done by the trigger system, which uses information from several LHCb subdetectors to select interesting events and reject backgrounds. The trigger system consists of a low-level hardware trigger (L0), and a high-level software trigger (HLT). Together they reduce the event rate from 40 MHz to a few kHz.

One of the characteristics of b and c hadrons are their long lifetimes, which combined with a high relativistic boost, results in an average flight distance of approximately 1 cm making it possible to separate the point where they were created (PV) from the point where they decayed (SV). To exploit this signature, an excellent vertex resolution is required.

Furthermore, to facilitate the event reconstruction and to reduce radiation damage, the number of *pp* collisions per bunch crossing is reduced. This is achieved by luminosity levelling, in particular, by displacing the beams at the LHCb interaction point resulting in a luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [51]. Additionally, a good momentum resolution and particle identification are required in order to accurately reconstruct particle masses, discard backgrounds consisting of random combinations of tracks and precisely identify decay products.

The LHCb detector is described in a coordinate system, where the z axis is coincident with the beam pipe and points from the interaction point at $z = 0$ into the forward direction of the experiment. The y axis points vertically upwards and the x axis points outwards with respect to the LHC ring. The LHCb detector is approximate $20\,\mathrm{m}$ long and $10\,\mathrm{m}$ tall. The various subdetectors arranged along the beam pipe are designed to meet the above mentioned requirements and can be divided into two main systems.

- **The tracking system** measures particle momenta and directions. It consists of the Vertex locator (VELO) surrounding the interaction point, and the TT and T (T1, T2 and T3) stations placed upstream and downstream of the spectrometer magnet, respectively.

- **The particle identification system** is used to assign particle hypotheses to final state particles. It consists of the two RICH detectors (RICH1 and RICH2) placed upstream and downstream of the magnet, the calorimeter system (SPD, PS, ECAL and HCAL) downstream of the magnet and the muon system placed in the very back of the detector.

Most subdetectors are assembled in two halves making it possible to move them out horizontally away from the beam. This is useful for assembly, maintenance purposes and for accessing the beam pipe [51, 109].

The following sections provide an overview of the two main systems with their respective subdetectors, as well as the trigger system and the LHCb software. In places where the experiment settings or performance differ between Run 1 and Run 2, only the latter will be explained, as this thesis only uses data collected during Run 2.

3.3   The tracking system

The main task of the tracking system is to provide an efficient reconstruction of charged particles, in particular, their trajectories, momenta and the sign of their electric charges.

The first tracking detector encountered by particles produced in the collision is the Vertex Locator (VELO) which surrounds the interaction point. After the VELO, particles travel through the Tracker Turicensis (TT) located before the magnet. As charged particles travel through the magnetic field their trajectories are bent according to their momenta.

After the magnet, they encounter the T stations (T1–T3), where particles travelling with small angles with respect to the beam pipe, traverse the Inner Tracker (IT) that occupies the central region of the T stations, while particles travelling with larger angles, traverse the Outer Tracker (OT) that occupies the outer regions of the T stations. For a track to reach the downstream stations its momentum has to be at least 1.5 GeV/c [110].

The tracking system is based on two different technologies [111]. The VELO, TT and IT detectors are based on the more expensive silicon technology, which offers the very fine resolution required in regions of high particle density, while the OT is based on the cheaper gaseous straw tube technology resulting in a coarser resolution, which is acceptable in the regions of low particle density. The material budget of the tracking system is kept as low as possible to reduce the number of interactions in the detector material [51]. In general, these interactions result in the production of photons and electrons, which makes it harder to identify photons and electrons actually originating from b- or c-hadron decays. Multiple scattering of particles degrades their momentum resolution, and hadronic interactions completely stop them from traversing the detector.

All parts of the LHCb detector are exposed to radiation, however, the subdetectors close to the beam pipe such as the silicon trackers receive the highest doses [112], which result in radiation damages. In particular, the hadronic flux causes displacements of atoms in the silicon from their lattice sites. This changes the leakage current as well as the effective doping concentration, and over time, the performance of the detectors degrade. For this reason, the LHCb experiment is monitoring the radiation damages to ensure good quality of data and to initiate investigation or exchange of detector parts [112–114].

3.3.1 The Vertex Locator

The VELO [115] provides precise measurements of track coordinates close to the pp collision point. These are used to reconstruct the primary vertex and identify displaced secondary vertices that are distinctive features of b- and c-hadron decays.

The VELO is a silicon micro-strip detector surrounding the pp interaction region as illustrated in Fig. 3.5. It consists of two sets of 21 half circular silicon sensors, each 300 μm thick, arranged along the beam axis. The VELO is designed in such a way that all tracks contained inside the LHCb acceptance and originating from a PV within $|z| < 10.6$ cm will traverse at least three silicon sensors [112]. Half of the silicon sensors measure the radial coordinate (r), while the other half measures the azimuthal coordinate (ϕ). They are arranged into pairs of r and ϕ sensors glued back to back. Together with the coordinate along the beam axis (z), the 3D position of the particle interaction with the detector, also known as the hit position, can be obtained. The detector hit resolution is a function of the sensor thickness and track angle, however, the best hit resolution measured is around 4 μm [112].

The VELO is operated in a vacuum that is separated from the LHC machine vacuum
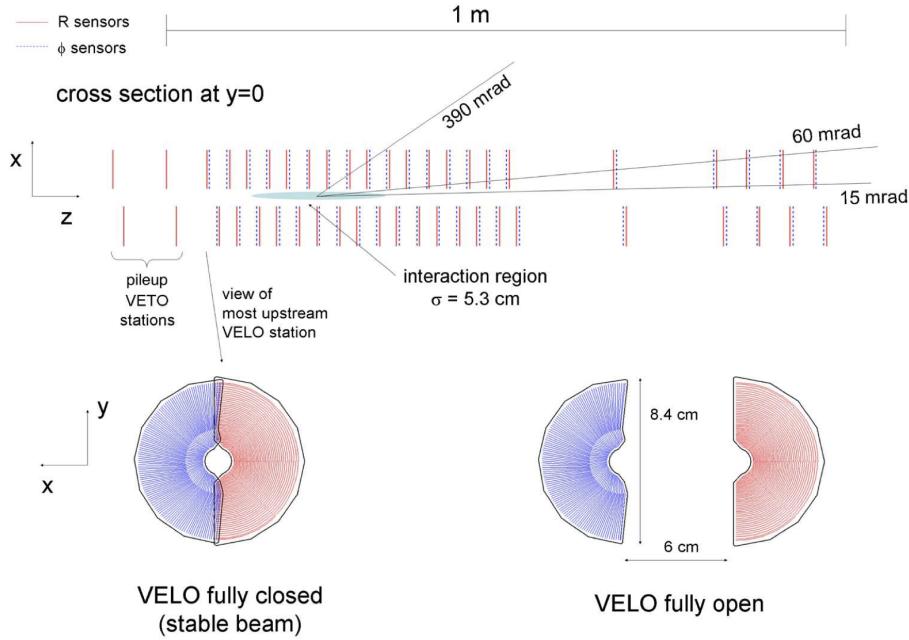
Figure 3.5 – The VELO consisting of two sets of half circular silicon modules arranged along the beam axis [51].

by an RF box [51]. It consists of sheets of corrugated aluminium separating the VELO sensors from the beam. This is to avoid interference between the beam and the VELO, as well as protecting the LHC machine vacuum from outgassing of the detector components. The RF foil contributes $\sim 43\%$ to the material budget of the VELO making it the second largest component after the silicon sensors. On average particles traverse 0.227 radiation lengths (X_0) of material before leaving the VELO at a z position of 83.5 cm [112]. The resolution of the IP is limited by the RF foil. This is mainly due to multiple scattering, but also due to the distance between sensors and beam enforced by the RF foil.

During operation the VELO is placed as close as possible to the beam, *i.e.* at 7 mm [112], where it is exposed to high levels of radiation. In Run 1 the particle fluence measured up to 5×10^{13} 1 MeV neutron equivalents $/ \, cm^2 / \, fb$. In order to make the VELO more resistant to radiation damages it is made of radiant tolerant oxygenated n^+-on-n sensors operating at a temperature of approximately 7°C. During beam injection where the beam size and the beam do not collide, the two sets of silicon sensors are moved apart to avoid damages.

Information from the VELO is used to compute several important quantities extensively used in LHCb physics analyses. One of these is the impact parameter (IP) defined as the shortest distance between the track and the PV. Since the decay vertices of long-lived particles such as B and D mesons are displaced from the PV, they are characterised by larger IPs compared to particles produced at the PV (prompt particles). By imposing requirements on the IP of signal candidates, prompt backgrounds can be efficiently rejected resulting in an increased signal-to-background ratio. This works well thanks to the precise measurements provided by the VELO where the IP resolution is less than 35 μm for tracks
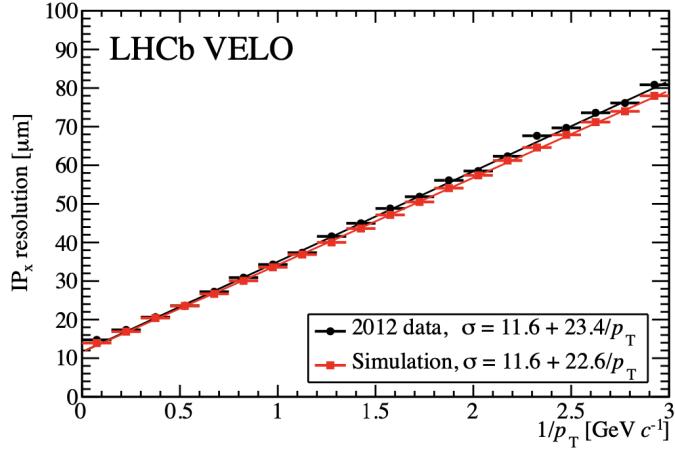
Figure 3.6 – IP resolution along the x axis as a function of $1/p_T$ for 2012 LHCb data (black) and simulation (red) [112].

with transverse momentum (p_T) greater than $1\,\text{GeV}/c$ [112], as illustrated in Fig. 3.6.

Other important quantities include the PV and SV positions, as well as decay lengths. The resolution of the PV which is strongly dependent on the number of tracks in the vertex, is $13\,\mu\text{m}$ [112] in the x and y coordinates and $71\,\mu\text{m}$ in z coordinate for 25 tracks. Using momentum information, decay lengths can be converted into proper decay times, and the typical decay time resolution is about $50\,\text{fs}$.

3.3.2 The dipole magnet

The LHCb Magnet [116] is a warm dipole magnet, which deflects charged particles in the horizontal plane. The curvatures of the trajectories together with the strength of the magnetic field is used to determine the momenta of particles and the sign of their charges.

Positive and negative particles are deflected in opposite directions in the horizontal plane. To minimise detection asymmetries caused by differences in the detector performance and material distribution in the left and right sides, the magnetic field direction is changed regularly. This results in two sets of data taken, each with their polarity, also known as the MagDown and MagUp samples.

The magnet consists of two saddle-shaped coils mounted mirror-symmetrically to each other inside of an iron window frame as illustrated in Fig. 3.7. Each coil consists of fifteen aluminum "pancakes" arranged in five triplets. Aluminum is chosen as conductor material for reasons of cost and reliability. The magnet produces a vertical field in the gap, and due to the shape of the coils that follow the LHCb acceptance, the magnetic field covers the entire forward acceptance of the LHCb detector. The integrated field is about 4 Tm, which combined with simulations and measurements with an array of Hall probes of the magnetic field inside the detector volume, ensures a good performance of the tracking
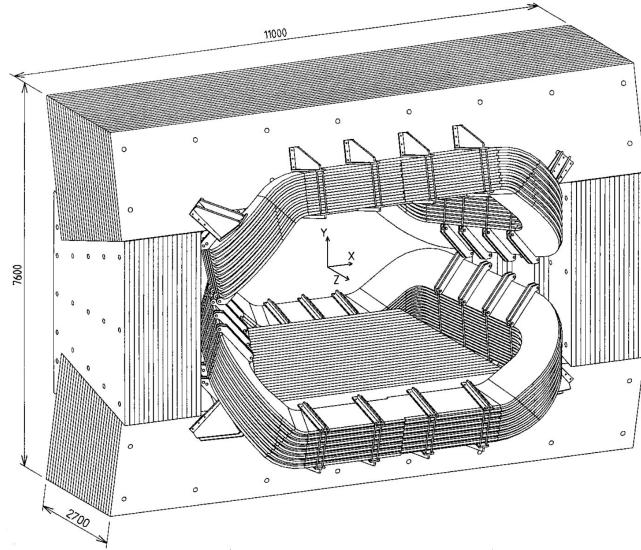
Figure 3.7 – Illustration of the LHCb dipole magnet [116].

system corresponding to a relative momentum resolution of less than 1.0% for momenta up to 200 GeV/c [116].

3.3.3 The Silicon Tracker

The Silicon Tracker [117] consists of two detectors: the Tracker Turicensis (TT) situated upstream of the magnet and the Inner Tracker (IT) situated downstream of the magnet. Both detectors use p^+–on–n silicon microstrip sensors that are well suited for high particle densities due to their fast response, high spatial resolution and radiation hardness. Together they detect hits of charged particles in the high particle-density regions after the VELO. As charged particles traverse the silicon, electrons are kicked loose from the atoms resulting in the creation of electron hole pairs. They drift in the presence of the applied electric field towards the strips, where the charges are collected resulting in an electric pulse that is amplified and detected. This information is then used to reconstruct the hit position of the traversing particle.

The strip pitch, which is the distance between the strips, is approximately 200 μm for both the TT and IT. This makes it possible to reach the required fine spatial hit resolution of around 50 μm [110], which is necessary for distinguishing hits from different particles in the high particle density regions. The silicon sensors provide a high hit efficiency close to 100% and a good signal-to-noise ratio, which reduces the risk of reconstructing tracks from noise. Both the TT and IT are operated at a temperature below 5°C inside a volume shielded from light as well as thermally and electrically insulated [51].
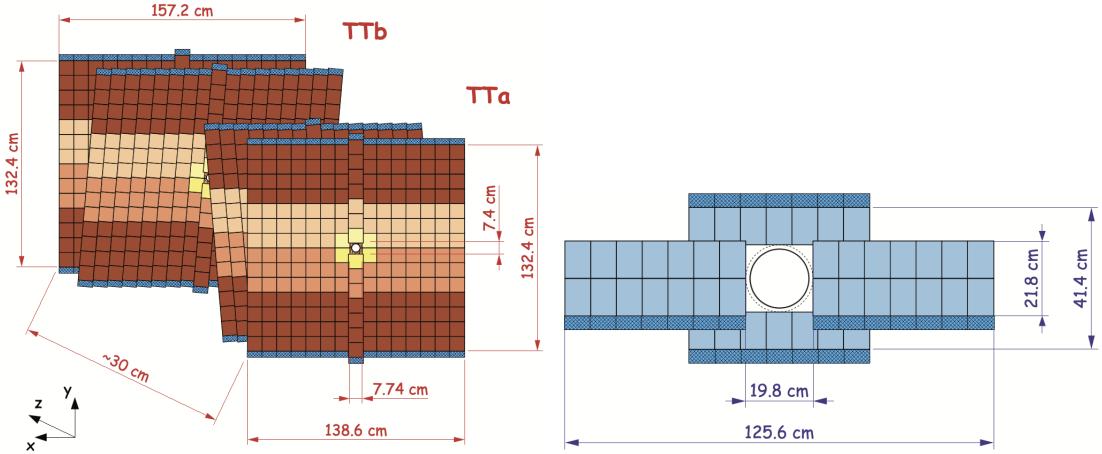
Figure 3.8 – Layout of the Silicon Tracker modules, showing the four layers of the TT (left) and one of the four IT detector boxes arranged around the LHC beampipe (right) [117].

The Tracker Turicensis

The TT[1] is placed before the dipole magnet and covers the full LHCb acceptance [51]. It consists of four approximately $150 \, \text{cm} \times 130 \, \text{cm}$ planar detection layers arranged in two pairs separated by approximately $30 \, \text{cm}$ along the beam axis. In order to measure the 2D coordinates of a track, the silicon strips of the first and fourth layer are aligned vertically, while the second and third layers are tilted by an angle of -5 and $+5$ degrees with respect to the vertical axis, as illustrated in Fig. 3.8. This arrangement is also known as the *xuvx* geometry.

Each detection layer is made of half-modules covering half of the height of the detection layer. A half-module consists of seven silicon sensors. To avoid acceptance gaps and to be able to perform alignment of the modules, neighbouring modules are staggered by approximately $1 \, \text{cm}$ in the z direction and overlap by a few millimeters in the x direction. The material distribution of the TT is rather uniform with $0.04X_0$ where 50% is due to the active material of the silicon sensors.

The Inner Tracker

The IT detector [119] covers a cross-shaped region of about $120 \, \text{cm} \times 40 \, \text{cm}$ in the centre of the three large tracking (T) stations downstream of the magnet, as illustrated in Fig 3.8. Despite covering only 1.3% of the active surface of the T stations, the IT measures about 20% of all charged particles due to the high density of particles close to the beampipe. There are four detection layers arranged in a *xuvx* geometry in each detector box [51]. All detection layers are made of seven detector modules. As for the TT, to avoid acceptance gaps and to perform module alignment, neighbouring modules are staggered by $4 \, \text{mm}$ in

---

[1]Its peculiar name "Turicensis" originates from the Gallo-Roman settlement once situated at today's city of Zurich [118], where the TT was produced.
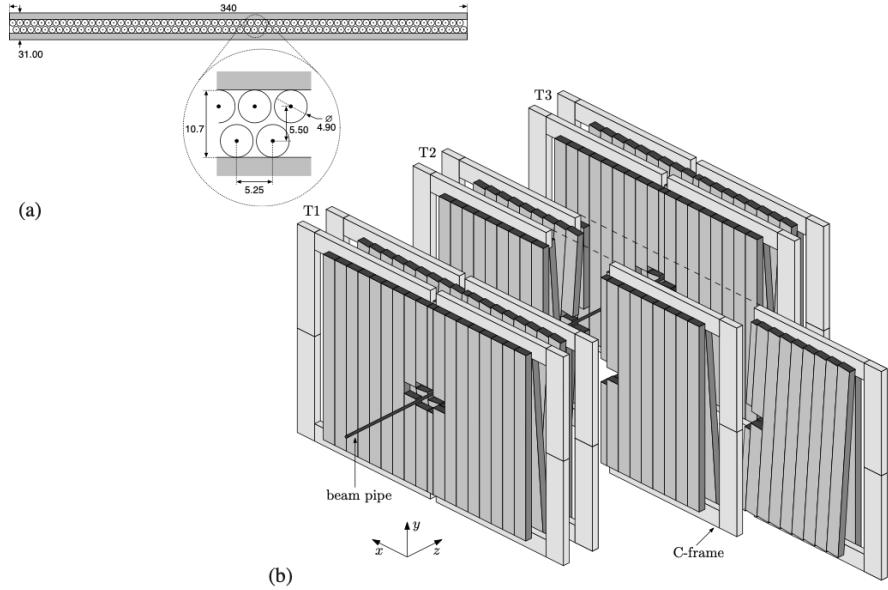
Figure 3.9 – Layout of the OT stations with the module cross section (a), and the arrangement of OT straw-tube modules in layers and stations (b) [120].

the z direction and overlap by 3 mm in the x direction. The material distribution for the IT is less uniform compared to the TT due to readout hybrids, mechanical supports, cooling pipes and cables placed inside the LHCb acceptance. Close to the beam pipe the material budget amounts to $0.035X_0$ per station where more than $0.015X_0$ is due to the active material of the silicon sensors.

3.3.4 The Outer Tracker

The objective of the OT [120] is to detect hits of charged particles in the low density region, thus it covers the active area of the three T stations that surrounds the IT, as illustrated in Fig. 3.9. This area is around 600 cm × 490 cm for the largest T station.

The OT is a gaseous straw-tube detector with a coarser resolution than the silicon detector, however, it is cheaper and therefore chosen to cover the large areas of low particle densities. As charged particles traverse the straw-tubes they ionise the gas. The positive ions and negative electrons drift each in their direction due to the applied electric field. Their drift-times are used to reconstruct the hit position [111].

The OT consists of three stations composed of four layers arranged in an *xuvx* geometry. Each OT station consists of 72 separate modules containing two layers of drift-tubes with an inner diameter of 4.9 mm. By using a mixture of Argon (70%), CO_2 (28.5%) and O_2 (1.5%) inside the tubes, a fast drift time below 50 ns and a good resolution of the position of single hits corresponding to 200 μm is obtained [51].
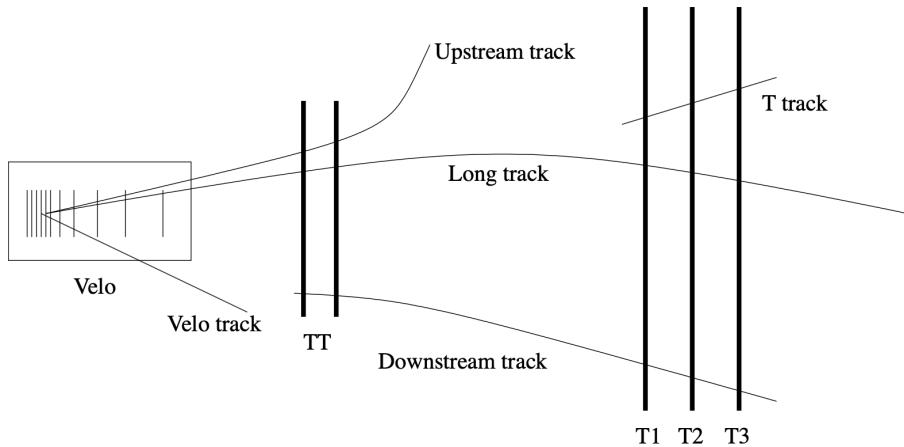
Figure 3.10 – Track types based on traversed subdetectors [121].

3.3.5 The track reconstruction

Reconstructed tracks fall into different categories depending on how the tracks are reconstructed in the various subdetectors as illustrated in Fig. 3.10. These track categories are defined in the following way:

- **VELO tracks** only leave hits in the VELO, and they are usually characterised by large angles with respect to the beam axis or a backward flight direction. They are mainly used to reconstruct the primary interaction point.

- **Long tracks** leave hits in the VELO, T stations and optionally the TT station. As they travel through the entire magnetic field their momentum can be determined with the highest precision, and therefore, they are the most important tracks for physics analyses.

- **Upstream tracks** leave hits in subdectors situated upstream of the magnet, *i.e.* the VELO and TT. In general, these are characterized by an insufficient momentum for traversing the magnetic field to reach T1-T3. They are mostly used in background studies related to the particle identification algorithm of the RICH1, however, they are also used in some physics analyses, *e.g.* $K_S^0 \to \pi^+\pi^-$.

- **Downstream tracks** leave hits in the TT and the T1–T3 stations. They are used to reconstruct long lived particles, for instance K_S^0 and Λ that decay outside of the VELO acceptance.

- **T tracks** only leave hits in the T1–T3 stations, and in general, they arise from secondary interactions. They are used in background studies related to the particle identification algorithm of the RICH2 [110].

To reconstruct different track types various algorithms are used [122]. In the first step, a pattern-recognition algorithm is used to find the track. There are two independent

algorithms: the VELO tracking [123] and T-seeding [124] algorithms. These provide information to the other algorithms that perform the next steps of the track reconstruction.

Long tracks can be reconstructed by the forward tracking algorithm [125] and the matching algorithm [126,127]. In the first case, a search for track candidates in the VELO is launched by the VELO tracking algorithm. The forward tracking algorithm extrapolates the VELO tracks through the magnet to the T stations. To avoid loss of track efficiency due to acceptance gaps in the TT, hits in this detector are not used by the forward tracking algorithm. The search window in the T stations is defined in such a way that hits from a maximally deflected particle of p_{T} above $80\,\mathrm{MeV}/c$ would be considered. In the second case, the VELO tracking and T-seeding algorithms search for track candidates in the VELO and T stations, respectively. The matching algorithm extrapolates the tracks and matches them in the magnet region. The forward tracking and the matching algorithms are complementary, and can therefore compensate each other in case one algorithm experiences a loss in efficiency. In the final step, the track is fitted by a Kalman filter [128].

The track reconstruction efficiencies for long tracks depend mainly on the particle kinematics and the number of charged particles in the event. The average track reconstruction efficiency is found to be above approximately 96% [122]. The relative uncertainty on the measured momentum ranges from 0.5% at low momentum up to 1.0% at $200\,\mathrm{GeV}/c$ [129].

3.4   The particle identification system

The objective of the particle identification (PID) system is to determine the identity of particles. This is essential for selecting signal processes with final states consisting of muons, electrons, pions, kaons, protons or photons and for suppressing backgrounds associated with misidentified particles. The PID system consists of three parts: the Ring Imaging Cherenkov detectors (RICH), the calorimeters (PS, SPD, ECAL, HCAL) and the muon system.

3.4.1   The RICH detector

The primary task of the two RICH detectors is to identify and distinguish charged hadrons such as pions, kaons and protons. This information is used by the software trigger and also in the offline physics analyses [51].

Particle identification in the RICH detectors is based on the Cherenkov effect. This effect occurs when a charged particle travels through a medium with refractive index n at a speed v_p greater than the velocity of light in that medium, c/n, corresponding to $c/n < v_p < c$. It emits light with an angle relative to the velocity direction given by [130]

$$\cos\theta = \frac{c}{n\,v_p}\,, \tag{3.3}$$
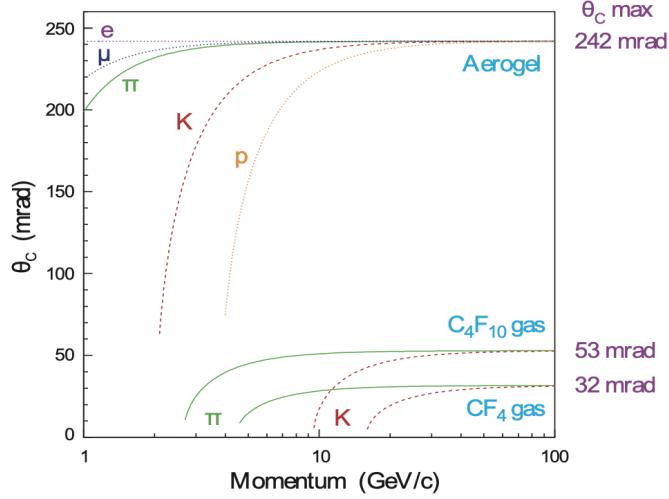
Figure 3.11 – Cherenkov angle versus track momentum for the RICH radiators. RICH1 uses a C_4F_{10} radiator, while RICH2 uses a CF_4 radiator (Run 2). Aerogel was only used in Run 1 [51].

where the relativistic momentum of the particle of mass m_p is $p_p = \gamma m_p v_p$ with γ being the Lorentz factor. Measuring the Cherenkov emission angle with the RICH detector and combining it with the already known refractive index and particle momentum measured by the tracking system, allows the determination of the particle mass. The ability to correctly identify a particle by this method depends on the momentum. It has to be high enough for the particle to emit Cherenkov light, however, as its momentum increases (the increase of the emission angle slows down and) the curves associated with different particles become more difficult to distinguish. The momentum range in which the relevant particles can be separated depends on the choice of the detector medium, also known as the radiator.

In general, particles of low-momentum have a larger angle with respect to the beam axis compared to high-momentum particles. Consequently, particles of low momentum emit Cherenkov light with larger polar emission angles than high momentum particles. In order for the RICH system to cover the full momentum range it is divided into two detectors, RICH1 and RICH2. They are placed upstream and downstream with respect to the magnet, see Fig. 3.4, and have different gas radiators.

In both RICH detectors the emission angle is measured by using a combination of spherical and flat mirrors to reflect the Cherenkov photons out of the LHCb acceptance where Hybrid Photon Detectors (HPDs) measure their spatial positions [51]. Inside the HPDs, a photon may convert on a photocathode and thereby release a photoelectron, which is accelerated by an electric field on to a reverse-biased silicon detector. Electron-hole pairs are created and the consequent energy deposit is measured resulting in a very high efficiency for detecting single photoelectrons. Information from the HPDs is used to reconstruct the Cherenkov light emission angle, which is an input for the particle identification (PID) algorithms.
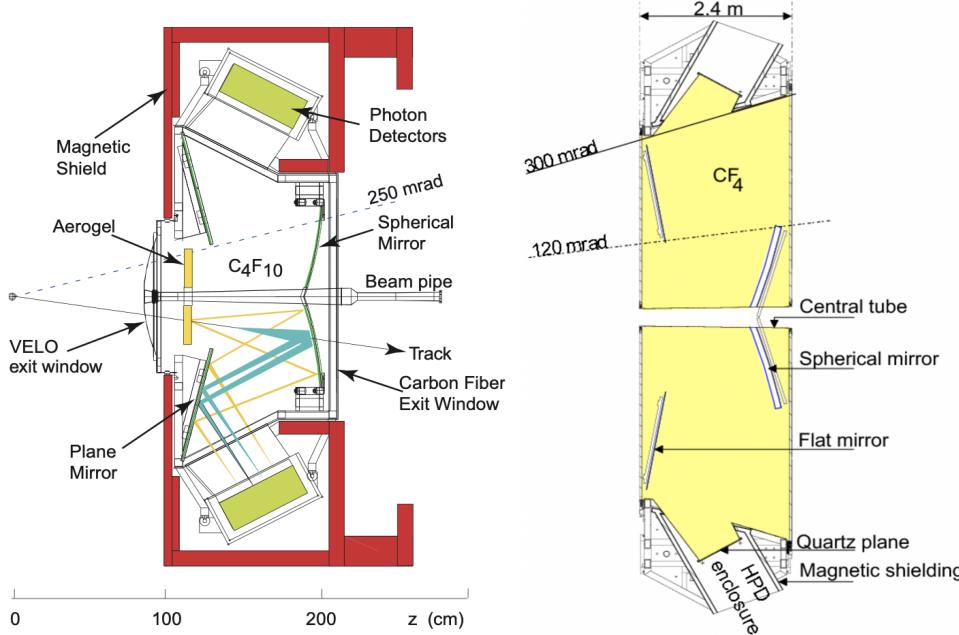
Figure 3.12 – Side view of the RICH1 detector (left), where the aerogel radiator was removed for Run 2, and RICH2 (right) [51].

The RICH1

The first detector, RICH1, provides PID information for low momentum particles. It is placed upstream of the magnet, between the VELO and the TT station, where it covers the full angular acceptance (see Fig. 3.12). It contains a C_4F_{10} gas radiator resulting in PID information for particles with momentum in the appropriate range 1–60 GeV/c [51], as illustrated in Fig. 3.11.

The RICH2

The second detector, RICH2, provides PID information for high momentum particles. It is placed downstream of the magnet just after the T stations. Although, its acceptance is limited, it covers the relevant region where high momentum particles are produced (see Fig. 3.12). It contains a CF_4 gas radiator resulting in PID information for particles with momentum in the appropriate range 15–100 GeV/c [51], as illustrated in Fig. 3.11.

3.4.2   The calorimeter system

The calorimeter system measures the energy and position of the final state particles. It has several purposes such as providing transverse energy measurements for hadron, electron and photon candidates for the first trigger level (L0). Furthermore, it distinguishes between electrons, photons and hadrons, and it is the only system that can provide information on photons.
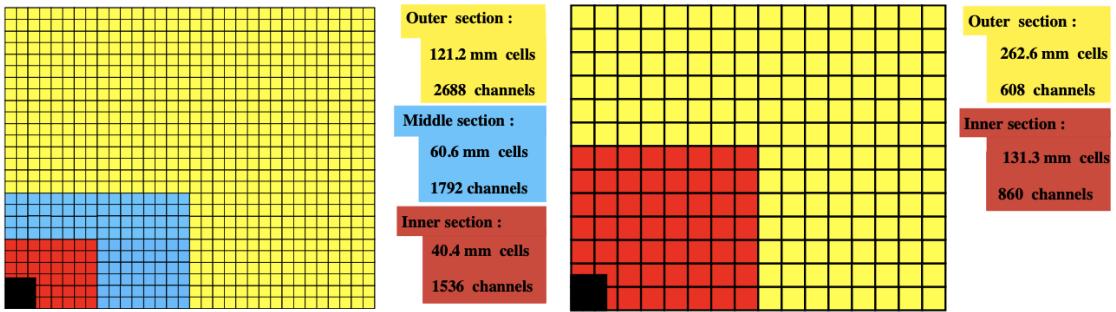
Figure 3.13 – Segmentation of SPD, PS and ECAL (left), and HCAL (right). The figure illustrates one quarter of the detector front face [51].

The calorimeter system is composed of four subdetectors placed after each other downstream of the RICH2. The first one is the Scintillating Pad Detector (SPD) that identifies charged particles, next there is the Pre-Shower detector (PS) that helps to distinguish between electrons and photons, and finally, there are the electromagnetic and hadronic calorimeters (ECAL and HCAL) that measure the energy and position of electromagnetic and hadronic showers.

All the subdetectors rely on the same technology where scintillating light is transmitted by optical fibres to photo multipliers. The hit density in the calorimeter varies by two orders of magnitude as one moves away from the beam pipe. This is why all subdetectors have varying segmentation in the transverse plane as illustrated in Fig. 3.13. The segmentation of the SPD, PS and ECAL is the same, and their dimensions follow a projective geometry with respect to the z axis, which is why the cell sizes increase with z.

The SPD/PS

The SPD and PS consist of two planes of scintillating pads separated by a 15 mm thick lead sheet corresponding to $2.5X_0$ [51]. The active surface of the detectors is about $7.6\,\mathrm{m} \times 6.2\,\mathrm{m}$, and is divided into three regions with different segmentation, with cells of dimensions of approximately $4\,\mathrm{cm} \times 4\,\mathrm{cm}$, $6\,\mathrm{cm} \times 6\,\mathrm{cm}$ and $12\,\mathrm{cm} \times 12\,\mathrm{cm}$, respectively. However, the dimensions of the SPD are about 0.45% smaller than those of the PS due to the projectivity requirement [51].

The SPD identifies charged particles, since only charged particles interact electromagnetically with the medium causing excitation or ionization followed by the emission of photons via scintillation. After the SPD, particles traverse a layer of lead. In the dense medium, electrons and photons will quickly loose energy as they collide and interact with electrons. This results in electromagnetic showers that can be measured by the PS. Hadrons, on the other hand, are much heavier than electrons and are therefore not affected much. To produce showers they will have to collide with nuclei, which due to the lower cross-sections, is less likely to happen. Hadrons will therefore travel a longer distance in

the calorimeter before producing hadronic showers. This is why the PS only identifies electrons and photons. The information from the SPD and PS is used by the L0 trigger lines to distinguish between hadron, photon and electron candidates and to reduce the complexity of events by requiring the number of hits in the SPD to be below a certain maximum limit [122].

The ECAL

The ECAL measures energies and positions of electrons and photons. To obtain an optimal energy resolution high-energy photon showers have to be contained in the ECAL. The thickness of the ECAL therefore corresponds to $25X_0$ [131]. The ECAL uses shashlik calorimeter technology, which consists of stacks of alternating layers of lead as absorber and scintillators as active material that are read out by wavelength shifting fibres. This type of calorimeter provides a good compromise between cost, fast response (approximately 25 ns) and stable operation in an environment of high radiation rate. The calorimeter is segmented and divided into the same inner, middle and outer regions as the SPD and the PS, illustrated in Fig. 3.13.

The relative energy resolution of the ECAL is given by [131]

$$\frac{\sigma(E)}{E} = \frac{(9.0 \pm 0.5)\%}{\sqrt{E}} \otimes (0.8 \pm 0.2)\% \otimes \frac{0.003}{E \sin \theta} \ , \tag{3.4}$$

where E is the energy of the particle in GeV, θ is the angle between the centre of the activated ECAL cell and the beam pipe. The first term is due to statistical fluctuations of the electromagnetic shower, the second term is a constant containing mis-calibrations, leakage, *etc.*, and the third term is due to noise. The maximum detectable transverse energy is set to 10 GeV, which is well above the typical energy deposits for *b*- and *c*-hadron decays in LHCb events.

The HCAL

The HCAL measures energies and positions of hadrons, such as pions, kaons and protons. This information is mostly used by the L0 trigger. The HCAL consists of alternating layers of iron acting as absorber and scintillating tiles acting as active material. The HCAL is segmented in the transverse plane, which enhance the collection of light compared to a perpendicular segmentation [131]. It is divided into an inner region with cells of 131.3 mm and an outer region with cells of 262.6 mm, as illustrated in Fig. 3.13. The thickness of the HCAL is limited by the space available in the LHCb cavern and is therefore only 5.6 hadronic interaction lengths (λ), where one λ is about 17 cm in iron. A leakage can occur for very energetic hadronic showers, *e.g.* showers produced by a 50 GeV pion. However, this does not affect the rejection of low E_T events, which is important for the hadronic trigger performance.

The relative energy resolution of the HCAL is given by [131]

$$\frac{\sigma(E)}{E} = \frac{(67 \pm 5)\%}{\sqrt{E}} \otimes (9 \pm 2)\% \,, \tag{3.5}$$

where E is the particle energy in GeV. The first term is due to statistical fluctuations of the shower, the second term is a constant containing mis-calibrations, leakage, *etc.* [132].

3.4.3 The muon system

The muon system provides essential information for the trigger system and for the identification of muons.

Compared to other final state particles that can be measured by LHCb, muons are the most penetrating ones. While electrons, pions, kaons, protons and photons are stopped in the calorimeter system, muons with momenta greater than $6\,\text{GeV}/c$ travel through the calorimeter system and the five muon stations. The reason for this is a combination of a long lifetime ($\tau_\mu \sim 2\,\mu\text{s}$) [15], low radiation loss due to the high mass ($m_\mu \sim 100\,\text{MeV}/c^2$) and low interaction probability. Muons do not interact strongly, and although they interact electromagnetically, and therefore collide with electrons in the detector medium, they only loose a small fraction of their energy thanks to a mass that is about 200 times higher than the one of the electron.

The muon system consists of five rectangular stations (M1–M5) which gradually increase in size along the beam axis, as illustrated in Fig. 3.14. M1 is placed in front of the SPD/PS to improve the p_T measurement, while M2–M5 are placed after the calorimeter. The L0 muon trigger uses information of the muon p_T by the track reconstructed from aligned hits in all five stations. The angular acceptances are 20–306 mrad in the vertical plane, and 16–258 mrad in the horizontal plane. With this acceptance, about 20% [51] of all muons produced in inclusive b semileptonic decays in 4π traverse the muon system.

Each station is composed of Multi Wire Proportional Chambers (MWPC) filled with a mix of carbon dioxide, argon and tetrafluoromethane, except the inner region of M1 where triple gas electron multiplier (GEM) detectors, which are more radiation hard, are used. Between the stations there is a 80 cm thick wall of iron to filter the muons, *i.e.* stop hadrons, electrons or photons which do enter the muon system. As the muon traverses the chamber, it interacts with the gas, and wire electrodes provide a space point measurement of the track. With information from M1–M3, which provide a high spatial resolution in the bending plane, the p_T of the muon candidate can be determined with a resolution of 20%. As M4 and M5 have a limited spatial resolution, their main purpose consist of identifying penetrating particles [51].
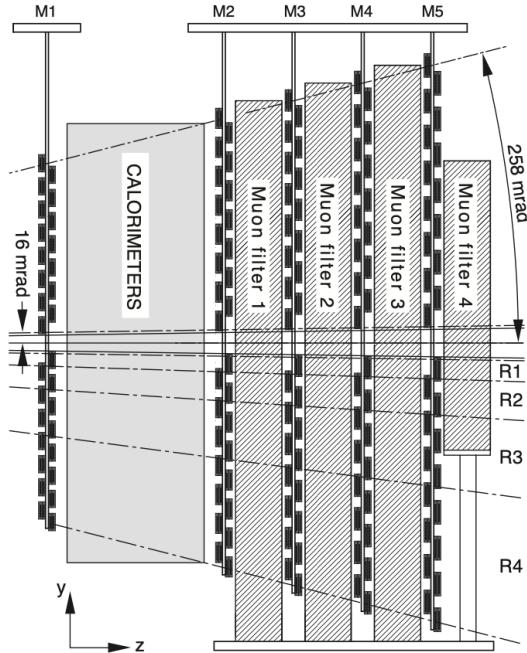
Figure 3.14 – Side view of the muon system [51].

3.4.4   The particle identification

As particles travel through the LHCb detector they interact with one or more subdetectors and leave signatures that can be used to determine their identities, as illustrated in Fig. 3.15. Charged particles such as electrons, muons, pions, protons and kaons are identified using information from the RICH, the muon system and the calorimeters, while the identification of neutral particles such as photons and neutral pions, use only information from the calorimeters [51].

The PID for charged particles

Each PID detector provides data to one or more algorithms that are used to associate a track with a particle species. Although these algorithms are very different, their likelihood information can be combined into a single powerful observable called the Combined Delta Log-Likelihood (CombDLL) [134], which is used in the trigger and offline analyses.

For instance, information from the RICH system on the Cherenkov angles and the track momenta measured by the tracking system is combined into an overall event log-likelihood, where all tracks of the event in the RICH system are considered simultaneously. The likelihood minimisation procedure starts by assuming all particles to be pions, as they are the most abundant particles in LHCb. For each track, the mass hypothesis is changed to electron, muon, kaon or proton and the likelihood is recomputed. The mass hypothesis that gives the largest increase in likelihood is assigned to that track. The RICH system
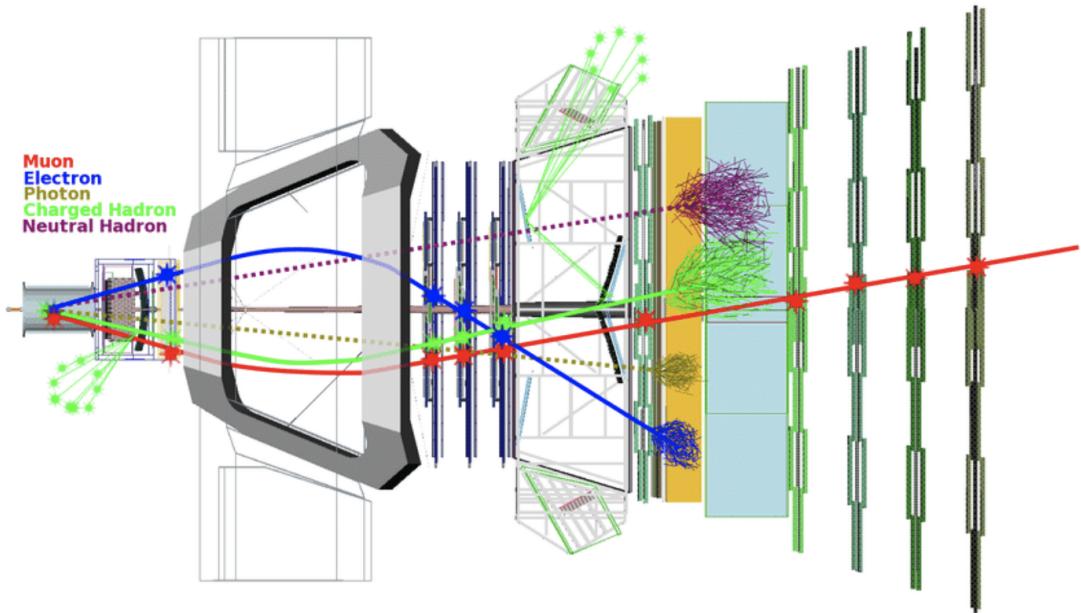
Figure 3.15 – Signatures of muons, electrons, photons, charged and neutral hadrons in the LHCb subdetectors [133]. Although not illustrated, the muons and electrons can produce Cherenkov light in the RICH detectors if their momenta are above the thresholds defined in Fig. 3.11).

proves to have a good particle identification over the full momentum range. In Run 1, the average efficiency for a kaon with momentum between 2 and $100 \, \mathrm{GeV}/c$ to be identified as a kaon was $\sim 95\%$, while the mis-identification of pions to kaons was $\sim 10\%$ [135]. In this way the algorithm associated with the RICH system results in likelihoods of each charged particle relative to the pion hypothesis, and likewise for the algorithms associated with the calorimeters and the muon system.

Finally, the likelihoods of the RICH, calorimeters and muon system are combined into the CombDLL. In this way, each track is assigned a set of CombDLL variables containing the global log-likelihood differences between each particle species and pion hypotheses [129]. In LHCb analyses, these variables are usually referred to as `PID` variables rather than CombDLL variables. The global log-likelihood difference between a particle species i and pion hypotheses for a track associated with particle j is therefore denoted `PID`$i(j)$. These variables are extensively used in LHCb analyses, together with another set of variables called `ProbNN` variables. These are the response of an artificial neural network trained on `PID` variables [134] and various track variables in order to determine the probability `ProbNN`$i(j)$ for a track j to be produced by a specific particle type i.

The PID for neutral particles

For neutral particles the PID information is only provided by the calorimeter system. Neutral particles are distinguished from charged particles in the calorimeter by the absence

of tracks that can be matched to the energy deposits.

To distinguish between photons and neutral pions (π^0) decaying into two photons the shapes of the clusters are considered. Moreover, photons can convert, *i.e.* the photon can interact with detector material upstream of the calorimeter and can convert into an electron-positron pair, which is identified using the PID detectors.

The most probable decay of a π^0 is into two photons. If the transverse momentum of the π^0 is below $2\,\mathrm{GeV}/c$ the two photon clusters can be separated. In this case, the obtained mass resolution of the π^0 is $8\,\mathrm{MeV}/c^2$ [131]. For π^0 with higher p_T the photon clusters overlap and a dedicated algorithm is used to disentangle a potential pair of photons resulting in a mass resolution of $20\,\mathrm{MeV}/c^2$ for these so-called merged π^0.

3.5   The trigger system

With a bunch crossing frequency of about $40\,\mathrm{MHz}$ delivered by the LHC, where only a few Hz consist of interesting b- and c-hadron decays, an efficient trigger system is required. The goal of the trigger system is to select events of interest, while discarding backgrounds, and thereby reducing the event rate to several kHz to meet both the data acquisition and computing capacities of the LHCb experiment. The trigger system selects events that pass certain requirements, for instance, those which contain decay particles of high p_T or E_T associated with b- and c-hadron decays. A collection of such requirements is called a trigger line.

The trigger system is divided into two stages as illustrated in Fig. 3.16. A low level hardware trigger (L0) based on Field Programmable Gate Arrays (FPGA) [51] uses raw detector information to perform a fast selection, and a High Level Trigger (HLT) based on software run on CPU-based computer farms that fully reconstructs and selects events.

3.5.1   The L0 trigger

The L0 trigger [136] reduces the event rate from $40\,\mathrm{MHz}$ to $1\,\mathrm{MHz}$ by selecting events with high p_T tracks in the muon system or high E_T clusters in the ECAL and HCAL. The read-out bandwidth of $1\,\mathrm{MHz}$ is shared between hadronic, muonic and electromagnetic trigger lines.

The hadronic and electromagnetic trigger lines search for clusters with high E_T in the ECAL or HCAL. For hadrons the HCAL cluster has to match a cluster in the ECAL. Furthermore, information from SPD and PS is used to distinguish between hadrons, photons and electrons [131].

The muonic trigger lines search for straight-line tracks in all five muon stations that are consistent with originating from the PV. Depending on the specific line (muon or di-muon),

LHCb Trigger Run 2
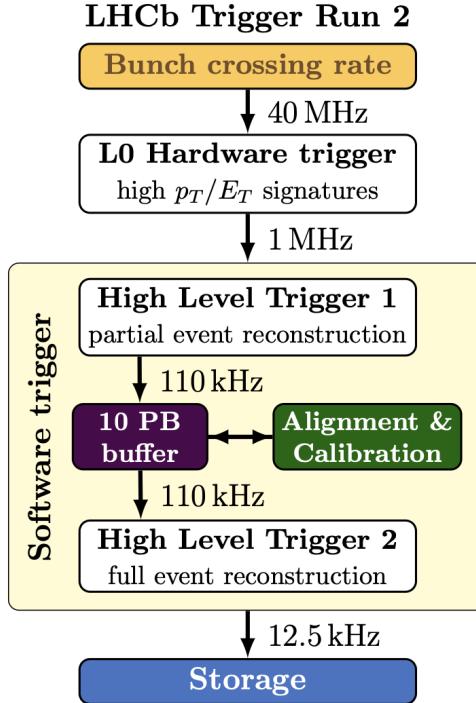
Figure 3.16 – The Run 2 trigger system, which reduces the event rate from 40 MHz to 12.5 kHz [122].

one or two muons with different p_T conditions are required.

3.5.2 The HLT

Events selected by the L0 trigger are transferred to the Event Filter Farm (EFF) consisting of 27,000 logical CPU cores with a total disk buffer of 10 PB [122]. Here the event rate is reduced from 1 MHz to around 12.5 kHz by the software-based HLT, which is divided into two different trigger stages called HLT1 and HLT2.

The HLT1 performs a partial event reconstruction. This consists of launching a sequence of algorithms that reconstruct vertices and long tracks. A pattern recognition algorithm is used to first reconstruct the VELO tracks. These are are then extrapolated to the TT detector, and finally, they are extended to the T stations to obtain long tracks. Only long tracks with a minimum p_T of 500 MeV/c are reconstructed. In addition to this, the fitted VELO tracks are reused to determine the position of the PVs. Due to time constraints PID algorithms are not executed, except for muons where the clean signature makes it possible to identify them already in this stage. The HLT1 trigger stage reduces the event rate from 1 MHz to around 150 kHz. The bandwidth is shared between several trigger lines optimised for selecting b- and c-hadron decays. A sub-sample of selected HLT1 events are used to perform real-time detector alignment and calibration.

The HLT2 provides offline precision by exploiting all available information from the tracking and PID systems [122]. It performs a full event reconstruction where tracks are now reconstructed without a minimum p_T requirement. The event rate is reduced to around 12.5 kHz and the remaining events are saved to permanent storage for physics analysis. The HLT2 bandwidth is shared between the inclusive trigger lines, which are mostly topological *b*-hadron, inclusive *c*-hadron and dimuon trigger lines, and the so-called TURBO stream, which consists of exclusive trigger lines that fully reconstruct specific decays. The inclusive topological *b* trigger lines look for decay vertices displaced from the PV consisting of two, three, or four tracks with high p_T. The lines are trained to recognise *b*-hadron decays that can be fully reconstructed inside the detector and reject those that cannot, as well as those formed by the decay products from *c* hadrons.

3.6 The LHCb software

The LHCb software [137] is developed within an object oriented framework called GAUDI [138]. It provides a common infrastructure and environment for software applications in all stages of data processing including simulation of physics and detector, high level software triggers, event reconstruction, physics analysis, *etc.*.

Monte-Calo (MC) simulation is a fundamental part of high-energy particle physics experiments like LHCb. It is used to understand experimental conditions and performances of detectors, and to validate physics analyses. At LHCb, simulations are provided by the GAUSS [139] application within the GAUDI framework. Initial particles are generated inside the GAUSS package. The subsequent tracking through the LHCb detectors is simulated using several external libraries. For example PYTHIA [140] is used to create inelastic *pp* collisions while the decay and time evolution of the produced particles is taken care of by EVTGEN [141], which is optimised for *B* meson decays. The final state QED radiation is managed by the PHOTOS [142] package. The physics interactions undergone by the particles travelling through the detector is handled by the GEANT4 toolkit [143]. The final stage of simulation is the BOOLE [144] digitization program. It generates detector responses according to the hits created in the previous step. It therefore provides a simulation of the detector response; in addition the program simulates the read-out electronics and the L0 trigger. The output has the same format as the real data.

In order to reconstruct events from real data as well as from simulation data, the BRUNEL [144] application is used. With this an identical processing of real data and simulated data is ensured. The HLT trigger lines are run by the MOORE software [145], and in the last stage of data processing, DAVINCI [144] provides the framework for the offline analysis. For instance, DAVINCI is used to select specific physics objects and compute a set of variables relevant for signal selecting and background rejection.

4 Signal selection

This chapter describes the selection of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay in data collected by the LHCb experiment in 2018. The signal selection is optimised with simulated samples of the signal and dominant background decays, and finally, it is verified through control studies. The general steps of the signal selection are:

- **Trigger selection** : a set of L0, HLT1 and HLT2 trigger lines are chosen to select events with signal candidates.

- **Stripping selection** : requirements are applied to select B mesons decaying into three tracks consistent with a final state of one muon and two charged pions.

- **Preselection** : simple requirements in one or two dimensions are applied to increase the ratio of signal to background, increase their separation in a discriminating variable or remove specific types of backgrounds.

- **MVA selection** : two multivariate classifier (MVAs) are developed, *i.e.* a charge isolation MVA is used to quantify the isolation of signal candidates, which is used by the main MVA, together with other properties of the candidates, to select signal in data.

- **Control studies** : a control channel is used to verify the selection, in particular, the MVA selection.

A detailed description of each step is provided in the following sections.

Table 4.1 – MC simulations used to develop the signal selection and verify the simulation of signal. Additional charged or neutral particles are represented with X and Y.

Decay	Objective	Type
$B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$	Signal	exclusive
$B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ with $\bar{D}^0 \to \pi^+\pi^- X$	Background	cocktail
$B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ with $D^- \to \pi^+\pi^- X$	Background	cocktail
$B^0 \to D^{*-}(\to \bar{D}^0\pi^-)\mu^+\nu_\mu$ with $\bar{D}^0 \to K^+2\pi^-\pi^+$	Background	exclusive
$B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0(\to \pi^+\pi^-)$	Control	exclusive
$B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$	Control Background	exclusive

4.1 Data and simulated samples

This analysis is based on data collected by the LHCb experiment at a centre-of-mass energy of 13 TeV in 2018 corresponding to an integrated luminosity of $2.19\,\mathrm{fb}^{-1}$.

Simulated events are used to develop the signal selection. In particular, the simulations of the signal and backgrounds are used to study their respective contributions in data, and to develop requirements, also known as *cuts*, that increase the ratio of signal with respect to background in data. The simulated samples are also used to train multivariate classifiers that further distinguish signal from background. Finally, to validate the signal selection, a simulation of the control channel is also needed. All simulated samples used in this chapter are listed in Table 4.1, more information on the software used to produce them can be found in Sec. 3.6.

The signal is modelled with BCL form factors [69] obtained from a combined fit to LCSR predictions [146] and measurements by Belle [62] and BaBar [63].

Three different simulations generated with 2018 data-taking conditions are used to describe the backgrounds. The first two backgrounds in Table 4.1 are so-called cocktail samples, which are used to describe the dominant backgrounds in data. They contain the sums of $B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ modes resulting in $\bar{D}^0 \to \pi^+\pi^- X$ and $D^- \to \pi^+\pi^- X$ decays (X and Y represent additional charged or neutral particles), respectively. The simulation of the background $B^0 \to D^{*-}(\to \bar{D}^0\pi^-)\mu^+\nu_\mu$ with $\bar{D}^0 \to K^+2\pi^-\pi^+$ is used to train the charge isolation MVA explained in Sec. 4.5.3.

In order to validate the signal selection for 2018, the signal and the control channel $B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0(\to \pi^+\pi^-)$ are generated with the data-taking conditions for that year. As a part of validating the signal selection, the control channel is fitted in data where a simulation generated with 2018 data-taking conditions of the $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)\mu^+\nu_\mu$ background is used to constrain the fit, as explained in Sec. 4.6.2.

4.2 Trigger selection

The trigger system explained in Sec. 3.5.1 reduces the event rate from $40\,\mathrm{MHz}$ to a few kHz. Each trigger stage, L0, HLT1 and HLT2, consists of multiple trigger decision lines. If a trigger line is fired due to a particle originating from the signal, the event is referred to as *triggered on signal* (TOS), while if it is fired due to a particle that does not originate from the signal, it is referred to as *triggered independently from signal* (TIS). In this analysis, signal candidates are required to pass one trigger line per trigger stage, and to be TOS:

- `muplus_L0MuonDecision_TOS`: selects events with one high-p_{T} muon with $p_{\mathrm{T}}(\mu) >$ $1.35\,\mathrm{GeV}/c$ and with less than 450 hits in the SPD detector, to reduce the complexity of the event [122].

- `Bplus_Hlt1TrackMVADecision_TOS`: selects events with a high-p_{T} track well separated from all PVs and with a good fit quality, *e.g.* originating from a B meson decay [147].

- `Bplus_Hlt2TopoMu3BodyDecision_TOS` : selects events with a displaced topology of three tracks, *e.g.* originating from a B meson decay. First, a search for three displaced tracks is performed, and subsequently, an MVA method trained on seven variables, *i.e.* the B candidate $\sum |p_{\mathrm{T}}|$, $p_{\mathrm{T}}^{\mathrm{min}}$, mass, $m_{\mathrm{corr}}(B)$, DOCA[1], χ^2_{IP} and flight distance, is used to perform the event selection [147].

The motivation for choosing these trigger lines is to retain a high signal efficiency, reduce the amount of background and to keep the trigger selection simple in order to reduce the systematic uncertainties.

4.3 Stripping selection

The stripping stage consists of a set of simple selection cuts compiled into stripping lines. In this analysis, different stripping lines are used to select a signal candidate, as well as a background candidate used to understand and describe the background present in data. All stripping lines used in this analysis are summarised in Table 4.2 along with their pre-scale information. The pre-scale is used to reduce the amount of recorded data as it determines how often a line is fired. If the pre-scale is one it is always fired, while if it is zero, it is never fired. The exact definitions of the stripping lines can be found in Ref. [148]. In general, the same requirements are applied in the background stripping lines as in the signal stripping line, however, there are few important differences as summarised below:

- The `B2XuMuNuBu2Rho_Line` is the signal stripping line that selects $B^+ \to \rho^0 \mu^+ \nu_\mu$ candidates. It requires three long tracks to form a vertex displaced from the primary

---

[1]DOCA is the distance of closest approach between two tracks.

Table 4.2 – Stripping lines for the $B^+ \to \rho^0 \mu^+ \nu_\mu$ analysis.

Lines	Objective	Pre-scale
B2XuMuNuBu2Rho_Line	Signal	1.0
B2XuMuNuBu2Rho_WSLine	Background	0.2
B2XuMuNuBu2Rho_NoPIDMuLine	μ misID	0.02
StrippingB2XuMuNuB2Pi_Line	π misID	0.2

vertex. One track must be consistent with a muon, while the remaining tracks must be consistent with two oppositely-charged pions. Note that $m(\pi^+\pi^-)$ is not restricted to allow for the normalisation channel to be selected as well.

- The `B2XuMuNuBu2Rho_WSLine` selects wrong-sign-mixed-sign (WS-MS) $(\mu^+\pi^-\pi^-)$ candidates and wrong-sign-same-sign (WS-SS) $(\mu^+\pi^+\pi^+)$ candidates. This line is, in particular, used for studying the combinatorial background.

- The `B2XuMuNuBu2Rho_NoPIDMuLine` selects signal candidates where no PID cuts have been applied to the μ^+ candidate. It is used to estimate the contribution of misidentified muons (μ misID) in data.

- A similar stripping line was developed for estimating the contribution of misidentified pions (π misID), however, an error in the line made it inadequate for this purpose. Instead, the `StrippingB2XuMuNuB2Pi_Line` is used to estimate the π misID contribution. This line selects $B^0 \to \pi^- \mu^+ \nu_\mu$ candidates. A π^+ without any PID requirements is added to form a $B^+ \to \rho^0 \mu^+ \nu_\mu$ candidate. It is ensured that the general selection is the same as in the other `B2XuMuNuBu2Rho` stripping lines.

The requirements of the `B2XuMuNuBu2Rho_Line` summarised in Table 4.3 are explained in greater detail below:

- Requirements on minimum track p_{T} and p are imposed to reduce the contribution of combinatorial background, particles produced in the PV (prompt particles) and misID backgrounds. In general, low momenta particles are more likely to be misidentified. For instance, if the momentum of a particle is below the threshold of the RICH detectors, explained in Sec. 3.4.1, Cherenkov light cannot be emitted, and if the momentum of a muon is too low it might not reach the muon system, described in Sec. 3.4.3, or only partially traverse it. In both cases, less information is available for the particle identification.

- Tracks that do not correspond to the flight path of a particle, also known as ghosts or fake tracks, are rejected by applying a cut on their probability to be a ghost. In addition to this, a good fit quality of the tracks, *i.e.* a small χ^2/ndof, is required. Likewise, a good fit quality of the decay vertices associated with the B^+ and ρ^0 mesons are required, *i.e.* a small χ^2_{vtx}/ndof.

- To reduce misID backgrounds, the muon and pion track candidates have to satisfy certain PID requirements. To select muons, requirements on the `PIDmu` and `IsMuon` variables are imposed. By requiring the `IsMuon` variable to be true, the track is required to have a least one hit in a certain number of muon stations (depending on its momentum) in an area around the extrapolated track [149]. To reject kaons and muons misidentified as pions cuts are applied on the pion `PIDmu` and `PIDK` variables.

- The DIRA is the cosine of the angle between the reconstructed momentum of the mother particle, *i.e.* the sum of the reconstructed daughter momenta, and its direction of flight given by the PV and SV. As the ρ^0 and B^+ mesons are highly boosted they are characterised by a DIRA close to one.

- The variable $\chi^2_{\rm IP}$ of a particle is the difference between the χ^2 of the PV fit with and without the particle. If the particle originates from the PV this difference should be small. Therefore the $\chi^2_{\rm IP}$ for the B^+ candidate is required to be small (although missing a neutrino, its $\chi^2_{\rm IP}$ is small, because of its high boost), while for the B^+ daughters it should be large.

- The invariant mass of the two pions, $m(\pi^+\pi^-)$, the visible B^+ mass, $m_{\rm vis}(B^+)$, and the corrected B^+ mass, $m_{\rm corr}(B^+)$, are required to be in mass regions consistent with the signal.

- The B^+ χ^2 PV separation, also known as the χ^2 distance from the PV, measures the compatibility of the PV and the B^+ decay vertex. To ensure that the B^+ candidate can be well separated from the PV, a high χ^2-distance is required.

4.4 Preselection

After trigger and stripping selection, data is still heavily dominated by background. This is clearly visible in Fig. 4.1 where the $m_{\rm corr}(B^+)$ and $m(\pi^+\pi^-)$ distributions are illustrated for signal MC and a subsample of 2018 data.

In data there is no peak in the $m_{\rm corr}(B^+)$ distribution at the known B mass of about $5279\,{\rm MeV}/c^2$ associated with the signal, however, there is a large hill of background peaking between 3500–4000 MeV/c^2. In the $m(\pi^+\pi^-)$ distribution in data there is a peak at the known ρ^0 mass of about $775\,{\rm MeV}/c^2$, however, this also includes ρ^0-meson decays from background processes. The peak at $1865\,{\rm MeV}/c^2$ corresponds to $D^0 \to \pi^+\pi^-$ and includes the decay $B^+ \to \bar{D}^0\mu^+\nu_\mu$ used as a normalisation channel for this analysis. The prominent bump at $m(\pi^+\pi^-) \sim 1350\,{\rm MeV}/c^2$ is partially reconstructed $D^0 \to K^0_{\rm S}\pi^+\pi^-$.

In order to reduce the large amount of background, while retaining a high signal efficiency, *preselection cuts* are applied on top of the trigger and stripping selection. To develop powerful preselection cuts, it is necessary to understand the dominant backgrounds in data and obtain a sample that can be used to describe them.

Table 4.3 – Requirements of the signal stripping line `B2XuMuNuBu2Rho_Line`.

Candidate	Variable	Requirement	Unit
μ^+	Track p_{T}	> 1.5	GeV/c
	Track p	> 6	GeV/c
	Track χ^2/ndof	< 4	
	Track χ^2_{IP}	> 25	
	Track ghost probability	< 0.35	
	`PIDmu`	> 3	
	`IsMuon`	True	
$\pi^\pm$	Track p_{T}	> 400	MeV/c
	Track p	> 3	GeV/c
	Track χ^2/ndof	< 4	
	Track χ^2_{IP}	> 36	
	Track ghost probability	< 0.5	
	`PIDmu`	< 2	
	`PIDK`	< -2	
ρ^0	Two-track p_{T}	> 1000	MeV/c
	Track p_{T} of one π	> 900	MeV/c
	Track p of one π	> 5	GeV/c
	$\rho^0\ \chi^2_{\mathrm{IP}}$	> 50	
	$\rho^0\ \chi^2_{\mathrm{vtx}}/\mathrm{ndof}$	< 4	
	ρ^0 DIRA	> 0.98	
	$m(\pi^+\pi^-) - m_{\rho^0}$	< 1500	MeV/c^2
B^+	$B^+\ \chi^2_{\mathrm{vtx}}$	< 18	
	$B^+\ \chi^2$ PV separation	> 120	
	B^+ DIRA	> 0.999	
	m_{corr}	$\in [2.5, 7.0]$	GeV/c^2
	m_{vis}	$\in [2.0, 5.5]$	GeV/c^2

4.4.1 Dominant backgrounds

Based on the distributions of $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+\pi^-)$ in data, the dominant background is expected to consist of partially reconstructed decay processes. The large background hill at low $m_{\mathrm{corr}}(B^+)$ is characteristic for semileptonic B decays with more than one missing particle in their final states. The reason for this is that $p_\perp$ and $m_{\mathrm{vis}}(B^+)$ in Eq. 2.19 are underestimated resulting in a lower $m_{\mathrm{corr}}(B^+)$. Moreover, semileptonic $b \to c$ decays are Cabibbo favoured resulting in an inclusive rate that is about 50 times higher than the one of semileptonic $b \to u$ decays [150]. Thus, the dominant background is expected to be semileptonic $b \to c$ decays with more than one missing particle in their final states.

Four different samples are used as background candidates. These are either simulations from Table 4.1 or obtained from data with the background stripping lines listed in Table 4.2:

- Simulated sample of $B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ decays with $\bar{D}^0 \to \pi^+\pi^- X$ (V_{cb}-D^0).
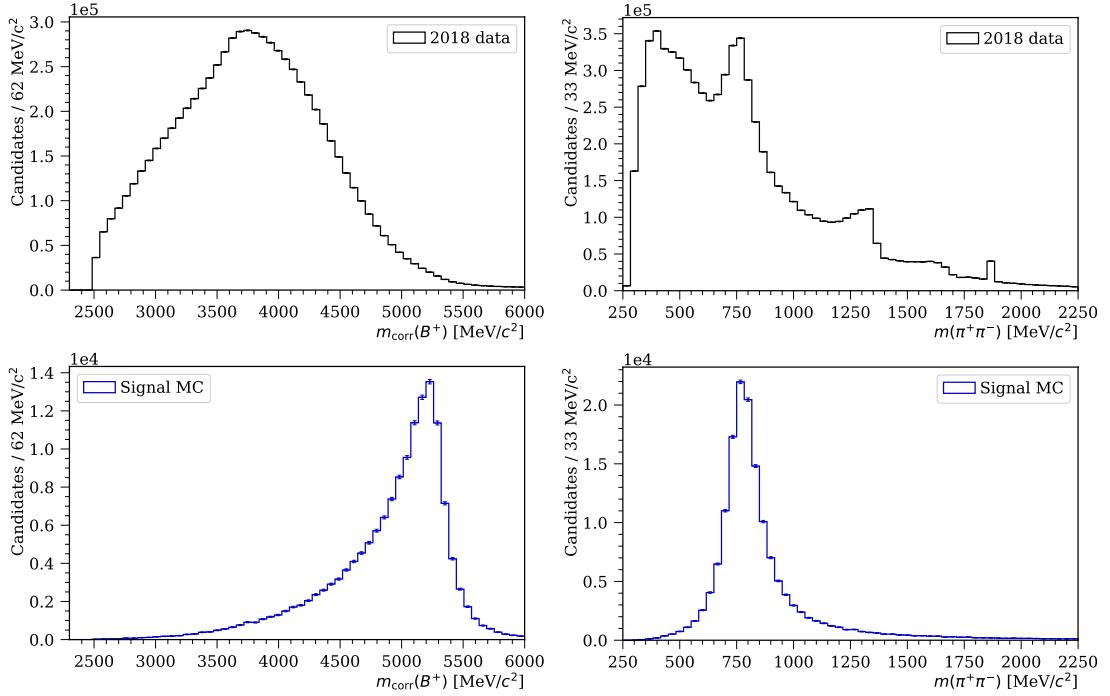
Figure 4.1 – Distributions of $m_{corr}(B)$ and $m(\pi^+\pi^-)$ in a subsample of 2018 data (top row) and signal MC (bottom row) after stripping and trigger selections.

- Simulated sample of $B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ decays with $\bar{D}^- \to \pi^+\pi^- X$ (V_{cb}-D^-).

- Data sample of $\mu^+\pi^-\pi^-$ candidates (WS-MS). It is expected to contain combinatorial background as well as semileptonic $b \to c$ decays with at least one additional charged particle in its final state.

- Data sample of $\mu^+\pi^+\pi^+$ candidates (WS-SS). It is expected to contain combinatorial background and semileptonic $b \to c$ decays with at least two additional charged particles in its final state.

The distributions of $m_{corr}(B^+)$ and $m(\pi^+\pi^-)$ are compared between the four types of background candidates and data, as illustrated in Fig. 4.2. As data is dominated by backgrounds, a good background sample is expected to have distributions similar to the ones in data. The WS-MS sample follows data well in $m_{corr}(B^+)$, but it does not describe the ρ^0 resonance and structures around $1300 \, \mathrm{MeV}/c^2$ and $1700 \, \mathrm{MeV}/c^2$ in the $m(\pi^+\pi^-)$ spectrum, this however, is well described by the V_{cb}-D^0 sample, which is therefore used to optimise the preselection.

4.4.2 Diagonal cuts

Although a 1D cut in $m_{corr}(B^+)$ could remove a large amount of background while retaining a high signal yield, it would make it difficult, if not impossible, to perform a fit in $m_{corr}(B^+)$,
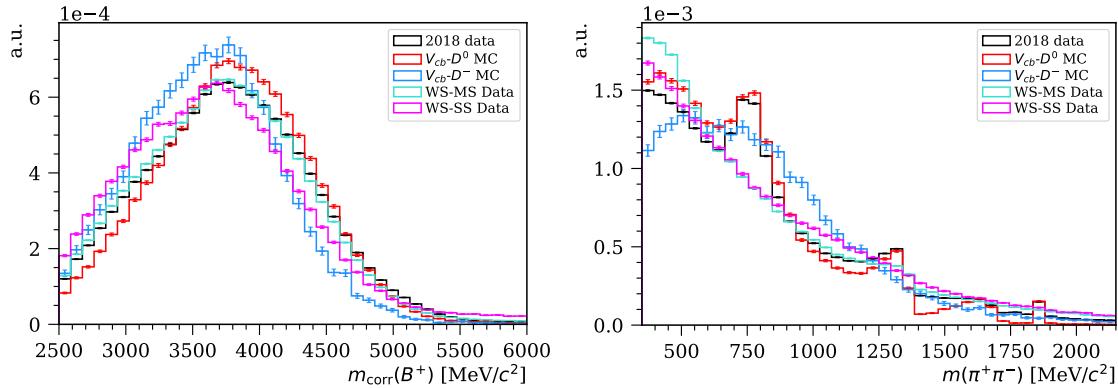
Figure 4.2 – $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right) distributions of data and background candidate samples.

because the shapes of backgrounds and signal would become hard to distinguish. Thus, the low $m_{\text{corr}}(B^+)$ region is needed to perform the fit of signal and backgrounds in data. However, the ratio of background with respect to signal in this region must be decreased to improve the precision on the extracted signal yield. To achieve this, two *diagonal cuts* are used. These are cuts in one variable X as a linear function of the variable $m_{\text{corr}}(B^+)$. The two diagonal cuts are chosen as an easy way to reject a reasonable fraction of the background without cutting into $m_{\text{corr}}(B^+)$, while the detailed signal versus background discrimination will be done by the MVA selection described in Sec. 4.5. In the next two paragraphs, the diagonal cuts, summarised in Table 4.4, are explained in greater detail.

First diagonal cut

The first diagonal cut is taken from Ref. [151]. It is a cut on the transverse momentum of the ρ^0 meson with respect to the B^+ flight direction, $p_\perp(\rho^0)$, as a function of $m_{\text{corr}}(B^+)$. The 2D distributions for the MC signal and V_{cb}-D^0 background are illustrated in Fig. 4.3. The signal is characterised by high $m_{\text{corr}}(B^+)$ over the full $p_\perp(\rho^0)$ range, while the background is characterised by a broader distribution peaking at lower $m_{\text{corr}}(B^+)$ and $p_\perp(\rho^0)$. A diagonal cut that removes 47% of the background and 11% of the signal is chosen.

Second diagonal cut

The second diagonal cut is on the μ^+ momentum in the approximate B^+ rest frame, $p_{B\text{rest}}(\mu^+)$, as a function of $m_{\text{corr}}(B^+)$. To approximate the B^+ momentum, an approach introduced in Ref. [152] is used. The proper velocity $\gamma\beta$ of the visible decay products (Y) along the z axis is assumed to be equal to the one of the B^+ meson along the same axis, resulting in the simple relation between the B^+ meson momentum and the visible
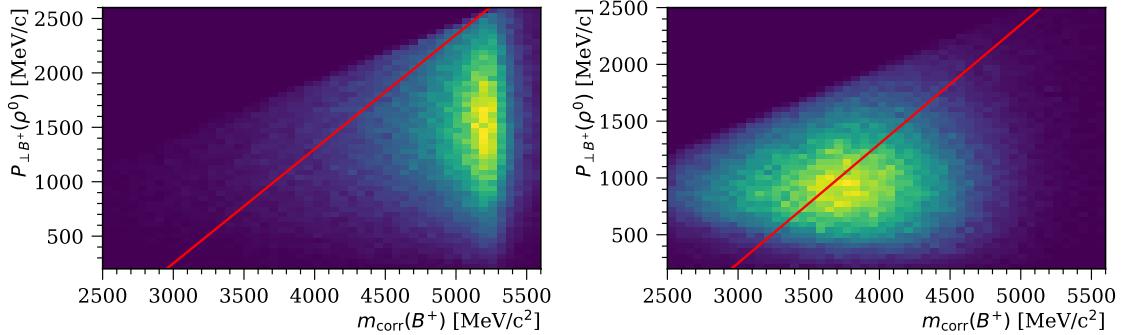
Figure 4.3 – First diagonal cut (red line) on signal MC (left) and V_{cb}-D^0 background (right).

momentum along the z coordinate:

$$(p_{B^+})_z = \frac{m_B}{m_Y}(p_Y)_z \tag{4.1}$$

Using the direction defined by the PV and the B^+ decay vertex with the angle θ with respect to the z axis, the following expression for the B^+ momentum is obtained:

$$|p_{B^+}| = \frac{m_B}{m_Y}(p_Y)_z\sqrt{1+\tan^2\theta} \tag{4.2}$$

This is then used to boost the μ^+ back into the approximate B rest frame where $p_{B\mathrm{rest}}(\mu^+)$ can be obtained.

The 2D cut in $p_{B\mathrm{rest}}(\mu^+)$ and $m_{\mathrm{corr}}(B^+)$ exploits the different spin structure of semileptonic decays into pseudoscalar mesons (like the D^0 meson) and vector mesons (like the ρ^0). As presented in Ref. [63], the V–A structure of the charged weak current leads to high μ^+ momenta in the B^+ rest frame for $B^+ \to \rho^0\mu^+\nu_\mu$, while semileptonic decays into pseudoscalar mesons peak at lower momenta. This is illustrated in Fig. 4.4, where signal MC is characterised by a distribution that peaks at high $p_{B\mathrm{rest}}(\mu^+)$ and $m_{\mathrm{corr}}(B^+)$, while the V_{cb}-D^0 background is characterised by a broader distribution at low $p_{B\mathrm{rest}}(\mu^+)$ and $m_{\mathrm{corr}}(B^+)$. Applying both diagonal cuts reduces the background with 76% and the signal by 20%.

4.4.3 Cut on the corrected B mass error

The corrected B^+ mass error, which depends on the errors of the PV, SV and final state momenta, is required to satisfy

$$\sigma_{m_{\mathrm{corr}}}(B^+) < 100\,\mathrm{MeV}/c^2\,. \tag{4.3}$$

The objective of this requirement is to improve the separation of the signal and background shapes in the $m_{\mathrm{corr}}(B^+)$ distribution. In particular, to improve the fit sensitivity in the
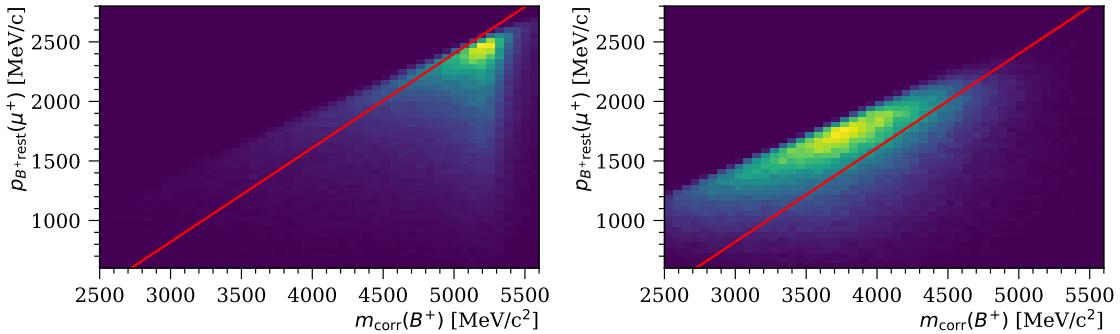
Figure 4.4 – Second diagonal cut (red line) on signal MC (left) and V_{cb}-D^0 background (right).
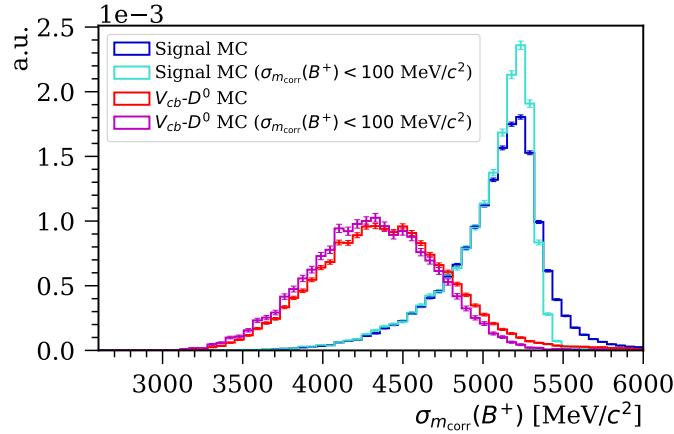
Figure 4.5 – Distributions of $m_{\mathrm{corr}}(B^+)$ with/without $\sigma_{m_{\mathrm{corr}}}(B^+) < 100\,\mathrm{MeV}/c^2$ for signal MC (cyan/blue) and V_{cb}-D^0 background (magenta/red).

region around the nominal B^+ mass. This is illustrated in Fig. 4.5, where the distributions of $m_{\mathrm{corr}}(B^+)$ with and without the requirement $\sigma_{m_{\mathrm{corr}}}(B^+) < 100\,\mathrm{MeV}/c^2$ are shown for signal MC and V_{cb}-D^0 background. As a result of the cut, the signal becomes more peaky, and the upper tail of the signal $m_{\mathrm{corr}}(B^+)$ distribution, which is purely due to resolution effects, is significantly reduced. At the same time, the background distribution is pushed towards lower $m_{\mathrm{corr}}(B^+)$ causing a larger separation between the signal and background.

4.4.4   Reducing backgrounds with misidentified daughters

A set of preselection cuts are used to reduce backgrounds with misidentified final state particles, also known as misID backgrounds.
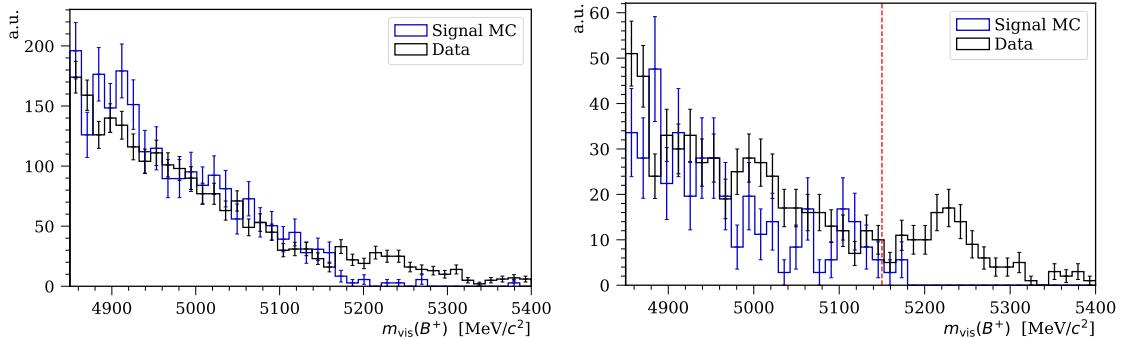
Figure 4.6 – $m_{\text{vis}}(B^+)$ distribution of data (black) and signal MC (blue) in the region $570\,\text{MeV}/c^2 < m(\pi^+\pi^-) < 1100\,\text{MeV}/c^2$ before (left) and after (right) the requirement $\texttt{PIDK}(\mu^+) > -3$. The preselection requirement $m_{\text{vis}}(B^+) < 5150\,\text{MeV}/c^2$ is illustrated by the dashed red line (right).

Cut on the visible B^+ mass

To remove the background consisting of $B^+ \to K^+_{\text{misID}\mu}\pi^+\pi^-$ decays where $K^+_{\text{misID}\mu}$ is a true K^+ misidentified as a μ^+, a cut is imposed on the visible B mass:

$$m_{\text{vis}}(B^+) < 5150\,\text{MeV}/c^2\,. \tag{4.4}$$

As the final state of $B^+ \to K^+_{\text{misID}\mu}\pi^+\pi^-$ is fully reconstructed, a peak is expected to form in the $m_{\text{vis}}(B^+)$ distribution. However, since the $K^+_{\text{misID}\mu}$ is assigned a mass that is roughly five times smaller than its nominal mass, the peak will be shifted below its true B^+ mass. This is illustrated in Fig. 4.6 where the distribution of $m_{\text{vis}}(B^+)$ is plotted for signal MC and data. By requiring the μ^+ to be kaon-like, *i.e.* $\texttt{PIDK}(\mu^+) > -3$, within the ρ^0 mass window of $570\,\text{MeV}/c^2 < m(\pi^+\pi^-) < 1100\,\text{MeV}/c^2$, it becomes possible to identify the $B^+ \to K^+_{\text{misID}\mu}\pi^+\pi^-$ background as a peak in data (not in signal MC) just below the nominal B^+ mass.

Cut on `PIDmu` and `IsMuon` of π^-

To reduce the charmonium backgrounds $B \to J/\psi(\to \mu^+\mu^-_{\text{misID}\pi})\pi^+ X$ where $\mu^-_{\text{misID}\pi}$ is a true μ^- misidentified as π^-, cuts are imposed on `PIDmu` (π^-) and `IsMuon` (π^-):

$$\texttt{PIDmu}(\pi^-) < 0\,, \tag{4.5}$$
$$\texttt{IsMuon}(\pi^-) = \text{False}\,.$$

By assigning the muon mass to the $\mu^-_{\text{misID}\pi}$, and computing the invariant mass of μ^+ and $\mu^-_{\text{misID}\pi}$, the charmonium misID background can be identified in data as a peak at the nominal J/ψ mass, as illustrated in Fig. 4.7. After applying the aforementioned cuts, the peak is significantly reduced. The two cuts remove about 8% of the signal.
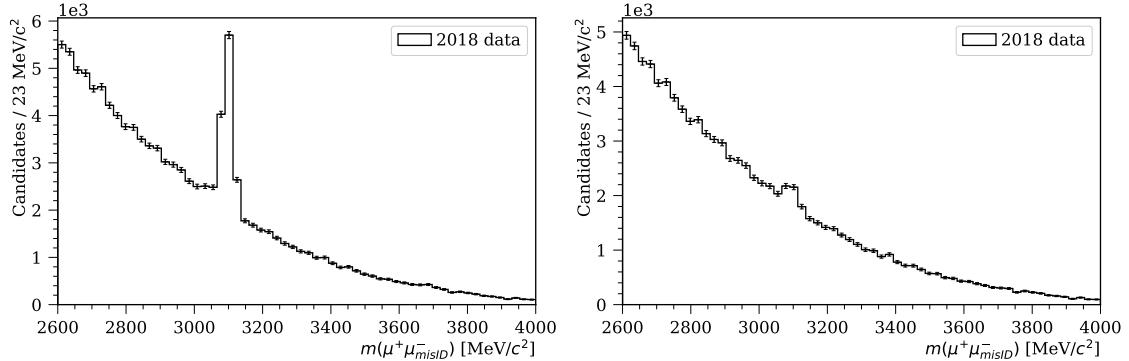
Figure 4.7 – Dimuon invariant mass distribution of $B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ candidates in data where the muon mass is reassigned to the π^- candidate before (left) and after (right) the requirements $\texttt{PIDmu}(\pi^-) < 0$ and $\texttt{IsMuon}(\pi^-) = \text{False}$.

Selection cut on `ProbNNp` of $\pi^\pm$

To reduce backgrounds where a true proton or antiproton is misidentified as a $\pi^\pm$, *e.g.* in the decay $\Lambda_b^0 \to \Lambda_c^+ \mu^- \bar{\nu}_\mu$ with $\Lambda_c^+ \to p K_S \pi^+ \pi^-$ or $\Lambda_c^+ \to p K^- \pi^+$, the following cut is imposed:

$$\texttt{ProbNNp}(\pi^\pm) < 0.7 \qquad\qquad\qquad (4.6)$$

This background is identified as a peak at $\texttt{ProbNNp}(\pi^-) \sim 1$ in data, as illustrated in Fig. 4.8. To confirm that this is the expected distribution for protons misidentified as pions, the PIDGEN tool, which is part of the PIDCALIB package [153], is applied. It computes 4D distributions in the PID variable, track p_T, track pseudorapidity η and the number of tracks in the event `nTracks` of the calibration samples, *i.e.* pure data samples with one single known particle species obtained with the *sPlot* technique described in Sec. 4.6.2. This is used to generate the PID response consistent with data given the track p_T, η and `nTracks` in simulation. The generation is performed with an unbinned approach, where the 4D distribution is computed with a kernel density estimation procedure based on the MEERKAT library [154]. With this tool the PID response of a proton with the same kinematics as the pion candidate is generated, as illustrated in Fig. 4.8. The $\texttt{ProbNNp}(\pi^-)$ in data and the re-sampled $\texttt{ProbNNp}(\pi^-)$ in signal MC based on proton calibration samples both peak at $\texttt{ProbNNp}(\pi^-) \sim 1$, while the re-sampled $\texttt{ProbNNp}(\pi^-)$ in signal MC based on pion calibration samples has a flat distribution. This confirms the presence of the proton misID background in data.

This cut removes about 80% of protons with the same kinematics as the signal pions. On the other hand, less than 1% of the signal is removed by applying the same cut on the generated $\texttt{ProbNNp}(\pi^\pm)$ distribution based on pion calibration samples. In conclusion, this cut removes the majority of protons misidentified as pions, while maintaining the same high signal efficiency.
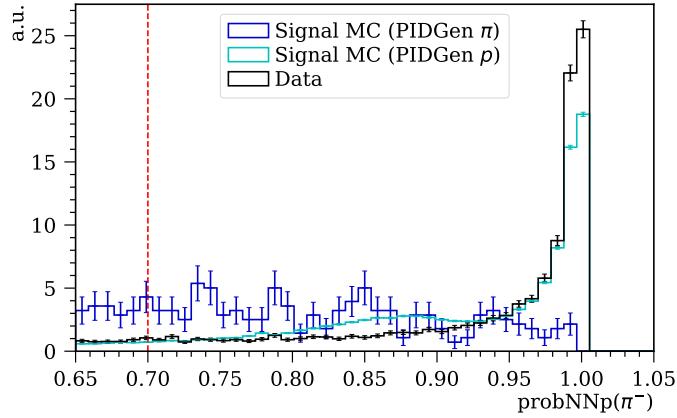
Figure 4.8 – Distributions of `ProbNNp`(π^-) in data (black), the generated `ProbNNp`(π^-) for signal MC based on proton calibration samples (cyan) and based on pion calibration samples (blue). The peak at `ProbNNp`$(\pi^-) \sim 1$ corresponds to proton misID background, which is removed by imposing `ProbNNp`$(\pi^\pm) < 0.7$ (red line).

Table 4.4 – Preselection cuts and accumulated cut efficiencies for the signal, ϵ_{sig}, and V_{cb}-D^0 background, ϵ_{bkg}, estimated from simulations.

Type	Preselection cut	ϵ_{sig}	ϵ_{bkg}
2D cut	$1.05 \times m_{\mathrm{corr}}(B^+)c^2 - 2900\,\mathrm{MeV} > p_\perp(\rho^0)c$	89%	53%
2D cut	$p_{B\mathrm{rest}}(\mu^+)c < 0.79 \times m_{\mathrm{corr}}(B^+)c^2 - 1550\,\mathrm{MeV}$	80%	24%
1D	$\sigma_{m_{\mathrm{corr}}}(B^+) < 100\,\mathrm{MeV}/c^2$	27%	9%
1D misID Bkg	$m_{\mathrm{vis}}(B^+) < 5150\,\mathrm{MeV}/c^2$	27%	9%
1D misID Bkg	`PIDmu`$(\pi^-) < 0$ and `IsMuon`$(\pi^-) =$ False	24%	8%
1D misID Bkg	`ProbNNp`$(\pi^\pm) < 0.7$	24%	8%

4.4.5 Preselection summary

The full preselection is summarised in Table 4.4. The cut efficiency for signal ϵ_{sig} is based on the 2018 signal MC sample, while the background efficiency ϵ_{bkg} is based on the 2018 V_{cb}-D^0 background sample. Note that the background composition of the fitted data is not represented by the V_{cb}-D^0 background sample, as the selection is optimised on this background component. The cuts are listed in the order of their application, and the quoted efficiency is the combined efficiency of the cut plus all the previous preselection cuts with respect to the samples after trigger and stripping selection. The total preselection efficiency of signal is estimated to be $\epsilon_{\mathrm{sig}} = 24\%$.

The effect of the preselection on the shapes of the $m_{\mathrm{corr}}(B)$ and $m(\pi^+\pi^-)$ distributions for data and the four types of background candidates are illustrated in Fig. 4.9. The backgrounds in data after the preselection are well described by the V_{cb}-D^0 background sample.
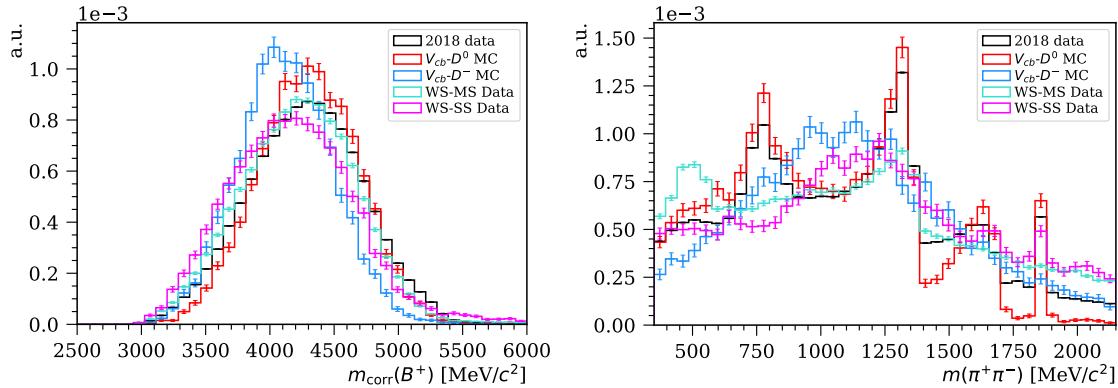
Figure 4.9 – Distributions of $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right) for candidates in data (black) and in four different background samples (other colours) after the preselection.

4.5 Multivariate selection

After the preselection, two multivariate classifiers (MVAs) based on supervised machine learning algorithms are used to further reduce the amount of background with respect to signal in data. To do this, the ROOT Toolkit for Multivariate Data Analysis (TMVA) [155, 156] is used.

4.5.1 MVA selection strategy

As discussed in Sec. 4.4.2, the dominant background is expected to consist of partially reconstructed semileptonic $b \to c$ decays. This background differs from the signal by having at least one additional particle in the final state, and it is therefore characterised by being *non-isolated*, *i.e.* having at least one other particle in the event compatible with originating from the decay vertex of the signal candidate. On the other hand, a true signal candidate is characterised by being *isolated*, *i.e.* having no other particles in the event compatible with originating from its decay vertex. This difference in *isolation* of signal and background can be exploited in the MVA selection.

The first MVA, referred to as the *charge isolation MVA*, is trained to distinguish isolated and non-isolated signal candidates in terms of charged particles in the event. The second MVA, known as the *main MVA*, uses this information together with kinematic and geometric properties of the candidates to distinguish signal from background.

The training, testing and evaluation of each MVA follows the same principle. First, it is trained on samples containing true signal candidates (signal training sample) and true background candidates (background training sample). Based on differences in physical properties (input variables), the MVA learns how to distinguish signal and background candidates. The MVA output is a number between zero and one corresponding to the probability of the candidate to be signal.
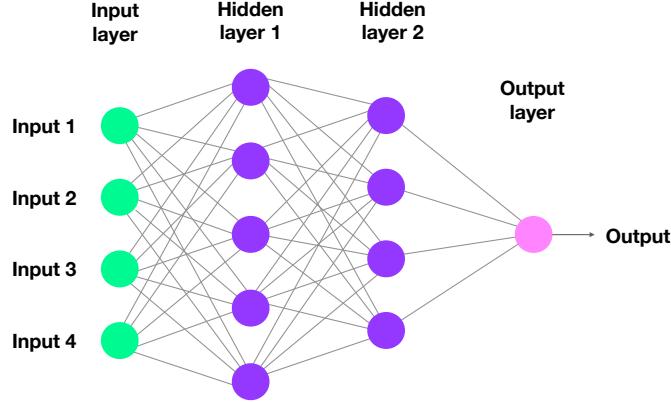
Figure 4.10 – Schematic illustration of an artificial neural network consisting of an input layer with four neurons (green), two hidden layers with five and four neurons, respectively (purple), and an output layer with a single neuron (red).

Next, the MVA performance is evaluated using new samples of true signal candidates (signal test sample) and true background candidates (background test sample). Based on this test, a receiver operating characteristic (ROC) curve [157], which is a graph of the background rejection $(1 - \epsilon_{\mathrm{background}})$ versus the signal efficiency $(\epsilon_{\mathrm{signal}})$, is computed. The ROC curve shows the MVA performance of all classification thresholds (all cuts on the MVA output value). The area under the ROC curve (AUC) is used as a figure of merit. It is an aggregate measure of the MVA performance over all possible classification thresholds, and it can be interpreted as the probability of the MVA to rate any true signal candidate higher than any true background candidate. If the MVA prediction is 100% correct the AUC is one, while if its prediction is 100% wrong it is zero. Finally, the MVA is applied to data and an optimal cut on the MVA output value is determined.

4.5.2 Artificial neural networks

Both MVAs used for this analysis are artificial Neural Networks (NN) [158]. The structure of NNs is inspired by the human brain where interconnected neurons are responsible for processing information [159]. Usually, the artificial neurons are organised into sequential layers, where the neurons in one layer receive information from the previous layer, as illustrated in Fig. 4.10.

In the first layer, known as the *input layer*, the number of neurons has to equal the number of input variables. The last layer is the *output layer* with one or more neurons depending on the classification problem. There can be multiple intermediate layers, known as *hidden layers*, and in that case the network is referred to as a *deep neural network* (DNN). The input for each neuron, except those in the input layer, is the output of every neuron in the previous layer, as illustrated in Fig. 4.11. Each input x_i is associated with a *weight w_i*.
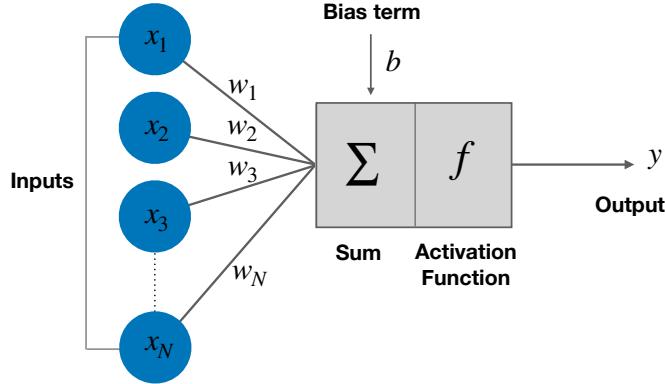
Figure 4.11 – Single artificial neuron.

The weighted sum of inputs is computed and a *bias term b* is added to it. This is then passed on to the *activation function f* that produces an *output response value y* [158]:

$$f(\vec{w} \times \vec{x} + b) = f(z) = y\,. \tag{4.7}$$

The choice of activation function plays a crucial role for the network to be able to model complex and non-linear relations in data. Among the popular activation functions used in hidden layers are [158]:

- Sigmoid, *i.e.* $\sigma(x) = 1/(1 + e^{-x})$ that outputs a number between 0 and 1 for any value x.

- Hyperbolic Tangent (Tanh), *i.e.* $\tanh(x) = (e^x - e^{-x})/(e^x - e^{-x})$ that outputs a number between -1 and 1 for any value x.

- Rectified linear unit (ReLU), *i.e.* $\text{ReLU}(x) = \max(0, x)$ that outputs x if $x > 0$ and zero if $x < 0$.

Historically, the Sigmoid and Tanh functions were first introduced [160]. However, the main problem related to these is that their derivatives tend to zero as $x \to \pm\infty$, which slows down the gradient-based learning (described later in this section). This problem is solved by the ReLU activation function, where the derivative is one for positive inputs and otherwise zero. In addition to this, it is faster to compute a ReLU compared to a Sigmoid and a Tanh. Today, the ReLU function is therefore the most common choice of activation function in NNs.

For the output layer, the popular activation functions are [158]:

- Linear, *i.e.* it does not change the weighted sum of the input, but returns directly the value.

- Sigmoid, *i.e.* $\sigma(x) = 1/(1 + e^{-x})$ that outputs a number between 0 and 1 for any value x.

- Softmax, *i.e.* $\text{sm}(x) = e^x / \sum e^x$ (x is a vector of outputs) that outputs a vector of values that sum to 1 representing the probabilities of belonging to each output class.

The training of a NN consists of finding the model parameters, *i.e.* weights and biases, resulting in the most accurate predictions. This is done by first defining a loss function that computes the error of the model. The most common loss function is the Mean Squared Error (MSE) and Cross-Entropy (CE) [158].

The objective of the training is to minimise the loss function. This is often done with an optimisation algorithm called Stochastic Gradient Descent (SGD) together with Back-Propagation (BP) [158]. Essentially, the BP algorithm computes the gradient of the loss function with respect to all model parameters. This is used by the SGD algorithm to perform a step in the direction of the steepest decent, thus updating the parameters of the model. The step size, also known as the learning rate, is used to control how much the parameters are changed with respect to the gradient.

The optimisation algorithm works through the entire training sample during one epoch. This is repeated until the minimum of the loss function has been reached while avoiding overtraining, *i.e.* the model stops generalising and starts learning the statistical fluctuations in the training sample. Overtraining can be reduced in two ways: by adding more training examples, or by changing the complexity of the NN, which is often achieved with regularization methods that aim at keeping the weights small [158]. In general, this results in a more stable NN that is less sensitive to statistical fluctuations in the training sample. A common method is *weight regularization* where a penalty term proportional to the magnitude of the weights is added to the loss function. Another method is *drop out* where a certain percentage of randomly chosen neurons in a given layer is ignored between each update. Effectively, a large number of NN with different architectures are trained in parallel resulting in a more robust NN. Finally, the method of *early stopping* monitors the model performance on a test sample and stops the training when the performance starts to degrade.

In the following sections, the MVA selection used in this analysis is presented, including the MVA training, testing and evaluation on data. Finally, the potential of using neutral isolation is reviewed.

4.5.3 Charge isolation MVA

The objective of the charge isolation MVA is to evaluate the isolation of signal candidates in terms of every charged particles (tracks) in the event.
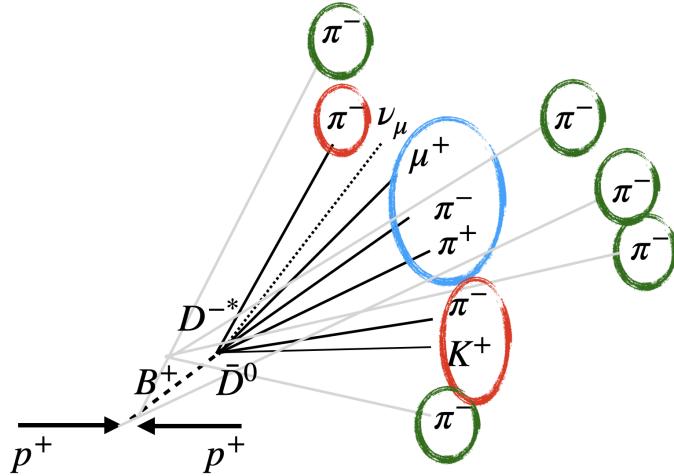
Figure 4.12 – The signal candidate (blue) originating from the background decay $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0(\rightarrow K^+ 2\pi^- \pi^+)\pi^-)\mu^+\nu_\mu$. The non-reconstructed decay products K^+ and $2\pi^-$ result in non-isolated tracks (red) and the charged particles produced in other processes of the event result in isolated tracks (green).

The simulation of the decay $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0(\rightarrow K^+ 2\pi^- \pi^+)\pi^-)\mu^+\nu_\mu$ (see Table 4.1) is used to train the charge isolation MVA. After applying the signal stripping selection, the sample contains $B^+ \rightarrow \rho^0(\rightarrow \pi^+\pi^-)\mu^+\nu_\mu$ candidates with three additional tracks coming from the decay vertices of the B^+- and ρ^0 candidates, as illustrated in Fig. 4.12. The additional tracks are said to be *non-isolated*, while all other tracks in the event are said to be *isolated*. The MVA has to learn how to distinguish non-isolated and isolated tracks. Thus, for each $B^+ \rightarrow \rho^0(\rightarrow \pi^+\pi^-)\mu^+\nu_\mu$ candidate, the non-isolated tracks, *i.e.* K^+, π^- and π^-, are used as the signal sample, while the isolated tracks in the event are used as the background sample.

A set of input variables are computed for the MVA training. The variables that are found to be most powerful in separating non-isolated and isolated tracks[2] are:

- p_T: transverse momentum of the track.

- $p_\perp$: transverse momentum of the track with respect to the B^+ flight direction.

- $\min(\chi^2_\mathrm{IP})$: minimum χ^2_IP of the track with respect to all PVs.

- $\Delta\chi^2_\mathrm{vtx}$: the χ^2_vtx difference between the original B^+ (ρ^0) vertex and the one with the additional track.

The distributions of the input variables for signal and background are illustrated in Fig. 4.13, where both long and upstream tracks (defined in Sec. 3.3.5) makes up the training samples.

---

[2]Other variables such as the track momentum and $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ with $\Delta\eta$ and $\Delta\phi$ being the difference in pseudorapidity and azimuthal angle between the track and the B^+ flight direction were also tested in the MVA training, however, they were not found to improve the performance of the MVA, and are therefore not used in the final charge isolation MVA.
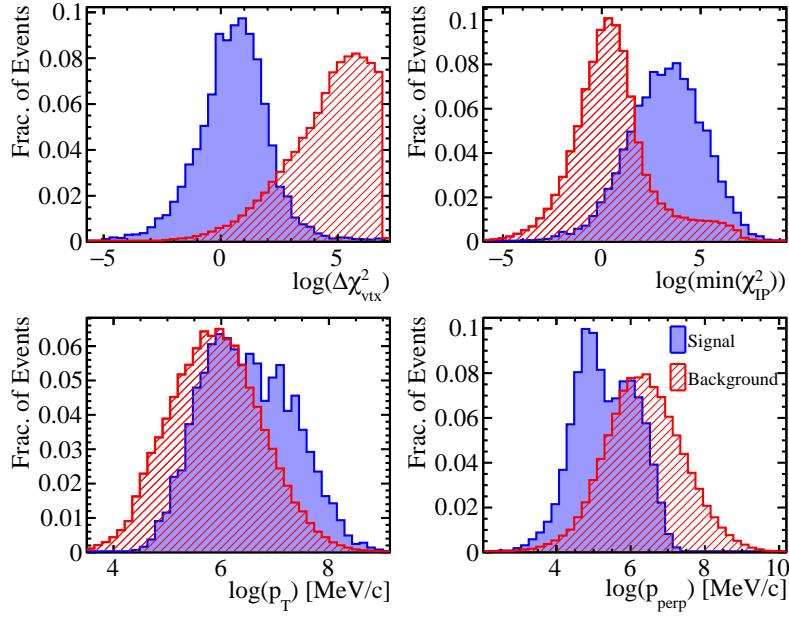
Figure 4.13 – Charge isolation MVA input variables for signal (blue) and background (red).

The logarithm is applied to the variables to make their distributions more uniform and to avoid extreme values. This was found to improve the performance of the charge isolation MVA.

The MVA method used is a NN referred to as a Multi-Layer Perceptron (MLP) [156]. The NN consist of an input layer with four neurons corresponding to the number of input variables, a hidden layer with nine neurons and an output layer with a single neuron. The activation function for each neuron in the hidden layers is Tanh (ReLU is not available with the MLP method in TMVA), while a linear activation function is used for the output neuron. The NN training uses SGD with a BP method called Broyden-Fletcher-Goldfarb-Shannon (BFGS) [156] to minimise the MSE loss function. The BFGS method differs from the standard BP method by using second derivatives of the loss function. The model parameters are updated for each training example, where the learning rate is set to 0.02.

The signal and background input samples consists of ~ 17k and ~ 187k tracks, respectively. These are long and upstream tracks with a pion hypothesis, *i.e.* they are assigned the pion mass. Each sample is split up in a training and testing sample of the same size. Since the MLP training performance can be sensitive to the relative amount of signal and background, the event weights are renormalised in such a way that the sum of all weighted signal training events are the same as the sum of all weighted background training events. The number of training epochs are set to 60. To avoid overtraining a Bayesian extension of the MLP method [156] is used. Essentially, this method reduces the NN model complexity by adding a term to the loss function that penalizes large weights. The NN architecture and model parameters are optimised to maximise the AUC value. The performance of the final NN is illustrated by the ROC curve in Fig. 4.14, where the AUC reaches 99%. The
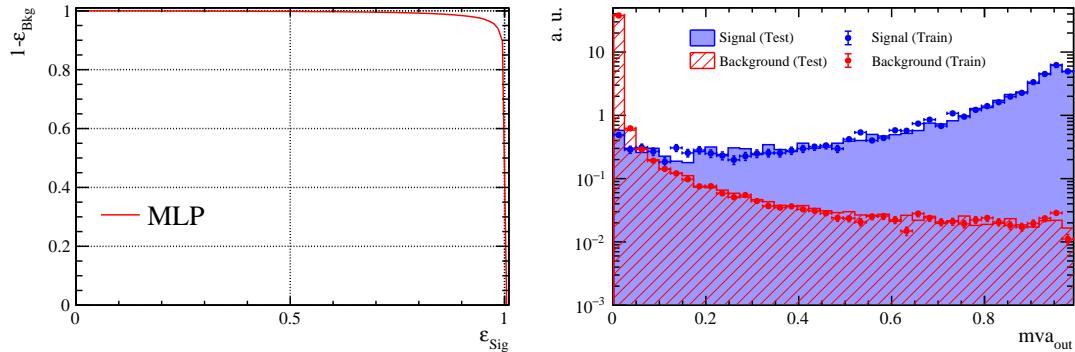
Figure 4.14 – Performance of charge isolation MVA trained on long and upstream tracks. The ROC curve (left) with an AUC = 99% and the MVA output (right) for signal (blue) and background (red) for the training (markers) and testing (filled histograms) samples.

distribution of the output value for signal and background from the training and testing sample are illustrated in log scale in Fig. 4.14. This shows a strong separation power as well as a good consistency of the NN output between the training and testing samples.

Different architectures and model parameters were tested, *i.e.* adding more neurons to the hidden layer, adding an additional hidden layer, changing the learning rate and neuron activation function as well as increasing the number of epochs. Moreover, another TMVA classifier, *i.e.* an adaptive boosted decision tree (AdaBDT) [156], was tested. In short, this MVA method combines the concept of Decision Trees [161] and Adaptive Boosting [162]. A Decision Tree consists of nodes and branches. At each node, one feature of data is evaluated, *i.e.* a requirement is imposed, which splits the path of data into two branches. Depending on the property of the training candidate, it is send to one or the other branch. The training consists of finding the best combination of features and requirements that correctly classifies signal and background in data. The Adaptive Boosting assigns a weight to each training candidate. If the candidate is wrongly classified, it is assigned a higher weight, such that the classifier becomes better at classifying difficult candidates in the next iteration. Although the AdaBDT method was found to achieve similar performance as the NN, the NN method is preferred, because its application is faster.

Charge isolation variables

The charge isolation MVA is used to compute a set of *isolation variables*. These variables quantify the isolation of signal candidates and are used by the main MVA to reject background.

For each signal candidate every long and upstream track of the event with $\Delta\chi^2_{\mathrm{vtx}} < 100$ and that results in a χ^2 of the refitted vertex that is less than 50 is evaluated by the charge isolation MVA. The most non-isolated track is added to the signal candidate, and the B^+ and ρ^0 vertices are refitted. Kinematic, geometric and particle identification variables are

computed for the four-track candidate. This procedure is repeated for the second and third most non-isolated track. The computed variables as well as the charge isolation MVA output for the three most non-isolated tracks make up the charge isolation variables. Examples of isolation variables are:

- $m(X + a^{\text{th}} \text{ track})$: invariant mass of the $X = B^+, \rho^0$ candidate plus the ath most non-isolated track of the event.

- $\text{mva}_{\text{out}}(a^{\text{th}} + \text{vtx}_X)$: the charge isolation MVA output value of the ath most non-isolated track with respect to the $X = B^+, \rho^0$ candidate decay vertex.

Charge isolation using VELO tracks

The potential of charge isolation using VELO tracks (defined in Sec. 3.3.5) is also assessed. The challenge, however, is that these tracks have no momentum information, since their trajectories are not bent by the LHCb magnet.

A charge isolation MVA based on VELO tracks is trained with the two variables $\min(\chi^2_{\text{IP}})$ and $\Delta\chi^2_{\text{vtx}}$. A reasonable performance corresponding to an AUC value of $\sim 97\%$ is achieved. However, the background rejection power of the MVA response variable is not found to be significant in the final MVA selection described in Sec. 4.5.4, and it is therefore not used in final the signal selection.

More information regarding the training and performance of the charge isolation MVA based on VELO tracks can be found in Appendix B.

Isolation against neutral particles

In addition to charge isolation, isolation against neutral particles is also relevant to reject typical background processes with large branching ratios like:

$$B^+ \to \bar{D}^0 \mu^+ \nu_\mu \text{ where } \bar{D}^0 \to K^0_S \pi^+ \pi^- \text{ or } \bar{D}^0 \to K^0_S \pi^+ \pi^- \pi^0 \text{ with } K^0_S \to \pi^0 \pi^0$$

$$B^+ \to \bar{D}^{*0} \mu^+ \nu_\mu \text{ where } \bar{D}^{*0} \to \bar{D}^0 \pi^0 \text{ or } \bar{D}^{*0} \to \bar{D}^0 \gamma$$

For instance, the branching ratios of the decays $\bar{D}^0 \to K^0_S \pi^+ \pi^-$ and $\bar{D}^0 \to K^0_S \pi^+ \pi^- \pi^0$ are 2.8% and 5.2% [15], respectively.

The MC cocktail of $B^+ \to D^{(*,**)} \mu^+ \nu_\mu(Y)$ decays with $\bar{D}^0 \to \pi^+ \pi^- X$ (see Table 4.1) is used to train and test a NN. Following the same approach as for the charge isolation, the signal and background input samples consist of isolated and non-isolated neutral particles. This is done separately for photons and neutral pions. The input variables p_{T}, p, $p_\perp$ and ΔR are used for the training, since neutral particles do not create tracks, and the variables $\min(\chi^2_{\text{IP}})$ and $\Delta\chi^2_{\text{vtx}}$ can therefore not be used.

The neutral isolation NNs based on photons and neutral pions result in reasonable performances with AUC values of $\sim 85\%$. However, the background rejection power based on information from the neutral isolation NNs is not found to be significant in the final MVA selection described in Sec. 4.5.4, and it is therefore not used in the final signal selection.

More information regarding the training and performance of the neutral isolation MVAs can be found in Appendix C.

4.5.4 Main MVA

The objective of the main MVA is to reject background based on the isolation, and on kinematic and on geometric properties of the signal candidates.

The MVA is trained on the 2018 simulation of signal and $V_{cb}\text{-}D^0$ background that, according to the comparison between data and background samples in Fig. 4.9, is found to describe the dominant background in data. After preselection, each sample contains ~ 11k candidates and is split up in a training and testing sample of equal size.

The MVA is a DNN based on the Tensorflow [163] deep learning library interfaced with TMVA via Keras [164]. The DNN consists of an input layer with twelve neurons corresponding to the number of input variables, two hidden layers with 100 and 50 neurons, respectively, and an output layer with two neurons assigning the probability of candidates to be signal and background[3], respectively. The activation function for each neuron in the hidden layers is a ReLU, while the activation function for the output layer is a Softmax. To avoid overtraining, each hidden layer is associated with a dropout of 30% and weight regularization. The optimisation uses SGD and BG with a learning rate of 0.01 to minimise the crossentropy loss function. The model parameters are updated after each batch.

Initially, all variables with potential discrimination power are used to train the DNN. Before transferring the input variables to the DNN, they are preprocessed by applying a Gaussian transformation [156]. This re-shapes and re-scales their distributions resulting in a more stable training and a better model performance. In the training, variables that can be removed without degrading the DNN performance, *i.e.* decreasing the AUC value, are excluded one by one. Simultaneously, overtraining is monitored and avoided with early stopping. The input variables resulting in the best performance of the DNN are found to be:

- $p_{\mathrm{T}}(\rho^0)$

- $\chi^2_{\mathrm{IP}}(B^+)$

- $\chi^2_{\mathrm{vtx}}(B^+)$

- $m(B^+ + 1^{\mathrm{st}}\text{ track})$

---

[3]The probability of signal, $p(s)$, and background, $p(b)$, are related by $p(s) = 1 - p(b)$.
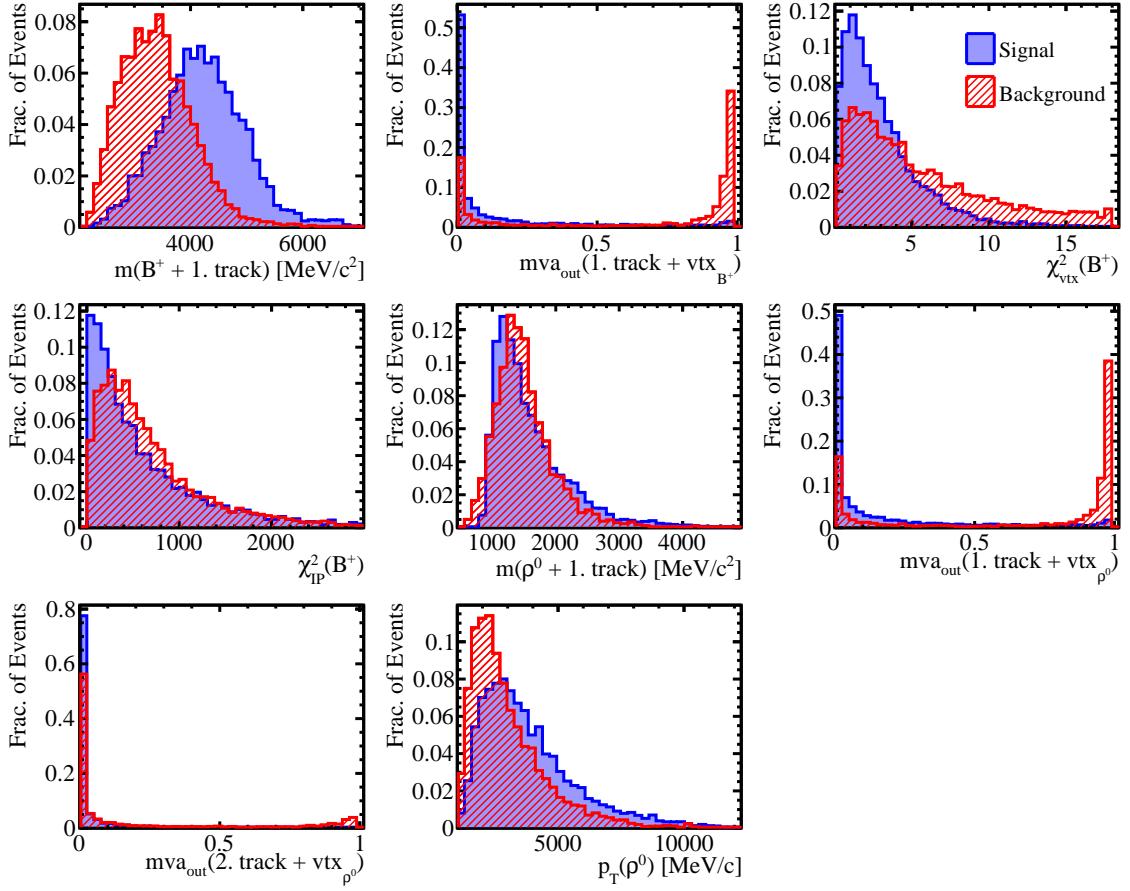
Figure 4.15 – Distributions of the input variables of the main MVA for signal (blue) and background (red).

- $\mathrm{mva_{out}}(1^{st}\text{ track} + \mathrm{vtx}_{B+})$

- $m(\rho^0 + 1^{st}\text{ track})$

- $\mathrm{mva_{out}}(1^{st}\text{ track} + \mathrm{vtx}_{\rho^0})$

- $\mathrm{mva_{out}}(2^{nd}\text{ track} + \mathrm{vtx}_{\rho^0})$

Where the first variable is kinematic, the following two are geometric and the last four are isolation variables. Their distributions in the signal and background training samples are illustrated in Fig. 4.15. In particular, isolation variables like $m(B^+ + 1^{st}\text{ track})$, $\mathrm{mva_{out}}(1^{st}\text{ track} + \mathrm{vtx}_{B+})$ and $\mathrm{mva_{out}}(1^{st}\text{ track} + \mathrm{vtx}_{\rho^0})$ are found to be powerful discriminating variables.

With a batch size of 32, the optimal number of epochs is found to be 175. This is illustrated by the performance curves in Fig. 4.16, where the model loss and accuracy on the training and test samples are plotted as a function of the number of epochs. The training is stable, *i.e.* accuracy increases and the loss decreases continuously without large fluctuations from
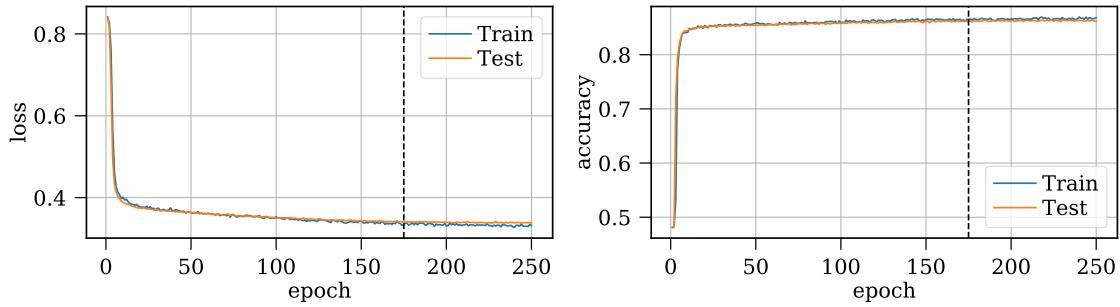
Figure 4.16 – DNN performance curves: model loss (left) and accuracy (right) versus number of epochs for the training (blue) and testing (orange) samples. The training is stopped after epoch = 175 (dashed black line) to avoid overtraining.
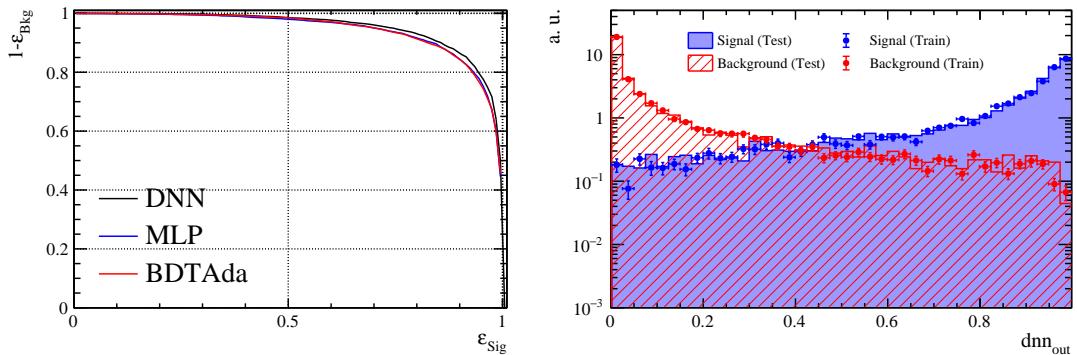
Figure 4.17 – ROC curve (left) for the DNN (black) compared to the BDTAda method (red) and the MLP (blue); DNN output distributions (right) for signal (blue) and background (red) for training (markers) and testing (filled histograms) samples.

one epoch to the next. After 175 epochs, the DNN starts to perform better on the training data than on test data, and the training is therefore stopped.

The final performance of the DNN is illustrated by the ROC curve in Fig. 4.17 resulting in an AUC value of 95.4%. This is about 1% better than the BDTAda and MLP classifiers. The distribution of the DNN output value for the signal and background training and testing samples are illustrated in log scale in Fig. 4.17. This shows a strong separation power as well as a good consistency between the training and testing samples.

In addition to optimising the DNN performance with different combinations of input variables, different architectures and model parameters were also tested, *i.e.* adding more neurons to the hidden layers, adding additional hidden layers, increasing/decreasing the drop out, changing the learning rate and neuron activation function as well as varying the batch size, however, the best performance was found with the aforementioned model architecture and parameters.
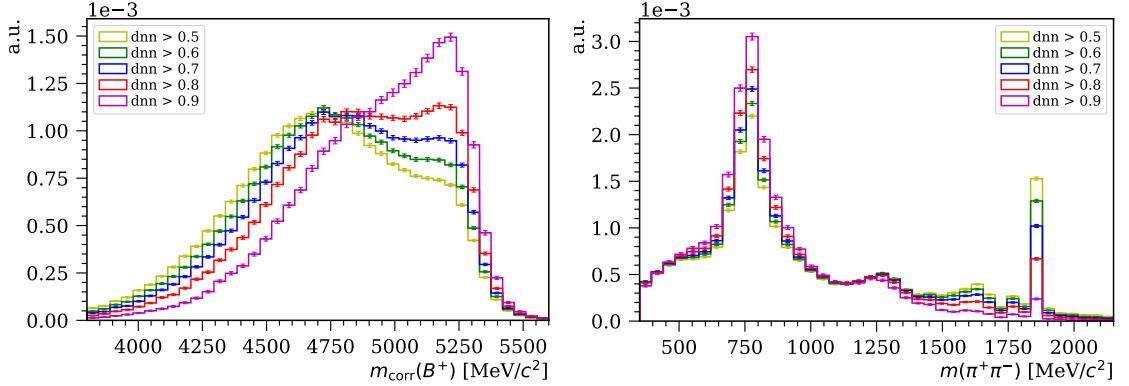
Figure 4.18 – Normalised distributions of $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right) in 2018 data for different $\mathrm{dnn_{out}}$ requirements. The $m_{\mathrm{corr}}(B^+)$ distribution is shown for $570 < m(\pi^+\pi^-) < 1100$ MeV/c^2, while the $m(\pi^+\pi^-)$ distribution is shown for $m_{\mathrm{corr}}(B^+) > 5000\,\mathrm{MeV}/c^2$.

4.5.5 DNN evaluated on data

Finally, the DNN is evaluated on the full 2018 data sample. The corrected B^+ mass distribution and invariant $\pi^+\pi^-$ mass distribution are shown in Fig. 4.18 for different selection requirements on the output value of the DNN ($\mathrm{dnn_{out}}$).

When plotting the corrected B^+ mass distribution, candidates are also required to be within $570 < m(\pi^+\pi^-) < 1100$ MeV/c^2. As the requirement on $\mathrm{dnn_{out}}$ becomes harder, the bump at $\sim 5250\,\mathrm{MeV}/c^2$ associated with signal increases with respect to the low $m_{\mathrm{corr}}(B^+)$ hill initially peaking at $\sim 4700\,\mathrm{MeV}/c^2$. To ensure that the bump at $\sim 5250\,\mathrm{MeV}/c^2$ in the $m_{\mathrm{corr}}(B^+)$ distribution is in fact signal, a cut corresponding to $m_{\mathrm{corr}}(B^+) > 5000\,\mathrm{MeV}/c^2$ is applied when plotting the $m(\pi^+\pi^-)$ distribution. The peak at $\sim 770\,\mathrm{MeV}/c^2$ associated with the ρ^0 resonance increases with respect to the lower and upper side bands as the requirements on $\mathrm{dnn_{out}}$ become harder. This confirms that the DNN selects signal while it rejects backgrounds.

It is clear that the MVA selection does not remove all backgrounds, and it is therefore necessary to identify, model and include all remaining components in the final fit. This part of the analysis as well as the final cut on $\mathrm{dnn_{out}}$ is presented in Chapter 7.

4.6 Control studies

To verify the MVA selection, consistency in the MVA variables between the simulated signal and the signal in data has to be ensured. This requires a control channel that can represent the signal and that can be fully reconstructed in data and cleanly selected.

For this analysis, the decay $B^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\rho^0$ is used. It is a suitable control channel because it has the same topology and same visible final state as the signal if one muon
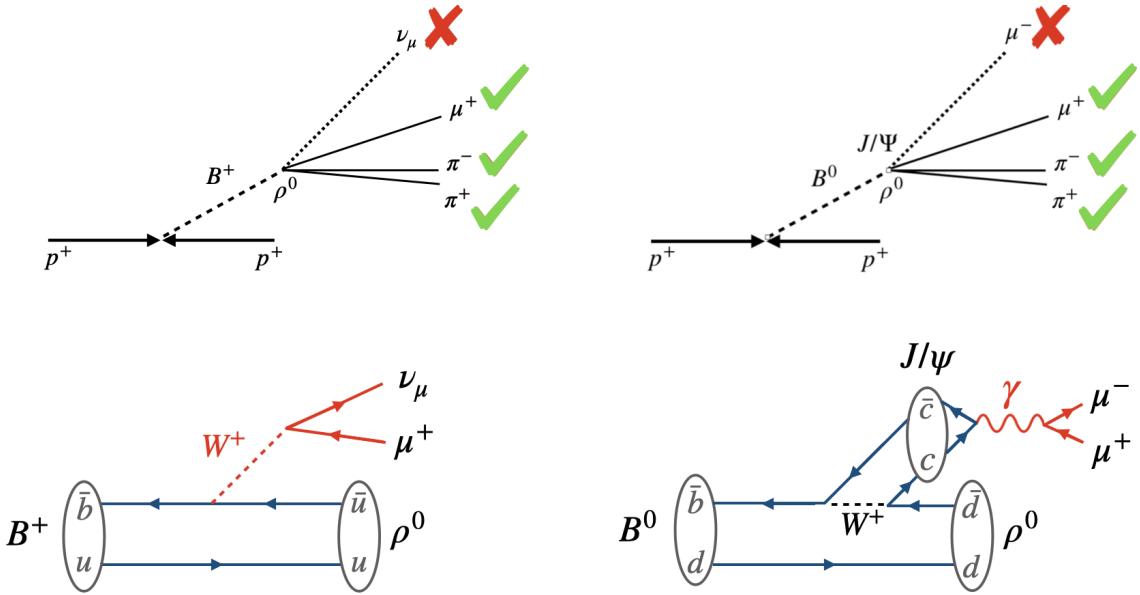
Figure 4.19 – Comparison of signal and control channel topology (top) and physical process (bottom).

is ignored, as illustrated in Fig. 4.19. The variables relevant for the MVA selection are therefore expected to be similar for the control and signal channel, and in particular, they are expected to have similar coverage[4]. Finally, the decay $B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0$ can be fully reconstructed by adding the missing muon to the $B^+ \to \rho^0\mu^+\nu_\mu$ candidate.

The main difference between the control channel and the signal is the physics of the leptonic system, as illustrated in Fig. 4.19. For signal, the muon and neutrino are produced in a weak decay mediated by the W^+ boson, while for the control channel, the two muons are produced in an electromagnetic decay of the J/ψ meson. Some differences in the kinematics of the two modes are therefore expected.

In the following sections the control studies will be presented. In particular, the choice of control channel is justified by comparing the control and signal MC variables as well as explaining their differences. This will be followed by the reconstruction and extraction of control channel decays in data. Finally, distributions in control MC will be corrected to data and the MVA selection verified.

4.6.1 Comparison of the control and signal channels

The first step is to ensure that the control channels can be used to represent the signal. In particular, it should have the same coverage as signal, since without proper coverage, differences between signal MC and data cannot be computed outside the coverage of the control channel. In addition to this, the differences between variables for control MC and

---

[4]The two decay channels have the same coverage if their variables cover the same ranges.

signal MC must be well understood, before the control channel can be used to verify the MVA selection.

The 2018 MC samples for the signal and control channels (see Table 4.1) are used. The trigger and stripping selections for signal are applied to both samples, and a set of relevant variables are used in the comparison.

The coverage of the DNN input variable distributions are found to be the same for the control channel and signal, however, discrepancies are observed in several variables, as illustrated in Appendix D.1. In order to understand the origin of these discrepancies, the difference in the leptonic system of the two modes is compensated with the three requirements illustrated in Fig. 4.20:

- The q^2 region of signal MC is restricted to a region around the $q^2 = (m(J/\psi))^2$ peak of the control channel.

- The helicity angle θ_ℓ of signal MC is reweighted to control MC.

- The missing muon μ_{miss} is excluded from the computation of isolation variables. To ensure that the μ_{miss} (found with the charge isolation MVA) originates from the J/ψ, the invariant mass of the two muons are required to be consistent with the known J/ψ mass.

After applying the three requirements, a good overall agreement is found between signal and control MC variables, as illustrated in Appendix D.1. For the few variables that do not agree, all differences are understood. For instance, after restricting the q^2 region of signal and reweighting θ_ℓ to the control channel the variable $\chi^2_{\mathrm{IP}}(B^+)$ agrees well, as illustrated in Fig. 4.21, however, after requiring $m(\mu^+\mu^-_{\mathrm{miss}})$ to be consistent with the J/ψ, a discrepancy is introduced. The reason is that the charge isolation MVA used to identify μ^-_{miss} selects a sub-sample of the control channel characterised by high $p_{\mathrm{T}}(\mu^-_{\mathrm{miss}})$ resulting in a worse $\chi^2_{\mathrm{IP}}(B^+)$.

Finally, the DNN is applied to signal and control MC, and their output distributions are compared in Fig. 4.22. After applying the three requirements to compensate for differences in their leptonic systems, the two distributions agree. In conclusion, it is found that the control channel can be used to represent the signal, because it has the same coverage as signal in all variables relevant to the selection and their physical differences are well understood.

4.6.2 Control channel reconstructed in data

After establishing that the control channel can be used to represent signal, consistency between control MC and data must be ensured.

To reconstruct the control channel in data the charge isolation MVA described in Sec. 4.5.3
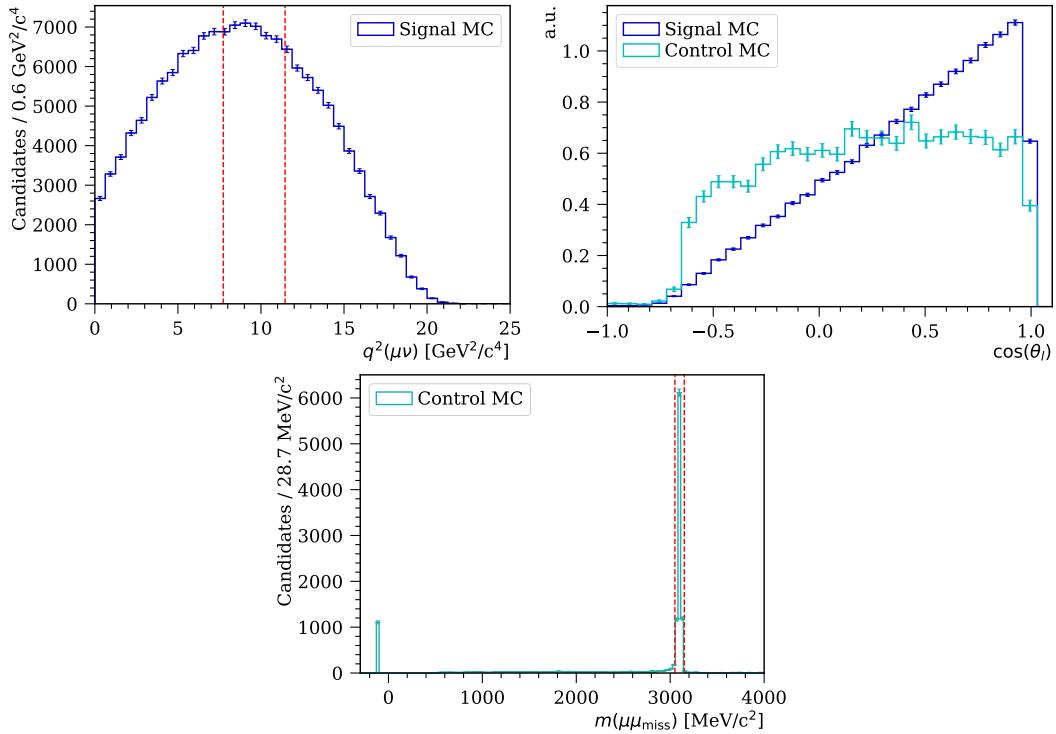
Figure 4.20 – The three requirements imposed on signal MC (blue) and control MC (cyan). The q^2 region (top left) for signal is restricted to a window (red lines) around the J/ψ mass peak. The leptonic helicity angle θ_ℓ (top right) for signal is reweighted to the control channel. The invariant mass of μ^+ and μ^-_{miss} (bottom) is restricted to be within $20\,\text{MeV}/c^2$ around the nominal J/ψ mass peak (red lines). The peak at $m(\mu^+\mu^-_{\text{miss}}) = -100\,\text{MeV}/c^2$ contains candidates where μ^-_{miss} is not reconstructed.

is used to find the most non-isolated long track with a muon hypothesis, *i.e.* μ_{miss}, which is added to the signal candidate. The control channel yield is then extracted from a fit to the invariant B mass. To reduce the backgrounds, the invariant mass of the two muons and two pions are restricted to a window around the J/ψ and ρ^0 mass, respectively. Moreover, a small $\Delta\chi^2_{\text{vtx}}$ value is required to ensure that μ_{miss} is compatible with the signal hypothesis. The specific preselection requirements are listed in Table 4.5.

Unbinned maximum likelihood fit

After the preselection, an *unbinned maximum extended likelihood fit* is performed to the invariant mass of the reconstructed $B^+ \to \rho^0\mu^+\nu_\mu$ candidates combined with μ^-_{miss}. This is done with the RooFit library [165].

As the data sample contains both the control channel and backgrounds, a composite model is needed to describe data. Given N distinct fit components described by the probability density functions (PDFs) $F_i(x|\theta)$, where $x = m(\mu^+\mu^-_{\text{miss}}\pi^+\pi^-)$ is the observable, and θ is
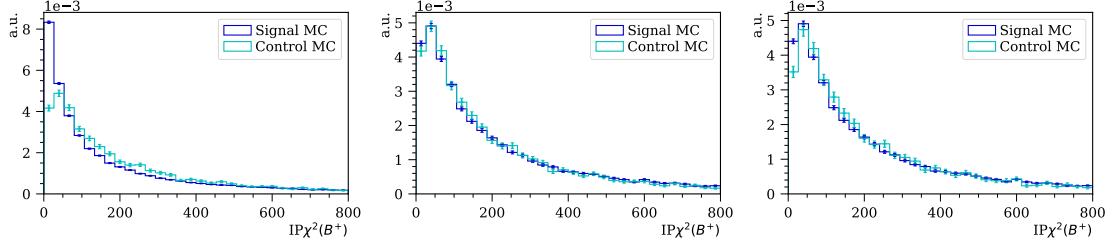
Figure 4.21 – $\chi^2_{\text{IP}}(B^+)$ distributions for signal MC (blue) and control MC (cyan). Before compensating for differences, there is a discrepancy (left). After restricting q^2 and reweighting θ_ℓ the discrepancy is gone (middle), however, a small discrepancy is introduced after restricting $m(\mu^+\mu^-{}_{\text{miss}})$ (right).
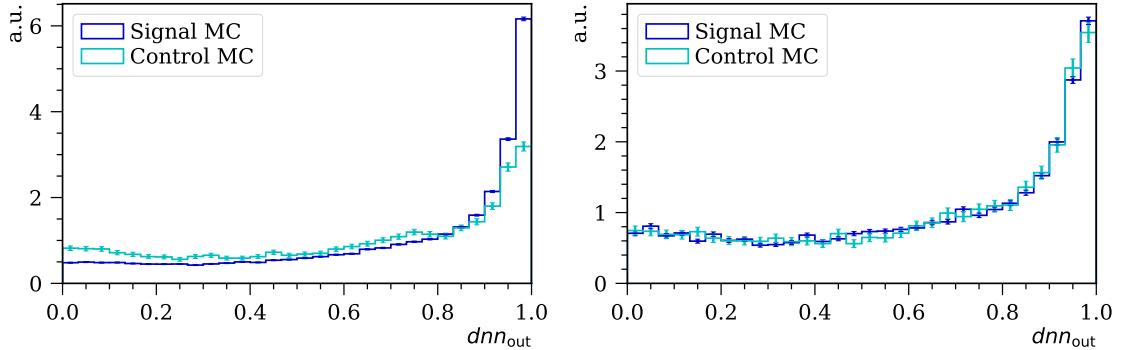
Figure 4.22 – DNN output distributions for signal MC (blue) and control MC (cyan) before (left) and after (right) applying the q^2 cut, θ_ℓ reweighting and excluding the missing μ^-.

a set of model parameters, the PDF of the composite model is given by:

$$F_{\text{tot}}(x|\theta, f) = \sum_{i=1}^{N-1} f_i F_i(x|\theta) + (1 - \sum_{i=1}^{N-1} f_i) F_N(x|\theta), \qquad (4.8)$$

where f_i is the fractional contribution of component F_i to the total PDF. The fractional contribution of F_N is given by $1 - \sum_{i=1}^{N-1} f_i$, and thus $f = (f_1, f_2, .., f_{N-1})$.

The parameter of interest is the yield of the control channel, and it is therefore convenient to express the composite PDF in terms of the control channel yield N_C and total background yield N_B as

$$F_{\text{tot}}(x|\theta, N_C, N_B) = \frac{N_C}{N_C + N_B} F_C(x|\theta_C) + \frac{N_B}{N_C + N_B} F_B(x|\theta_B), \qquad (4.9)$$

where F_C and F_B are the PDFs of the control channel and the background components with their respective model parameters θ_C and θ_B. The obtained yield $N_C + N_B$ from the minimisation of the negative log-likelihood (NLL) fluctuates according to a Poisson distribution. For this reason, a Poisson term that depends on $N_C + N_B$ and the observed

Table 4.5 – Preselection requirements applied to the data for the control channel reconstruction.

Requirement	Unit
$3077 < m(\mu^+\mu^-_{\text{miss}}) < 3117$	MeV/c^2
$570 < m(\pi^+\pi^-) < 1100$	MeV/c^2
$\Delta\chi^2_{\text{vtx}} < 25$	-

yield N_{data} is added to the likelihood resulting in the *extended likelihood*:

$$\mathcal{L}(\theta, N_C, N_B | x) = \frac{e^{N_C + N_B}(N_C + N_B)^{N_{\text{data}}}}{N_{\text{data}}!} \prod_{j=1}^{N_{\text{data}}} F_{\text{tot}}(x_j | \theta) \,, \tag{4.10}$$

and the *logarithmic extended likelihood*:

$$-\log(\mathcal{L}(\theta, N_C, N_B | x)) = - \sum_{j=1}^{N_{\text{data}}} \log(F_{\text{tot}}(x_j | \theta, N_C, N_B)) \tag{4.11}$$
$$- \ln(N_{\text{data}}!) - (N_C + N_B) + N_{\text{data}} \ln(N_C + N_B) \,,$$

which is minimised with the MINUIT algorithm [166].

Fit of the control channel in data

Before performing the unbinned maximum likelihood fit to the invariant mass distribution in data, a composite PDF describing both the contribution from the control channel and the backgrounds is built by applying Eq. 4.8.

The control channel (C) is modelled as the weighted sum of two Gaussian PDFs with a shared mean μ_C and different widths σ_{C1} and σ_{C2}:

$$F_C(x | \theta_C) = f_{C1} g_1(x | \mu_C, \sigma_{C1}) + (1 - f_{C1}) g_2(x | \mu_C, \sigma_{C2}) \,. \tag{4.12}$$

The parameters to be optimised are $\theta_C = \{\mu_C, \sigma_{C1}, \sigma_{C2}, f_{C1}\}$.

The main backgrounds are identified as the decay $B_s^0 \to J/\psi(\to \mu^+\mu^-)\pi^+\pi^-$ (Bs), $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to \pi^\pm K^\pm)$ where the $K^\pm$ is misidentified as a $\pi^\pm$ (misID) and combinatorial background (CombBg). The background model is the weighted sum of the PDFs associated with each of the backgrounds:

$$F_B(x | \theta_B) = f_{Bs} F_{Bs}(x | \theta_{Bs}) + f_{\text{misID}} F_{\text{misID}}(x | \theta_{\text{misID}}) \tag{4.13}$$
$$+ (1 - f_{Bs} - f_{\text{misID}}) F_{\text{CombBg}}(x | \theta_{\text{CombBg}}) \,.$$
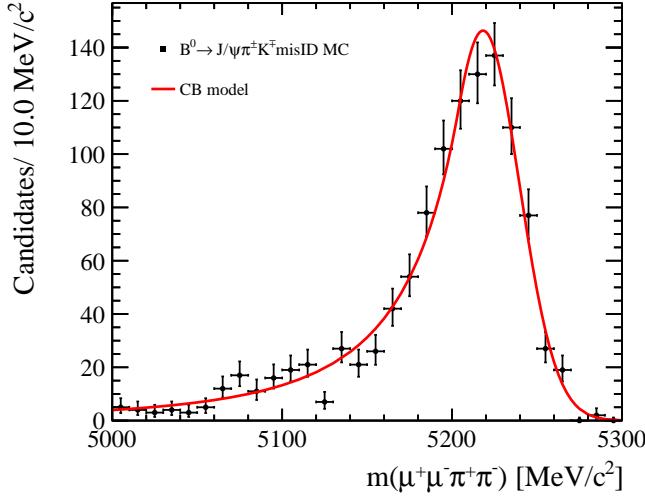
Figure 4.23 – Fit of the invariant mass distribution of the reconstructed $B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0$ candidates in a sample of the simulated $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ decays with a CB function.

Like the control channel, the $B_s^0 \to J/\psi(\to \mu^+\mu^-)\pi^+\pi^-$ background is modelled with two Gaussian PDFs, however, the mean μ_{Bs} is constrained by fixing the mass difference $\Delta m \equiv \mu_{Bs} - \mu_C = 87.40 \, \mathrm{MeV}/c^2$ [15], *i.e.* the parameters to be optimised are $\theta_{Bs} = \{\mu_C, \sigma_{Bs1}, \sigma_{Bs2}, f_{Bs1}\}$.

The $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ background is modelled with a Crystal Ball (CB) function [167], which is essentially a Gaussian with an exponential low-end tail. It is parameterised by a Gaussian mean μ_{misID} and width σ_{misID} and tail parameters n_{misID} and α_{misID}. To make the fit stable, the three last parameters are fixed to the 2018 simulation of $B^0 \to J/\psi(\to \mu^+\mu^-)K^{*0}(\to K^+\pi^-)$ (see Table 4.1), which is fitted in Fig. 4.23 with a CB function, *i.e.* the parameter to be optimised is $\theta_{\mathrm{misID}} = \{\mu_{\mathrm{misID}}\}$.

The combinatorial background is modelled with an exponential function with a slope parameter a_{CB}, *i.e.* the parameter to be optimised is therefore $\theta_{\mathrm{CB}} = \{a_{\mathrm{CB}}\}$.

Finally, the composite PDF of the control channel and backgrounds with thirteen free parameters can be expressed with Eq. 4.9 where

$$\theta = \theta_C + \theta_B = \{\mu_C, \sigma_{C1}, \sigma_{C2}, f_{C1}, \sigma_{Bs1}, \sigma_{Bs2}, f_{Bs1}, f_{Bs}, \mu_{\mathrm{misID}}, f_{\mathrm{misID}}, a_{\mathrm{CB}}\}. \quad (4.14)$$

This model is fitted to the full 2018 data sample in the invariant mass range from $5000 \, \mathrm{MeV}/c^2$ to $5560 \, \mathrm{MeV}/c^2$ by minimising the extended NLL defined in Eq. 4.12. The result of the fit is illustrated in Fig. 4.24. The quality of the fit is found to be good ($\chi^2/\mathrm{ndof} = 1.4$ with $\mathrm{ndof} = 87$) with no observed bias as indicated by the pull distribution.
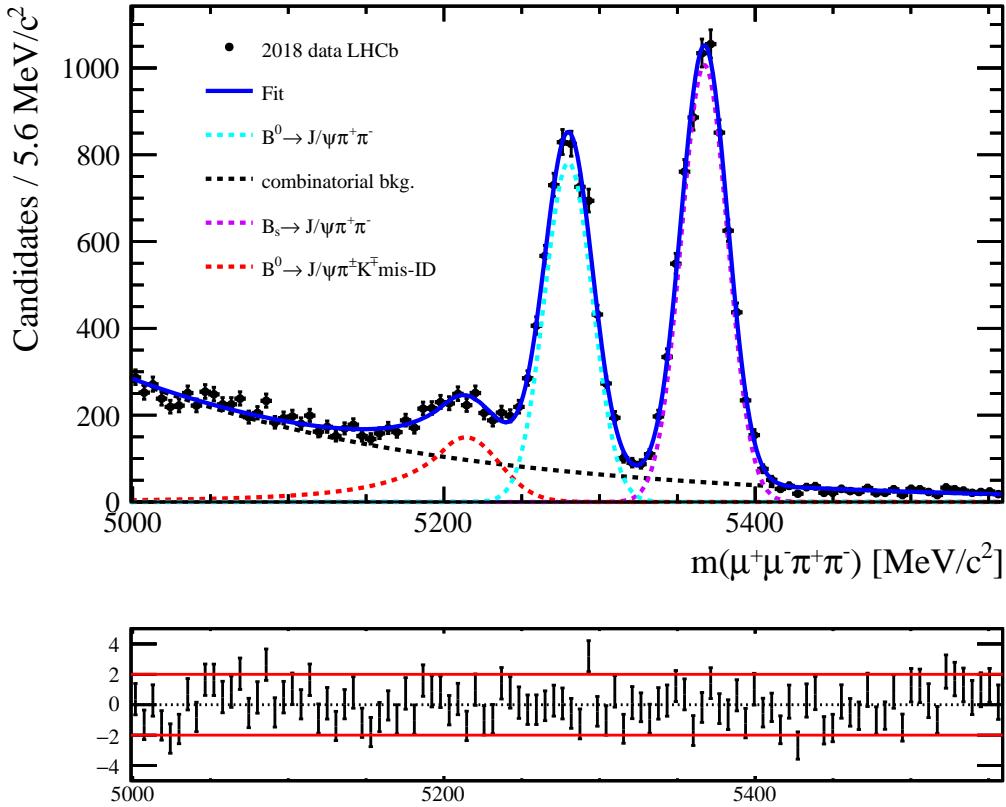
Figure 4.24 – Invariant mass distribution of the reconstructed $B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0(\to \pi^+\pi^-)$ candidates in 2018 data (black markers) with the fit components: the control mode (cyan), combinatorial background (black), $B_s^0 \to J/\psi(\to \mu^+\mu^-)\pi^+\pi^-$ background (purple) and $B^0 \to J/\psi(\to \mu^+\mu^-)\pi^\pm K^\pm$ background (red).

In the mass range from $5230\,\mathrm{MeV}/c^2$ to $5330\,\mathrm{MeV}/c^2$, the yields of the control channel and the background contribution are found to be

$$N_C = 5355 \pm 86\,,$$
$$N_B = 3304 \pm 26\,,$$

(4.15)

corresponding to a signal significance of $N_C/\sqrt{N_C + N_B} = 58$.

The *sPlot* technique

To extract the distributions of the control channel variables in data the *sPlot* technique [168] is applied. This technique consists of using the information from a maximum likelihood fit to the variable x to infer the distributions of target variables y for each of the components present in data. However, for this technique to work, y must be uncorrelated with x.

The PDF $f_i(x)$ describing the shape of component i and its yield N_i are obtained from a

maximum likelihood fit. This is used to assign an *sWeight* s_i to each event e that allows to reconstruct the distribution of y for any component i. The sWeight is defined as:

$$s_i(x_e) = \frac{\sum_{j=1}^{n_c} V_{ij} f_j(x_e)}{\sum_{k=1}^{n_c} N_k f_k(x_e)} \,, \tag{4.16}$$

where n_c is the number of fit components and V_{ij} is the covariance matrix

$$V_{ij}^{-1} = \frac{\partial^2 (-\mathcal{L})}{\partial N_i \partial N_j} = \sum_{e=1}^{N_{\text{data}}} \frac{f_i(x_e) f_j(x_e)}{(\sum_{k=1}^{n_c} N_k(x_e))^2} \,. \tag{4.17}$$

with N_{data} being the total number of events in the data sample.

For this study, the discriminating variable is the invariant mass, and the target variables are those relevant for the MVA selection. In the invariant mass range of the fit, sWeights for the control channel are assigned to all candidates.

4.6.3 Comparison of control MC and data

The distributions of variables relevant for the selection are extracted in data for the control channel by applying the sWeights, and a comparison between variable distributions in data and MC are performed, as illustrated in Appendix D.2. Discrepancies are observed in certain variables, *i.e.* some charge isolation variables and the number of tracks in the event, `nTracks`. The latter variable is not directly used in the preselection or MVA selection, however, since many variables such as PID variables depend on the event multiplicity, which is well described by `nTracks`, it is important that this variable is correctly modelled in MC.

In order to correct for the observed discrepancies, the two variables `nTracks` and $\text{mva}_{\text{out}}(1^{\text{st}}$ track $+ \text{vtx}_{B+})$ illustrated in Fig. 4.25 are used. For each variable weights are computed by dividing the histogram of the variable in MC with the one in data. Since the two variables are found to be uncorrelated, the product of the weights $w_{tot} = w_{\texttt{nTracks}} \times w_{\text{mva}}$ can be used to reweight MC. After reweighting, all input variables for the DNN and close to all variables used for the preselection are found to agree reasonably well between MC and data, as illustrated in Appendix D.2. A small unexplained discrepancy remains for high $p_{B\text{rest}}(\mu^+)$. However, for the signal decay this variable is highly correlated with the form factors, and it can therefore not be assumed that the observed discrepancy in the control channel between MC and data is the same for the signal channel.

4.6.4 Verification of MVA selection

Finally, the MVA selection is verified by training the DNN with the reweighted signal MC sample, and then applied to the control channel in MC and data. The output distribution
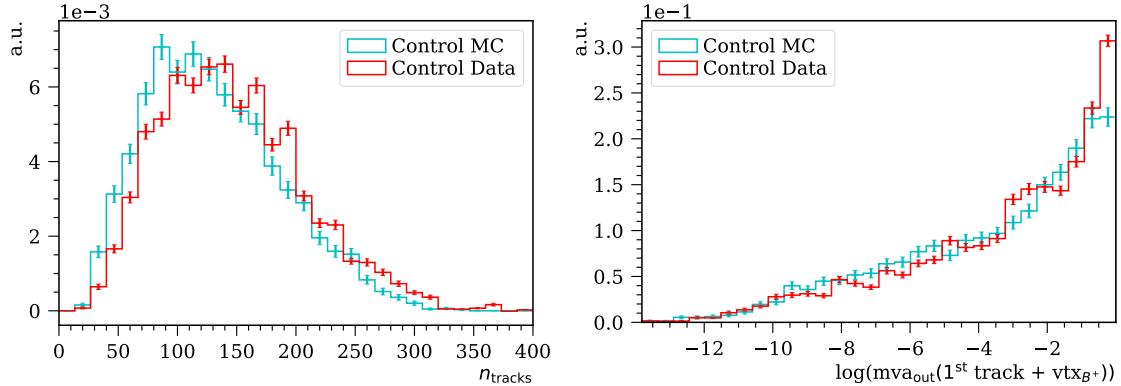
Figure 4.25 – Differences between control MC (cyan) and control data (red) in the two variables `nTracks` (left) and $\mathrm{mva}_{\mathrm{out}}(1^{\mathrm{st}}$ track $+$ vtx$_{B^+})$ (right).
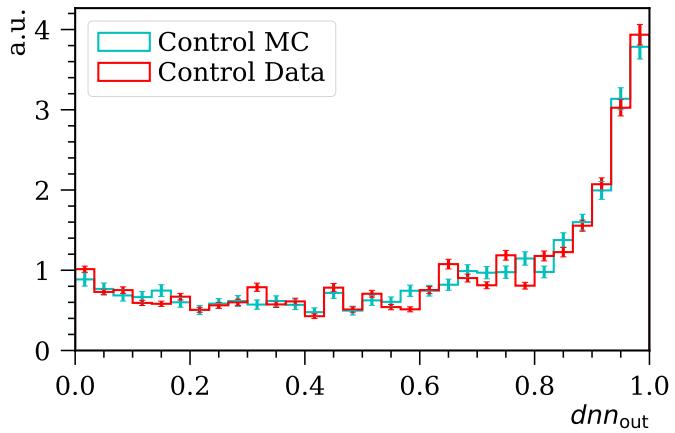
Figure 4.26 – Output of the DNN trained with $w_{\mathtt{nTracks}} \times w_{\mathrm{mva}}$ weights for the control channel in MC (cyan) and data (red).

of the DNN for MC and data is found to be in a good agreement, as illustrated in Fig. 4.26. In conclusion, after using the control channel to correct for discrepancies between MC and data, the simulated signal has been verified together with the MVA selection.

5 Neutrino reconstruction

In this analysis, the measurement of the differential branching fraction of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay is performed in bins of the dilepton invariant mass squared, q^2. However, since the momentum of the B^+ candidate is unknown and the neutrino is unmeasured, the momentum of the neutrino can only be determined up to a twofold ambiguity leading to two possible $q^2_\pm$ solutions. The correct solution is not known a priori, and a choice which q^2 solution to use must therefore be made. Moreover, due to detector resolution effects a non-negligible fraction of the signal decays has imaginary q^2 solutions, and these decays can only be used in the measurement if their real q^2 value can be estimated.

In this chapter, the derivation of q^2 is presented in Sec. 5.1, and the strategies used to choose one of the two q^2 solutions are studied in Secs. 5.2–5.4. The latter section presents the final approach where the choice of the q^2 solution is optimised by using a linear regression model. The q^2 binning scheme used to obtain the differential branching fraction of the signal is defined in Sec. 5.5, and finally, the resolution and bias of the reconstructed q^2 in each q^2 bin is computed in Sec. 5.5.1.

5.1   Dilepton mass squared q^2

As described in Appendix A, the transverse momentum of the neutrino with respect to the B^+ flight direction, $\vec{p}_{\nu_\mu \perp}$, can be determined by constraining the direction of the B^+ meson momentum to that defined by the PV and SV, as illustrated in Fig. A.1. However, its momentum parallel to the B^+ flight direction, $\vec{p}_{\nu_\mu \parallel}$, can only be determined up to a twofold ambiguity resulting in two solutions for q^2, which will be demonstrated in the following.

Using the same kinematic constraints and coordinate system as in Appendix A, and considering the visible final state particles, $i.e.$ the ρ^0 meson and μ^+ lepton, as a combined system with the energy, momentum and invariant mass denoted E_{vis}, $\vec{p}_{\text{vis}}$ and m_{vis}, respectively, the conservation of momentum and energy, leads to the following relations for

the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay in the laboratory frame:

$$E_{B^+} = E_{\text{vis}} + E_{\nu_\mu} \,, \tag{5.1}$$

$$\vec{p}_{B^+} = \vec{p}_{\text{vis}\|} + \vec{p}_{\nu_\mu \|} \,, \tag{5.2}$$

$$\vec{0} = \vec{p}_{\text{vis}\perp} + \vec{p}_{\nu_\mu \perp} \,, \tag{5.3}$$

where $\vec{p}_{\text{vis}\perp} = -\vec{p}_{\nu_\mu \perp} \equiv \vec{p}_\perp$. Assuming the neutrino mass to be negligible, *i.e.* $E_{\nu_\mu} = |\vec{p}_{\nu_\mu}|$, and using the speed of light $c = 1$, one can write:

$$m_{B^+}^2 = E_{B^+}^2 - p_{B^+}^2 \tag{5.4}$$

$$= (E_{\text{vis}} + \sqrt{p_{\nu_\mu \|}^2 + p_\perp^2})^2 - (p_{\text{vis}\|} + p_{\nu_\mu \|})^2 \tag{5.5}$$

$$= m_{\text{vis}}^2 + 2p_\perp^2 + 2E_{\text{vis}}\sqrt{p_{\nu_\mu \|}^2 + p_\perp^2} - 2p_{\text{vis}\|}p_{\nu_\mu \|} \,. \tag{5.6}$$

Note that $p_{\nu_\mu \|}$, $p_{\text{vis}\|}$ and $p_\perp$ are the components of the momenta illustrated in Fig. A.1 and they are in principle allowed to take positive as well as negative values. Rearranging the terms and squaring both sides of the equation results in:

$$(2E_{\text{vis}}\sqrt{p_{\nu_\mu \|}^2 + p_\perp^2})^2 = (m_{B^+}^2 - m_{\text{vis}}^2 - 2p_\perp^2 + 2p_{\text{vis}\|}p_{\nu_\mu \|})^2 \tag{5.7}$$

$$0 = (m_{\text{vis}}^2 + p_\perp^2)p_{\nu_\mu \|}^2 - (m_{B^+}^2 - m_{\text{vis}}^2 - 2p_\perp^2)p_{\text{vis}\|}p_{\nu_\mu \|} \tag{5.8}$$

$$+ E_{\text{vis}}^2 p_\perp^2 - 1/4(m_{B^+}^2 - m_{\text{vis}}^2 - 2p_\perp^2)^2 \,.$$

This results in a second order equation in $p_{\nu_\mu \|}$:

$$a \times p_{\nu_\mu \|}^2 + b \times p_{\nu_\mu \|} + c = 0 \,, \tag{5.9}$$

with the solution:

$$p_{\nu_\mu \|}^{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{5.10}$$

where:

$$a = (m_{\text{vis}}^2 + p_\perp^2) \tag{5.11}$$

$$b = -(m_B^2 - m_{\text{vis}}^2 - 2p_\perp^2)p_{\text{vis}\|} \tag{5.12}$$

$$c = E_{\text{vis}}^2 p_\perp^2 - \frac{1}{4}(m_B^2 - m_{\text{vis}}^2 - 2p_\perp^2)^2 \,, \tag{5.13}$$

Since the neutrino momentum is $\vec{p}_{\nu_\mu,\pm} = \vec{p}_{\nu_\mu \perp} + \vec{p}_{\nu_\mu \|}^{\pm}$ the computation of q^2 results in the two solutions $q_\pm^2 = (P_{\mu^+} + P_{\nu_\mu,\pm})^2$ (here P is the four momentum), where the correct one is not known a priori.
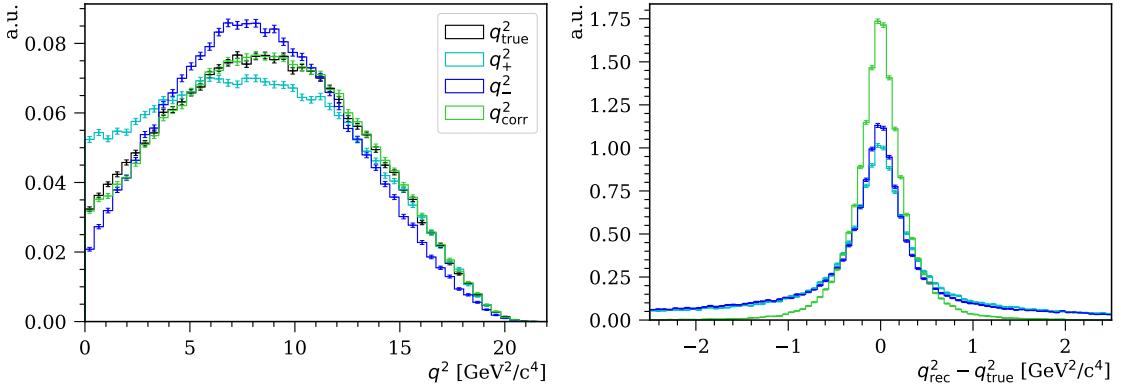
Figure 5.1 – Distributions of q^2 solutions in signal MC (left), *i.e.* q_{true}^2 reconstructed from the true μ and ν_μ four-momenta, the two real $q_\pm^2$ solutions and q_{corr}^2 consisting of only correctly chosen $q_\pm^2$ solutions, and their corresponding resolutions (right).

5.2 Two real $q_\pm^2$ solutions

After applying the trigger, stripping and preselection requirements, approximately 86% of the simulated signal decays have reconstructed q^2 solutions that are real. Knowing the true value of q^2 in the simulation, q_{true}^2, computed without detector resolution effects, the correct reconstructed solution, q_{corr}^2, can be determined as the q_+^2/q_-^2 solution closest to q_{true}^2. It is found to consist of approximately 51% of q_-^2 and 49% of q_+^2 solutions. Figure 5.1 shows the distributions of the reconstructed q^2 solutions, $q_{\text{rec}}^2 = q_+^2, q_-^2, q_{\text{corr}}^2$ together with the true q^2.

The resolution in a reconstructed q^2 solution is quantified by the variable:

$$\sqrt{\Delta^2}(q_{\text{rec}}^2) = \sqrt{\frac{\sum_i^{N_{\text{tot}}}(q_{\text{rec,i}}^2 - q_{\text{true,i}}^2)^2}{N_{\text{tot}}}} , \tag{5.14}$$

where N_{tot} is the total number of candidates.

The resolution of the reconstructed q^2 solution is illustrated in Fig. 5.1. The best q^2 resolution achievable within the experimental setup of this analysis is represented by q_{corr}^2 with an $\sqrt{\Delta^2}$ of $0.48\,\text{GeV}^2/c^4$. However, this requires that the correct q^2 solution is selected every time, which is not achievable in data. Selecting only q_-^2 or q_+^2 solutions results in degraded $\sqrt{\Delta^2}$ resolutions of $2.06\,\text{GeV}^2/c^4$ or $2.51\,\text{GeV}^2/c^4$, respectively. Moreover, selecting only one of the solutions introduces a bias in q^2, *i.e.* the mean value of $q_{\text{rec,i}}^2 - q_{\text{true,i}}^2$ equal to $\hat{\mu} = -0.50\,\text{GeV}^2/c^4$ and $-0.32\,\text{GeV}^2/c^4$ for q_+^2 and q_-^2, respectively, to be compared to $\hat{\mu} = -0.01\,\text{GeV}^2/c^4$ for q_{corr}^2. In order to improve the resolution and the bias in q^2, a linear regression model based on machine learning is used to choose between q_+^2 and q_-^2. However, before explaining the details of this method, the treatment of imaginary solutions will be discussed.
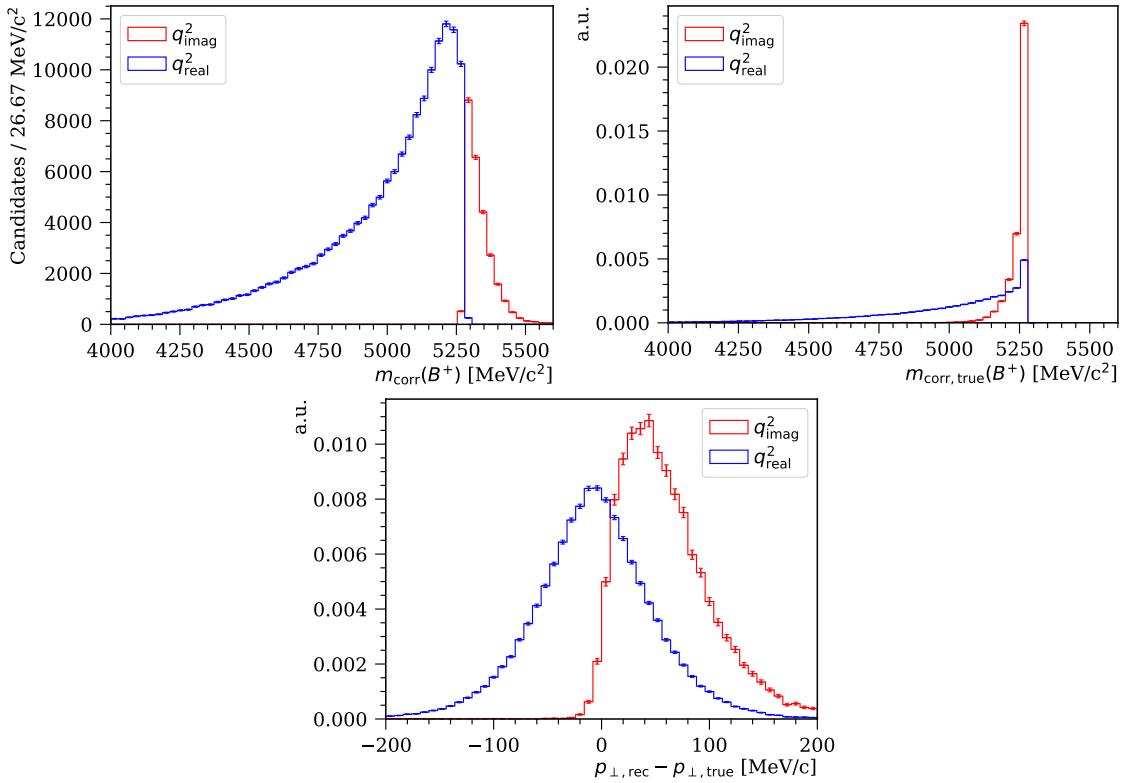
Figure 5.2 – Comparison of signal candidates with real (blue) and imaginary (red) q^2 solutions for the variables: $m_{\text{corr}}(B^+)$ (top left), $m_{\text{corr,true}}(B^+)$ (top right) and $p_{\perp,\text{rec}} - p_{\perp,\text{true}}$ (bottom).

5.3 Imaginary q^2 solutions

Approximately 14% of all signal candidates have imaginary q^2 solutions. They arise from detector resolution effects that cause the discriminant in Eq. 2.21 to become negative. In order to use these signal candidates in the analysis, a real q^2 solution must be estimated.

The simplest approach is to force the discriminant to be zero. This results in a q^2 distribution, $q^2_{\text{imag,zero}}$, with a resolution of $\sqrt{\Delta^2} = 2.0\,\text{GeV}^2/c^4$ and a bias of $\hat{\mu} = 0.45\,\text{GeV}^2/c^4$, as illustrated in Fig. 5.3. Another approach is to correct for detector resolution effects on the visible $p_{\perp,\text{rec}}$. The motivation is based on the following arguments and illustrations in Fig. 5.2. The signal candidates with imaginary q^2 solutions are characterised by an unphysical $m_{\text{corr}}(B^+)$ that is greater than the nominal B^+ mass. In general, this is caused by their true $m_{\text{corr}}(B^+)$ being close to the nominal B^+ mass, and their visible $p_{\perp,\text{rec}}$ being overestimated due to detector resolution effects, where $p_{\perp,\text{rec}} - p_{\perp,\text{true}}$ for signal candidates with imaginary q^2 solutions is mostly positive due to the fact that $m_{\text{corr}}(B^+) = m(B^+)$ corresponds to $p_{\perp,\text{true}} = 0$. For signal candidates with real q^2 solutions, $p_{\perp,\text{rec}} - p_{\perp,\text{true}}$ is symmetric around zero. Consequently, resolution effects force $m_{\text{corr}}(B^+) > m(B^+)$ for candidates with imaginary q^2 solutions.
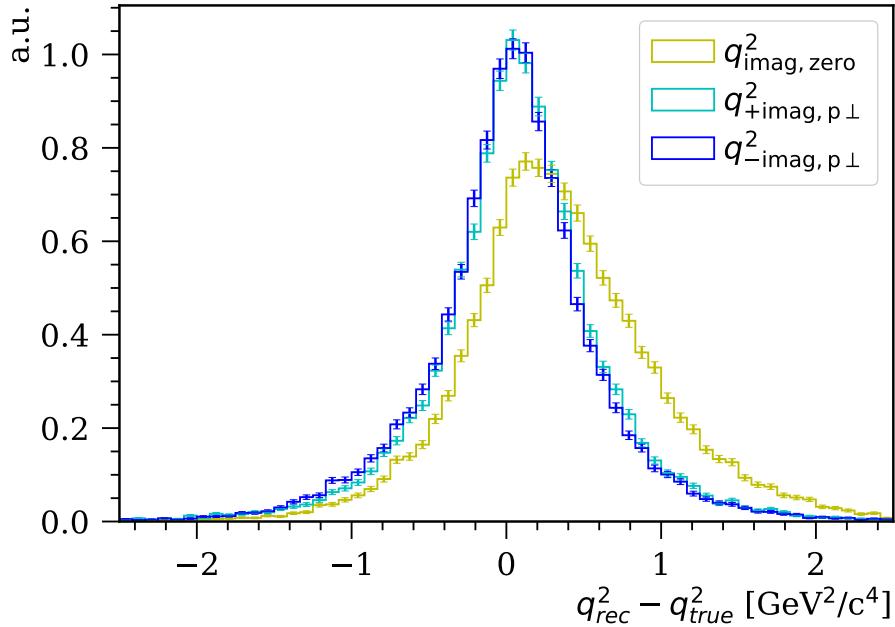
Figure 5.3 – Resolution of the corrected imaginary q^2 solutions: $q^2_{\text{imag,zero}}$ (yellow), $q^2_{+\text{imag,p}\perp}$ (cyan) and $q^2_{-\text{imag,p}\perp}$ (blue).

Based on these arguments, the visible $p_\perp$ of candidates with imaginary q^2 solutions is corrected by assuming $m_{\text{corr}}(B^+) \sim m(B^+)$ in Eq. 2.19 leading to:

$$p_{\perp,\text{corr}} = \frac{m(B^+)^2 - m^2_{\text{vis}}}{2m(B^+)} \ . \tag{5.15}$$

This is used to recompute the energy and invariant mass of the visible final state, which is then used to obtain two real q^2 solutions, $q^2_{-\text{imag,p}\perp}$ and $q^2_{+\text{imag,p}\perp}$. Compared to the first approach, this results in a better q^2 resolution corresponding to $\sqrt{\Delta^2} \sim 0.8\,\text{GeV}^2/c^4$ for both solutions, as illustrated in Fig. 5.3. Moreover, the bias in q^2 is reduced to $\hat{\mu} = 0.07\,\text{GeV}^2/c^4$ and $0.03\,\text{GeV}^2/c^4$ for $q^2_{+\text{imag,p}\perp}$ and $q^2_{-\text{imag,p}\perp}$, respectively. Consequently, this approach is used to approximate the real q^2 solutions of the candidates with imaginary q^2 solutions.

5.4 Selecting q^2 solutions with regression method

At this stage, the best resolution in q^2 is achieved by using the q^2_- and $q^2_{-\text{imag,p}\perp}$ solutions only. However, as discussed in Sec. 5.2, this choice introduces a bias. To reduce the bias and improve the resolution in q^2, a method developed in Ref. [169] is applied. The basic principle of this method is to use a multivariate linear regression model [170] to predict the true B momentum and choose the q^2 solution most consistent with the predicted value. The prediction is based on two *regression variables*[1], that are correlated with the B

---

[1]In Ref. [170] more regression variables are tested, however, only $1/\sin\theta_{\text{flight}}$ and $|\vec{F}|$ are found to be useful for predicting the B momentum.

momentum, p, but independent of the B decay properties.

The first variable is the polar angle, θ_{flight}, of the flight vector, $\vec{F}$, defined between the PV and SV. Since the p_{T} and θ_{flight} is found to be only weakly correlated, the B momentum, p, can be approximated as:

$$p = \frac{\bar{p}_{\text{T}}}{\sin\theta_{\text{flight}}} \; , \tag{5.16}$$

where $\bar{p}_{\text{T}}$ is the average transverse momentum as a function of pseudo rapidity $\eta = -\ln(\tan(\theta_{\text{flight}}/2))$.

The second variable is the flight distance, $|\vec{F}|$, that is related to p via

$$p = \frac{M|\vec{F}|}{t} \; , \tag{5.17}$$

where M and t are the B mass and decay time, respectively.

Then the B momentum can be predicted from the following linear regression model:

$$p_{\text{pred}} = \beta_0 + \frac{\beta_1}{\sin\theta_{\text{flight}}} + \beta_2|\vec{F}| \; , \tag{5.18}$$

where the coefficients $\beta_{0,1,2}$ are to be determined.

In this analysis, a linear discriminant algorithm (LD) provided by TMVA [156] is used to determine the coefficients. It uses the standard approach of *least squares* where the sum of the squared residuals, *i.e.* the difference between p_{true} and p_{pred}, is minimised.

A sample consisting of $\sim$ 15k signal MC candidates is split into two samples used for training and testing the LD regression algorithm. Both samples have undergone trigger, stripping and preselection, and they are weighted to correct for data/MC differences as described in Sec. 4.6.3. The distributions of the regression variables, *i.e.* $1/\sin\theta_{\text{flight}}$ and $|\vec{F}|$ versus p_{true} are illustrated in Fig. 5.4. The correlations between p_{true} and $1/\sin\theta_{\text{flight}}$ is 52% and between $|\vec{F}|$ it is 77%.

The LD regression algorithm is trained, and no significant overtraining is observed, *i.e.* the average quadratic deviation between p_{true} and p_{pred} differs less than 2% for the training and testing sample. The algorithm is then applied to a new signal MC sample, which will be used in the final fit of the signal channel. The q^2 solution, q^2_{reg}, most consistent with p_{pred} is selected. Considering only candidates with real q^2 solutions, this method results in 60% correct q^2 solutions compared to 51% when selecting only q^2_-. Considering only candidates with corrected imaginary q^2 solutions, the regression method selects the q^2 solution closest to the true q^2 value in 51% of the cases. The resolution of q^2_{reg} and q^2_-
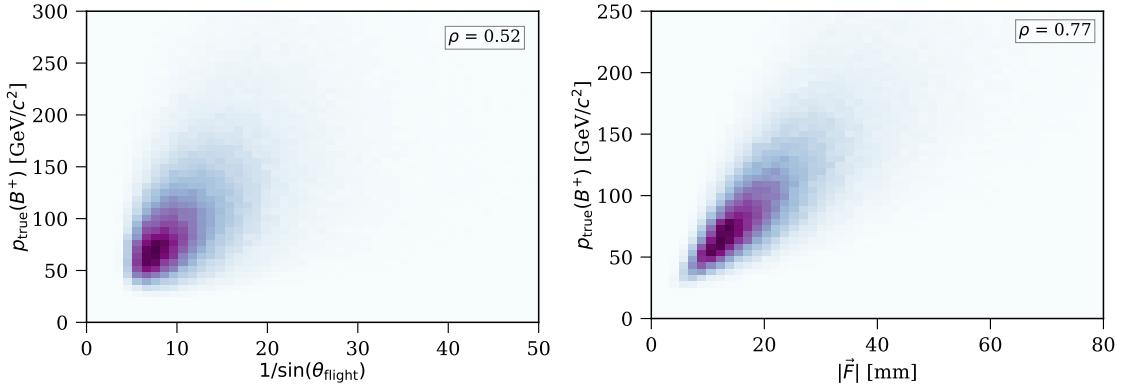
Figure 5.4 – Distributions in signal MC of the regression variables $1/\sin\theta_{\text{flight}}$ (left) and $|\vec{F}|$ (right) versus $p_{\text{true}}(B^+)$.
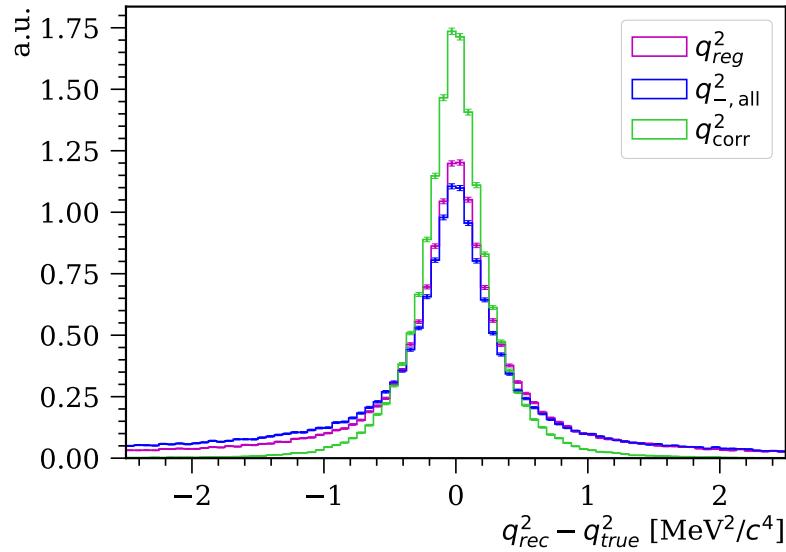
Figure 5.5 – Resolution of q^2_{reg} (magenta) and $q^2_{-,\text{all}}$ (blue), where signal candidates with corrected imaginary q^2 solutions are included, and q^2_{corr} (green).

(including the corrected imaginary q^2 solutions) are illustrated in Fig. 5.5 together with the resolution of q^2_{corr}. The resolution of q^2_{reg} is $\sqrt{\Delta^2} = 1.68\,\text{GeV}^2/c^4$. This corresponds to a $\sim 14\%$ improvement in the resolution compared to q^2_-. Moreover, the bias in q^2_{reg} is $\hat{\mu} = 0.11\,\text{GeV}^2/c^4$, which is about 60% smaller than in q^2_-. Consequently, for this analysis q^2_{reg} is used to reconstruct q^2. A summary of the $\sqrt{\Delta^2}$ and bias for different reconstructed q^2 solutions is given in Table 5.1.

5.5 q^2 binning scheme

A q^2 binning scheme is defined for measuring the differential branching fraction of the signal. To obtain a precise measurement of $|V_{ub}|$ and $B^+ \to \rho^0$ form factors a fine q^2

Table 5.1 – Summary of the $\sqrt{\Delta^2}$ and $\hat{\mu}$ of the reconstructed q^2 solutions: q^2_{corr}, q^2_+, q^2_- and q^2_{reg}. For the three latter q^2 solutions, both real and corrected imaginary q^2 solutions are included.

Solution	$\sqrt{\Delta^2}\,[\,\mathrm{GeV}^2/c^4]$	$\hat{\mu}\,[\,\mathrm{GeV}^2/c^4]$
q^2_{corr}	0.48	0.02
q^2_+	2.34	0.42
q^2_-	1.93	0.27
q^2_{reg}	1.68	0.11
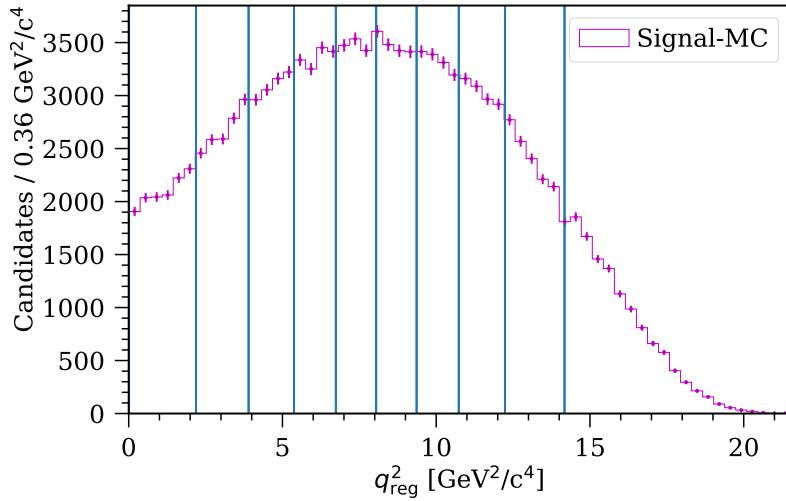
Figure 5.6 – Binning scheme (blue vertical lines) of q^2_{reg} (purple) for the signal channel fit.

binning is desired, however, the binning cannot be too fine, as this will result in a unstable fit. A q^2 binning scheme with ten bins of approximately equal signal statistics (using the BCL form factor scheme) is chosen for the signal fit. This is illustrated in Fig. 5.6 together with the distribution of q^2_{reg} in signal MC. The bin boundaries are defined in Table 5.2.

5.5.1 Resolution and bias in bins of q^2

Since the final measurement is performed in bins of q^2, the resolution and bias of the reconstructed q^2 is computed for each bin. This is done in Fig. 5.7, which shows the $\sqrt{\Delta^2}$ and $\hat{\mu}$ of q^2_{reg} for signal MC in the ten bins defined in Table 5.2. The plot shows the resolution and bias after the full selection defined in Sec. 7.1. In general, the $\sqrt{\Delta^2}$ decreases with increasing q^2 bin. The best resolution is obtained in the last bin corresponding to $\sqrt{\Delta^2} = 1.23\,\mathrm{GeV}^2/c^4$, while the worst resolution is obtained in the first bin corresponding to $\sqrt{\Delta^2} = 1.74\,\mathrm{GeV}^2/c^4$. The bias starts out at a maximum negative value of $\hat{\mu} = -0.68\,\mathrm{GeV}^2/c^4$ in bin 1 and turns positive in bins 7–10 with the largest positive bias being $\hat{\mu} = 0.31\,\mathrm{GeV}^2/c^4$. The effect of the resolution and bias as a function of q^2 on the final result is assessed in Sec. 7.4.

Table 5.2 – Bin boundaries q^2_{min} and q^2_{max} defining the q^2 binning scheme for the signal channel fit.

Bin	q^2_{min} [GeV$^2/c^4$]	q^2_{max} [GeV$^2/c^4$]
1	0.0	2.2
2	2.2	3.9
3	3.9	5.3
4	5.3	6.7
5	6.7	8.0
6	8.0	9.3
7	9.3	10.7
8	10.7	12.2
9	12.2	14.2
10	14.2	21.5
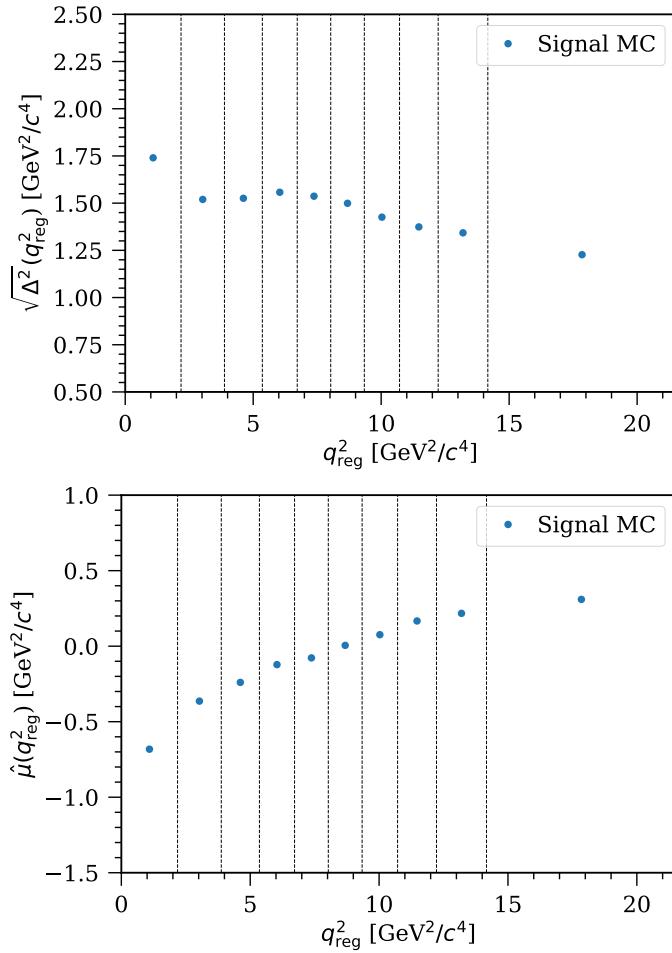
Figure 5.7 – $\sqrt{\Delta^2}(q^2_{\text{reg}})$ (top) and $\hat{\mu}(q^2_{\text{reg}})$ (bottom) in q^2 bins (dashed vertical lines) for signal MC (blue markers) after the full selection.

6 Normalisation channel fit

Since the B^+-meson production rate cannot be precisely determined at LHCb, the measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction is performed relative to another decay, also known as a *normalisation channel*. This results in a cancellation of the uncertainty associated with the B^+-meson production rate and to a large extent to a cancellation of uncertainties related to the event selection efficiencies.

In this analysis, the normalisation channel is the semileptonic $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ decay with $\bar{D}^0 \to \pi^+ \pi^-$. Its branching fraction is about 22% of the one of the signal [15],

$$\mathcal{B}(B^+ \to \bar{D}^0 \mu^+ \nu_\mu) \times \mathcal{B}(\bar{D}^0 \to \pi^+ \pi^-) = (3.34 \pm 0.14) \times 10^{-5}\,, \tag{6.1}$$

with a relative uncertainty of 4.2%.

Like the signal, the normalisation channel is a B^+ meson decay with a final state consisting of $\mu^+ \pi^+ \pi^-$ and one unmeasured ν_μ. Due to the strict $\chi^2_{\mathrm{vtx}}(B^+)$ requirement in the `B2XuMuNuBu2Rho_Line` (see Table 4.3) the flight distances of the D^0 mesons are in general short such that a similar topology between the signal and normalisation channel is obtained.

Since the q^2 distribution of the normalisation channel does not cover the same range as that of the signal, the normalised differential branching fraction of the signal is obtained by measuring the signal yield in bins of q^2 relative to the yield of the normalisation channel integrated over the full q^2 region. The yield of the normalisation channel is obtained from a template fit to $m_{\mathrm{corr}}(B^+)$. In the following sections the details of the fit and the different background contributions will be explained.

6.1 Selection of the normalisation mode

The selection of the normalisation mode aims at satisfying two requirements. On the one hand, it tries to be as similar to the signal selection as possible in order to reduce the systematic uncertainty of the final measurement of the differential branching fraction of the

Table 6.1 – Summary of the selection requirements specific to the normalisation channel.

Selection requirement
$\text{mva}_{\text{out}}(1^{\text{st}}\ \text{track} + \text{vtx}_{\rho^0}) < 0.7$
$\sigma_{m_{\text{corr}}}(B^+) < 150\,\text{MeV}/c^2$
$1835\,\text{MeV}/c^2 < m(\pi^+\pi^-) < 1905\,\text{MeV}/c^2$

signal decay. On the other hand, since the branching fraction of the normalisation channel is about 1/5 of the one of the signal, some selection requirements that are inefficient for the normalisation channel have to be modified in order to keep the statistical uncertainty of the extracted yield at a reasonable level.

While the trigger and stripping selections are the same for the signal and normalisation channels (see Secs. 4.2–4.3), some requirements of the preselection and MVA selection are found to increase the relative uncertainty of the yield extracted in the normalisation channel fit, as discussed in Sec. 6.4, and they are therefore modified as summarised in Table 6.1. For the preselection (see Sec. 4.4.5) the requirement on $\sigma_{m_{\text{corr}}}(B^+)$ is relaxed, while for the MVA selection the requirement on dnn_{out} is completely dropped. Instead, a requirement on the charge isolation variable $\text{mva}_{\text{out}}(1^{\text{st}}\ \text{track} + \text{vtx}_{\rho^0})$ is introduced. As explained in Sec. 6.2.2, this requirement results in a better rejection of backgrounds, in particular combinatorial background (CombBg), while retaining a high efficiency of the normalisation channel. Finally, the invariant mass of the two pions is required to fall within a window around the D^0 mass of $1864.84 \pm 0.05\,\text{MeV}/c^2$ [15] defined as $1835.0\,\text{MeV}/c^2 < m(\pi^+\pi^-) < 1905.0\,\text{MeV}/c^2$. The choice of an asymmetric window around the D^0 mass is motivated in Sec. 6.2.2.

6.2   Fit components

The yield of the normalisation channel is extracted from a fit to $m_{\text{corr}}(B^+)$ using the full 2018 LHCb data sample. As the dipion mass is restricted to a narrow region around the D^0 mass, the physics backgrounds are semileptonic $B \to D^{(*,**)}\mu\nu_\mu(X)$ decays (with X being a charged or neutral π) where $D^{(*,**)}$ cascades into a D^0 that subsequently decays into $\pi^+\pi^-$. In addition to this, CombBg and misIDμ background, *i.e.* normalisation candidates with a misidentified μ, are expected to contribute to the fit. Since the $m_{\text{corr}}(B^+)$ distribution cannot easily be described by an analytic expression, histograms, also known as *templates*, are obtained from simulations or data-driven methods, and are used to describe the $m_{\text{corr}}(B^+)$ shape of the fit components. The HISTFACTORY framework, described in Sec. 6.3, is then used to perform the template fit of the normalisation channel. In the following sections the different contributions to the fit as well as the extraction of their templates are explained.

Table 6.2 – Simulations of the normalisation channel and physics backgrounds, where X represents $\pi^\pm/\pi^0$ and $N_{B,\text{gen}}$ is the total number of generated B^+ or B^0.

Name	MC cocktail	$N_{B,\text{gen}}$
NORMBu	$B^+ \rightarrow \bar{D}^{(*,**)0}\mu^+\nu_\mu(X), \bar{D}^0 \rightarrow \pi^+\pi^-$	214964207
NORMBd	$B^0 \rightarrow D^{(*,**)-}\mu^+\nu_\mu(X), D^0 \rightarrow \pi^+\pi^-$	275086439

6.2.1 Normalisation channel and physics backgrounds

The normalisation channel and physics backgrounds are modelled with the two simulation samples defined in Table 6.2. They consist of $B^+ \rightarrow \bar{D}^{(*,**)0}\mu^+\nu_\mu(X)$ (NORMBu) and $B^0 \rightarrow \bar{D}^{*,**)}\mu^+\nu_\mu(X)$ (NORMBd) decays, respectively. For both samples $X = \pi^\pm/\pi^0$ and the $D^{(*,**)}$ cascades into a D^0 that is subsequently forced to decay into $\pi^+\pi^-$.

Using the true identity of the final state particles, it is possible to extract fourteen decay processes from NORMBu, specified in Table 6.3, and twelve decay processes from NORMBd, specified in Table 6.4. Except for the decay $B^0 \rightarrow D^*(2010)^-\mu^+\nu_\mu$ that is modelled with CLN form factors [171] (described below), most of the semileptonic decays in NORMBu and NORMBd are modelled with ISGW2 form factors [71], however, for the nonresonant $B \rightarrow D^{(*)}\pi\mu\nu_\mu$ decays (Bu11, Bu12, Bd10 and Bd11) the Goity-Roberts model [172] is used.

The dominant processes are $B^+ \rightarrow \bar{D}^*(2007)^0\mu^+\nu_\mu$ with $\mathcal{B} \sim 5.6\%$, $B^0 \rightarrow D^*(2010)^-\mu^+\nu_\mu$ with $\mathcal{B} \sim 3.3\%$ and the normalisation channel $B^+ \rightarrow \bar{D}^0\mu^+\nu_\mu$ with $\mathcal{B} \sim 2.3\%$. The remaining processes, often involving higher D resonances, have small branching fractions of the order of $\sim 10^{-3}$–10^{-5}. The branching fractions listed in Tables 6.3 and 6.4 are obtained from Ref. [15]. The branching fractions of the nonresonant decays (Bu11, Bu12, Bd10 and Bd11) have not been properly measured. In certain theoretical studies [173], the branching fraction for $B \rightarrow D\pi\ell\bar{\nu}$ decays has been argued to be as big as 0.5%–1%, however, in a more recent measurements from BaBar the nonresonant contribution has been found to be consistent with zero [174].

For modelling semileptonic $B \rightarrow D^{(*)}\ell\bar{\nu}$ decays a more accurate form factor parameterisation than ISGW2 exists. This is known as the CLN form factor parameterisation [171][1]. The decay rate is modelled with two quantities, *i.e.* $\eta_{\text{EW}}G(1)|V_{cb}|$ and ρ^2, where η_{EW} is the electroweak correction, $|V_{cb}|$ is the CKM matrix element and $G(1)$ and ρ^2 are slope parameters. Both quantities have been precisely measured by several experiments [57]:

$$\eta_{\text{EW}}G(1)|V_{cb}| = 41.53 \pm 0.44 \pm 0.88\,, \tag{6.2}$$

$$\rho^2 = 1.129 \pm 0.024 \pm 0.023\,, \tag{6.3}$$

$$\text{correlation} = 0.96/0.72/0.76\ \text{stat/syst/total}\,. \tag{6.4}$$

---

[1] Developed by I. Caprini, L. Lellouch and M. Neubert.
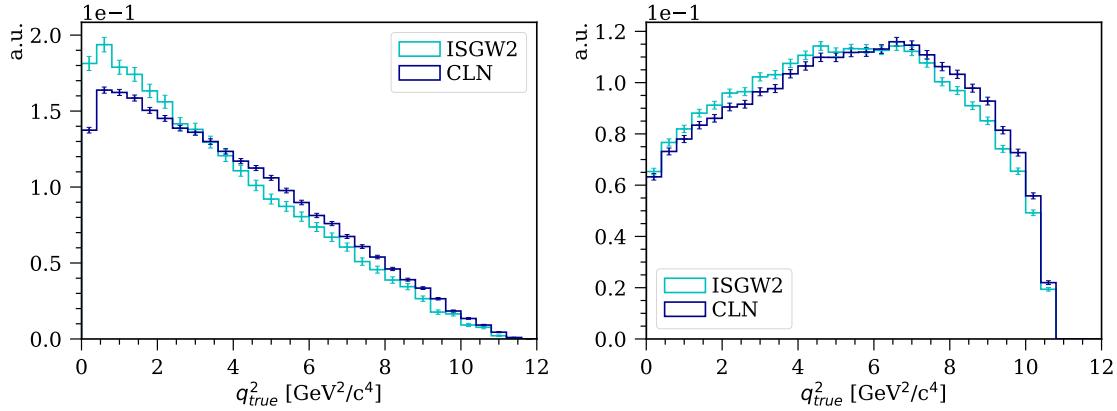
Figure 6.1 – True q^2 distribution of the normalisation channel (left) and of the $B^+ \rightarrow \bar{D}^{0^*}\mu^+\nu_\mu$ decay (right) based on the ISGW2 (cyan) and CLN (blue) form factors.

While the $m_{\text{corr}}(B^+)$ distribution is essentially unaffected by changing the form factor parameterisation, the q^2 distribution is strongly affected. Thus to compare the expected and measured q^2 distribution in Sec. 6.5, the q^2 distribution of the two processes, *i.e.* Bu1 and Bu2, are re-weighted according to the CLN form factor parameterisation. In practice, a correction factor is obtained by comparing the true q^2 distributions based on ISGW2 and CLN form factors generated with EVTGEN [141]. These distributions are illustrated for the normalisation channel and the $B^+ \rightarrow \bar{D}^{0^*}\mu^+\nu_\mu$ decay in Fig. 6.1.

To reduce the complexity of the fit, the less prominent decay processes with similar $m_{\text{corr}}(B^+)$ distributions are grouped into the same templates, as illustrated in Fig. 6.2. This results in the eight templates illustrated in Fig. 6.3. The first three templates describe the dominant processes corresponding to the normalisation channel (NORM), the $B^+ \rightarrow \bar{D}^{*0}\mu^+\nu_\mu$ (Bu2) and $B^0 \rightarrow D^{*-}\mu^+\nu_\mu$ (Bd1) decays, while the remaining five grouped templates, denoted GR1, GR2, GR3, GR4 and GR5, describe different collections of physics backgrounds. The nonresonant $B \rightarrow D^{(*)}\pi\mu\nu_\mu$ decays for which the branching fractions are not known are contained in GR5, and all $B \rightarrow D^{(*)}\tau^+\nu_\tau$ decays involving τ leptons are contained in GR4.

The expected template yield, N_{T}, depends on the branching fraction of its components, $\mathcal{B}_i$, and their selection efficiencies, ϵ_i, by:

$$N_{\text{T}} = N_B \times \sum_i^n \mathcal{B}_i \times \epsilon_i \,, \qquad (6.5)$$

where N_B is the total number of B^+ or B^0 mesons produced in the measurement, *i.e.* $N_B = N_{B^+} = N_{B^0}$, and n is the number of processes making up the template T. Since N_B is not well known at LHCb, the yield of each template cannot be precisely determined, however, by computing the ratio of Eq. 6.5 between two templates, T$_1$ and T$_2$, with n_1 and

Table 6.3 – Truth matched $B^+ \to \bar{D}^{*0}\mu^+\nu_\mu(X)$ processes with $\bar{D}^0 \to \pi^+\pi^-$ in the NORMBu MC cocktail defined in Table 6.2. The total branching fractions $\mathcal{B}_{tot}$ of each processes with ID Bui (i is in the range 1–14) are listed together with the branching fractions $\mathcal{B}$ for each subdecay retrieved from the PDG [15].

ID	Decay process	$\mathcal{B}$	$\mathcal{B}_{tot}$
Bu1	$B^+ \to \bar{D}^0\mu^+\nu_\mu$	$(2.30 \pm 0.09)\%$	$(2.30 \pm 0.09)\%$
Bu2	$B^+ \to \bar{D}^*(2007)^0\mu^+\nu_\mu$	$(5.58 \pm 0.22)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(5.58 \pm 0.22)\%$
Bu3	$B^+ \to \bar{D}_0^*(2420)^0\mu^+\nu_\mu$	$(3.8 \pm 0.8) \times 10^{-3}$	
	$\bar{D}^*(2420)^0 \to \bar{D}^0\pi^0$	$1/3$	$(1.3 \pm 0.3) \times 10^{-3}$
Bu4	$B^+ \to \bar{D}_1(2420)^0\mu^+\nu_\mu$	$(4.55 \pm 0.30) \times 10^{-3}$	
	$\bar{D}_1(2420)^0 \to \bar{D}^*(2010)^+\pi^-$	$2/3$	
	$\bar{D}^*(2010)^+ \to \bar{D}^0\pi^+$	$(67.7 \pm 0.5)\%$	$(2.05 \pm 0.14) \times 10^{-3}$
Bu5	$B^+ \to \bar{D}_1(2420)^0\mu^+\nu_\mu$	$(4.55 \pm 0.30) \times 10^{-3}$	
	$\bar{D}_1(2420)^0 \to \bar{D}^*(2007)^0\pi^0$	$1/3$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(1.52 \pm 0.10) \times 10^{-3}$
Bu6	$B^+ \to \bar{D}_1(2430)^0\mu^+\nu_\mu$	$(4.1 \pm 0.9) \times 10^{-3}$	
	$\bar{D}_1(2430)^0 \to \bar{D}^*(2010)^+\pi^-$	$2/3$	
	$\bar{D}^*(2010)^+ \to \bar{D}^0\pi^+$	$(67.7 \pm 0.5)\%$	$(1.9 \pm 0.4) \times 10^{-3}$
Bu7	$B^+ \to \bar{D}_1(2430)^0\mu^+\nu_\mu$	$(4.1 \pm 0.9) \times 10^{-3}$	
	$\bar{D}_1(2430)^0 \to \bar{D}^*(2007)^0\pi^0$	$1/3$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(1.4 \pm 0.3) \times 10^{-3}$
Bu8	$B^+ \to \bar{D}_2(2460)^0\mu^+\nu_\mu$	$(2.21 \pm 0.53) \times 10^{-3}$	
	$D_2(2460)^0 \to \bar{D}^*(2010)^+\pi^-$	0.2090	
	$\bar{D}^*(2010)^+ \to \bar{D}^0\pi^+$	$(67.7 \pm 0.5)\%$	$(3.13 \pm 0.75) \times 10^{-4}$
Bu9	$B^+ \to \bar{D}_2(2460)^0\mu^+\nu_\mu$	$(2.21 \pm 0.53) \times 10^{-3}$	
	$D_2(2460)^0 \to \bar{D}^0\pi^0$	0.2290	$(5.06 \pm 0.12) \times 10^{-4}$
Bu10	$B^+ \to \bar{D}_2(2460)^0\mu^+\nu_\mu$	$(2.21 \pm 0.53) \times 10^{-3}$	
	$D_2(2460)^0 \to \bar{D}^*(2007)^0\pi^0$	0.1030	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(2.28 \pm 0.55) \times 10^{-4}$
Bu11	$B^+ \to \bar{D}^0\mu^+\nu_\mu\pi^0$		
Bu12	$B^+ \to \bar{D}^*(2007)^0\mu^+\nu_\mu\pi^0$		
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	
Bu13	$B^+ \to \bar{D}^0\tau^+\nu_\tau$	$(7.7 \pm 2.5) \times 10^{-3}$	
	$\tau^+ \to \mu^+\bar{\nu}_\mu\nu_\tau$	$(17.39 \pm 0.04)\%$	$(1.3 \pm 0.4) \times 10^{-3}$
Bu14	$B^+ \to \bar{D}^*(2007)^0\tau^+\nu_\tau$	$(1.88 \pm 0.20)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	
	$\tau^+ \to \mu^+\bar{\nu}_\mu\nu_\tau$	$(17.39 \pm 0.04)\%$	$(3.27 \pm 0.35) \times 10^{-3}$

Table 6.4 – Truth matched $B^0 \to D^{(*,**)-}\mu^+\nu_\mu(X)$ processes with $D^0 \to \pi^+\pi^-$ in the NORMBd MC cocktail defined in Table 6.2. The total branching fractions $\mathcal{B}_{\text{tot}}$ of each processes with ID Bui (i is in the range 1–12) are listed together with the branching fractions $\mathcal{B}$ for each subdecay retrieved from Ref. [15].

ID	Decay process	$\mathcal{B}$	$\mathcal{B}_{\text{tot}}$
Bd1	$B^0 \to D^*(2010)^-\mu^+\nu_\mu$	$(4.97 \pm 0.12)\%$	
	$D^*(2010)^- \to \bar{D}^0\pi^-$	$(67.7 \pm 0.5)\%$	$(3.31 \pm 0.08)\%$
Bd2	$B^0 \to D^*(2300)^-\mu^+\nu_\mu$	$(4.5 \pm 1.8) \times 10^{-3}$	
	$D^*(2300)^- \to \bar{D}^0\pi^-$	$2/3$	$(3.0 \pm 1.2) \times 10^{-3}$
Bd3	$B^0 \to D_1(2420)^-\mu^+\nu_\mu$	$(4.20 \pm 0.42) \times 10^{-3}$	
	$D_1(2420)^- \to \bar{D}^*(2007)^0\pi^-$	$2/3$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(2.80 \pm 0.28) \times 10^{-3}$
Bd4	$B^0 \to D_1(2420)^-\mu^+\nu_\mu$	$(4.20 \pm 0.42) \times 10^{-3}$	
	$D_1(2420)^- \to \bar{D}^*(2010)^-\pi^0$	$1/3$	
	$\bar{D}^*(2010)^- \to \bar{D}^0\pi^-$	$(67.7 \pm 0.5)\%$	$(9.48 \pm 0.95) \times 10^{-4}$
Bd5	$B^0 \to D_1'(2430)^-\mu^+\nu_\mu$	$(4.7 \pm 1.4) \times 10^{-3}$	
	$D_1'(2430)^- \to \bar{D}^*(2007)^0\pi^-$	$2/3$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(3.1 \pm 0.9) \times 10^{-3}$
Bd6	$B^0 \to D_1'(2430)^-\mu^+\nu_\mu$	$(4.7 \pm 1.4) \times 10^{-3}$	
	$D_1'(2430)^- \to \bar{D}^*(2010)^-\pi^0\pi^-$	$1/3$	
	$\bar{D}^*(2010)^- \to \bar{D}^0\pi^-$	$(67.7 \pm 0.5)\%$	$(1.1 \pm 0.3) \times 10^{-3}$
Bd7	$B^0 \to D_2^*(2460)^-\mu^+\nu_\mu$	$(2.6 \pm 0.7) \times 10^{-3}$	
	$D_2^*(2460)^- \to \bar{D}^0\pi^-$	0.4590	$(1.2 \pm 0.3) \times 10^{-3}$
Bd8	$B^0 \to D_2^*(2460)^-\mu^+\nu_\mu$	$(2.6 \pm 0.7) \times 10^{-3}$	
	$D_2^*(2460)^- \to \bar{D}^*(2007)^0\pi^-$	0.2090	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\pi^0$	$(64.7 \pm 0.9)\%$	
	$\bar{D}^*(2007)^0 \to \bar{D}^0\gamma$	$(35.3 \pm 0.9)\%$	$(5.4 \pm 1.5) \times 10^{-4}$
Bd9	$B^0 \to D_2^*(2460)^-\mu^+\nu_\mu$	$(2.6 \pm 0.7) \times 10^{-3}$	
	$D_2^*(2460)^- \to \bar{D}^*(2010)^-\pi^0$	0.1030	
	$\bar{D}^*(2010)^- \to \bar{D}^0\pi^-$	$(67.7 \pm 0.5)\%$	$(1.8 \pm 0.4) \times 10^{-4}$
Bd10	$B^0 \to \bar{D}^0\mu^+\nu_\mu\pi^-$		
Bd11	$B^0 \to D^*(2010)^-\mu^+\nu_\mu\pi^0$		
	$D^*(2010)^- \to \bar{D}^0\pi^-$	$(67.7 \pm 0.5)\%$	
Bd12	$B^0 \to \bar{D}^*(2010)^-\tau^+\nu_\tau$	$(1.58 \pm 0.09)\%$	
	$\bar{D}^*(2010)^- \to \bar{D}^0\pi^-$	$(67.7 \pm 0.5)\%$	
	$\tau^+ \to \mu^+\bar{\nu_\mu}\nu_\tau$	$(17.39 \pm 0.04)\%$	$(1.86 \pm 0.11) \times 10^{-3}$
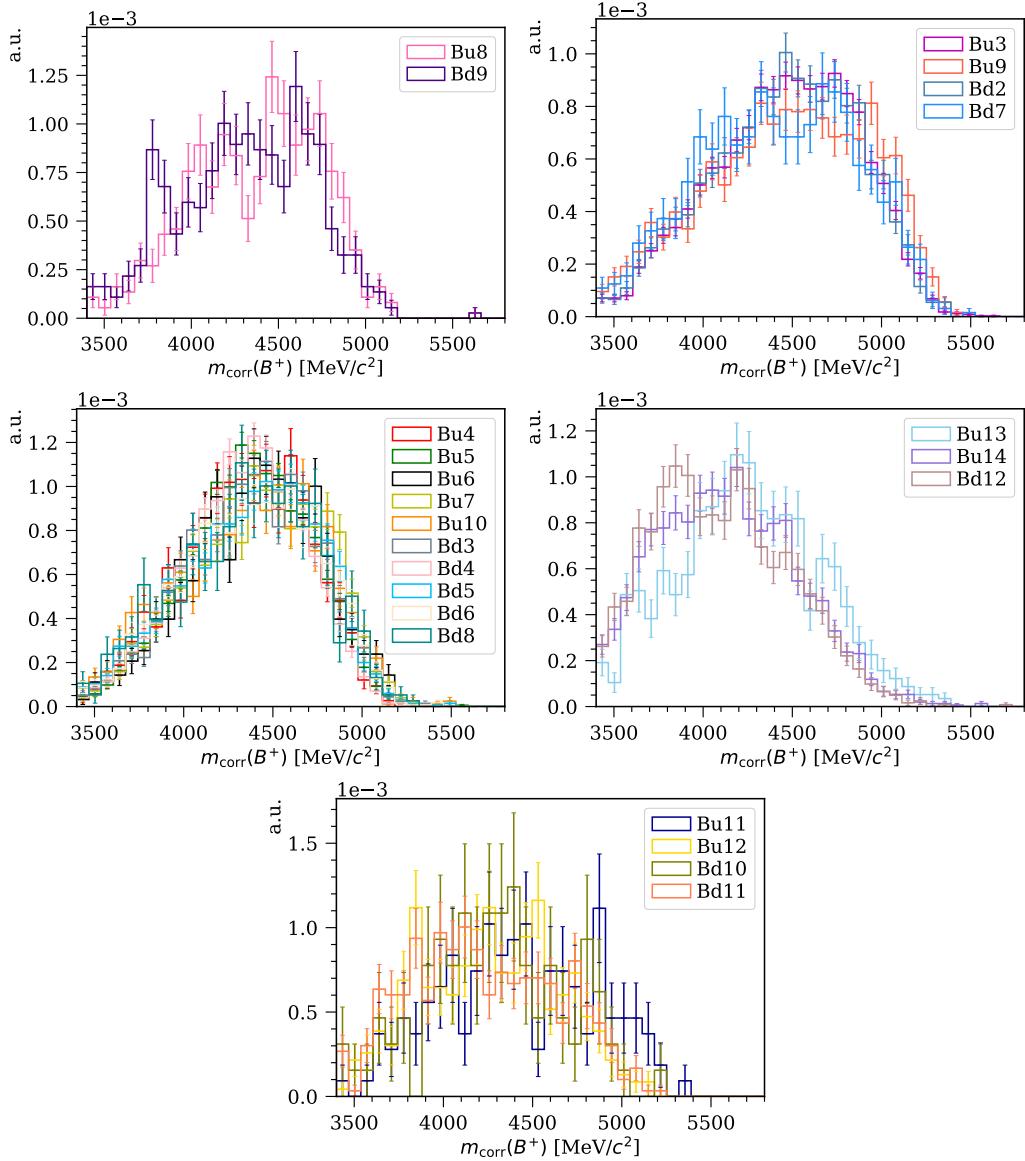
Figure 6.2 – Decay processes with similar $m_{\mathrm{corr}}(B^+)$ shapes are grouped together resulting in the templates GR1 (top left), GR2 (top right), GR3 (middle left), GR4 (middle right) and GR5 (bottom).

n_2 components, respectively, the relative contribution of the two can be determined by:

$$R_{\mathrm{T_1/T_2}} = \frac{\sum_i^{n_1} \mathcal{B}_i \times \epsilon_i}{\sum_j^{n_2} \mathcal{B}_j \times \epsilon_j} \, . \tag{6.6}$$

The relevant quantity for comparing template contributions is therefore $\sum_i^n \mathcal{B}_i \times \epsilon_i$, which is listed in Table 6.5 for each of the templates (except GR5 where the branching fraction is not known). The quoted uncertainty is computed by propagating[2] the branching fraction

---

[2]The possible correlations between the measured $\mathcal{B}s$ are not taken into account.
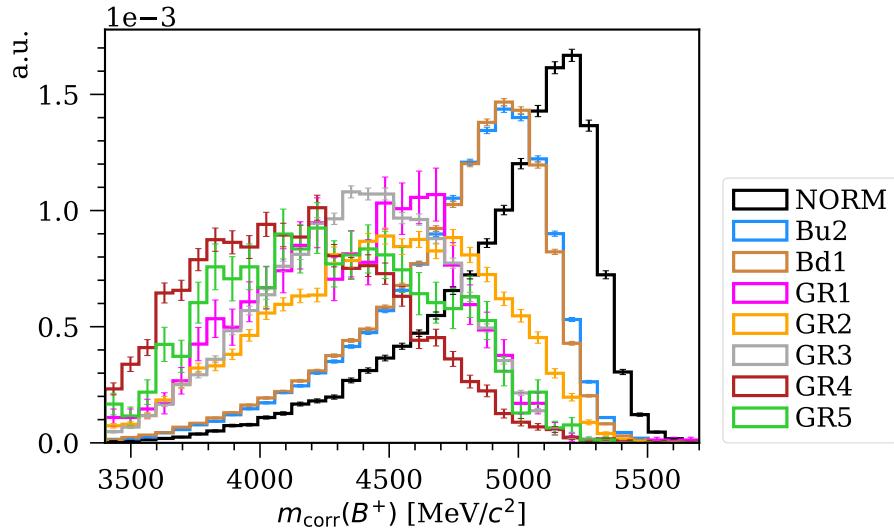
Figure 6.3 – Templates of $m_{\mathrm{corr}}(B)$ for the normalisation channel and physics backgrounds.

uncertainties of the template components listed in Tables 6.3 and 6.4. The uncertainties of the selection efficiencies are assumed to be negligible compared to the ones of the branching fractions and are therefore not included in the error propagation. The ratios of each template T with respect to the Bu2 template denoted $R_{\mathrm{T/Bu2}}$ are also listed. This shows that the expected contribution of the normalisation channel and Bd1 relative to the dominant component Bu2 is about 35.5% and 34.0%, respectively. For the remaining components GR1–GR4, the expected relative contributions are between 0.2–8.2%.

Table 6.5 – The relative contributions of templates are obtained by comparing the quantity $\sum_i^n \mathcal{B}_i \times \epsilon_i$ where $\mathcal{B}_i$ is the branching fraction of a template component, ϵ_i is its selection efficiency and n is the number of components making up the template. The ratio of each template T with respect to the Bu2 template is denoted $R_{\mathrm{T/Bu2}}$ and defined in Eq. 6.6.

Template (T)	$\sum_i \mathcal{B}_{\mathrm{tot,i}} \times \epsilon_i$	$R_{\mathrm{T/Bu2}}$
NORM	$(1.91 \pm 0.07) \times 10^{-5}$	$(3.55 \pm 0.20) \times 10^{-1}$
Bu2	$(5.38 \pm 0.23) \times 10^{-5}$	1.00
Bd1	$(3.08 \pm 0.13) \times 10^{-5}$	$(3.40 \pm 0.17) \times 10^{-1}$
GR1	$(1.33 \pm 0.25) \times 10^{-7}$	$(2.47 \pm 0.48) \times 10^{-3}$
GR2	$(2.09 \pm 0.33) \times 10^{-6}$	$(3.88 \pm 0.63) \times 10^{-2}$
GR3	$(4.41 \pm 0.35) \times 10^{-6}$	$(8.20 \pm 0.73) \times 10^{-2}$
GR4	$(1.89 \pm 0.18) \times 10^{-6}$	$(3.52 \pm 0.37) \times 10^{-2}$

6.2.2 Estimating combinatorial background

The combinatorial background consists of normalisation candidates formed by tracks originating from different processes. These are referred to as *random tracks*. Since the pions are not originating from $D^0 \to \pi^+\pi^-$ decays, the CombBg has a non-peaking and
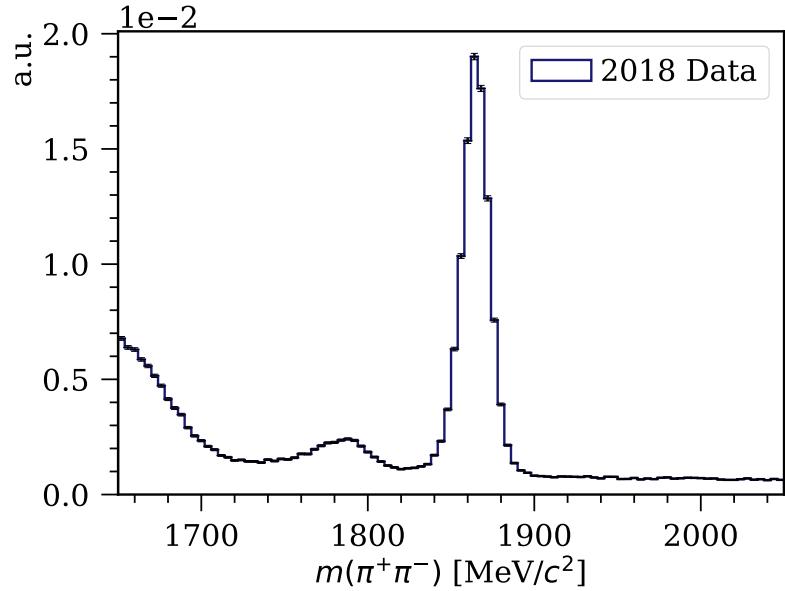
Figure 6.4 – Distribution of $m(\pi^+\pi^-)$ in data after the full selection. The structure at $m(\pi^+\pi^-) \sim 1790 \, \mathrm{MeV}/c^2$ is the $D^0 \to \pi K_{\mathrm{misID}\pi}$ background, while the one with $m(\pi^+\pi^-) < 1700 \, \mathrm{MeV}/c^2$ is associated with $D_s^+ \to \pi\pi\pi$ backgrounds.

Table 6.6 – Definitions of the D^0 peak region used to select normalisation candidates and the upper SB region used to model the CombBg.

Name	Range in $m(\pi^+\pi^-)$ $[\mathrm{MeV}/c^2]$
D^0 peak	$1835 < m(\pi^+\pi^-) < 1905$
Upper SB	$1905 < m(\pi^+\pi^-) < 1950$

approximately flat $m(\pi^+\pi^-)$ distribution. If the regions below and above the D^0 peak region, also known as *side bands* (SB), are approximately flat, they can be assumed to consist of combinatorial background only, and the normalisation candidates from these regions can be used to model the CombBg template.

However, as illustrated in Fig. 6.4 a structure is visible in the lower SB region of data at $m(\pi^+\pi^-) \sim 1790 \, \mathrm{MeV}/c^2$. This structure is associated with the $D^0 \to \pi K_{\mathrm{misID}\pi}$ background, while the structure at $m(\pi^+\pi^-) < 1700 \, \mathrm{MeV}/c^2$ is associated with $D_s^+ \to \pi\pi\pi$ backgrounds. Both backgrounds leak into the lower SB region, and consequently, only the upper SB region is used to model the CombBg, which results in the $m_{\mathrm{corr}}(B^+)$ template illustrated in Fig. 6.5. To minimise the contribution from the aforementioned backgrounds in the fit, the $m(\pi^+\pi^-)$ selection requirement is chosen to be more restrictive when $m(\pi^+\pi^-) < m(D^0)$ than for $m(\pi^+\pi^-) > m(D^0)$ leading to a slightly asymmetric selection window around the D^0 peak. The exact definitions of the upper SB and D^0 peak region are given in Table 6.6.

In order to estimate the expected yield of the CombBg in the fit, the distribution of
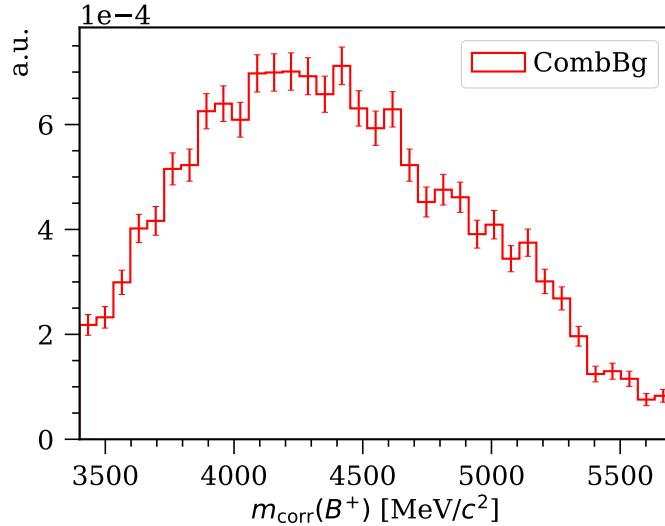
Figure 6.5 – $m_{\text{corr}}(B^+)$ distribution of $B^+ \to \bar{D}^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ candidates in the upper side band of the $\pi^+\pi^-$ mass, used as template for the combinatorial background.

$m(\pi^+\pi^-)$ in the upper SB is fitted with a linear function. By integrating the function over the D^0 mass peak region, the CombBg contribution can be estimated. Before applying the MVA selection, the CombBg constitutes approximately 34% of data in the D^0 peak region. To reduce this contribution the output variable of the DNN described in Sec. 4.5.4 is considered, however, since the distributions of some variables used to train the DNN, *e.g.* $m(\rho^0+1^{\text{st}}$ track), are different for the signal and for the normalisation channel, the selection efficiency of dnn$_{\text{out}}$ is reduced for the normalisation channel. Instead, the potential of other discriminating variables, *i.e.* the ghost probability of the π^+ track probGhost(π^+_{track}), $\chi^2_{\text{IP}}(\rho^0)$, dnn$_{\text{out}}$ and mva$_{\text{out}}(1^{\text{st}}$ track + vtx$_{\rho^0}$), is assessed in Fig. 6.6, where the selection efficiency of the normalisation channel is plotted against the rejection efficiency of the CombBg.

The best performing variable is found to be mva$_{\text{out}}(1^{\text{st}}$ track + vtx$_{\rho^0}$), which quantifies how likely the D^0 candidate is to have a third track originating from its decay vertex. This variable is a good discriminator against CombBg due to the fact that this background consists of normalisation candidates where the two $\pi^\pm$ tracks do not originate from the same $D^0 \to \pi^+\pi^-$ decay, but from various processes often involving the production of other charged particles that can be associated with the reconstructed $\mu^+\pi^+\pi^-$ vertex. By requiring mva$_{\text{out}}(1^{\text{st}}$ track + vtx$_{\rho^0}) < 0.7$ the combinatorial background is reduced to 13% of data in the D^0 peak region, while the normalisation channel is only reduced by 7%.

After applying the full selection defined in Table 6.1 to data, the fit to the upper SB region results in a linear regression curve with gradient $\alpha = -0.224 \pm 0.128$, interception $\beta = 621 \pm 247$ and covariance $cov(\alpha, \beta) = -31.6$. The CombBg yield, N_{CombBg}, is obtained from the integral of the D^0 peak region, as illustrated in Fig. 6.7. Assuming Poisson statistics, the statistical uncertainty of the yield is $\sigma_{\text{pois}} = \sqrt{N_{\text{CombBg}}}$. In addition to
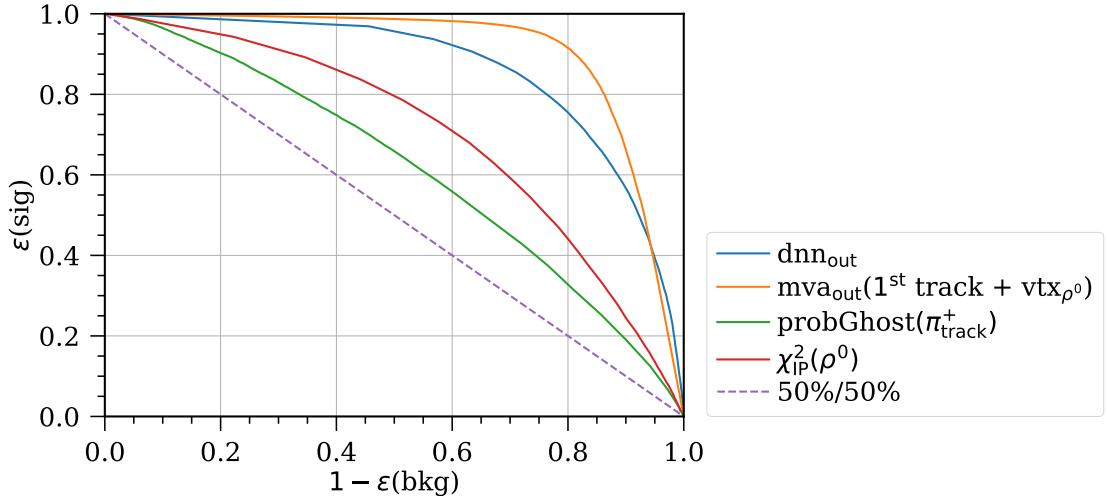
Figure 6.6 – Selection efficiency of the normalisation channel versus rejection efficiency of the combinatorial background. The best performing variable is $\mathrm{mva_{out}}(1^{\mathrm{st}}\ \mathrm{track} + \mathrm{vtx}_{\rho^0})$ (orange curve).

this, the uncertainty related to the fitting procedure, σ_{fit}, is estimated by propagating the uncertainties of the regression parameters in the integral. Finally, the total uncertainty of the CombBg yield is obtained by adding σ_{pois} and σ_{fit} in quadrature[3]. This results in the following estimate of the CombBg yield and its uncertainty:

$$N_{\mathrm{CombBg}} = 14083 \pm 565\,. \tag{6.7}$$

6.2.3   Estimating misID background

Despite an overall high performance of the particle identification at LHCb, a certain fraction of normalisation candidates are formed with a misidentified μ or π, also referred to as misIDi background with $i = \mu, \pi$. Candidates formed with more than one misidentified particle are assumed to be very rare and are therefore not considered.

For the normalisation channel, the contribution of misIDπ background is the same in the D^0 peak region as in the upper SB region, and it is therefore accounted for by the CombBg template. However, the misIDμ background is composed of two different components, *i.e.* one where the misidentified μ^+ is combined with two random $\pi^\pm$ tracks (misID$\mu^{\pi^\pm}$), and another where the misidentified μ^+ is combined with a $D^0 \to \pi^+\pi^-$ decay (misIDμ^{D^0}). The contribution of misID$\mu^{\pi^\pm}$ is, like for the misIDπ background, the same in the D^0 peak region as in the upper SB region. Thus, this component is accounted for by the CombBg template. However, the contribution of misIDμ^{D^0} is limited to the D^0 peak region, where

---

[3]This assumes zero correlation between σ_{pois} and σ_{fit}, which is not completely correct. The uncertainty is therefore slightly over or underestimated, however, the affect on the final fit result is assumed to be negligible.
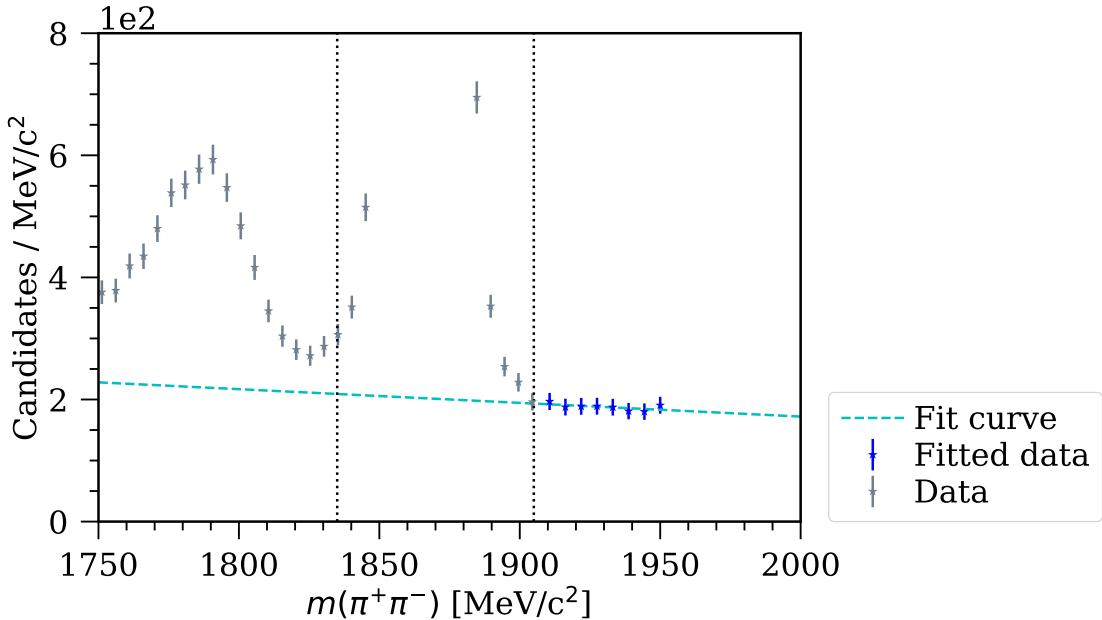
Figure 6.7 – The yield of the CombBg in the D^0 peak region (dashed vertical lines) is estimated by integrating the linear model (dashed cyan line) fitted to the data in the upper SB (blue markers).

both components are present. Therefore, to estimate the misIDμ^{D^0} background in this region, the misID$\mu^{\pi^\pm}$ contribution to the CombBg template is subtracted from the total misIDμ^{tot} contribution in the D^0 peak region.

Formulating the misID problem

The misIDμ background is expected to consist of π, K and p, while the contribution from electrons (that never reach the Muon System) is negligible. The contribution from ghosts is also found to be negligible based on the studies in Sec. 7.1.4. Since the measured μ candidates in data are composed by true μ, π, K and p, one can write the number of measured μ candidates as linear combination of the number of true particles, *i.e.*:

$$N_{\mu_{\text{meas}}} = \epsilon_{\mu_\mu} N_{\mu_{\text{true}}} + \epsilon_{\mu_\pi} N_{\pi_{\text{true}}} + \epsilon_{\mu_K} N_{K_{\text{true}}} + \epsilon_{\mu_p} N_{p_{\text{true}}}, \tag{6.8}$$

where the coefficient ϵ_{i_j} is the efficiency for a true particle j to be measured as a particle i, and $N_{j_{\text{true}}}$ is the number of true particle j in data. However, a priori, neither the efficiencies nor the number of true particles are known.

With the PIDCALIB package [153] it is possible to estimate the efficiencies using data driven methods, which will be further explained in the next paragraph. Assuming efficiencies, there are still four unknowns in Eq. 6.8. In order to solve this, three equations expressing $N_{j_{\text{meas}}}$ with $j = \pi, K, p$ as linear combinations of the corresponding true particles are constructed in addition to Eq. 6.8. Then the number of measured particle species in $\vec{N}_{\text{meas}}$

Table 6.7 – Definitions of the particle regions defined with PID requirements. In addition to the PID requirements, the muon is required to be within the acceptance of the Muon System, and the number of hits in the SPD detector is required to be `nSPDHits < 450`.

Particle	Region
μ	`IsMuon = 1, PIDmu > 3`
π	`ProbNNpi > 0.1, PIDK < 0, PIDp < 0, IsMuon ≠ 1, PIDmu < 3`
K	`ProbNNK > 0.1, PIDK > 0, (PIDp − PIDK) < 0, IsMuon ≠ 1, PIDmu < 3`
p	`ProbNNp > 0.1, PIDp > 0, (PIDp − PIDK) > 2, IsMuon ≠ 1, PIDmu < 3`

can be related to the number of true particle species in $\vec{N}_{\text{true}}$ by the efficiency matrix $\mathbf{E}$, *i.e.*:

$$
\begin{pmatrix} N_{\mu_{\text{meas}}} \\ N_{\pi_{\text{meas}}} \\ N_{K_{\text{meas}}} \\ N_{p_{\text{meas}}} \end{pmatrix} = \begin{pmatrix} \epsilon_{\mu_\mu} & \epsilon_{\mu_\pi} & \epsilon_{\mu_K} & \epsilon_{\mu_p} \\ \epsilon_{\pi_\mu} & \epsilon_{\pi_\pi} & \epsilon_{\pi_K} & \epsilon_{\pi_p} \\ \epsilon_{K_\mu} & \epsilon_{K_\pi} & \epsilon_{K_K} & \epsilon_{K_p} \\ \epsilon_{p_\mu} & \epsilon_{p_\pi} & \epsilon_{p_K} & \epsilon_{p_p} \end{pmatrix} \begin{pmatrix} N_{\mu_{\text{true}}} \\ N_{\pi_{\text{true}}} \\ N_{K_{\text{true}}} \\ N_{p_{\text{true}}} \end{pmatrix} .
\tag{6.9}
$$

Since the goal is to determine the true particle composition of the measured μ candidates, the above matrix equation has to be inverted, *i.e.*:

$$
\vec{N}_{\text{true}} = \mathbf{E}^{-1} \vec{N}_{\text{meas}} .
\tag{6.10}
$$

Knowing $\vec{N}_{\text{true}}$ the number of μ candidates with true particle identity j is given by:

$$
N_{j_{\text{misID}\mu}} = \epsilon_{\mu_j} N_{j_{\text{true}}} .
\tag{6.11}
$$

Before PIDCALIB can be used to compute ϵ_{i_j} and thereby the efficiency matrix in Eq. 6.9 can be populated, requirements for a particle to be measured as a μ, π, K or p candidate have to be defined. To do this, PID information is used. Four sets of PID requirements, each favouring the selection of one particle species, are defined in Table 6.7. These are referred to as particle regions.

The μ region is defined by the PID requirements imposed on μ candidates by the `B2XuMuNuBu2Rho_Line` stripping line (see Table 4.3). All particle regions are disjoint and favour the selection of their respective particles.

Computing efficiencies with PIDCalib

The PIDCALIB package provides access to data *calibration samples* with particles of one known species. These samples are obtained with the *sPlot* technique described in Sec. 4.6.2 and without any use of PID information.

Table 6.8 – Binning of the variables p, η and `nSPDHits` used to compute PID efficiencies.

Variable	Standard binning scheme
p	$[6100, 29575, 53050, 76525, 100000, 330500]$ MeV/c
η	$[1.7, 2.5, 3.3, 4.1, 4.9]$
`nSPDHits`	$[0, 150, 300, 450]$
Variable	Small binning scheme
p	$[6100, 53050, 100000, 330500]$ MeV/c
η	$[1.7, 3.3, 4.9]$
`nSPDHits`	$[0, 225, 450]$

Applying the PID requirements of the particle regions defined in Table 6.7 to the calibration samples, the efficiencies for any true particles to be measured as any of the four particle candidates are obtained. However, the PID efficiencies of the calibration samples are not necessarily the same as for the so-called *reference sample* for which the misID contribution has to be determined. The reason for this is that the PID efficiency depends on the track kinematics, pseudorapidity and the event multiplicity, and these properties are most likely different for the calibration and reference samples.

However, the PID efficiency can be very well parameterised by the track p, η and number of hits in the SPD detector, `nSPDHits`. Thus, for the reference sample, the average efficiency, $\bar{\epsilon}$, of a PID requirement is obtained by assigning the calibration sample efficiency, ϵ_i, of the ith bin defined by p, η and `nSPDHits` to reference tracks belonging to the same bin [153]:

$$\bar{\epsilon} = \frac{1}{R} \sum_i \epsilon_i R_i \,, \tag{6.12}$$

where R_i is the number of reference tracks in bin i and R is the total number of tracks in the reference sample.

The binning in p, η and `nSPDHits` should be defined in such a way that changes in the PID efficiency is accurately described for a given requirement. The binning should therefore aim for constant PID efficiency in each bin, however, it should also aim at having sufficient number of events per bin. In this analysis, two binning schemes are used to accommodate the large differences in PID efficiencies for the different combinations of particle regions and calibration samples. The *standard binning* scheme has a slightly finer binning than the *small binning* scheme as defined in Table 6.8. Using these two binning schemes, all bins of the calibration samples are populated with tracks, and the PID efficiency can therefore be computed in all bins. Moreover, the binning ensures that at least 99% of the tracks in the reference sample is within the bin limits, and they can therefore be assigned a PID efficiency.

In order to compute the efficiency matrix in Eq. 6.9, the non-muonic candidates π_{meas},

K_{meas} and p_{meas} must be present in the reference sample. For this reason, the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line with no PID requirements on the μ candidate is therefore used (see Table 4.2). Moreover, since the `muplus_L0MuonDecision_TOS` trigger line is correlated with the `IsMuon` requirement, the trigger selection is not aplied to the reference sample. Besides this, the selection applied to the reference sample is the same as the one used to select normalisation candidates, as described in Sec. 6.1. Using the particle regions defined in Table 6.7 and the binning scheme in Table 6.8, the efficiencies ϵ_{i_j} are computed with PIDCALIB in the D^0 peak region for the estimation of the total misIDμ^{tot} contribution, and in the upper SB region for the estimation of the misID$\mu^{\pi^{\pm}}$ contribution. This results in the two efficiency matrices:

$$\mathbf{E}^{\mathrm{D^0peak}} = \begin{pmatrix} 0.9896 & 0.0030 & 0.0067 & 0.0007 \\ 0.0051 & 0.9475 & 0.0378 & 0.0096 \\ 0.0007 & 0.0607 & 0.9072 & 0.0315 \\ 0.0012 & 0.0215 & 0.0313 & 0.9461 \end{pmatrix}, \tag{6.13}$$

$$\mathbf{E}^{\mathrm{SB}} = \begin{pmatrix} 0.9893 & 0.0031 & 0.0068 & 0.0008 \\ 0.0050 & 0.9515 & 0.0339 & 0.0095 \\ 0.0007 & 0.0547 & 0.9123 & 0.0324 \\ 0.0012 & 0.0219 & 0.0317 & 0.9453 \end{pmatrix}. \tag{6.14}$$

This confirms that the particle regions favour the selection of their associated particle, since the efficiencies on the diagonal are large (91%–99%) compared to the off-diagonal efficiencies between 0.07%–6%.

The number of measured μ, π, K and p candidates in the D^0 peak and SB region of the reference sample are given by:

$$\vec{N}^{\mathrm{D^0peak}}_{\text{meas}} = \begin{pmatrix} 2608 \\ 6427 \\ 5710 \\ 803 \end{pmatrix}, \qquad \vec{N}^{\mathrm{SB}}_{\text{meas}} = \begin{pmatrix} 213 \\ 3154 \\ 3413 \\ 439 \end{pmatrix}. \tag{6.15}$$

Using these and the computed efficiency matrices in Eqs. 6.13 and 6.14, the corresponding true particles can be found by solving Eq. 6.10. However, due to the statistical and systematic uncertainties associated with determining the elements of $\mathbf{E}$, the matrix is, in general, not invertible. To overcome this problem, the ROOUNFOLD package [175] is used.

Unfolding the misID background with RooUnfold

The RooUnfold package provides various unfolding algorithms. For this analysis, the D'Agostini method [176], *i.e.* the RooUnfoldBayes algorithm, is used.

The goal is to unfold the measured particle vector $\vec{N}_{\text{meas}}$ to obtain the underlying true particle vector $\vec{N}_{\text{true}}$. Finally, this can be used together with the efficiency matrix $\mathbf{E}$ to estimate the contribution from misID background by using Eq. 6.11.

The matrix $\mathbf{E}$ can be seen as the *folding matrix* since multiplying it with $\vec{N}_{\text{true}}$ gives $\vec{N}_{\text{meas}}$, *i.e.* the matrix $\mathbf{E}$ folds the true distribution. To perform the reverse action, the inverse matrix $\mathbf{E^{-1}}$ is required. However, as previously stated, measurement uncertainties result in $\mathbf{E}$ not being invertible. Therefore one has to estimate the inverse matrix $\tilde{\mathbf{E}}$, which is done by the D'Agostini method.

The estimated number $\hat{N}_{j_{\text{true}}}$ of true particles j is given by:

$$\hat{N}_{j_{\text{true}}} = \frac{1}{\epsilon_j} \sum_i \tilde{\epsilon}_{ij} N_{i_{\text{meas}}} \tag{6.16}$$

where $\epsilon_j = \sum_i \epsilon_{ij}$ is total efficiency of the true particle j, and $N_{i_{\text{meas}}}$ is the number of measured particles i. The inverse matrix element $\tilde{\epsilon}_{ij}$ can be written as the conditional probability following Bayes theorem [176]:

$$\tilde{\epsilon}_{ij} = P(j|i) \tag{6.17}$$
$$= \frac{P(i|j)P(j)}{P(i)}, \tag{6.18}$$

where $P(j|i)$ is the probability of having a true particle j given the measured particle i, $P(i|j)$ is the probability of measuring a particle i given the true particle j, $P(j)$ and $P(i)$ are the probabilities of having a true particle j and measuring particle i, respectively. Using that ϵ_{ij} can be understood as a conditional probability to measure a particle as i given that its true identity is j corresponding to $\epsilon_{ij} = P(i|j)$, one can rewrite Eq. 6.18 as:

$$\tilde{\epsilon}_{ij} = \frac{\epsilon_{ij}P(j)}{\sum_{j'} \epsilon_{ij'} P(j')}, \tag{6.19}$$

where $P(j')$ are the unknown true particle probabilities.

The basic idea behind Bayes unfolding is to start with the best initial guess of the true particle spectra. In case of complete ignorance, a uniform probability of $P(j)^0 = 1/N_{\text{bins}}$, where N_{bins} is the number of particle species, is used. This leads to an initial expected number of true particles j by $\hat{N}_{j_{\text{true}}}^0 = P(j)^0 N_{\text{meas}}$. The initial guess $P(j)^0$ is used in Eq. 6.19 to get the first estimate of $\tilde{\epsilon}_{ij}$, which can be used to estimate the number of

Table 6.9 – True composition of the measured μ in the D^0 peak and SB region of the reference sample.

Measured μ	true μ	true π	true K	true p	total misID
D^0 peak region	2544.9	20.3	38.3	0.4	2.27%
SB region	183.9	10.2	23.6	0.25	15.66%

true particles $\hat{N}_{j_{\text{true}}}$ in Eq. 6.16. This is then used to estimate the total number of true particles $\hat{N}_{\text{true}} = \sum_j \hat{N}_{j_{\text{true}}}$ and the true spectrum $\hat{P}(j) = \hat{N}_{j_{\text{true}}}/\hat{N}_{\text{true}}$. A χ^2 comparison between the new $\hat{N}_{j_{\text{true}}}$ and the initial guess $\hat{N}^0_{j_{\text{true}}}$ is performed. This procedure is repeated where $P(j)^0$ is replaced by the updated estimate of $\hat{P}(j)$, and the iteration is stopped when the χ^2 value is small enough. The Bayes unfolding method has been improved as described in Ref. [177], however, the basic ideas and assumptions are substantially unchanged. The RooUnfold package accounts for propagation of error through multiple iterations. Regularization is achieved by stopping the iterations before reaching the true inverse solution, which will suffer from high fluctuations.

Applying the Baysian unfolding algorithm with the efficiency matrices in Eq. 6.13 and Eq. 6.14 together with the measured particle compositions in Eq. 6.15, the true particle composition of the reference sample in the D^0 peak and SB regions are found to be:

$$\vec{N}_{\text{true}}^{D^0 \text{peak}} = \begin{pmatrix} 2572 \\ 6724 \\ 5695 \\ 557 \end{pmatrix}, \qquad \vec{N}_{\text{true}}^{\text{SB}} = \begin{pmatrix} 186 \\ 3272 \\ 3456 \\ 305 \end{pmatrix}, \tag{6.20}$$

The true compositions of measured μ candidates in the D^0 peak and SB region are obtained with Eq. 6.11 and summarised in Table 6.9.

In order to obtain template shapes of the misIDμ backgrounds, the true composition of the measured μ candidates is computed in bins of $m_{\text{corr}}(B^+)$. The resulting distributions of misIDμ^{tot} and misID$\mu^{\pi^\pm}$ are illustrated in Fig. 6.8. The compositions of misID backgrounds are similar in the two regions.

The final misIDμ^{D^0} template used in the fit of the normalisation channel is obtained by subtracting the misID$\mu^{\pi^\pm}$ template (right in Fig. 6.8) from the misIDμ^{tot} template (left in Fig. 6.8) However, in order to do this, the two misID contributions must be correctly scaled. The total yield of misIDμ^{tot} is 2.27% of the fitted data corresponding to $\sim 2414 \pm 7$ candidates. The yield of the misID$\mu^{\pi^\pm}$ in the CombBg template is 15.66% of the expected template yield computed in Sec. 6.2.2 resulting in $\sim 2204 \pm 87$ candidates. Consequently, the yield of the final misIDμ^{D^0} template is:

$$N_{\text{misID}\mu^{\text{D}^0}} = 210 \pm 87, \tag{6.21}$$
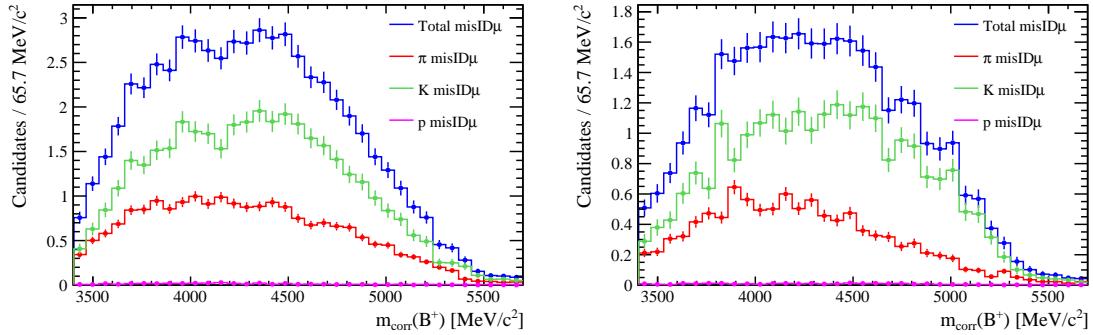
Figure 6.8 – Templates for the total misIDμ background (blue) consisting of true π (red), K (green) and p (pink) misidentified as μ for the D^0 peak (left) and SB (right) denoted misIDμ^{tot} and misID$\mu^{\pi^{\pm}}$, respectively.
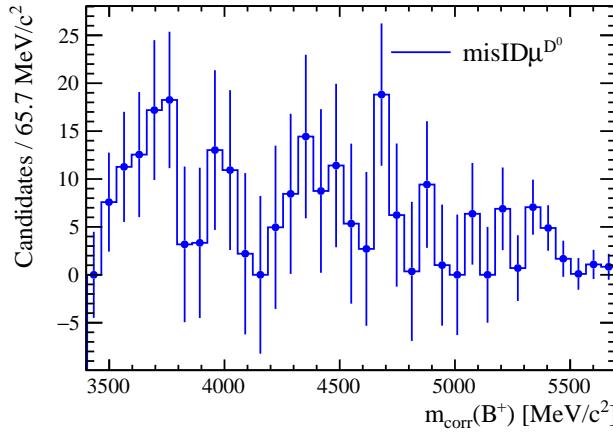
Figure 6.9 – MisID template resulting from subtracting the misID contribution of the CombBg from the total misID contribution under the D^0 peak.

where the uncertainties associated with misIDμ^{tot} and misID$\mu^{\pi^{\pm}}$ are added in quadrature to estimate the uncertainty of $N_{\text{misID}\mu^{D^0}}$. The $m_{\text{corr}}(B^+)$ template after subtracting the correctly scaled templates is illustrated in Fig. 6.9.

6.3   Template fitting framework

The yield of the normalisation channel is extracted from an extended maximum likelihood fit to $m_{\text{corr}}(B^+)$. However, as previously mentioned, the distributions of $m_{\text{corr}}(B^+)$ cannot be described analytically. Instead, histograms obtained from simulations, as described in Sec. 6.2.1, or data driven methods, as described in Sec. 6.2.2 and Sec. 6.2.3, are used to model the shape of $m_{\text{corr}}(B^+)$ for all fit components.

To perform the fit with histogram templates, the dedicated HISTFACTORY toolkit [178] based on the ROOFIT/ROOSTATS framework [165] is applied. Parameterised PDFs are

created from the histograms and are used to build a negative log likelihood function which is minimised with the MINUIT package [166].

To account for the uncertainty related to the finite statistics of the templates, HISTFACTORY uses a modified version of the Beeston-Barlow method [179], also known as the *Beeston-Barlow light* method. The original Beeston-Barlow method predicts the number of data events in bin i to be

$$f_i = N_D \sum_{j=1}^{m} \frac{P_j A_{ji}}{N_j} \,, \tag{6.22}$$

where N_D is the total number of data events, N_j is the number of events in the template of component j and P_j is the relative proportion of component j with $\sum_{j}^{m} P_j = 1$. The uncertainty associated with the limited template statistics is contained in A_{ji}. This is the unknown expected number of events in the template for each component j in bin i. The observed number of events in template j in bin i denoted a_{ji} is generated from A_{ji}, which can be assumed to follow a Poisson distribution. The expected number of events in a bin can therefore be expressed as $f_i = \sum_{j=1}^{m} p_j A_{ji}$ where $p_j = N_D P_j / N_j$.

The likelihood to be maximised is the combined probability of the number of observed data events d_i and observed events a_{ji} in template j:

$$\ln(\mathcal{L}) = \sum_{i=1}^{n} d_i \ln(f_i) - f_i + \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ji} \ln(A_{ji}) - A_{ji} \,. \tag{6.23}$$

This results in a estimation of p_j and A_{ji}. However, to accomplish this, one has to deal with a maximisation problem of $m \times (n+1)$ unknowns. Although solvable, it is often computationally demanding leading to slowly converging fits.

For this reason HISTFACTORY uses the so-called *Beeston-Barlow light* method [178]. Instead of considering the nuisance parameter A_{ji} for each template in each bin, only one single nuisance parameter γ_i per bin is considered. This parameter accounts for the total template estimate and the total statistical uncertainty in that bin. An analytic solution to γ_i is found by solving a simple quadratic expression, which is included in the total likelihood. Thus, in each bin the combined model is assigned a statistical uncertainty based on the sum of the relative uncertainties from each template and each bin can then fluctuate around its expected value by a fractional amount γ_i.

Based on external information obtained from previous measurements or theoretical predictions, the overall normalisation of templates can be constrained. This can be useful, *e.g.* when the shapes of two templates, T_1 and T_2, are not distinguishable by the fit. Given the branching fractions and selection efficiencies of the processes making up the two templates, their relative contribution, $R \equiv R_{T_1/T_2}^{\text{expect}}$, defined in Eq. 6.6, can be used to create a constraint term that is multiplied with the total likelihood function. The measured

relative contribution, $r \equiv R^{\mathrm{meas}}_{\mathrm{T}_1/\mathrm{T}_2}$, can be constrained by [178]:

$$G(r|R, \sigma_R) = \frac{1}{\sqrt{2\pi}\sigma_R} e^{-\frac{(r-R)^2}{2\sigma_R^2}} \ , \tag{6.24}$$

where the uncertainty σ_R is associated with the external measurements.

External information related to the shape uncertainties of the templates can also be accounted for by HISTFACTORY. Usually, these are caused by imperfect theoretical predictions, for instance, of form factors which affect the shape of the templates. In practice, two templates are computed corresponding to a $\pm 1\sigma$ parametric shape uncertainty, respectively. These are then used by HISTFACTORY to allow for correlated and coherent bin-by-bin changes in the template shape. The shape variation is included in the total likelihood as a constraint term.

6.4   Normalisation channel fit

To extract the yield of the normalisation channel a binned maximum likelihood fit to $m_{\mathrm{corr}}(B^+)$ of the full 2018 data sample is performed with HISTFACTORY.

The fit model is build with ten different templates:

- The normalisation channel (NORM) obtained from simulation, as described in Sec. 6.2.1.

- Simulated physics backgrounds (Bu2, Bd1 and GRi with i in the range 0–5), consisting of the decays $B^+ \rightarrow \bar{D}^{(*,**)0}\mu^+\nu_\mu(X)$ and $B^0 \rightarrow D^{(*,**)-}\mu^+\nu_\mu(X)$ grouped into templates based on their $m_{\mathrm{corr}}(B^+)$ shapes, as described in Sec. 6.2.1.

- Combinatorial background (CombBg) modelled with the upper SB of the D^0 peak, as described in Sec. 6.2.2.

- Misidentified μ candidates combined with $D^0 \rightarrow \pi^+\pi^-$ decays (misIDμ) modelled with data driven methods relying on the PIDCALIB and ROOUNFOLD tools, as described in Sec. 6.2.3.

All templates based on simulations are re-weighted with the data/MC weights $w_{\mathtt{nTracks}} \times w_{\mathrm{mva}}$ computed in Sec. 4.6.3. Although these weights are developed for the signal channel, the mis-modelling is expected to be similar for all simulations. In addition to this, the simulations of the normalisation channel and the $B^+ \rightarrow \bar{D}^{*0}\mu^+\nu_\mu$ decay are corrected to the CLN form factor parameterisation, as described in Sec. 6.2.1. While the effect on the Bu2 template is negligible, there is a small, however, non-negligible effect on the normalisation channel template. The shape uncertainty of this correction is estimated by generating the true q^2 distribution of the normalisation channel with $\rho^2 \pm 1\sigma$. Thus,

Table 6.10 – Summary of the fit components and constraints. The relative normalisation of template T with respect to Bu2 as defined in Eq. 6.6 is denoted $R_{\mathrm{T/Bu2}}$, and the yields of the CombBg and misIDμ background are denoted N_{CombBg} and $N_{\mathrm{misID}\mu}$.

Fit components (T)	Constraints
NORM	Free floating yield + shape uncertainty
Bu2	Used in Bd1 and GR1,2,3,4 constraints
Bd1	$R_{\mathrm{Bd1/Bu2}} = (3.40 \pm 0.17) \times 10^{-1}$
GR1	$R_{\mathrm{GR1/Bu2}} = (2.47 \pm 0.48) \times 10^{-3}$
GR2	$R_{\mathrm{GR2/Bu2}} = (3.88 \pm 0.63) \times 10^{-2}$
GR3	$R_{\mathrm{GR3/Bu2}} = (8.20 \pm 0.73) \times 10^{-2}$
GR4	$R_{\mathrm{GR4/Bu2}} = (3.52 \pm 0.37) \times 10^{-2}$
GR5	Free floating yield
CombBg	$N_{\mathrm{CombBg}} = 14083 \pm 565$
MisIDμ	$N_{\mathrm{misID}\mu} = 210 \pm 87$

the $\pm 1\sigma$ parametric shape uncertainty corresponds to the normalisation channel template corrected to the q^2 distributions based on $\rho^2 \pm 1\sigma$.

The CombBg (see Sec. 6.2.2) and misID background (see Sec. 6.2.3) yields are Gaussianly constrained in the fit. The uncertainties on their estimates are used to define the $\pm 1\sigma$ variation of the constraint. The contributions from GR1–GR4 to the fit are small with respect to the dominant processes (see Sec. 6.2.1) and their template shapes are not well distinguished by the fit. For this reason, the relative contributions $R_{\mathrm{GR}i/\mathrm{Bu2}}$ of GR1–GR4 with respect to Bu2, given in Table 6.5, are Gaussianly constrained in the fit. The corresponding shape variation of $\pm 1\sigma$ is defined by propagating the uncertainties of the measured branching fractions used to compute $R_{\mathrm{GR}i/\mathrm{Bu2}}$ in Eq. 6.6. In addition to this, the fit cannot properly distinguish the Bu2 and Bd1 templates, and thus, their relative contribution is also Gaussianly constrained in the fit. Since the nonresonant $B \to D^{(*)}\pi\mu^+\nu_\mu$ decays have not been measured, the GR5 template is floating freely in the fit. All templates and Gaussian constraints of the fit are summarised in Table 6.10.

As described in Sec. 6.1 the final selection requirement is optimised while fitting the normalisation channel. The figure of merit is the relative uncertainty of the yield of the normalisation channel. Since any requirement on dnn$_{\mathrm{out}}$ is found to increase the relative uncertainty of the extracted yield, it is not used in the final requirement. For $\sigma_{m_{\mathrm{corr}}}(B^+)$ any cut harder than $< 150\,\mathrm{MeV}/c^2$ is also found to increase the relative uncertainty of the extracted yield, and the final requirement is therefore $\sigma_{m_{\mathrm{corr}}}(B^+) < 150\,\mathrm{MeV}/c^2$. The fit is performed under different requirements on mva$_{\mathrm{out}}$(1$^{\mathrm{st}}$ track + vtx$_{\rho^0}$) and each time the constraints listed in Table 6.10 are adjusted to account for changes in selection efficiencies and expected yields of the CombBg and misIDμ backgrounds. The relative uncertainty is improved by cutting on mva$_{\mathrm{out}}$(1$^{\mathrm{st}}$ track + vtx$_{\rho^0}$), and in general, the fit and relative uncertainty of the yield is rather stable between mva$_{\mathrm{out}}$(1$^{\mathrm{st}}$ track + vtx$_{\rho^0}$) < 0.9 and mva$_{\mathrm{out}}$(1$^{\mathrm{st}}$ track + vtx$_{\rho^0}$) < 0.5. A requirement in the middle corresponding to mva$_{\mathrm{out}}$(1$^{\mathrm{st}}$
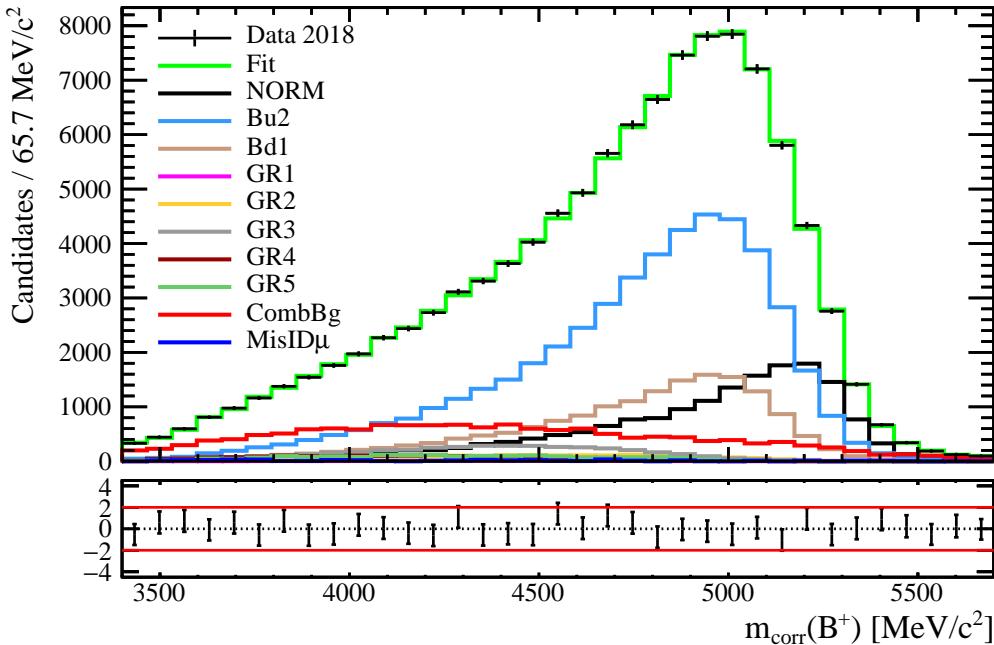
Figure 6.10 – $m_{\mathrm{corr}}(B^+)$ distribution of the template fit (green) of the normalisation channel and backgrounds in 2018 data (black markers). The main components are the normalisation channel (black), the $B^+ \to \bar{D}^{*0}\mu^+\nu_\mu$ (blue) and $B^0 \to D^{*-}\mu^+\nu_\mu$ (brown) decays and combinatorial background (red).

track + vtx$_{\rho^0}$) < 0.7 is chosen, which also significantly reduces the CombBg in the fit as described in Sec. 6.1.

The final fit of the normalisation channel is performed with the Beeston-Barlow light treatment of bin-by-bin statistical uncertainties for all templates. The result of the fit is illustrated in Fig. 6.10, and summarised in Table 6.11. The main components of the fit are the NORM, Bu2, Bd1 and CombBg templates. The yield of the normalisation channel is found to be 16957 ± 560 corresponding to a relative statistical uncertainty of 3.3%. The fit model describes data well and the constrained templates are all within $\pm 1\sigma$. Moreover, the fit finds a relative contribution of the normalisation channel with respect to Bu2 of $R_{\mathrm{NORM/Bu2}} = (3.50 \pm 0.2) \times 10^{-1}$ that is within one standard deviation from the expected relative contribution given in Table 6.5.

6.5   Measuring the q^2 distribution

To validate the normalisation channel fit in Sec. 6.4, the q^2 distributions of the normalisation channel and the $B^+ \to D^{*0}\mu^+\nu_\mu$ decay are measured and compared to the expected q^2 distributions based on the well measured CLN parameters [57], as described in Sec. 6.2.1.

Table 6.11 – Normalisation channel fit result. The yield of template T obtained with the fit is $N_\mathrm{T} \pm \sigma_{N_T}$ and $\Delta\sigma_\mathrm{C}/\sigma_\mathrm{C}$ is the relative shift of the imposed constraint. For the NORM template $\Delta\sigma_\mathrm{C}/\sigma_\mathrm{C}$ is the relative shift in the shape uncertainty caused by the CLN form factor correction.

Fit component (T)	N_T	σ_{N_T}/N_T	$\Delta\sigma_\mathrm{C}/\sigma_\mathrm{C}$
NORM	16957 ± 560	3.3%	-0.01
Bu2	48471 ± 824	1.7%	
Bd1	16459 ± 793	4.8%	-0.03
GR1	120 ± 23	19.3%	$+0.02$
GR2	1937 ± 297	15.3%	$+0.18$
GR3	4055 ± 334	8.2%	$+0.22$
GR4	1732 ± 175	10.1%	$+0.15$
GR5	1846 ± 725	39.3%	
CombBg	14527 ± 517	3.6%	$+0.20$
MisIDμ	402 ± 135	33.7%	$+0.59$

6.5.1 Fitting the normalisation channel in q^2 bins

The normalisation channel fit described in Sec. 6.4 is performed in bins of q^2_reg. To validate the q^2 regression method described in Sec. 5.4, the first three bins correspond to the q^2 binning scheme used for the signal channel, as defined in Sec. 5.5. However, for the upper end of the spectrum, the signal binning scheme is not adequate. While the contribution from the normalisation channel decreases with q^2, the contributions from the dominant backgrounds (Bu2 and Bd1) increase with q^2. Thus, in high q^2 bins the fit does not distinguish well between the normalisation channel and these backgrounds. To overcome this problem, the fourth to eight bins of the signal binning scheme are merged into one single bin. The final binning scheme consists of four bins with bin boundaries defined in Table 6.12.

The same approach used to fit the normalisation channel in the full q^2 region is used to perform the fit in the four q^2 bins, *i.e.* the selection defined in Table 6.1 is applied in each bin, the same templates are constrained, however, for each bin all constraints are updated[4], and the Beeston-Barlow light treatment is applied to all templates. The resulting fits in each q^2 bin are illustrated in Fig. 6.11 and the extracted yields of the normalisation channel and the $B^+ \to D^{*0}\mu^+\nu_\mu$ decay are summarised in Table 6.12.

The fit model describes data well in all four bins. The relative uncertainty of the extracted yield increases from 3.4% in the first bin to 9.6% in the last bin. In contrast to this, the relative uncertainty of the yield of Bu2 is between 2.3% and 3.3% in all four bins. Adding the extracted yields for each bin results in 6989 ± 673 for the normalisation channel and 48460 ± 793 for Bu2. Both yields are in agreement with the ones extracted from the fit to

---

[4]For each bin the $R_\mathrm{T/Bu2}$ is recomputed to account for changes in the efficiency, the CombBg yield is re-estimated by fitting the upper SB and the misIDμ yield and template shape is re-computed.
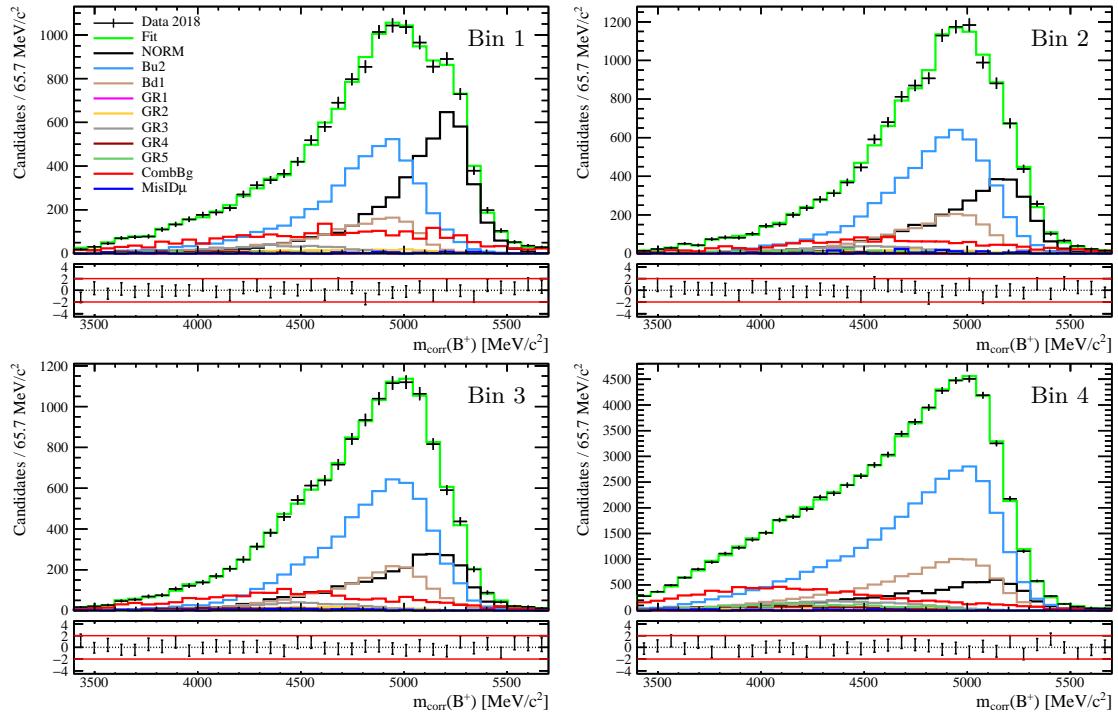
Figure 6.11 – Template fits (green) of the normalisation channel and backgrounds in 2018 data (black markers) performed in four regions of q^2_{reg} defined in Table 6.12. The main components are the normalisation channel (black), the $B^+ \to \bar{D}^{*0}\mu^+\nu_\mu$ (blue) and $B^0 \to D^{*-}\mu^+\nu_\mu$ (brown) decays and combinatorial background (red).

Table 6.12 – Fit result in bins of q^2 with bin boundaries q^2_{min} and q^2_{max}. The extracted yields of the normalisation channel and Bu2 are listed for each q^2 bin.

Bin	q^2_{min} [GeV$^2/c^4$]	q^2_{max} [GeV$^2/c^4$]	$N^{\text{Fit}}_{\text{NORM}}$	$N^{\text{Fit}}_{\text{Bu2}}$
1	0.0	2.2	4422 ± 152	4597 ± 148
2	2.2	3.9	3205 ± 185	5797 ± 190
3	3.9	5.3	2690 ± 216	6076 ± 198
4	5.3	12.2	6989 ± 673	31990 ± 729
Sum	0.0	12.2	17307 ± 746	48460 ± 793

the full q^2 region listed in Table 6.11.

The extracted yields listed in Table 6.12 divided by their respective bin widths correspond to the measured q^2 distributions of the normalisation channel and Bu2. However, to compare these to the expected q^2 distributions based on the CLN form factor parameterisation, the extracted yields in bins of q^2_{reg} have to be unfolded to get the true distribution q^2_{true} in data.

6.5.2 Comparing Bayes and SVD unfolding methods

To obtain the true q^2 distribution in data the ROOUNFOLD tool, introduced in Sec. 6.2.3, is used. In practice, 4×4 response matrices relating q^2_{true} and q^2_{reg} for the normalisation channel and Bu2 are constructed from their respective simulations. Since both q^2_{true} and q^2_{reg} are available in the simulation, a comparison and validation of different unfolding algorithms can be performed. Thus, the Bayes unfolding algorithm, described in Sec. 6.2.3, is compared to the Singular Value Decomposition (SVD) method [180] that is also provided by the ROOUNFOLD package. Here the unfolding procedure is based on the SVD of the response matrix and a regularisation is implemented by imposing a minimum-curvature condition. The unfolding can be seen as Fourier expansion in the expected spectrum versus the true spectrum based on simulation. The low frequencies are then associated with the systematic differences between the simulation and data, while high frequencies are associated with statistical fluctuations in data that without the proper regularisation get numerically enhanced. Thus, the regularisation parameter r determines how many of the terms in the expansion are included in the unfolding.

For both unfolding methods the optimal regularisation strength must be determined. Following the recommendations of ROOUNFOLD [175], the important properties are the bias, Δ, defined as the difference between the unfolded yield N_{unf} and the true yield N_{true}, the uncertainty of the unfolded yield σ_{unf} and the coverage defined as the probability of N_{true} to fall between $N_{\text{unf}} \pm 1\sigma_{\text{unf}}$. Thus, to optimise the regularisation strength a series of N_{runs} pseudo-experiments are performed for a range of integer values of r. For the Bayes method, $r = 1$ is the minimum regularisation, while no upper limit on r exists, however, the best performance is always found to be well below $r = 10$, and thus, all r between 1 and 10 are tested for the Bayes method. For the SVD method, $r = 2$ is the minimum regularisation, while the maximum regularisation is r equal to the number of bins, and thus, all r between 2 and 4 are tested for the SVD method. In each run, the simulated q^2_{reg} distribution is scaled to the measured number of signal candidates extracted from the fit to data (N_{NORM} or N_{Bu2}) and it is unfolded with the response matrix. However, to perform the unfolding with N_{runs} statistically independent q^2_{reg} distributions the bin contents are fluctuated according to a Gaussian where the width is defined by the uncertainties of the yields extracted from the fit in data. The fluctuated yields are denoted N_{meas} with uncertainty σ_{meas}. For each regularisation, the following quantities averaged over runs from $j = 1$ to $j = N_{\text{runs}}$ and over bins from $i = 1$ to $i = N_{\text{bins}}$ are considered:

Table 6.13 – The unfolded yields of the normalisation channel and Bu2 in bins of q^2_{true}.

Bin	$N^{\text{Unfolded}}_{\text{NORM}}$	$N^{\text{Unfolded}}_{\text{Bu2}}$
1	4837 ± 192	6753 ± 238
2	3498 ± 208	7419 ± 222
3	2878 ± 235	7350 ± 152
4	6094 ± 710	26938 ± 741
Sum	17307 ± 799	48461 ± 824

$$\sqrt{\Delta^2} = \frac{1}{N_{\text{bins}}} \sum_i^{N_{\text{bins}}} \sqrt{\frac{\sum_j^{N_{\text{runs}}} (N_{\text{unf,i,j}} - N_{\text{true,i,j}})^2}{N_{\text{runs}}}} \, , \tag{6.25}$$

$$R_\sigma = \frac{1}{N_{\text{bins}}} \frac{1}{N_{\text{runs}}} \sum_i^{N_{\text{bins}}} \frac{\sum_j^{N_{\text{runs}}} \sigma_{\text{unf,i,j}}}{\sum_j^{N_{\text{runs}}} \sigma_{\text{meas,i,j}}} \, , \tag{6.26}$$

$$p_{\text{cov}} = \frac{1}{N_{\text{bins}}} \frac{1}{N_{\text{runs}}} \sum_i^{N_{\text{bins}}} \sum_j^{N_{\text{runs}}} \frac{N_{\text{true,i,j}} \in N_{\text{unf}} \pm 1\sigma_{\text{unf}}}{N_{\text{true,i,j}} \in N_{\text{unf}} \pm 1\sigma_{\text{unf}} + N_{\text{true,i,j}} \notin N_{\text{unf}} \pm 1\sigma_{\text{unf}}} \, , \tag{6.27}$$

where $N_{\text{true,i,j}} \in (\notin) N_{\text{unf}} \pm 1\sigma_{\text{unf}}$ is the number of times N_{true} falls within (outside) $N_{\text{unf}} \pm 1\sigma_{\text{unf}}$. The optimal r should minimise $\sqrt{\Delta^2}$ that is used to quantify the bin-averaged bias, it should also satisfy $R_\sigma \approx 1$ without $R_\sigma < 1$, which means that the uncertainties σ_{unf} are not under- or overestimated with respect to σ_{meas}, and finally, the coverage p_{cov} should be at least 68.3% within one percent. In addition to this, the run-averaged bias Δ is checked for each bin to ensure that the bias is not isolated to specific bins.

For the normalisation channel and Bu2, a series of $N_{\text{runs}} = 5000$ pseudo-experiments are carried out for each regularisation strength of each unfolding method, and the results are shown in Appendix I. For the normalisation channel, the best performance is achieved with the Bayes unfolding method with $r = 3$ resulting in $\sqrt{\Delta^2} = 310$, $R_\sigma = 1.10$ and $p_{\text{cov}} = 68\%$. For Bu2, the best performance is achieved with the Bayes unfolding method with $r = 4$ resulting in $\sqrt{\Delta^2} = 245$, $R_\sigma = 1.07$ and $p_{\text{cov}} = 79\%$. For the normalisation channel, the biases between the unfolded and true yields are always below 7%, while for Bu2, they are always below 3%.

6.5.3 Unfolded q^2 distributions

Based on the comparison of the unfolding methods, described in Sec. 6.5.2, the Bayes unfolding algorithm with regularisation strengths $r = 3$ and $r = 4$ are used to unfold the measured q^2 distributions of the normalisation channel and Bu2, respectively. The results are illustrated in Fig. 6.12 and summarised in Table 6.13. For both decays, a good agreement is found between the unfolded and expected q^2 distributions.
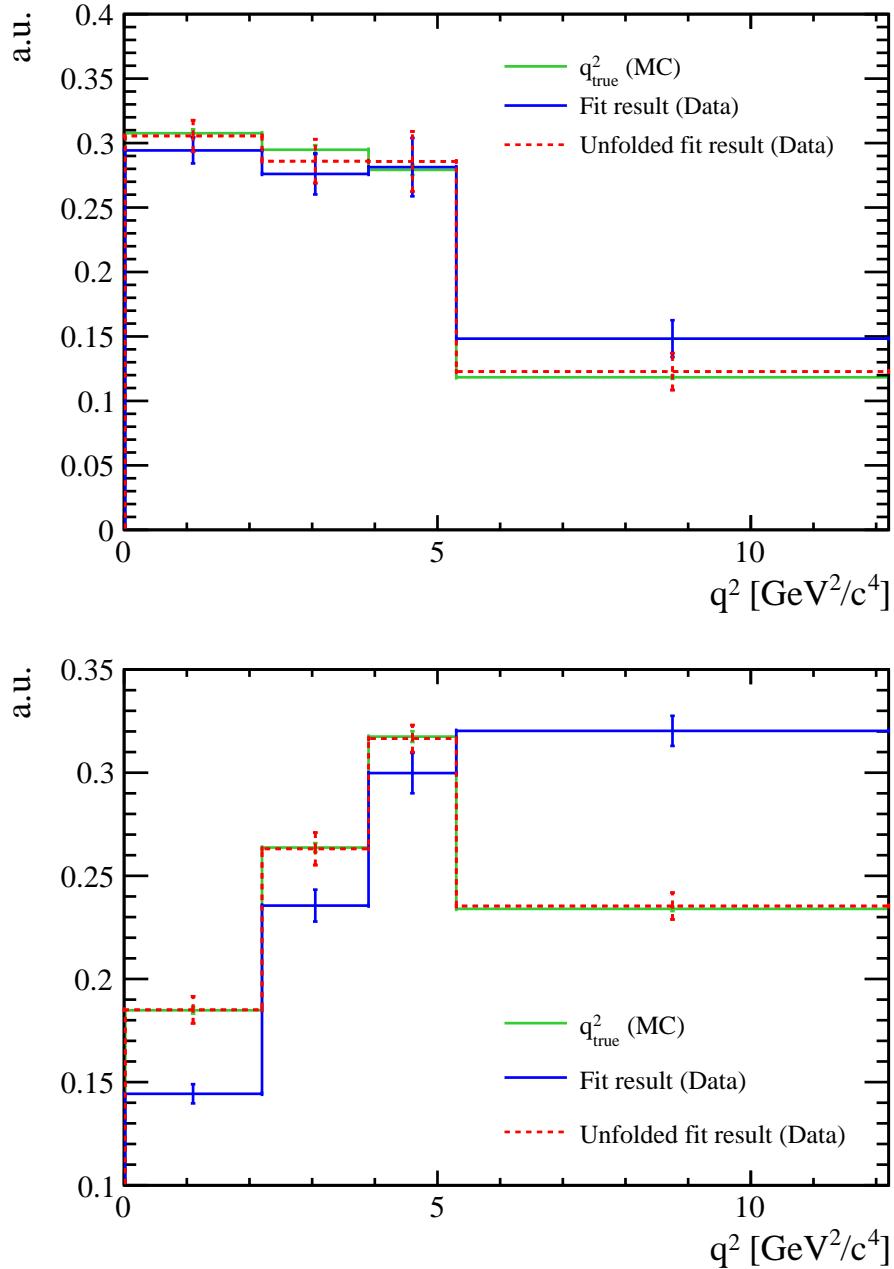
Figure 6.12 – Normalised q^2 distributions for the normalisation channel (top) and Bu2 (bottom), *i.e.* the reconstructed q^2_{reg} distribution in data (blue), the unfolded q^2_{true} distribution in data (dashed red), and the expected q^2_{true} distribution (green) based on the simulation with the CLN form factors.

7 Signal channel fit

This chapter describes the 2D template fit to $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+\pi^-)$ used to extract the signal yield in bins of q^2, the subsequent unfolding of the true q^2 distribution and the systematic uncertainties associated with the fitting and unfolding procedures. In Sec. 7.1, the modellings of the various processes contributing to the fit are described and the resulting fitting templates are illustrated. This is followed by Sec. 7.2, where template corrections and shape uncertainties are explained. The signal fit and the extracted signal yields in bins of q^2 are presented in Sec. 7.3, the unfolding of the true q^2 distribution is described in Sec. 7.4, and finally, the sources of systematic uncertainties are discussed and evaluated in Sec. 7.5.

7.1 Fit components

The final signal selection, summarised in Table 7.1, restricts signal candidates to a region around the peak of the ρ^0 resonance. In addition to this, a requirement on the output variable of the DNN is imposed to improve the signal purity in data. This requirement discriminates, in particular, against the partially reconstructed backgrounds with additional charged particles.

After applying the full signal selection, the remaining backgrounds include decays with the same final state as the signal or with additional neutral particles since these are not rejected by the charge isolation. However, as some of the charged particles of the event

Table 7.1 – Final signal selection after applying the stripping, trigger selections, and preselection.

Final signal selection
$\mathrm{dnn}_{\mathrm{out}} > 0.7$
$570\,\mathrm{MeV}/c^2 < m(\pi^+\pi^-) < 1100\,\mathrm{MeV}/c^2$
$3800\,\mathrm{MeV}/c^2 < m_{\mathrm{corr}}(B^+) < 5500\,\mathrm{MeV}/c^2$

are often not reconstructed, *e.g.* because they are outside the detector acceptance or their momenta are below detection thresholds, some decays with additional charged particles are not removed by the charge isolation and end up contributing to the selected sample. In addition to these physics backgrounds, there are also combinatorial and misID backgrounds, where the latter consists of signal candidates combined with either a misidentified μ or π. The most important processes expected to contribute to the sample are divided into the following categories[1]:

1. Signal, *i.e.* $B^+ \to \rho^0 \mu^+ \nu_\mu$ with $\rho^0 \to \pi^+ \pi^-$.

2. Exclusive semileptonic $|V_{cb}|$ decays[2] (excl. $|V_{cb}|$), *i.e.* $B_{(s)} \to D_{(s)}^{(*,**)} \mu \nu_\mu (X)$ and $\Lambda_b^0 \to \Lambda_c^{(*,**)+} \mu^- \bar{\nu}_\mu (X)$ decays where $D_{(s)}^{(*,**)}$ and $\Lambda_c^{(*,**)+}$ cascade into $D_{(s)}$ and Λ_c^+, respectively. The subsequent decays of $D_{(s)}$ and Λ_c^+ produce final states with $\pi^+ \pi^-$.

3. Exclusive semileptonic $|V_{ub}|$ decays (excl. $|V_{ub}|$), *i.e.* $B^+ \to X_u^0 \mu^+ \nu_\mu$ decays where $X_u^0 \to \pi^+ \pi^- (X)$ and with $X_u^0 = \omega, \eta^{'}, \rho(1450), f_0(500), f_0(980)$ and $f_2(1270)$.

4. Inclusive semileptonic $|V_{ub}|$ decays (incl. $|V_{ub}|$), *i.e.* $B \to X_u \mu^+ \nu_\mu$ decays where X_u represents a quasi state containing a u quark that produces a final state with $\pi^+ \pi^-$.

5. Charmonium decays, *i.e.* $B_{(s)} \to X_{c\bar{c}} Y$ decays where the charmonium state $X_{c\bar{c}}$ decays into $\mu^+ \mu^-$ and the hadronic state Y decays into a final state with $\pi^+ \pi^-$.

6. D_s^+ leptonic, *i.e.* the decay $B_s^0 \to D_s^+ \pi^+ \pi^- \pi^-$ with the leptonic decay $D_s^+ \to \mu^+ \nu_\mu$.

7. CombBg, *i.e.* a random combination of $\pi^+ \pi^+ \mu^+$ particles forming a common decay vertex.

8. MisIDμ and misIDπ, *i.e.* candidates formed with a single misidentified μ or π.

These processes are modelled by 2D templates in $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+ \pi^-)$. For the signal and physics backgrounds (1–6 in the above list), the templates are extracted from simulations, which will be explained in Sec. 7.1.1. For the CombBg and misID backgrounds (7–8 in the above list), the templates are obtained from data driven methods, which will be explained in Sec. 7.1.2 and 7.1.3, respectively.

7.1.1   Signal and physics backgrounds

The fitting templates based on simulations are summarised in Table 7.2 together with their associated branching fractions, while their estimated selection efficiencies in bins of q^2 are given in Appendix E. These templates are expected to capture the largest part of the physics backgrounds of the selected data sample.

---

[1]Here X represents one or more charged or neutral particles and $H^{*,**}$ represent higher resonances of hadron H.

[2]"Semileptonic $|V_{Qq}|$ decays" refers to $X_Q \to X_q \mu^+ \nu_\mu$ decays where X_Q and X_q are hadrons containing a Q and q quark, respectively.

Table 7.2 – Summary of the fitting templates based on simulations of physics processes. The combined branching fraction and uncertainty of all the processes making up each template is denoted $\mathcal{B} \pm \sigma_{\mathcal{B}}$. Additional charged or neutral particles are represented by X and Y, and higher resonances of a hadron H are represented by H^* or H^{**}. The hadronic states containing a u quark are denoted X_u, while charmonium states are denoted $X_{c\bar{c}}$.

Category	ID	Template processes	$\mathcal{B} \pm \sigma_{\mathcal{B}}$
Signal	T$_1$	$B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$	$(1.58 \pm 0.11) \times 10^{-4}$
		$B^+ \to \bar{D}^{(*,**)0,-}\mu^+\nu_\mu(X)$	
	T$_2$	$\bar{D}^0 \to K^0\pi^+\pi^-$	$(6.62 \pm 0.47) \times 10^{-3}$
	T$_3$	$\bar{D}^0 \to K^0\pi^+\pi^-\pi^0$	$(1.23 \pm 0.15) \times 10^{-2}$
	T$_4$	$\bar{D}^0 \to \pi^+\pi^-\pi^0$	$(1.76 \pm 0.09) \times 10^{-3}$
Excl. $|V_{cb}|$	T$_5$	$\bar{D}^0 \to \pi^+\pi^-\pi^0\pi^0$	$(1.21 \pm 0.11) \times 10^{-3}$
	T$_6$	$\bar{D}^0 \to K^+\pi^-\pi^+\pi^-$	$(9.72 \pm 0.34) \times 10^{-3}$
		$B^0 \to D^{(*,**)-}\mu^+\nu_\mu(X)$	
	T$_7$	$D^- \to \pi^+\pi^-\pi^-$	$(1.74 \pm 0.13) \times 10^{-2}$
	T$_8$	$B_s^0 \to D_s^{(*,**)-}\mu^+\nu_\mu$	$(3.36 \pm 0.64) \times 10^{-2}$
	T$_9$	$\Lambda_b^0 \to \Lambda_c^{(*,**)+}\mu^-\bar{\nu}_\mu(X)$	$(6.54 \pm 1.64) \times 10^{-2}$
	T$_{10}$	$B^+ \to \omega\mu^+\nu_\mu$	$(1.06 \pm 0.08) \times 10^{-4}$
	T$_{11}$	$B^+ \to \eta'\mu^+\nu_\mu$	$(1.85 \pm 0.65) \times 10^{-5}$
Excl. $|V_{ub}|$	T$_{12}$	$B^+ \to \rho(1450)\mu^+\nu_\mu$	$(0.60 \pm 1.50) \times 10^{-5}$
	T$_{13}$	$B^+ \to f_0(500)\mu^+\nu_\mu$	$(1.30 \pm 0.91) \times 10^{-5}$
	T$_{14}$	$B^+ \to f_0(980)\mu^+\nu_\mu$	$(1.70 \pm 1.02) \times 10^{-5}$
	T$_{15}$	$B^+ \to f_2(1270)\mu^+\nu_\mu$	$(2.03 \pm 1.20) \times 10^{-5}$
	T$_{16}$	$B^+ \to X_u^0\mu^+\nu_\mu$	$(1.16 \pm 0.21) \times 10^{-3}$
Incl. $|V_{ub}|$	T$_{17}$	$B^+ \to \pi^+\pi^-\mu^+\nu_\mu$	$(0.9 \pm 3.1) \times 10^{-5}$
	T$_{18}$	$B^0 \to X_u^-\mu^+\nu_\mu$	$(1.07 \pm 0.19) \times 10^{-3}$
	T$_{19}$	$B^+ \to J/\psi\rho^0$	$(1.52 \pm 0.11) \times 10^{-6}$
Charmonium	T$_{20}$	$B^+ \to X_{c\bar{c}}Y$	$(2.44 \pm 0.35) \times 10^{-4}$
	T$_{21}$	$B^0 \to X_{c\bar{c}}Y$	$(1.48 \pm 0.39) \times 10^{-4}$
	T$_{22}$	$B_s^0 \to X_{c\bar{c}}Y$	$(6.14 \pm 0.31) \times 10^{-5}$
D_s^+ leptonic	T$_{23}$	$B_s^0 \to D_s^+(\to \mu^+\nu_\mu)\pi^+\pi^-\pi^-$	$(3.31 \pm 0.55) \times 10^{-5}$

In the following paragraphs, the processes making up the templates are described in more detail, the extracted templates are illustrated as projections in $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+\pi^-)$ for the integrated q^2 distribution and the branching fractions associated with the templates are estimated. In addition to this, the reconstructed q^2 distribution associated with each template is illustrated to show its relative contribution in bins of q^2 defined according to the non-uniform binning scheme in Table 5.2.
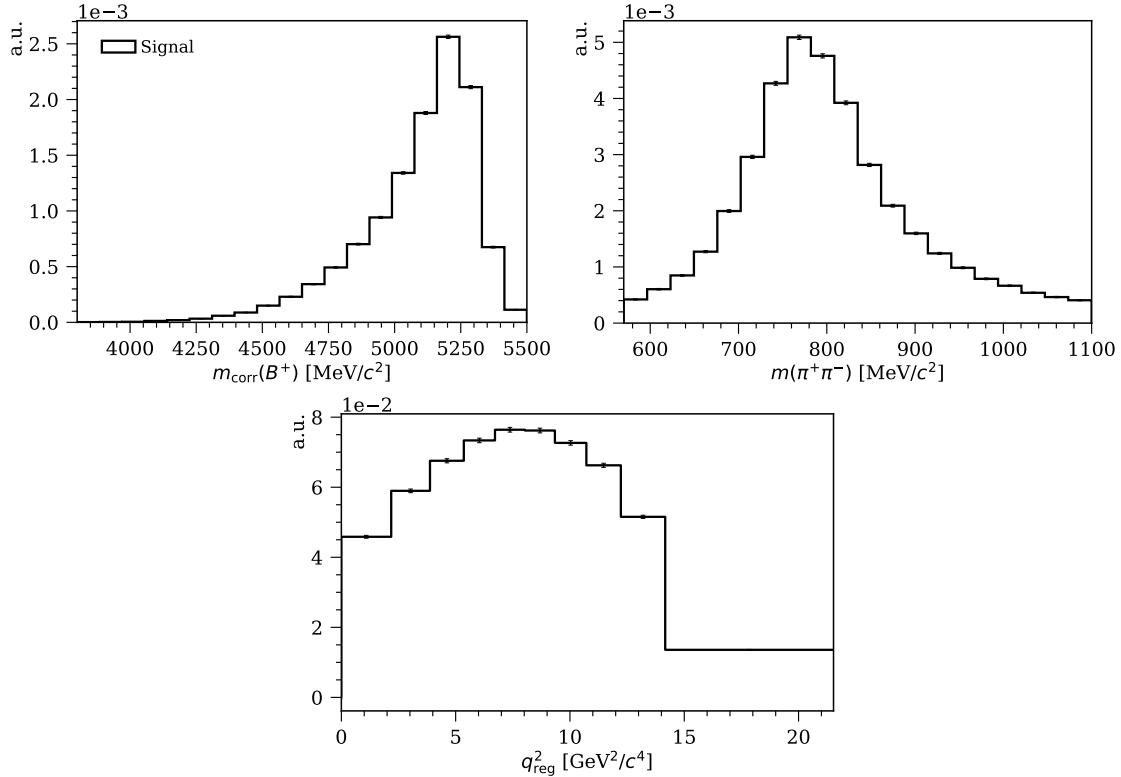
Figure 7.1 – Signal template projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

Signal

The signal template is extracted from the simulation described in Sec. 4.1 based on the BCL form factor parameterisation [69]. However, signal candidates used for developing the signal selection, described in Chap. 4, as well as the regression method, described in Chap. 5, are not reused for the signal template. The resulting template is illustrated in Fig. 7.1 along with the reconstructed q^2 distribution of the signal using the q^2 binning scheme[3] defined in Table 5.2.

Exclusive semileptonic $|V_{cb}|$ decays

The exclusive semileptonic $|V_{cb}|$ decays listed in Table 7.2 fall into four categories:

1. **Excl. $|V_{cb}|^{B^+}$:** $B^+ \to \bar{D}^{(*,**)0,-}\mu^+\nu_\mu(X)$.

2. **Excl. $|V_{cb}|^{B^0}$:** $B^0 \to D^{(*,**)-}\mu^+\nu_\mu(X)$.

---

[3]All q^2 plots shown with a non-uniform binning are displayed as *distributions*, *i.e.* each bin content is divided by the bin width.

3. **Excl. $|V_{cb}|^{B_s^0}$:** $B_s^0 \to D_s^{(*,**)-}\mu^+\nu_\mu$.

4. **Excl. $|V_{cb}|^{\Lambda_b^0}$:** $\Lambda_b^0 \to \Lambda_c^{+(*,**)}\mu^-\bar{\nu}_\mu(X)$.

The excl. $|V_{cb}|^{B^+}$ decays are simulated as a cocktail of $B^+ \to \bar{D}^{(*,**)0,-}\mu^+\nu_\mu(X)$ processes where the excited states $\bar{D}^{(*,**)0,-}$ cascade into $\bar{D}^0$ mesons that are forced to decay into one of the following decay channels [15]:

1. $\bar{D}^0 \to K_S^0\pi^+\pi^-$, $\mathcal{B} = (2.80 \pm 0.18)\%$.

2. $\bar{D}^0 \to K_S^0\pi^+\pi^-\pi^0$, $\mathcal{B} = (5.2 \pm 0.6)\%$.

3. $\bar{D}^0 \to \pi^+\pi^-\pi^0$, $\mathcal{B} = (1.49 \pm 0.06)\%$.

4. $\bar{D}^0 \to \pi^+\pi^-\pi^0\pi^0$, $\mathcal{B} = (1.02 \pm 0.09)\%$.

5. $\bar{D}^0 \to K^+\pi^-\pi^+\pi^-$, $\mathcal{B} = (8.22 \pm 0.14)\%$.

The decays involving K_S^0 mesons are assumed to have the same contribution to the fit as the decays where K_S^0 is replaced by K_L^0. Thus, the templates based on these decay channels describe both contributions which are assumed to be 50%/50% [15], and they are therefore denoted $\bar{D}^0 \to K^0\pi^+\pi^-$ and $\bar{D}^0 \to K^0\pi^+\pi^-\pi^0$ with the branching fractions $\mathcal{B} = (5.60 \pm 0.36)\%$ and $\mathcal{B} = (10.4 \pm 1.2)\%$, respectively.

The semileptonic decays making up the cocktail are the same as the ones used in the NORMBu simulation defined in Table 6.3 with additional processes of the type:

- $B^+ \to \bar{X}_c X_u \mu^+\nu_\mu$ with $X_c = D^0$, D^{*0}, D^{*-} and $X_u = \pi^+\pi^-$, $\pi^0\pi^0$, π^-, $\pi^0\pi^-$.

- $B^+ \to \bar{X}_c\tau^+\nu_\tau$ with $X_c = D_0(2420)^0$, $D_1(2420)^0$, $D_1(2430)^0$, $D_2(2460)^0$.

The semileptonic decays of the cocktail are modelled with ISGW2 form factors, except for the nonresonant $B \to D^{(*)}\pi(\pi)\mu\nu_\mu$ decays that are modelled with Goity-Roberts. The sum of all processes making up the cocktail constitutes about $(9.86 \pm 0.30)\%$ of the total B^+ decay width, where the dominant decays are $B^+ \to \bar{D}^*(2007)^0(\to \bar{D}^0\pi^0/\gamma)\mu^+\nu_\mu$ with $\mathcal{B} \sim 5.6\%$ and $B^+ \to \bar{D}^0\mu^+\nu_\mu$ with $\mathcal{B} \sim 2.3\%$.

However, this MC cocktail does not include D^0 mesons produced in semileptonic B^0 decays, which would result in $m_{\mathrm{corr}}(B^+)$ and q^2 distributions similar to the ones of $B^+ \to \bar{D}^{(*,**)0,-}\mu^+\nu_\mu(X)$ decays. To estimate this contribution, the cocktail simulations NORMBu and NORMBd used for the normalisation channel are compared after applying the signal selection, except for the requirement on $m(\pi^+\pi^-)$. The number of $D^0 \to \pi^+\pi^+$ candidates in the two samples are found to be $N_{\mathrm{NORMBu}} = 14001$ and $N_{\mathrm{NORMBd}} = 5868$ corresponding to the selection efficiencies $\epsilon_{\mathrm{NORMBu}} = 6.5 \times 10^{-5}$ and $\epsilon_{\mathrm{NORMBd}} = 2.1 \times 10^{-5}$. Given their branching fractions $\mathcal{B}_{\mathrm{NORMBu}} = (9.4 \pm 0.3)\%$ and $\mathcal{B}_{\mathrm{NORMBd}} = (4.8 \pm 0.2)\%$
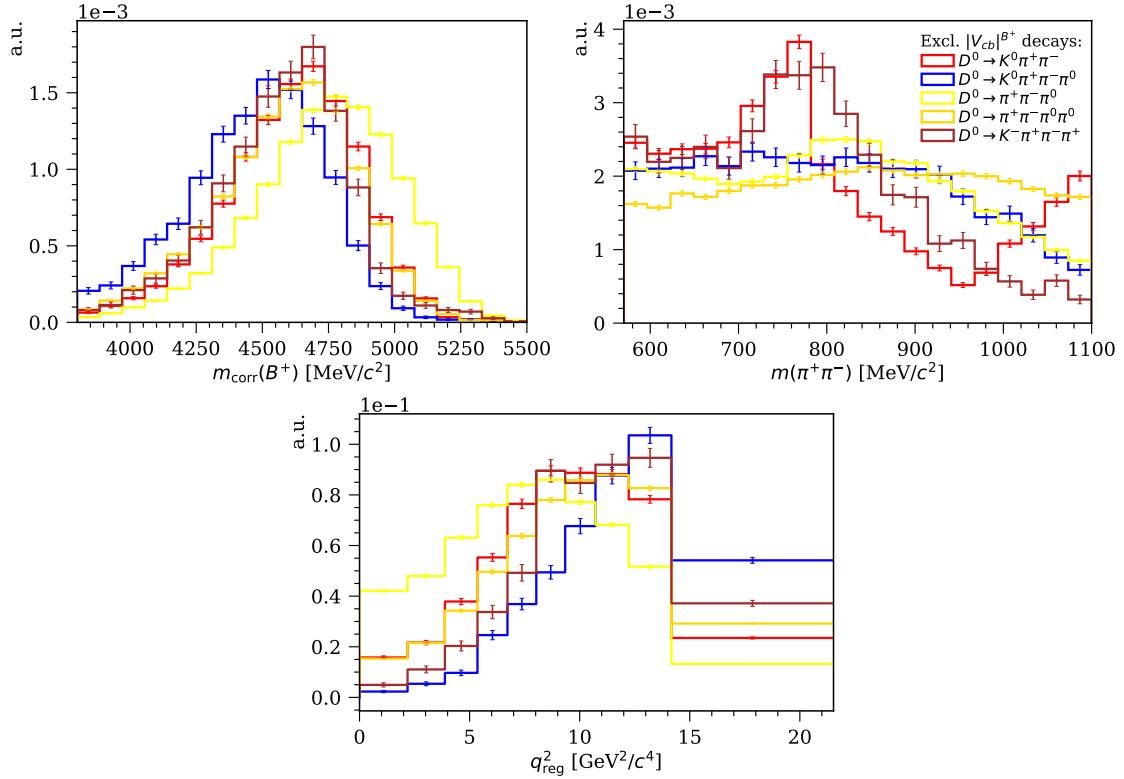
Figure 7.2 – Templates for the exclusive semileptonic $B^+ \to \bar{D}^{(*,**)0,+}\mu^+\nu_\mu(X)$ decays projected in $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

obtained by summing the decays in Tables 6.3 and 6.4, respectively, and assuming equal B^+ and B^0 production rates, a correction factor can be estimated to:

$$f_{\text{corr}} = 1 + \frac{\mathcal{B}_{\text{NORMBd}} \times \epsilon_{\text{NORMBd}}}{\mathcal{B}_{\text{NORMBu}} \times \epsilon_{\text{NORMBu}}} = 1.2\,. \tag{7.1}$$

Assuming the correction to be mostly independent of the kinematics of the two pions, it can be applied to the signal backgrounds. Thus, the branching fractions associated with the excl. $|V_{cb}|^{B^+}$ templates are multiplied with f_{corr} to account for the feed-down production of D^0 from semileptonic B^0 decays. Their final branching fractions are listed in Table 7.2.

The five excl. $|V_{cb}|^{B^+}$ templates are illustrated in Fig. 7.2. These decays are characterised by low $m_{\text{corr}}(B^+)$ and high q^2 compared to the signal. However, for the $\bar{D}^0 \to \pi^+\pi^-\pi^0$ decay, where only the light final state particle π^0 is not reconstructed, the q^2 distribution is more similar to the one of the signal. In the $m(\pi^+\pi^-)$ distribution structures consistent with a ρ^0 meson are observed for the $\bar{D}^0 \to K^0\pi^+\pi^-$ and $\bar{D}^0 \to K^+\pi^-\pi^+\pi^-$ channels. However, it should be noted that only for the three-body D^0 decays the effect of the $\rho^0 - \omega$ interference, which will be discussed in Sec. 7.2.1, is simulated.
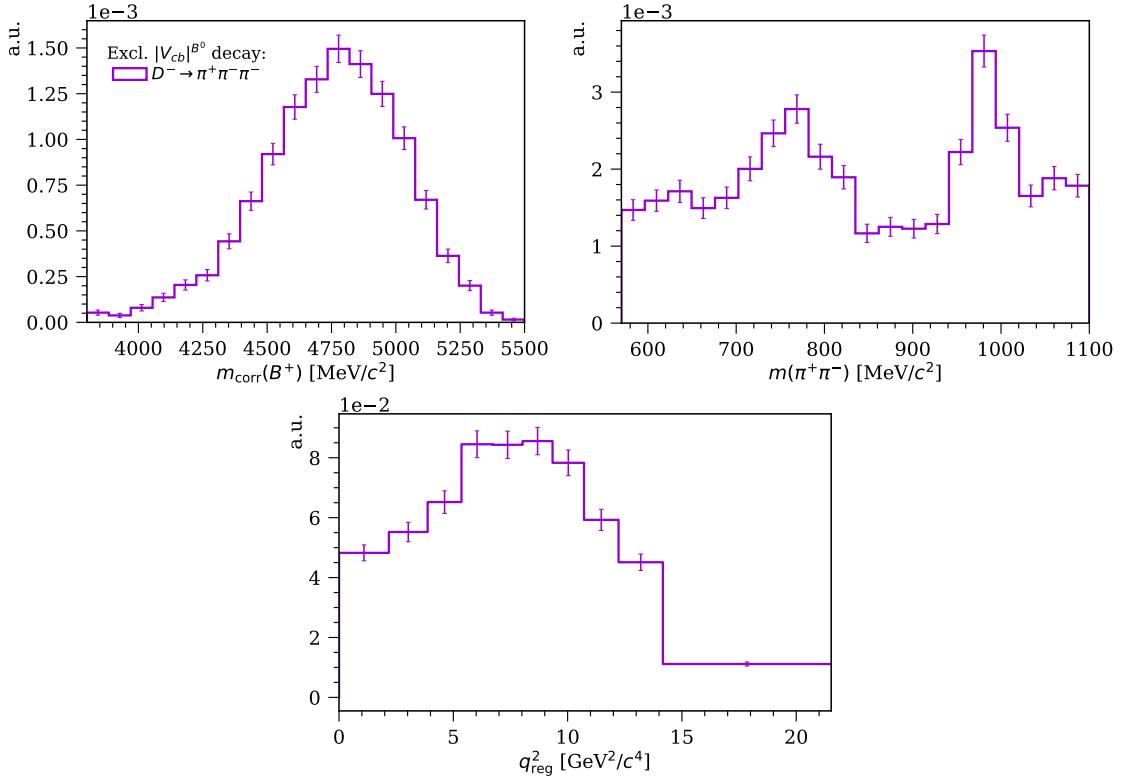
Figure 7.3 – Template for the exclusive semileptonic $B^0 \to D^{(*,**)-}\mu^+\nu_\mu(X)$ decays projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

The excl. $|V_{cb}|^{B^0}$ decays are simulated as a cocktail of $B^0 \to D^{(*,**)-}\mu^+\nu_\mu(X)$ processes where the excited states $D^{(*,**)-}$ cascade into D^- mesons that are forced to decay into [15]:

1. $D^- \to \pi^+\pi^-\pi^-$, $\quad \mathcal{B} = (3.27 \pm 0.18) \times 10^{-3}$.

The semileptonic decays making up the cocktail are the same as the ones used in the NORMBd simulation defined in Table 6.4 with the additional processes $B^0 \to D^-\mu^+\nu_\mu$, $B^0 \to D^-\tau^+\nu_\tau$ and $B^0 \to D^-\pi^0\mu^+\nu_\mu$. The sum of all processes making up the cocktail is about 5% of the total B^0 decay width. The dominant decays are $B^0 \to D^-\mu^+\nu_\mu$ with $\mathcal{B} = (2.24 \pm 0.09)\%$ and $B^0 \to D^*(2010)^-(\to D^-\pi^0/\gamma)\mu^+\nu_\mu$ with $\mathcal{B} = (1.61 \pm 0.05)\%$. Both of these decays are modelled with CLN form factors. The remaining decays are modelled with ISGW2 form factors, except for the nonresonant $B^0 \to D^{(*)}\pi\mu\nu_\mu$ decays that are modelled with Goity-Roberts.

The resulting excl. $|V_{cb}|^{B^0}$ template is illustrated in Fig. 7.3. Like the excl. $|V_{cb}|^{B^+}$ templates, it is characterised by low $m_{\mathrm{corr}}(B^+)$ compared to the signal, while the q^2 distribution is more similar to the one of the signal. In the $m(\pi^+\pi^-)$ distribution structures associated with ρ^0 and $f_0(980)$ resonances are visible.

The excl. $|V_{cb}|^{B_s^0}$ decays are simulated as a cocktail of $B_s^0 \to D_s^{(*,**)-}\mu^+\nu_\mu$ processes, where the excited states $D_s^{(*,**)-}$ cascade into D_s^- mesons that can decay into various channels as long as the final state of the full decay chain contains $\mu^+\pi^+\pi^-$. After applying the full selection, the single most dominant decay channels of the sample are:

1. $B_s^0 \to D_s^-(\to \eta'(\to \rho^0(\to \pi^+\pi^-)\gamma)X)\mu^+\nu_\mu$ (12.3%) .

2. $B_s^0 \to D_s^{*-}(\to D_s^-(\to \pi^+\pi^-X)\gamma/\pi^0)\mu^+\nu_\mu$ (8.9%) .

3. $B_s^0 \to D_s^{*-}(\to D_s^-(\to f_0(980)(\to \pi^+\pi^-)X)\gamma/\pi^0)\mu^+\nu_\mu$ (5.6%) .

The remaining part of the sample is composed by a large number of processes with small individual contributions. The semileptonic $B_s^0 \to D_s^-\mu^+\nu_\mu$ and $B_s^0 \to D_s^{*-}\mu^+\nu_\mu$ decays are modelled with CLN form factors, while the decays involving higher D_s^{**} resonances are modelled with ISGW2 form factors. The sum of all processes making up the cocktail is about 3% of the total B_s^0 decay width. In order to compare this to B^+ or B^0 decays, the average production fraction of B_s^0 mesons at LHCb ($\sqrt{s} = 13\,\text{TeV}$) with respect to B^0 mesons of 0.2539 ± 0.0079 [181] is used.

The resulting excl. $|V_{cb}|^{B_s^0}$ template is illustrated in Fig. 7.4. It is characterised by low $m_{\text{corr}}(B^+)$ and high q^2 compared to the signal.

The excl. $|V_{cb}|^{\Lambda_b^0}$ decays are simulated as a cocktail of $\Lambda_b^0 \to \Lambda_c^{(*,**)+}\mu^-\bar{\nu}_\mu(X)$ processes, where the excited states $\Lambda_c^{(*,**)+}$ cascade into Λ_c^+ that can decay into various channels as long as the final state of the full decay chain contains $\mu^+\pi^+\pi^-$. After applying the full selection, the single most dominant decay channels of the sample are:

1. $\Lambda_b^0 \to \Lambda_c^+(\to \pi^+\pi^-X)\mu^-\bar{\nu}_\mu$ (66.4%) .

2. $\Lambda_b^0 \to \Lambda_c^+(\to \rho^0(\to \pi^+\pi^-)X)\mu^-\bar{\nu}_\mu$ (5.3%) .

3. $\Lambda_b^0 \to \Lambda_c^+(\to \phi(1020)(\to \pi^+\pi^-\pi^0)X)\mu^-\bar{\nu}_\mu$ (4.6%) .

The sum of all processes making up the cocktail is about 7% of the total Λ_b^0 decay width. To compare this to B^+ or B^0 decays, the integrated production rate of Λ_b^0 mesons at LHCb ($\sqrt{s} = 13\,\text{TeV}$) with respect to B^0 mesons of 0.518 ± 0.036 [182], is used. The semileptonic $\Lambda_b^0 \to \Lambda_c^{(*,**)+}\mu^-\bar{\nu}_\mu(Y)$ decays are modelled with form factor predictions based on the quark model by Ref. [183].

The resulting excl. $|V_{cb}|^{\Lambda_b^0}$ template is illustrated in Fig. 7.5. It is characterised by low $m_{\text{corr}}(B^+)$ and high q^2 compared to the signal.

Exclusive semileptonic $|V_{ub}|$ decays

The exclusive semileptonic $|V_{ub}|$ templates are extracted from the six simulations listed in Table 7.2. The processes and sub-processes making up each of these simulations are
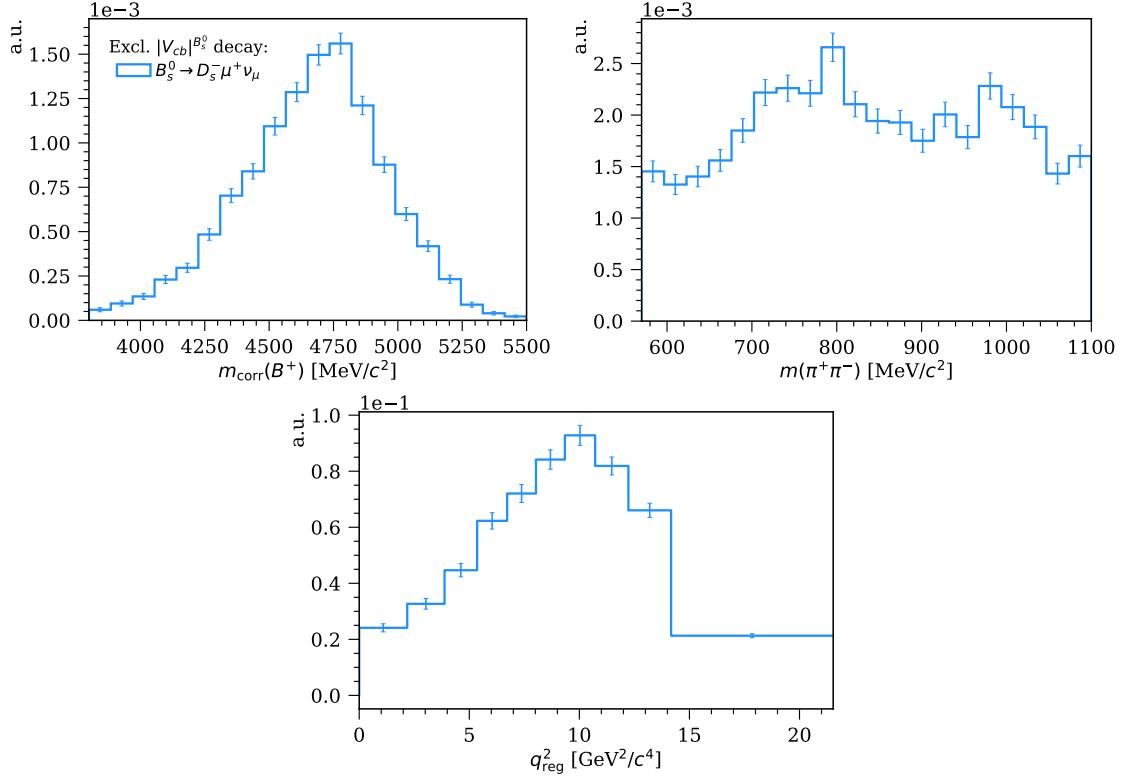
Figure 7.4 – Templates for the exclusive semileptonic $B_s^0 \to D_s^{(*,**)-}\mu^+\nu_\mu$ decays projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

specified below together with their respective branching fractions [15] (if measured) and the applied form factor model:

1. $B^+ \to \omega\mu^+\nu_\mu$, $\mathcal{B} = (1.19 \pm 0.09) \times 10^{-4}$, BCL form factors .

   (a) $\omega \to \pi^+\pi^-\pi^0$, $\mathcal{B} = (89.2 \pm 0.7)\%$.

   (b) $\omega \to \pi^+\pi^-$, $\mathcal{B} = (1.53^{+0.11}_{-0.13})\%$.

   $\mathcal{B}_{\mathrm{tot}} = (1.08 \pm 0.08) \times 10^{-4}$.

2. $B^+ \to \eta'\mu^+\nu_\mu$, $\mathcal{B} = (2.3 \pm 0.8) \times 10^{-5}$, ISGW2 form factors .

   (a) $\eta' \to \pi^+\pi^-\eta$, $\mathcal{B} = (42.5 \pm 0.5)\%$.

   (b) $\eta' \to \pi^0\pi^0\eta$, $\mathcal{B} = (22.4 \pm 0.5)\%$.

       i. $\eta \to \pi^-\pi^+\pi^0$, $\mathcal{B} = (23.02 \pm 0.25)\%$.

       ii. $\eta \to \pi^-\pi^+\gamma$, $\mathcal{B} = (4.28 \pm 0.07)\%$.

   (c) $\eta' \to \rho^0\gamma$, $\mathcal{B} = (29.5 \pm 0.4)\%$.

   (d) $\eta' \to \omega\gamma$, $\mathcal{B} = (2.52 \pm 0.07)\%$.

       i. $\omega \to \pi^-\pi^+\pi^0$, $\mathcal{B} = (89.2 \pm 0.7)\%$.
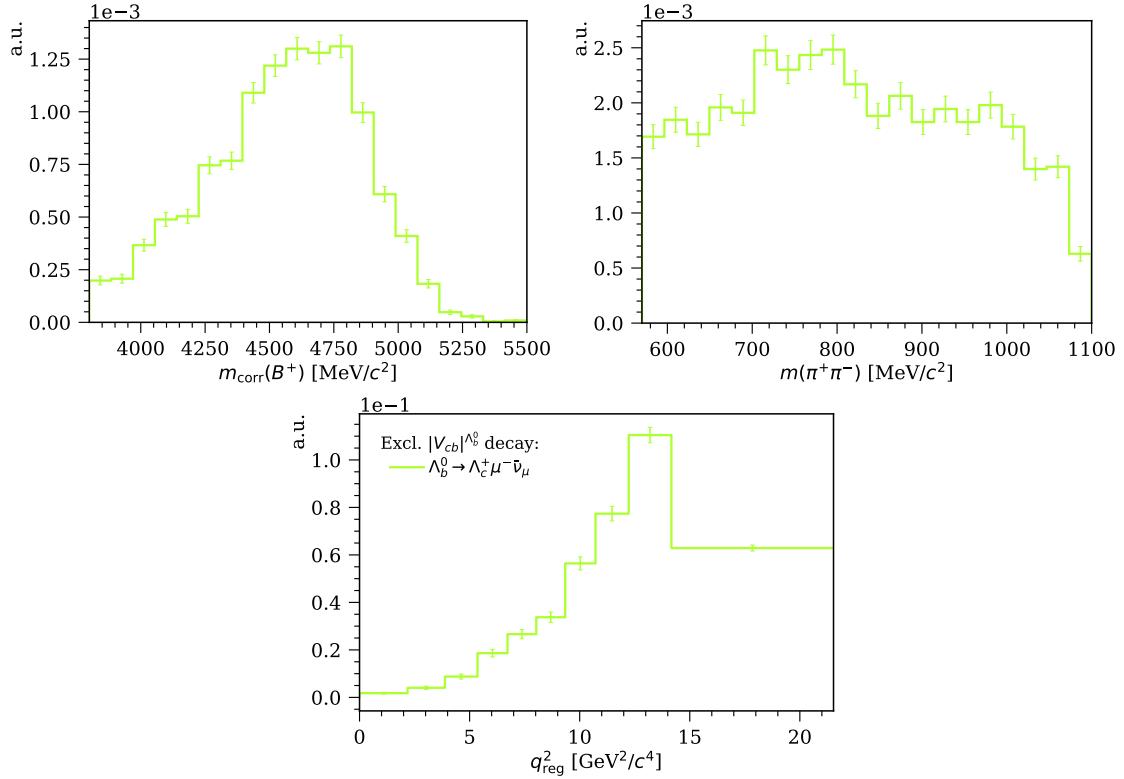
Figure 7.5 – Templates for exclusive semileptonic $\Lambda_b^0 \to \Lambda_c^{(*,**)+}\mu^-\bar{\nu}_\mu(X)$ decays projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

    ii. $\omega \to \pi^-\pi^+$, $\quad \mathcal{B} = (1.53^{+0.11}_{-0.13})\%$.

(e) $\eta' \to \pi^+\pi^-e^+e^-$, $\quad \mathcal{B} = (2.42 \pm 0.10) \times 10^{-3}$.

$\mathcal{B}_{\mathrm{tot}} = (1.85 \pm 0.65) \times 10^{-5}$.

3. $B^+ \to f_2(1270)\mu^+\nu_\mu$, $\quad \mathcal{B} = (2.14 \pm 1.10) \times 10^{-5}$, $\quad$ ISGW2 form factors.

    (a) $f_2(1270) \to \pi^+\pi^-$, $\quad \mathcal{B} = (84.2^{+2.9}_{-0.9})\%$.

    (b) $f_2(1270) \to \pi^+\pi^-\pi^0\pi^0$, $\quad \mathcal{B} = (7.7^{+1.1}_{-3.2})\%$.

    (c) $f_2(1270) \to \pi^+\pi^-\pi^+\pi^-$, $\quad \mathcal{B} = (2.8 \pm 0.4)\%$.

    (d) $f_2(1270) \to \eta\eta$, $\quad \mathcal{B} = (4.0 \pm 0.8) \times 10^{-3}$.

$\mathcal{B}_{\mathrm{tot}} = (2.03 \pm 1.20) \times 10^{-5}$.

4. $B^+ \to f_0(500)(\to \pi^+\pi^-)\mu^+\nu_\mu$, $\quad \mathcal{B} = (1.30 \pm 0.91) \times 10^{-5}$, $\quad$ ISGW2 form factors.

5. $B^+ \to f_0(980)(\to \pi^+\pi^-)\mu^+\nu_\mu$, $\quad \mathcal{B} = (1.70 \pm 1.02) \times 10^{-5}$, $\quad$ ISGW2 form factors.

6. $B^+ \to \rho(1450)(\to \pi^+\pi^-)\mu^+\nu_\mu$, $\quad \mathcal{B} = (0.60 \pm 1.50) \times 10^{-5}$, $\quad$ ISGW2 form factors.

Only the branching fractions of the $B^+ \to \omega(\to \pi^+\pi^- X)\mu^+\nu_\mu$, $B^+ \to \eta'(\to \pi^+\pi^- X)\mu^+\nu_\mu$ and $f_2(1270) \to \pi^+\pi^- X$ decays have been measured and published [15]. The remaining decays are expected to exist, and some of them have been observed. The listed branching fractions for the $B^+ \to X_u^0(\to \pi^+\pi^-)\ell^+\nu_\ell$ decays with $X_u = f_0(500), f_0(980), f_2(1270), \rho(1450)$ are extracted from a study [184] of the $B^+ \to \pi^+\pi^-\ell^+\nu_\ell$ decay process by Belle [185]. However, these branching fractions have very large uncertainties. The possible contribution from the decay $B^+ \to \rho^0\tau^+\nu_\tau$ with $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\mu$ is also assessed. Using the prediction of the ratio $R_{\rho^0} = 0.535 \pm 0.008$ [186] of the total decay rates of $B^+ \to \rho^0\tau^+\nu_\tau$ and $B^+ \to \rho^0\mu^+\nu_\mu$ to estimate $\mathcal{B}(B^+ \to \rho^0\tau^+\nu_\tau)$ and its selection efficiency based on simulation, its contribution to the fit relative to the signal is found to be 0.3%. In addition, its $m_{\text{corr}}(B^+)$ and q^2 distributions are broad. Thus, the contribution from $B^+ \to \rho^0\tau^+\nu_\tau$ to the fit is assumed to be negligible and is therefore not included in the fit.

The resulting excl. $|V_{ub}|$ templates are illustrated in Fig. 7.6. The small contribution from the $\omega \to \pi^+\pi^-$ channel is not included in the $B^+ \to \omega\mu^+\nu_\mu$ template. Instead, this contribution is taken into account in the signal by the $m(\pi^+\pi^-)$ line-shape correction described in Sec. 7.2.1. Thus, both the decays $B^+ \to \omega\mu^+\nu_\mu$ and $B^+ \to \eta'\mu^+\nu_\mu$ miss at least one additional particle with respect to the signal, and their $m_{\text{corr}}(B^+)$ distributions are therefore peaking at lower $m_{\text{corr}}(B^+)$ than the signal. The q^2 distribution of the $B^+ \to \omega\mu^+\nu_\mu$ decay is peaking at high q^2, while the q^2 distribution of $B^+ \to \eta'\mu^+\nu_\mu$ is rather similar to the signal. The decays involving the $\rho(1450)$, $f_0(500)$, $f_0(980)$ and $f_2(1270)$ mesons miss mostly only a ν_μ like the signal, and their $m_{\text{corr}}(B^+)$ distributions are therefore almost indistinguishable from the signal. However, they can be distinguished in the $m(\pi^+\pi^-)$ distribution. In addition, assuming the ISGW2 form factors provide a reasonable q^2 description, these decays are characterised by low q^2 compared to the signal.

Inclusive semileptonic $|V_{ub}|$ decays

The reported measurements of exclusive semileptonic $b \to u\ell\nu_\ell$ ($\ell = e, \mu$) decays constitute only about 25% of the total inclusive rate [185]. The remaining modes include semileptonic decays of heavier hadronic states X_u as well as nonresonant decays. The composition and nature of this background has not been precisely determined. In order to describe this background, the DeFazio-Neubert (DFN) model [187] is used. In this model, the fully differential decay rate of inclusive semileptonic $|V_{ub}|$ decays is computed with the heavy-quark expansion at leading order with given shape function parameters. The model is used by EVTGEN to generate $B \to X_u\mu\nu_\mu$ decays, while PYTHIA takes care of the hadronisation of X_u. This results in two inclusive semileptonic $|V_{ub}|$ simulations, where the excl. $|V_{ub}|$ decays defined in the previous paragraph are not included:

1. **Incl. $|V_{ub}|^{B^+}$** : $B^+ \to X_u^0\mu^+\nu_\mu$.

2. **Incl. $|V_{ub}|^{B^0}$** : $B^0 \to X_u^-\mu^+\nu_\mu$.

After the full selection, the incl. $|V_{ub}|^{B^+}$ simulation is composed of 66% of decays where
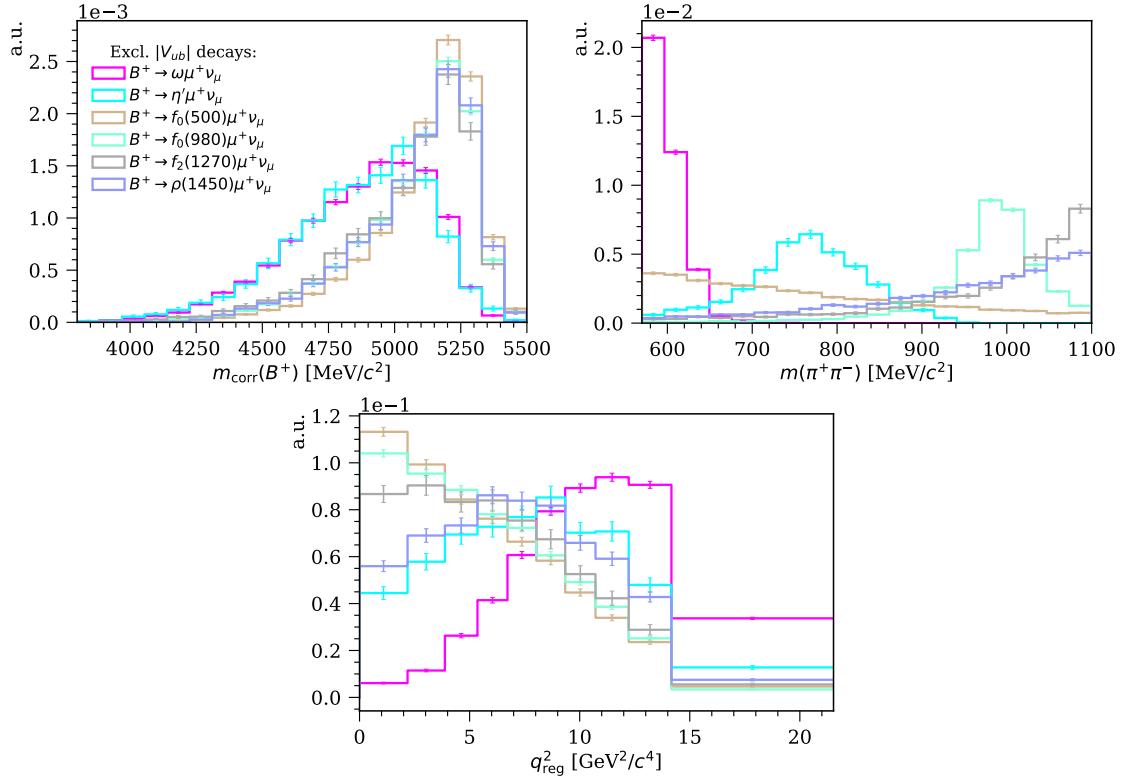
Figure 7.6 – Templates for the exclusive semileptonic $|V_{ub}|$ decays projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

the final state pions are originating from the nonresonant $X_u^0 \to \pi^+\pi^-$ process. These are referred to as $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ decays. The remaining 34% of the sample consist of decays where the final state pions are produced in the decay chain of the resonant X_u^0 state into combinations of light unflavoured mesons, *e.g.* $\rho^{0,\pm}$, ω, η, η' and $\pi^{0,\pm}$. These are referred to as $B^+ \to X_u^0\mu^+\nu_\mu$ decays. As the relative proportion of the two components are not well measured, and their associated $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+\pi^-)$ shapes are rather different, they are represented by two different templates in the fit. The incl. $|V_{ub}|^{B^0}$ simulation consists of semileptonic $B^0 \to X_u^-\mu^+\nu_\mu$ decays where the state X_u^- decays into combinations of light unflavoured mesons.

The resulting templates are illustrated in Fig. 7.7. The nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ decays are characterised by a $m_{\mathrm{corr}}(B^+)$ distribution that is indistinguishable from the signal as well as a q^2 distribution that is rather similar to the signal. However, the nonresonant contribution can be distinguished from the signal by its $m(\pi^+\pi^-)$ distribution. The incl. $B^+ \to X_u^0\mu^+\nu_\mu$ and $B^0 \to X_u^-\mu^+\nu_\mu$ decays are characterised by low $m_{\mathrm{corr}}(B^+)$ and high q^2 with respect to the signal. They consist of approximately 32% and 47% of $\rho^0 \to \pi^+\pi^-$ final state pions, and thus, their $m(\pi^+\pi^-)$ distributions are dominated by the ρ^0 resonance.
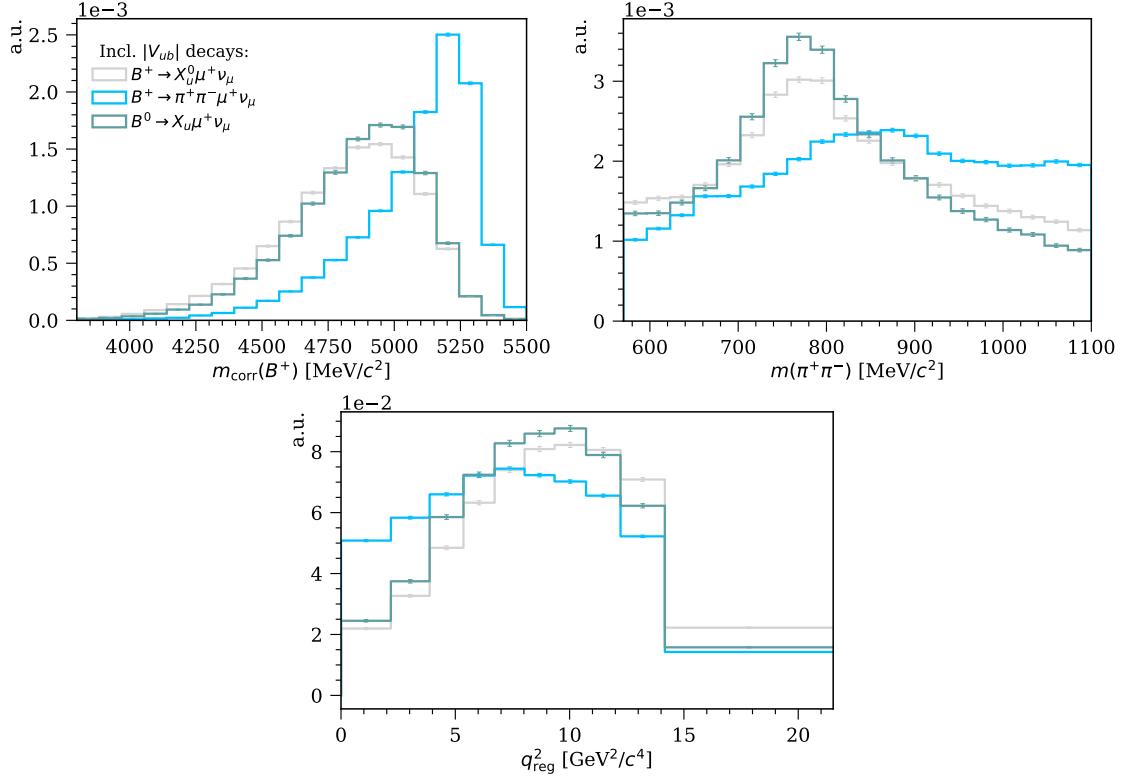
Figure 7.7 – Templates for inclusive semileptonic $|V_{ub}|$ decays projected in $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

The branching fractions of the processes making up the $B^+ \rightarrow X_u^0 \mu^+ \nu_\mu$ and $B^0 \rightarrow X_u^- \mu^+ \nu_\mu$ templates are not well known. However, an estimate can be made by subtracting the measured exclusive contributions [15] from the measured total inclusive rates corresponding to $\mathcal{B}(B^+ \rightarrow X_u^0 \ell \nu_\ell) = (1.65 \pm 0.21) \times 10^{-3}$ and $\mathcal{B}(B^0 \rightarrow X_u^- \ell \nu_\ell) = (1.51 \pm 0.19) \times 10^{-3}$ [15]. Similarly, the branching fraction of the nonresonant process making up the $B^+ \rightarrow \pi^+\pi^-\mu^+\nu_\mu$ template is not known. However, the combined resonant and nonresonant branching fraction of $B^+ \rightarrow \pi^+\pi^-\mu^+\nu_\mu$ decays has been measured to be $(2.3 \pm 0.4) \times 10^{-4}$ by the Belle collaboration [185]. Based on this result a rough estimate of the nonresonant contribution can be obtained, as described in Appendix G. This results in the following estimates for the branching fractions associated with the incl. $|V_{ub}|$ templates:

1. $\mathcal{B}(B^+ \rightarrow X_u^0 \mu^+ \nu_\mu) = (1.16 \pm 0.21) \times 10^{-3}$.

2. $\mathcal{B}(B^0 \rightarrow X_u^- \mu^+ \nu_\mu) = (1.07 \pm 0.19) \times 10^{-3}$.

3. $\mathcal{B}(B^+ \rightarrow \pi^+\pi^-\mu^+\nu_\mu) = (0.9 \pm 3.1) \times 10^{-5}$.

The effect of implicitly assuming the branching fraction of the signal in the estimation of the above branching fractions will be addressed in the evaluation of the systematic uncertainties.

Charmonium decays

The charmonium background consists of decays like $B_{(s)} \to X_{c\bar{c}}Y$ where the charmonium state $X_{c\bar{c}}$ cascades or decays directly into a $\mu^+\mu^-$ pair, while Y is either a nonresonant $\pi^+\pi^-$ pair or a hadronic state decaying into a final state containing a $\pi^+\pi^-$ pair. The contribution from these decays to the selected sample is expected to be relatively small, since they are associated with small branching fractions and small selection efficiencies due to the charge isolation. However, the impact from charmonium backgrounds is non-negligible because the fit is performed in bins of q^2, and their contribution is concentrated in the bins around and above the squared J/ψ mass.

The templates for the charmonium backgrounds are extracted from the four simulations listed in Table 7.2. The first simulation describes the control channel $B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0$ and is defined in Table 4.1. The three remaining simulations are cocktails of $B_{(s)} \to X_{c\bar{c}}Y$ with $X_{c\bar{c}} = J/\psi$, $\psi(2S)$, $\chi_{c1}(1P)$ or $\chi_{c2}(1P)$ that are forced to decay into channels that produces a $\mu^+\mu^-$ pair, usually, this proceeds via $J/\psi \to \mu^+\mu^-$. After the full selection the dominant contributions to each of the three templates are:

1. $B^+ \to X_{c\bar{c}}Y$.

    (a) $B^+ \to J/\psi(\to \mu^+\mu^-)\pi^+\pi^-$ (30.1%).
    (b) $B^+ \to J/\psi(\to \mu^+\mu^-)K_1(1270)^+(\to \rho^0(\to \pi^+\pi^-)K^+)$ (25.4%).
    (c) $B^+ \to J/\psi(\to \mu^+\mu^-)K_1(1270)^+(\to K^*(1430)^0(\to K^+\pi^-)\pi^+)$ (20.1%).

2. $B^0 \to X_{c\bar{c}}Y$.

    (a) $B^0 \to J/\psi(\to \mu^+\mu^-)K_1(1270)^0(\to \rho^0(\to \pi^+\pi^-)K^0)$ (39.3%).
    (b) $B^0 \to J/\psi(\to \mu^+\mu^-)K^*(892)^+(\to K^0\pi^+)\pi^-$ (38.7%).
    (c) $B^0 \to J/\psi(\to \mu^+\mu^-)K_1(1270)^0(\to K^*(1430)^+(\to K^0\pi^+)\pi^-)$ (10.7%).

3. $B_s^0 \to X_{c\bar{c}}Y$.

    (a) $B_s^0 \to J/\psi(\to \mu^+\mu^-)\phi(1020)(\to \pi^+\pi^-\pi^0)$ (54.4%).
    (b) $B_s^0 \to J/\psi(\to \mu^+\mu^-)f^0(980)(\to \pi^+\pi^-)$ (25.0%).
    (c) $B_s^0 \to J/\psi(\to \mu^+\mu^-)\eta'(\to \rho^0(\to \pi^+\pi^-)\gamma)$ (11.4%).

The sum of all processes making up the $B^0 \to X_{c\bar{c}}Y$ cocktail is $(1.48 \pm 0.39) \times 10^{-4}$ of the total B^0 decay width. For the $B^+ \to X_{c\bar{c}}Y$ cocktail, it is $(2.44 \pm 0.35) \times 10^{-4}$ of the total B^+ decay width, and for the $B_s^0 \to X_{c\bar{c}}Y$ cocktail, it is $(6.14 \pm 0.31) \times 10^{-5}$ of the total B_s^0 decay width.

The resulting templates are illustrated in Fig. 7.8. The control channel is almost indistinguishable from the signal channel in the $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$ distributions, however, as its branching fraction is approximately six times smaller than the one of the signal, its
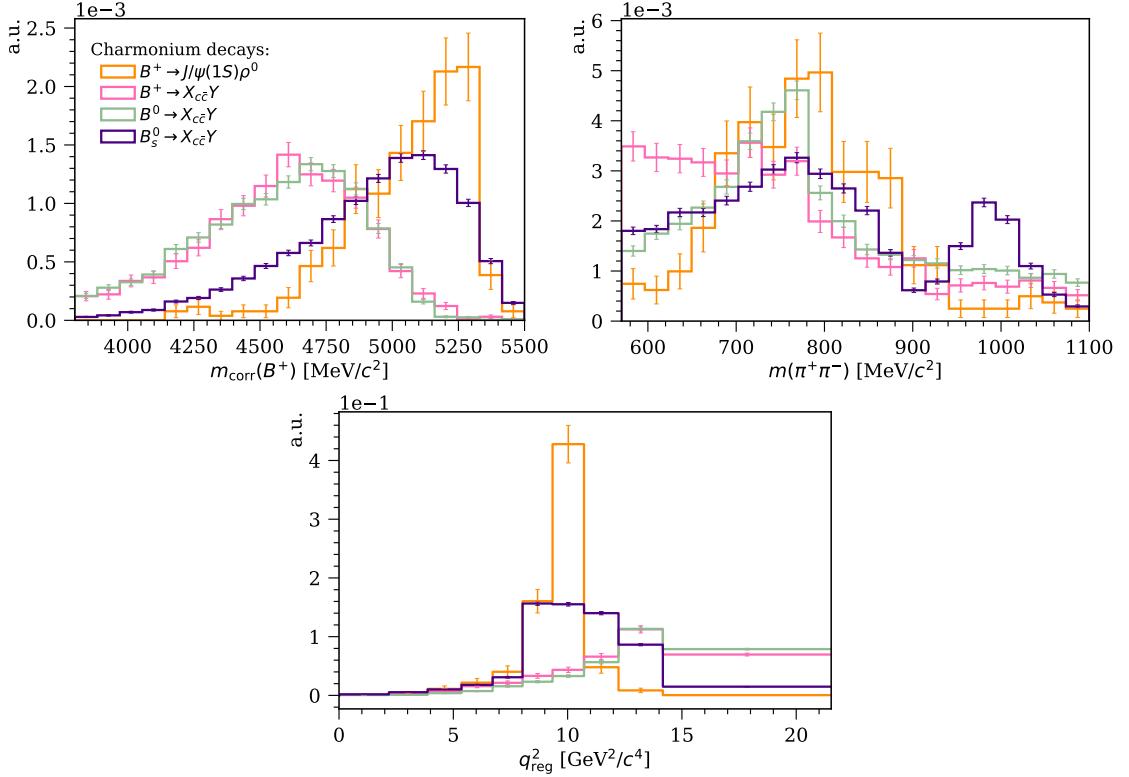
Figure 7.8 – Templates for the charmonium decays projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

contribution is expected to be small. The other templates can be distinguished from the signal in either the $m_{\mathrm{corr}}(B^+)$ or $m(\pi^+\pi^-)$ distribution. Moreover, their q^2 distributions are concentrated around and above $q^2 \sim 8 - 9 \, \mathrm{GeV^2/c^4}$.

Leptonic D_s^+ decay

The leptonic D_s^+ template is extracted from a simulation of the decay processes specified below together with their respective branching fractions [15]:

$$B_s^0 \to D_s^+ \pi^+ \pi^- \pi^-, \quad \mathcal{B} = (6.1 \pm 1.0) \times 10^{-3}\,.$$

$$D_s^+ \to \mu^+ \nu_\mu\,, \quad \mathcal{B} = (5.43 \pm 0.15) \times 10^{-3}\,.$$

$$\mathcal{B}_{\mathrm{tot}} = (3.31 \pm 0.55) \times 10^{-5}\,.$$

While most of the considered background processes, except for the charmonium backgrounds, are semileptonic B decays, this background is a hadronic B_s^0 decay where the μ^+ is produced in the leptonic decay of the D_s^+ meson and its q^2 distribution is therefore concentrated around $m(D_s^+)^2$. The resulting template is illustrated in Fig. 7.9. Since the final state is
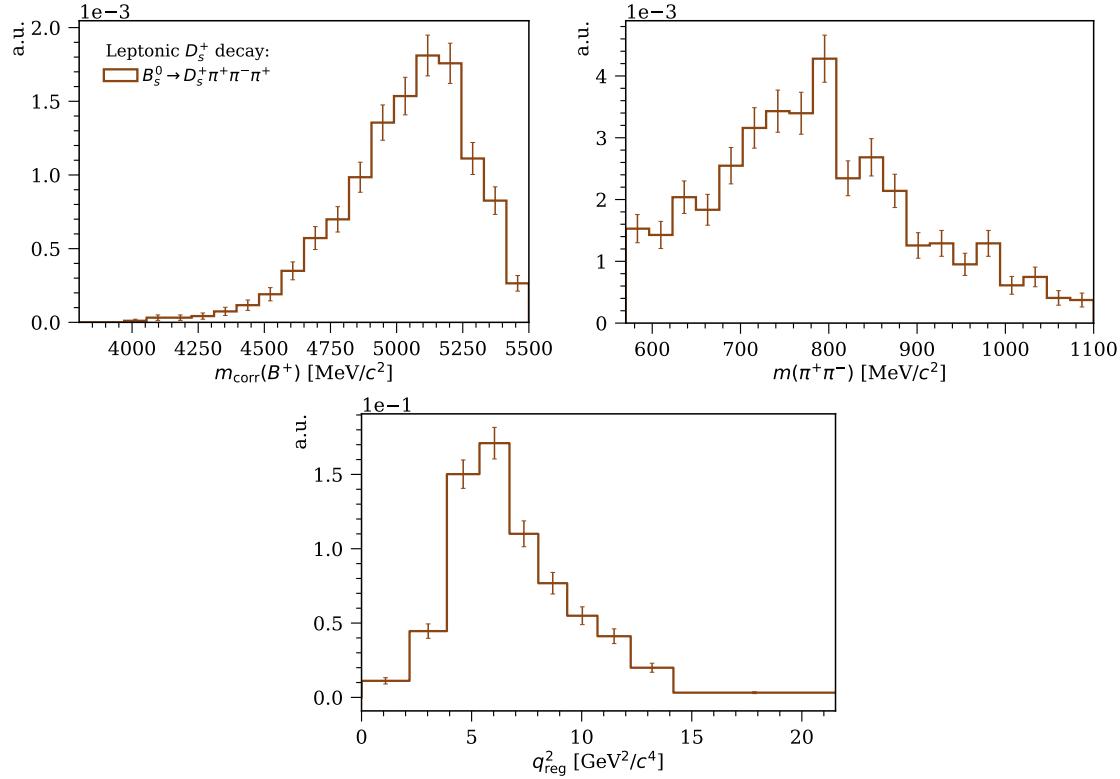
Figure 7.9 – Templates for the leptonic D_s^+ decay projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

produced by the heavier B_s^0 and only one additional particle is missing with respect to the signal, its $m_{\mathrm{corr}}(B^+)$ distribution is similar to the one of the signal. In addition to this, the decay $B_s^0 \to D_s^+\pi^+\pi^-\pi^-$ often proceeds via the process $B_s^0 \to D_s^+\rho^0\pi^-$, and its $m(\pi^+\pi^-)$ distribution is therefore also rather similar to the one of the signal, while its q^2 distribution is concentrated around 3–$11\,\mathrm{GeV}^2/c^4$.

7.1.2 Combinatorial background

In contrast to the normalisation channel, the side-band regions above and below the ρ^0-mass window contain various physics processes and can therefore not be used to model the CombBg contribution. Instead, wrong-sign candidates selected in data with the `B2XuMuNuBu2Rho_WSLine` stripping line defined in Table 4.2 are used. However, the WS-MS candidates, *i.e.* $\mu^+\pi^-\pi^-$, are not purely CombBg. They can originate from physics processes associated with at least one extra charged particle, for instance, the decay $B^0 \to D^-(\to K^+2\pi^-)\mu^+\nu_\mu X$ with a rather large branching fraction of about 1%. Although charge isolation discriminates against these backgrounds, they cannot be completely eliminated. On the other hand, WS-SS candidates, *i.e.* $\mu^+\pi^+\pi^+$, are less likely to contain physics processes, in particular, after applying charge isolation. The
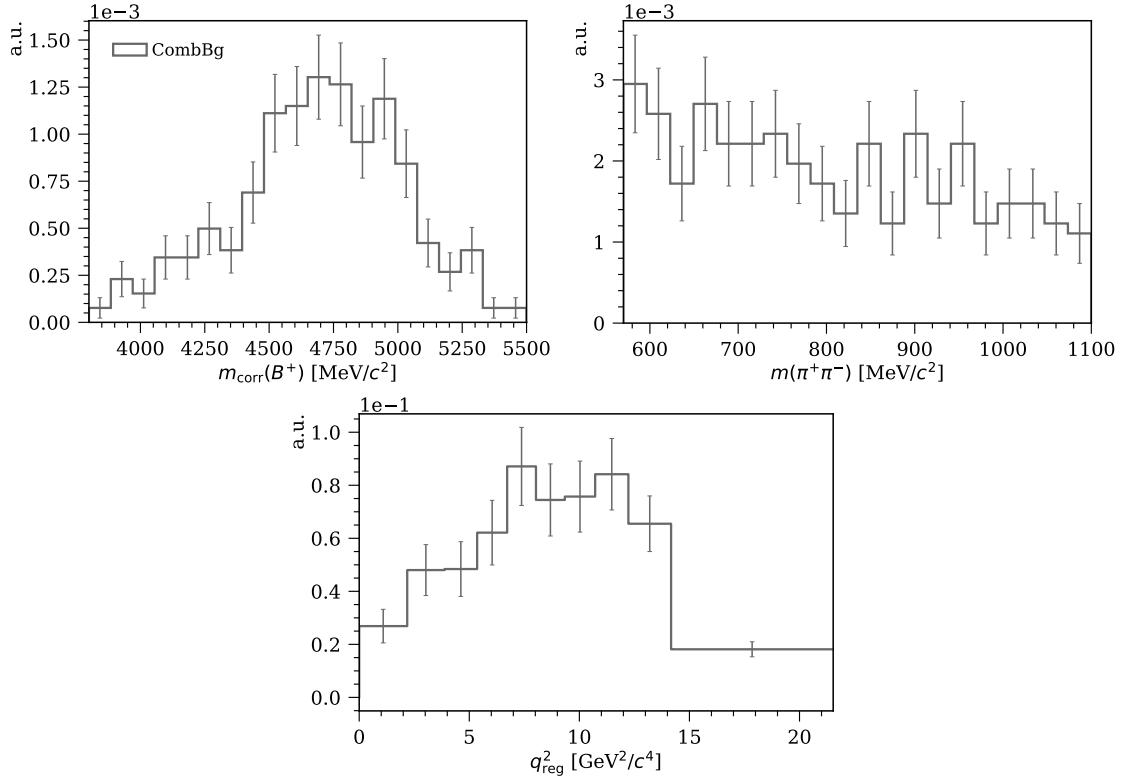
Figure 7.10 – Combinatorial background template projected in $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The reconstructed q^2 distribution (bottom) is illustrated with the non-uniform q^2 binning scheme defined in Table 5.2.

physics backgrounds in the WS-SS sample must have at least two additional charged particles due to the surplus charge of ± 3 of the final state, *e.g.* originating from the decay $B^+ \to D^0(\to 2\pi^+2\pi^-)\mu^+\nu_\mu X$ with a branching fraction of about 0.1%. In general, the branching fractions of these decays are small, and they are less likely to survive charge isolation. For this reason the WS-SS sample is used to model the CombBg. However, due to the pre-scaling of the `B2XuMuNuBu2Rho_WSLine` stripping line, the number of WS-SS candidates after the full selection is not sufficient for extracting a fitting template. Thus, both the 2017 and 2018 WS-SS data are used to extract the CombBg template illustrated in Fig. 7.10.

7.1.3 MisID backgrounds

For the signal channel, both the misIDμ and misIDπ backgrounds must be taken into account. In order to estimate their contributions and extract their templates, the same approach as developed for the normalisation channel, described in Sec. 6.2.3, is used. The details of the misID estimation, *i.e.* the efficiency matrices, binning schemes, number of measured and unfolded true particle species can be found in Appendix F. In the following two paragraphs, the misID background estimation is illustrated for the full q^2 region,

Table 7.3 – True composition of the measured μ candidates in data. The values given are the individual contributions to the measured μ candidates from true μ, π, K and p and from the total misID background consisting of true π, K and p.

measured	true μ	true π	true K	true p	total misID
μ	98.99%	0.46%	0.54%	0.01%	1.01%

however, to obtain correct misID templates for the binned signal fit this procedure is repeated in each q^2 bin. Moreover, the statistical uncertainties of the misID yields are taken to be the Poisson uncertainties, which is not entirely correct, since several sources contribute to the statistical uncertainty, *i.e.* the limited statistics of the reference sample as well as the calibration sample and the unfolding procedure. This makes the estimation of the statistical uncertainty non-trivial. However, in any case, the statistical uncertainties of the misID yields are assumed to be negligible with respect to their systematic uncertainties estimated in Appendix H.

MisIDμ

The efficiency matrix defined in Eq. 6.9 is computed with PIDCALIB by applying the same particle regions as used for the normalisation channel (see Table 6.7). The binning in p, η and `nSPDHits` is slightly adjusted (see Table F.1) to include high momentum candidates that fall outside the binning scheme used for the normalisation channel. The efficiencies are assigned to the reference sample that consists of signal candidates from data passing the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line (see Table 4.2) and the full signal selection (see Tables 4.4 and 7.1), without applying the requirements on the PID information of the μ^+ candidate. Of all the μ^+ candidates in the reference sample, 9% do not fall into any of the defined particle regions. The effect of this on the estimated misIDμ background is estimated in Appendix H, while the final statistical uncertainty is estimated in Sec. 7.5.4.

The Baysian unfolding algorithm of ROOUNFOLD uses the efficiency matrix in Eq. F.1 and the measured number of particle species in Eq. F.2 to unfold the true particle composition of the reference sample in Eq. F.3. The obtained true composition of the measured μ candidates is shown in Table 7.3, and the resulting misIDμ template is illustrated in Fig. 7.11. The main contribution is coming from misidentified π and K, while the contribution from p is negligible. The total misIDμ contribution to the selected sample is determined to be:

$$N_{\mathrm{misID}\mu} = 1555 \pm 39\,, \tag{7.2}$$

where the uncertainty is taken to be the Poisson uncertainty of the misIDμ yield.
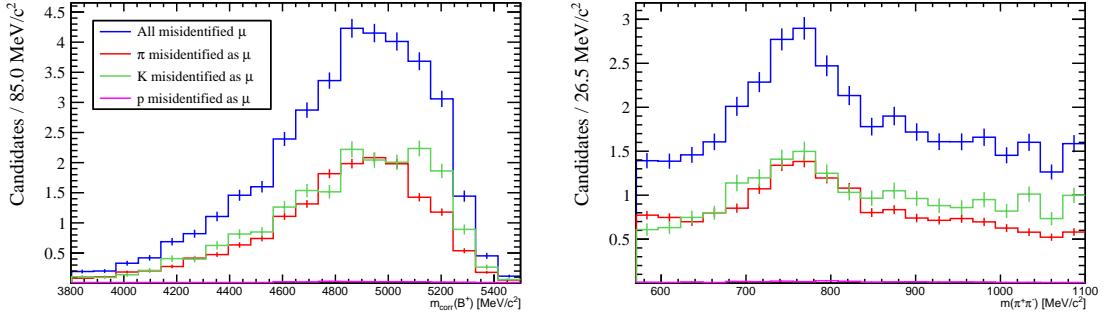
Figure 7.11 – Template for the misIDμ background projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). The total misIDμ background (blue) consists of π (red), K (green) and p (pink) misidentified as μ.

Table 7.4 – Particle regions used to estimate the misIDπ^+ background. In addition to the PID requirements, the muon is required to be within the acceptance of the Muon System, and the number of hits in the SPD detector is required to be `nSPDHits` < 450.

Particle	Region
π^+	`PIDmu` < 2, `PIDK` < -2.0, `ProbNNp` < 0.7
μ^+	`IsMuon` $= 1$, `PIDmu` > 2
K^+	`ProbNNK` > 0.1, (`PIDp` − `PIDK`) < 0.1, `IsMuon` $\neq 1$, `PIDK` > -2.0
p^+	`ProbNNp` > 0.1, (`PIDp` − `PIDK`) > 0.1, `IsMuon` $\neq 1$, `PIDK` > -2.0

MisIDπ

For the signal candidates $B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$, a stricter PID selection is applied to π^- than to π^+, as described in Sec. 4.4.4. The misIDπ^+ and misIDπ^- backgrounds are therefore estimated individually and later combined into one misIDπ template.

The particle regions used to estimate the misIDπ^+ and misIDπ^- backgrounds are summarised in Tables 7.4 and 7.5, respectively. The $\pi^\pm$ regions are defined by the PID requirements imposed on the $\pi^\pm$ candidates by the `B2XuMuNuBu2Rho_Line` stripping line (see Table 4.3) and by the preselection (see Table 4.4).

The efficiency matrices for the misIDπ^+ and misIDπ^- backgrounds are computed with the particle regions defined in Tables 7.4 and 7.5, respectively, together with the binning schemes defined in Table F.2. The reference samples are obtained by applying the `StrippingB2XuMuNuB2Pi_Line` to data (see Table 4.3). This line selects $B^0 \to \pi^-\mu^+\nu_\mu$ candidates, and by adding a $\pi^\pm$ without any PID requirements, a $B^+ \to \rho^0\mu^+\nu_\mu$ candidate is formed. Moreover, the full signal selection (see Tables 4.4 and 7.1) is applied, except for the requirements on the PID information of the considered $\pi^\pm$ candidate. In the reference sample used to estimate the misIDπ^+ background, about 5% of all π^+ candidates do not fall into any of the particle regions, while for the reference sample used to estimate the misIDπ^- background, about 12% of all π^- candidates do not fall into any of the particle

Table 7.5 – Particle regions used to estimate the misIDπ^- background. In addition to the PID requirements, the muon is required to be within the acceptance of the Muon System, and the number of hits in the SPD detector is required to be `nSPDHits` < 450.

Particle	Region
π^-	`IsMuon` $\neq 1$, `PIDmu` < 0, `PIDK` < -2.0, `ProbNNp` < 0.7
μ^-	`IsMuon` $= 1$, `PIDmu` > 0
K^-	`ProbNNK` > 0.1, (`PIDp` $-$ `PIDK`) < 0.1, `IsMuon` $\neq 1$, `PIDK` > -2.0
p^-	`ProbNNp` > 0.1, (`PIDp` $-$ `PIDK`) > 0.1, `IsMuon` $\neq 1$, `PIDK` > -2.0

Table 7.6 – True composition of the measured π^+ and π^- candidates in data. The values given are the individual contributions to the measured π^+ and π^- candidates from true π, μ, K and p and from the total misID background consisting of true μ, K and p.

measured π	true π	true μ	true K	true p	total misID
π^+	98.90%	0.19%	0.79%	0.12%	$1.10\ \%$
π^-	99.37%	0.47%	0.05%	0.11%	0.63%

regions. The maximal effect of this is estimated in Appendix H, while the final systematic uncertainty is estimated in Sec. 7.5.4.

The Bayesian unfolding algorithm of RooUnfold uses the efficiency matrices in Eqs. F.4 and F.5 and the measured number of particle species in Eqs. F.6 and F.7 to unfold the true particle composition of the reference sample in Eqs. F.8 and F.9.

The obtained true compositions of the measured π^+ and π^- candidates are shown in Table 7.6, and the resulting misIDπ^+ and misIDπ^- templates are illustrated in Fig. 7.12. For the misIDπ^+ background, the main contribution is coming from true K^+, while for the misIDπ^- background the main contribution is coming from true μ^-, which is due to the charmonium backgrounds $B \to J/\psi(\to \mu^+\mu^-_{\text{misID}\pi})\pi^+ X$. Taking into account the relative proportions of misIDπ^+ and misIDπ^- in data of 64% and 36%, respectively, they are combined into one misIDπ template with a total contribution to the selected sample of:

$$N_{\text{misID}\pi} = 2675 \pm 52 \,, \tag{7.3}$$

where the uncertainty is the Poisson uncertainty of the misIDπ yield.

7.1.4 Estimating background from fake tracks

A potential background in the signal channel as well as in the normalisation channel consists of decay candidates formed with fake tracks, also known as ghosts. In order to estimate the contribution from this background, the simulations of the signal and normalisation channel are studied. In simulation, a ghost track is assigned a true particle identity (`TRUEID`) of
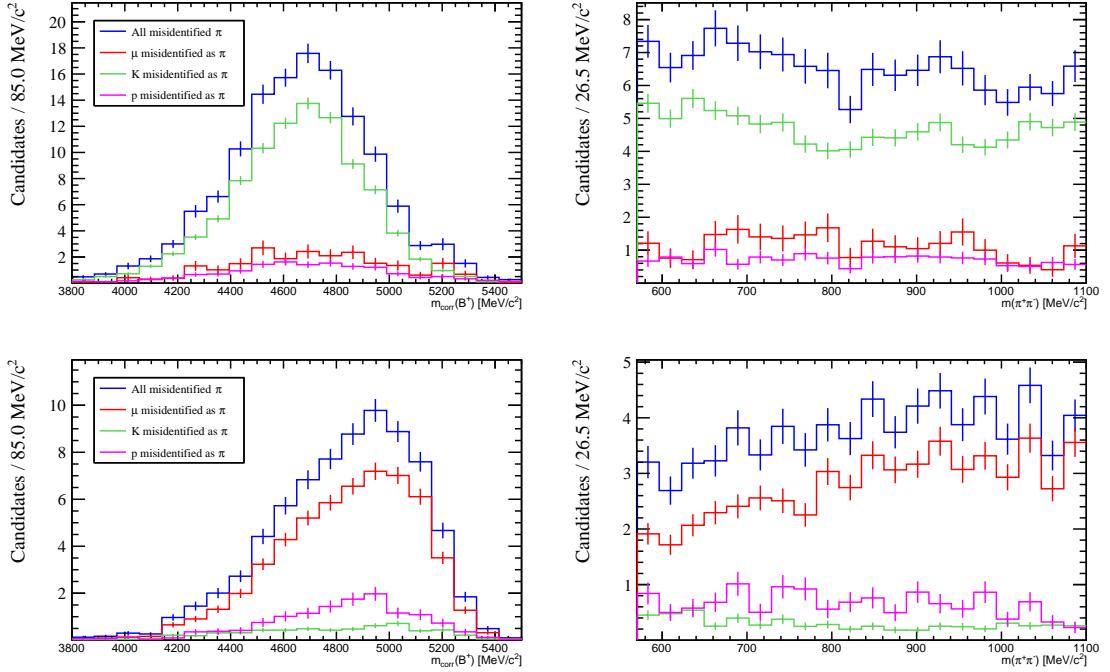
Figure 7.12 – Template for the misIDπ^+ (top) and misIDπ^- (bottom). The total misIDπ background (blue) consists of true μ (red), K (green) and p (pink) misidentified as π.

zero, however, it can happen that a true particle is, in fact, misidentified as a ghost. For instance, if the track of a particle is reconstructed with some hits made by another nearby passing particle. It is therefore necessary to asses the contribution from real ghosts.

After applying the full signal selection, 1.89% of the signal sample consists of ghost candidates, *i.e.* signal candidates with at least one particle with `TRUEID` = 0. Of these, 3.6% are μ^+ candidates, 52.1% are π^+ candidates and 44.3% are π^- candidates. Thus, most ghost candidates are formed with a $\pi^\pm$ candidate with `TRUEID` = 0. As expected, ghost candidates and truth-matched signal candidates can be distinguish by their ghost probability (GhostProb) distribution. This is illustrated in Fig. 7.13 where the truth-matched signal candidates peak at zero, while the distribution of ghost candidates with a π^+ candidate of `TRUEID` = 0, has a much broader distribution.

However, the $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$ distributions of the ghost candidates follow the ones of the truth-matched signal candidates, as illustrated in Fig. 7.14. Even when requiring GhostProb(π^+_{track}) > 0.05 where the distribution of the true signal candidates drops off, the $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$ distributions still agree with the true signal distributions. This indicates that the majority of ghost candidates are, in fact, true signal candidates.

In order to test this hypothesis, the contribution from ghosts in the normalisation channel is also considered. After applying the full selection of the normalisation channel, the NORMBu simulation, defined in Table 6.3, contains 1.71% ghost candidates with 6.5%
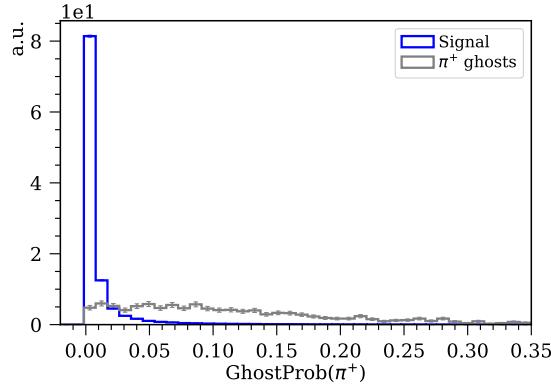
Figure 7.13 – Comparison of the GhostProb(π^+) distributions of truth-matched signal candidates (blue) and ghost candidates where the π^+ candidate has `TRUEID = 0` (grey).
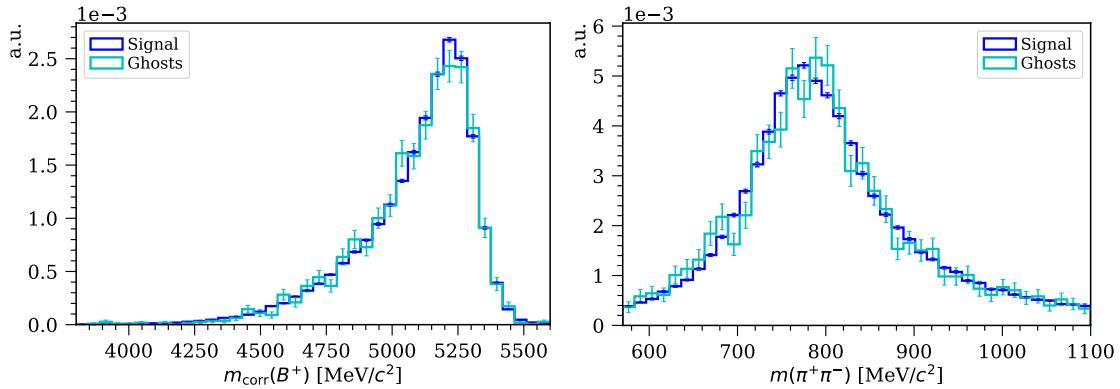
Figure 7.14 – Comparison of the $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right) distributions of truth-matched signal candidates (blue) and ghost candidates (cyan).

being μ^+ candidates, 50.4% being π^+ candidates and 43.0% being π^- candidates. The $m_{\text{corr}}(B^+)$ distribution of the ghost candidates is the same as for the truth-matched candidates, however, the $m(\pi^+\pi^-)$ distribution of the ghost candidates is broader than for the truth-matched candidates, as illustrated in Fig. 7.15. To estimate the proportion of real ghosts and fake ghosts, a fit is performed to the $m(\pi^+\pi^-)$ distribution of the ghost candidates in the NORMBu sample. The contribution from real ghosts are modelled with a first degree polynomial, while the contribution from fake ghosts is modelled with two Gaussians. The fit is an unbinned extended maximum likelihood fit provided by the ROOFIT framework, described in Sec. 4.6.2. The result of the fit, illustrated in Fig. 7.15, is:

$$N_{\text{true ghost}} = 28 \pm 38 \,, \tag{7.4}$$

$$N_{\text{fake ghost}} = 2888 \pm 103 \,. \tag{7.5}$$

Thus, the contribution from real ghosts to the sample is consistent with zero. Within one standard deviation the maximum contribution from real ghosts corresponds to $\sim 2\%$,
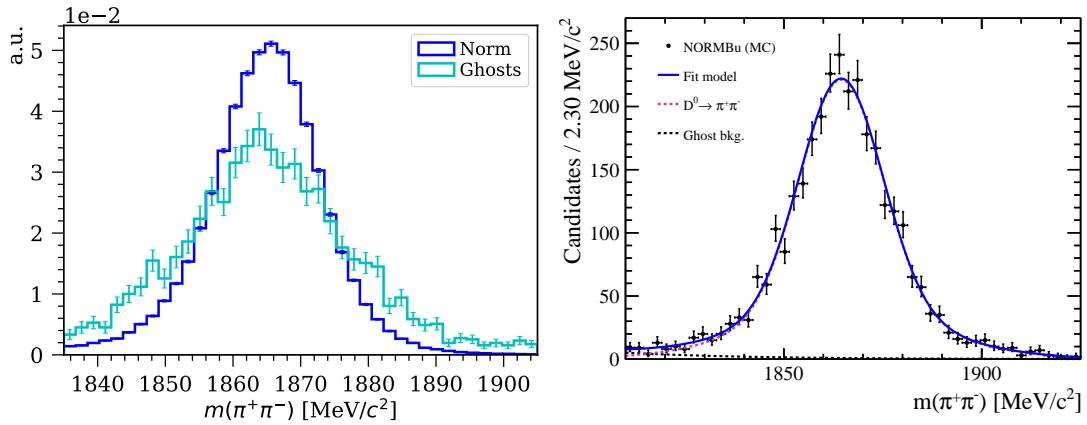
Figure 7.15 – Comparison between the $m(\pi^+\pi^-)$ distributions (left) of truth-matched NORMBu candidates (blue) and ghost candidates (cyan). A fit (blue) to the $m(\pi^+\pi^-)$ distribution (right) of the ghost candidates in the NORMBu simulation (black markers) with real ghost candidates modelled with a polynomial (dashed black) and fake ghost candidates modelled with two Gaussians (dashed pink).

however, as ghost candidates contribute with only 1.71% to the total NORMBu sample, their actual contribution to the normalisation channel is less than 0.1%. In conclusion, the ghost background contributions to the signal and normalisation channel are found to be negligible.

7.2 Template corrections

For some of the simulated decay processes, the distributions of $m_{\mathrm{corr}}(B^+)$, $m(\pi^+\pi^-)$ and q^2 are not well modelled leading to wrong template shapes and incorrect Gaussian constraints. In these situations, corrections based on data or improved theoretical models are developed and applied to the relevant templates.

All fitting templates extracted from simulations are re-weighted with the data/MC weights $w_{\mathtt{nTracks}} \times w_{\mathrm{mva}}$ computed in Sec. 4.6.3. Although these weights are developed for the signal channel, the mis-modelling of $\mathtt{nTracks}$ and $\mathrm{mva_{out}}(1^{\mathrm{st}}\ \mathrm{track} + \mathrm{vtx}_{\rho^0})$ is expected to be similar for all the simulated decays. Like for the normalisation channel, the true q^2 distributions of the decays $B^+ \to \bar{D}^0\mu^+\nu_\mu$ and $B^+ \to \bar{D}^{*0}\mu^+\nu_\mu$ are re-weighted according to the CLN form factor parameterisation as described in Sec. 6.2.1. These corrections and their associated uncertainties have a very small effect on the shapes of the five excl. $|V_{cb}|^{B^+}$ templates, however, they affect the relative contribution of the templates in each q^2 bin, as shown in Fig. 6.1.

In addition to this, the $m(\pi^+\pi^-)$ line-shapes of the signal and incl. $|V_{ub}|^{B^{+,0}}$ samples are corrected to a model where the ρ^0 line-shape is described with a Gounaris-Sakurai function and where the interference between the ρ^0 and ω mesons is taken into account. This will

149

be explained in Sec. 7.2.1. Furthermore, the $B^+ \to \eta' \mu^+ \nu_\mu$ fitting template is validated by comparing the decay $\eta' \to \pi^+ \pi^- \gamma$ in data and in simulation, as described in Sec. 7.2.2. Finally, the excl. $|V_{cb}|^{B^+}$ $D^0 \to K_S^0 \pi^+ \pi^- \pi^0$ decay, modelled without taking intermediate resonances of the four-body D^0 decay into account, is corrected to data, as described in Sec. 7.2.3.

7.2.1   Correcting the $m(\pi^+\pi^-)$ line-shape

The parameterisation of the ρ^0 line-shape is complicated by its large width and its dependence on the production process [15]. As it is not well described by a relativistic Breit-Wigner (BW), models with more shape parameters have been developed. In addition to this, interference can arise between the ρ^0 and ω mesons affecting the $m(\pi^+\pi^-)$ line-shape. Thus, the interference between the narrow ($\Gamma_{\omega \to \pi\pi} \sim 9\,\mathrm{MeV}$) resonant $\omega \to \pi^+\pi^-$ and broad ($\Gamma_{\rho^0 \to \pi\pi} \sim 147.4\,\mathrm{MeV}$) resonant $\rho^0 \to \pi^+\pi^-$ amplitudes results in an abrupt drop in the $m(\pi^+\pi^-)$ distribution in the region between the nominal ρ^0 and ω masses, as illustrated in Fig. 7.16. The interference process is not entirely understood on a fundamental level [188], and its effect has mostly been modelled from measurements of the $e^+e^- \to \pi^+\pi^-$ cross section, see *e.g.* Refs. [189, 190].

The effect of the $\rho^0 - \omega$ interference on the $m(\pi^+\pi^-)$ line-shape is only accounted for in the simulations of the excl. $|V_{cb}|^{B^+,B^0}$ processes involving a D meson three-body decay and in the simulation of the $|V_{cb}|^{B_s^0}$ process with the $D_s^+ \to \pi^+\pi^-\pi^+, K^+\pi^+\pi^-, K^+K^+\pi^-$ decays. For all other simulated decay processes used in this analysis, the ρ^0 meson is modelled with a BW line-shape without taking the $\rho^0 - \omega$ interference into account. To obtain a better description of the $m(\pi^+\pi^-)$ line-shape, the model presented in Ref. [189] is used, which describes a measurement of the $e^+e^- \to \pi^+\pi^-$ cross section performed by the CMD-2 experiment. To model the ρ^0 line-shape the commonly used Gounaris-Sakurai (GS) parameterisation [191] is employed:

$$\mathrm{GS}_{\rho^0}(m_{12}) = \frac{M_{\rho^0}^2 (1 + d \times \Gamma_{\rho^0}/M_\rho)}{M_{\rho^0}^2 - m_{12}^2 + f(m_{12}) - iM_{\rho^0}\Gamma_{\rho^0}(m_{12})}\,, \tag{7.6}$$

where m_{12} is the invariant mass of the two-pion final state, and M_{ρ^0} and Γ_{ρ^0} are the nominal mass and width of the ρ^0 meson, respectively. The constant denoted d depends on M_{ρ^0}, the nominal π mass and its momentum at resonance, $p_\pi(M_{\rho^0})$. The function $f(m_{12})$ depends on the same parameters as d as well as Γ_{ρ^0} and the mass-dependent momentum $p_\pi(m_{12})$ in the center of mass frame of the resonance. Finally, the mass-dependent ρ^0 width is taken from Ref. [192]:

$$\Gamma_{\rho^0}(m_{12}) = \Gamma_{\rho^0} \frac{M_{\rho^0}}{m_{12}} \left(\frac{p_\pi(m_{12})}{p_\pi(M_{\rho^0})} \right)^{2L_{\mathrm{eff}}+1} B_{L_{\mathrm{eff}}}(p_\pi(M_{\rho^0}), p_\pi(m_{12}), R)^2\,, \tag{7.7}$$

where R is a barrier factor radius and L_{eff} is the relative angular momentum of the ρ^0 meson with respect to the $\mu\nu_\mu$ system. The allowed values are $L_{\text{eff}} = 0, 1, 2$. The Blatt–Weisskopf function $B_{L_{\text{eff}}}$ [193, 194] accounts for the finite size of the decaying hadron. It depends on the value of L_{eff}, *i.e.*:

$$B_{L_{\text{eff}}=0} = 1 \,, \tag{7.8}$$

$$B_{L_{\text{eff}}=1} = \sqrt{\frac{1 + R^2 p_\pi(M_{\rho^0})^2}{1 + R^2 p_\pi(m_{12})^2}} \,, \tag{7.9}$$

$$B_{L_{\text{eff}}=2} = \sqrt{\frac{R^4 p_\pi(M_{\rho^0})^4 + 3R^2 p_\pi(M_{\rho^0})^2 + 9}{R^4 p_\pi(m_{12})^4 + 3R^2 p_\pi(m_{12})^2 + 9}} \,. \tag{7.10}$$

$$\tag{7.11}$$

Including the interference between the ρ^0 and ω meson, the resulting $m(\pi^+\pi^-)$ line-shape is [189]:

$$\text{GS}_{\rho^0-\omega}(m_{12}) = \text{GS}_{\rho^0}(m_{12})(1 + \delta \frac{m_{12}^2}{M_\omega^2} \text{BW}_\omega(m_{12})) \,, \tag{7.12}$$

where δ is the relative contribution of ω with respect to ρ^0. The ω parameterisation BW_ω is a simple BW model that depends on the nominal mass and width of the ω meson.

To set δ, M_{ρ^0} and Γ_{ρ^0}, the results of Ref. [189] based on the GS model are used. The values of Γ_ω, M_ω and M_π are taken from the PDG [15], while R is taken from Ref. [192]. The fraction of decays with any of the three L_{eff} values are not known. Initially, $m(\pi^+\pi^-)$ line-shape is defined with $L_{\text{eff}} = 1$ and denoted $\text{GS}_{\rho^0-\omega}^{\text{L}_{\text{eff}}=1}$. The resulting line-shape is compared to the BW line-shape without $\rho^0 - \omega$ interference in Fig. 7.16. The effect of the interference is a steep drop on the right side of the ρ^0 nominal mass.

The $m(\pi^+\pi^-)$ distribution of the simulated signal is corrected to the GS line-shape model with $\rho^0 - \omega$ interference by weights. Moreover, background templates with a significant $\rho^0 \to \pi^+\pi^-$ contribution are also corrected. This includes the two inclusive $B \to X_u\mu^+\nu_\mu$ templates. It should be noted that it is not necessarily true that the same correction applies to the different decay processes, however, due to the limited knowledge regarding these differences, the correction is assumed to be the same in this analysis. Since the amplitude of the $\omega \to \pi^+\pi^-$ process is included in the $\rho^0 - \omega$ interference description, these decays are removed from the incl. $|V_{ub}|$ simulations as well as in the exclusive simulation of the $B^+ \to \omega\mu^+\nu_\mu$ decay.

As explained in Sec. 7.3.1, the final $\text{GS}_{\rho^0-\omega}$ line-shape correction is optimised in the fit. The relative angular momentum is fixed to the optimal value determined by the fit corresponding to $L_{\text{eff}} = 1.44$, and instead of fixing the mixing parameter δ to the CMD-2
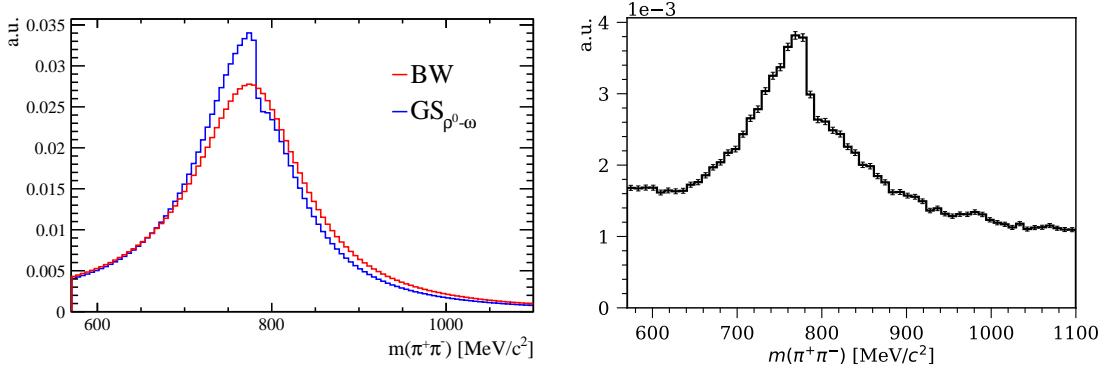
Figure 7.16 – $m(\pi^+\pi^-)$ line-shape (left) resulting from modelling the ρ^0 line-shape with a BW (red) and from modelling it with a GS with $L_{\text{eff}} = 1$ where $\rho^0 - \omega$ interference is taken into account (blue), and the $m(\pi^+\pi^-)$ line-shape in data including background (right).
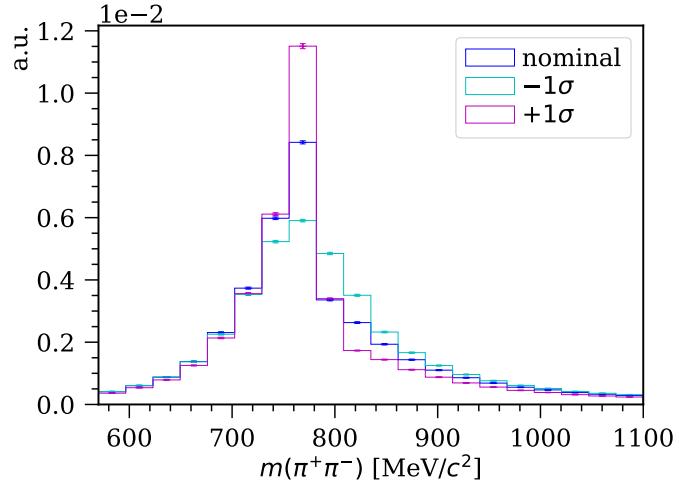
Figure 7.17 – Signal template after correcting $m(\pi^+\pi^-)$ to the $\text{GS}_{\rho^0-\omega}$ model with $L_{\text{eff}} = 1.44$ and $\delta = 0.01$ (blue) and the corresponding $\pm 1\sigma$ (magenta/cyan) shape variation templates defined by $\delta = 0.0$ and $\delta = 0.02$, respectively.

result, it is allowed to vary in the fit. Its nominal value is $\delta = 0.01$ with a shape variation of $\pm 1\sigma$ defined by $\delta = 0.0$ and $\delta = 0.02$. This is the final $m(\pi^+\pi^-)$ line-shape correction denoted $\text{GS}_{\rho^0-\omega}^{L_{\text{eff}}=1.44}$. The corrected $m(\pi^+\pi^-)$ line-shape of the signal template as well as its $\pm 1\sigma$ shape variation are shown in Fig. 7.17.

7.2.2   Verifying the $B^+ \to \eta'(\to \pi^+\pi^-\gamma)\mu^+\nu_\mu$ simulation

From available measurements of the $\eta' \to \pi^+\pi^-\gamma$ decay, the relative contributions from nonresonant $\eta' \to \pi^+\pi^-\gamma$ and resonant $\eta' \to \rho^0\gamma$ processes are not known [15]. However, in the simulation of $B^+ \to \eta'(\to \pi^+\pi^-X)\mu^+\nu_\mu$ only the resonant $\eta' \to \rho^0\gamma$ decay channel is included. Thus, to verify the simulation, a comparison to data is performed.
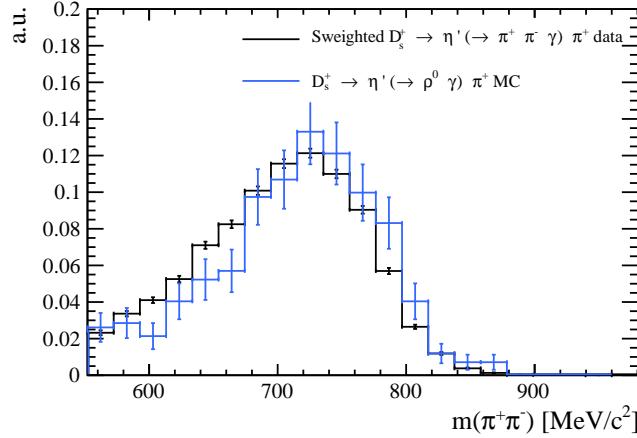
Figure 7.18 – Distributions of $m(\pi^+\pi^-)$ in sWeighted $\eta' \to \pi\pi\gamma$ data (black) and simulated $\eta' \to \rho^0\gamma$ decays (blue).

In practice, a simulation of the decay $D_s^+ \to \eta'\pi^+$ with $\eta' \to \rho^0\gamma$ is compared to sWeighted (see Sec. 4.6.2) data selected with a dedicated stripping line. The resulting $m(\pi^+\pi^-)$ distributions in simulation and in data are illustrated in Fig. 7.18. As the two distributions follow each other, one can conclude that most $\eta' \to \pi^+\pi^-\gamma$ decays proceed by the resonant $\eta' \to \rho^0\gamma$ process, and thus, the simulation is assumed to describe data reasonably well.

7.2.3 Correcting the $B^+ \to \bar{D}^0(\to K_S^0\pi^+\pi^-\pi^0)\mu^+\nu_\mu$ simulation

The resonant structure of the decay $D^0 \to K_S^0\pi^+\pi^-\pi^0$ has not been determined by previous measurements. For this reason, the simulation of the decay $B^+ \to \bar{D}^0(\to K_S^0\pi^+\pi^-\pi^0)\mu^+\nu_\mu$ does not include any resonantly produced $\pi^+\pi^-$ pair. To asses the impact of this, a comparison between simulation and data is performed.

In practice, a simulation of the decay $B^0 \to D^{*-}\mu^+\nu_\mu$ with $D^{*-} \to D^0\pi^+$ where $D^0 \to K_S^0\pi^+\pi^-\pi^0$ is compared to sWeighted (see Sec. 4.6.2) data selected with a dedicated stripping line. The resulting $m(\pi^+\pi^-)$ distributions in simulation and in data are illustrated in Fig. 7.19. The distribution in data does not indicate a strong contribution from resonantly produced $\pi^+\pi^-$ pairs. However, the two distributions are slightly different. The simulated $m(\pi^+\pi^-)$ distribution is therefore corrected to the one of data. This procedure introduces an uncertainty in the $m(\pi^+\pi^-)$ mass shape. To model this uncertainty the difference between the weighted and unweighted $m(\pi^+\pi^-)$ distribution is used to define a shape variation of $\pm 1\sigma$, as illustrated in Fig. 7.17.
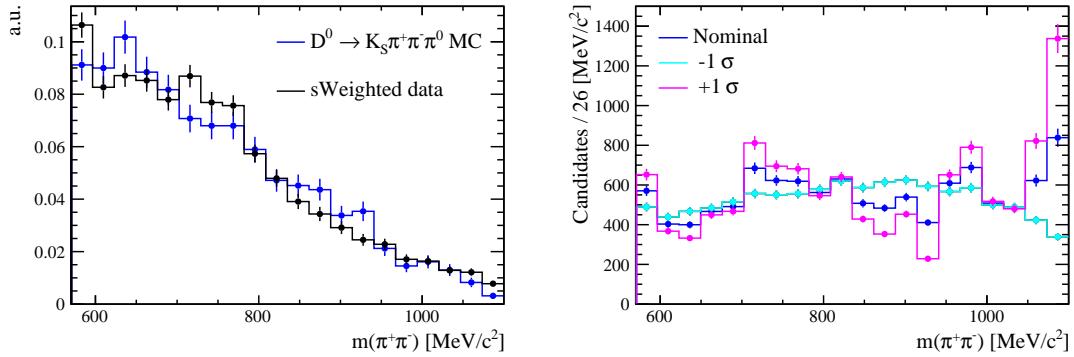
Figure 7.19 – Distributions of $m(\pi^+\pi^-)$ (left) in sWeighted data (black) and simulation (blue) for the decay $B^0 \to D^{*-}(\to D^0(\to K_S^0\pi^+\pi^-\pi^0)\pi^-)\mu^+\nu_\mu$ used to compute data/MC weights. The $|V_{cb}|^{B^+}$ $D^0 \to K_S^0\pi^+\pi^-\pi^0$ template (right) with the nominal (blue), $+1\sigma$ (pink) and -1σ (cyan) weights applied.

7.3   Signal channel fit

To extract the yield of the signal channel, a binned maximum likelihood fit is performed with HISTFACTORY to the $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$ distributions of the full 2018 data sample. The fit is performed in ten bins of q^2 defined in Sec. 5.5. For comparison and validation the fit is also performed in the full q^2 region. The fit model is build with 26 different templates presented in Secs. 7.1.1, 7.1.2 and 7.1.3, and their statistical uncertainties associated with the limited MC statistics are propagated to the final fit uncertainties via the Beeston-Barlow light method.

7.3.1   Establishing the nominal fit

The fitting templates described in Sec. 7.2 are corrected and their associated shape uncertainties are included in the fit. For the $m(\pi^+\pi^-)$ line-shape correction defined in Sec. 7.2.1 the model parameters are varied within their reported uncertainties and the largest impact on the shape is found to be caused by the relative angular momentum L_{eff}. The composition of decay candidates with $L_{\text{eff}} = 0, 1$ or 2 for the considered decays are not known. For this reason, the shape uncertainty of $\pm 1\sigma$ is initially defined by the two extreme situations $L_{\text{eff}} = 0$ and $L_{\text{eff}} = 2$, while $L_{\text{eff}} = 1$ corresponds to the nominal correction. This $m(\pi^+\pi^-)$ line-shape correction is denoted $\text{GS}_{\rho^0-\omega}^{L_{\text{eff}}=1}$. Initially, only the yields of the misID templates are constrained in the fit. This is referred to as *the initial fit*. The fit converges in the full q^2 region as well as in all ten bins of q^2. The result of the fit in the full q^2 region is projected in $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$, as illustrated in Fig. 7.20.

To improve and validate the fit, different studies are carried out. This includes *post fit comparisons* where a distribution of a variable X obtained with the fit model is compared to the same distribution in data, *i.e.* the yields obtained in the fit are used to scale the distribution of X for each fit component, and by stacking all distributions the fit model
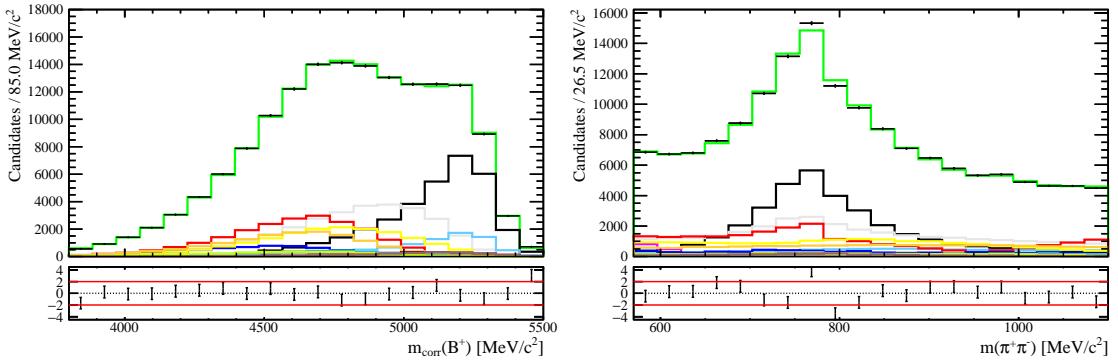
Figure 7.20 – Initial fit (green) of the signal (black) and backgrounds (other colours) in the full q^2 region projected in $m_{\text{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right). All templates, except the misID templates, are free floating in the fit and the ρ^0 line-shape uncertainty is defined by the variation of L_{eff}. A larger version of the legend is shown in Fig. 7.24.

prediction of X is obtained. In addition to this, the q^2 distributions obtained by the fit of certain background templates are compared to their expected q^2 distribution in simulation. In some cases it is found necessary to constrain the contribution of a template T_1 to a *reference template* T_{ref}. The reference template is chosen to be $|V_{cb}|^{B^+} D^0 \to K^0\pi^+\pi^-$ since it has a well measured branching fraction and populates all q^2 bins. The relative contribution $R_{T_1/T_{\text{ref}}}$, defined in Eq. 6.6, is computed with the branching fractions of the processes making up the templates, listed in Table 7.2, and their selection efficiencies. The uncertainties of the branching fractions are propagated to obtain the $\pm 1\sigma$ uncertainty of the relative contribution. A summary of the constraints and shape uncertainties included in the nominal fit of the full q^2 region are listed in Table 7.7. More details of their motivations are given in the following.

The $B^+ \to \eta'\mu^+\nu_\mu$, $B_s^0 \to D_s^+\pi^+\pi^-\pi^-$ and incl. $|V_{ub}|^{B^0}$ templates are not picked up by the fit leading to zero contribution in the full q^2 region and/or in certain bins of q^2. For this reason, their contributions with respect to the reference template are constrained. The same is true for the $B^+ \to f_0(500)\mu^+\nu_\mu$, $B^+ \to f_0(980)\mu^+\nu_\mu$, $B^+ \to f_2(1270)\mu^+\nu_\mu$ and $B^+ \to \rho(1450)^0\mu^+\nu_\mu$ templates, however, as described in Sec. 7.1.1, their branching fractions have not been precisely measured, and a systematic uncertainty is therefore assigned if they are used to constrain any of the templates, as described in Sec. 7.5.1. Constraining the $B^+ \to f_0(500)\mu^+\nu_\mu$ and $B^+ \to f_2(1270)\mu^+\nu_\mu$ templates works well in the fit, *i.e.* in the binned fits, their contributions with respect to the reference template are always within $\pm 1\sigma$ of the constraints. For this reason, they are constrained in the nominal fit. The $B^+ \to \rho(1450)^0\mu^+\nu_\mu$ decay is not constrained in the fit due to the large relative uncertainty of its branching fraction of $\sim 250\%$ combined with its small expected contribution of less than $\sim 1\%$ of the signal, which is consistent with the fit result. Constraining the $B^+ \to f_0(980)\mu^+\nu_\mu$ decay results in large negative yields of that template. Thus, its expected contribution of $\sim 11\%$ of the signal is assumed to be significantly overestimated. For this reason, the $B^+ \to f_0(980)\mu^+\nu_\mu$ decay is also not constrained in the fit.
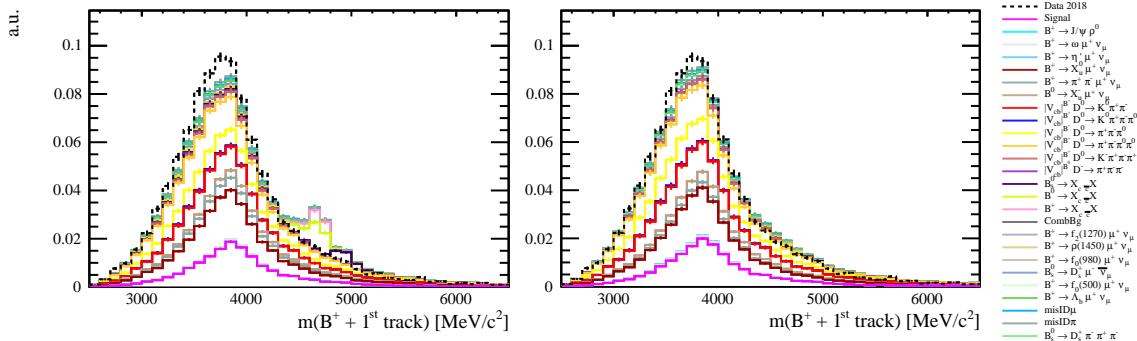
Figure 7.21 – Post fit comparison of $m(B^+ + 1^{st}$ track$)$ in data (black) and for the stacked fit components (other colours) in q^2 bin 9 before (left) and after (right) constraining the charmonium backgrounds.

Based on the post fit comparison of the isolation variable $m(B^+ + 1^{st}$ track$)$ the contributions from charmonium backgrounds are found to be overestimated by the fit, in particular, in the high q^2 bins. This is illustrated in Fig. 7.21 for bin 9 where an excess of partially reconstructed charmonium decays is visible. For this reason the relative contribution of each charmonium background is constrained in the fit. This essentially eliminates the aforementioned excess, as demonstrated in Fig. 7.21. Similarly, based on the post fit comparison of $m(\rho^0 + 1^{st}$ track$)$, the $B^0 \rightarrow D^-(\rightarrow \pi^-\pi^+\pi^-)\mu^+\nu_\mu$ template is found to be overestimated in certain q^2 bins, and its relative contribution is therefore constrained in the fit.

The q^2 distribution obtained by the fit for the prominent background templates $B^+ \rightarrow D^0\mu^+\nu_\mu$ with $D^0 \rightarrow K^0\pi^+\pi^-$, $D^0 \rightarrow \pi^+\pi^-\pi^0\pi^0$ and $D^0 \rightarrow \pi^+\pi^-\pi^0$ as well as $B^+ \rightarrow X_u^0\mu^+\nu_\mu$ and $B^+ \rightarrow \pi^+\pi^-\mu^+\nu_\mu$ are compared to their q^2 distribution in simulation. A reasonable agreement is found for the first three templates, as shown in Fig. 7.22, however, for the latter two templates the agreement is less good, as shown in Fig. 7.23, and they are therefore constrained in the fit. In addition, the contribution of the nonresonant $B^+ \rightarrow \pi^+\pi^-\mu^+\nu_\mu$ decay is found to be significantly overestimated with respect to its estimated branching fraction in Appendix G. In some bins its contribution is more than 3σ larger than expected, and it is also for this reason constrained in the fit. In general constraining the templates, improves the agreement between the expected and measured q^2 distributions, as illustrated in Fig. 7.23. However, in the last bin, the fit does not distinguish well between the signal and the nonresonant contribution, and since the nonresonant contribution is only loosely constrained in the fit due to its large branching fraction uncertainty, it becomes underestimated in this bin. Under the assumption that the simulation describes the q^2 distribution of the nonresonant contribution reasonably well, supported by the observed agreement in bin 1–9, the nonresonant contribution in bin 10 is constrained to the simulated q^2 distribution. In practice, the nonresonant yield obtained by the fit in bin 1–9 ($N_{\text{bin } 1–9} = 3287$) and its predicted contribution with respect to the
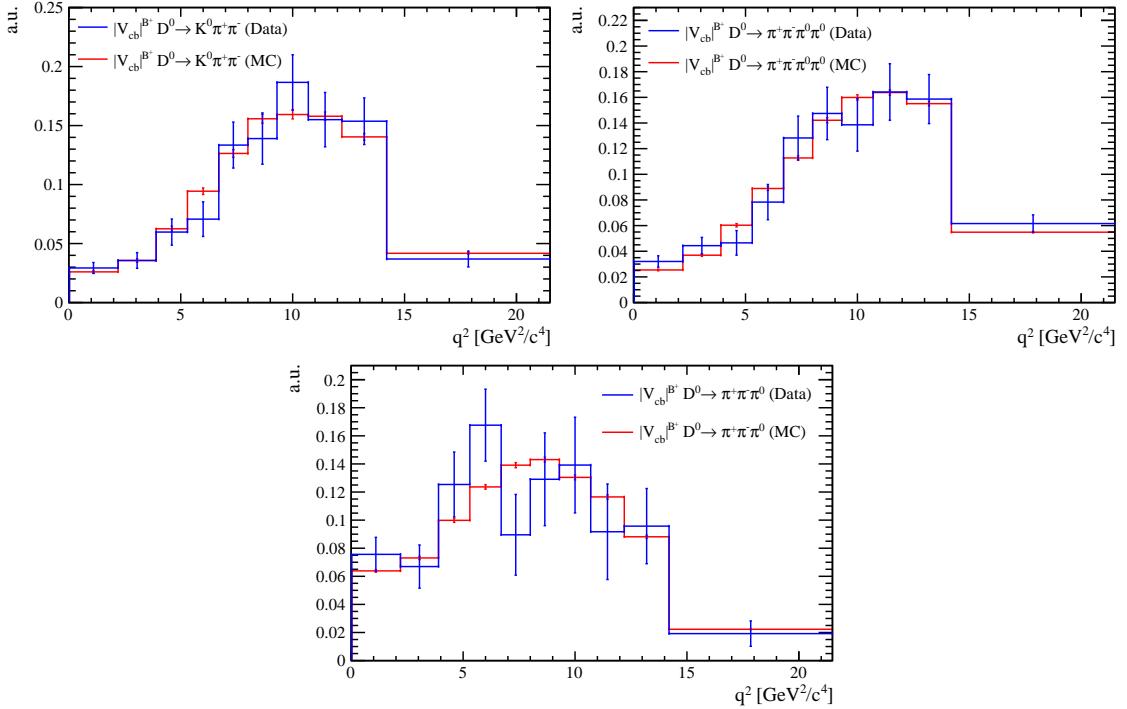
Figure 7.22 – q^2 distributions of the $B^+ \rightarrow D^0\mu^+\nu_\mu$ with $D^0 \rightarrow K^0\pi^+\pi^-$ (top left), $D^0 \rightarrow \pi^+\pi^-\pi^0\pi^0$ (top right) and $D^0 \rightarrow \pi^+\pi^-\pi^0$ (bottom) decays obtained in the fit (blue) and in simulation (red).

full q^2 region ($f_{\text{bin } 1-9} = 0.897$), is used to constrain the nonresonant yield in bin 10 to:

$$N_{\text{bin } 10} = N_{\text{bin } 1-9}(1/f_{\text{bin } 1-9} - 1) = 376\,. \tag{7.13}$$

The systematic uncertainty related to this procedure is estimated in Sec. 7.5.2.

As illustrated in Fig. 7.20, a discrepancy is observed under the ρ^0-meson peak between the fit and data suggesting that the $\text{GS}^{\text{L}_{\text{eff}}=1}_{\rho^0-\omega}$ line-shape correction is not optimal. This is not surprising, since the applied correction is based on $e^+e^- \rightarrow \pi^+\pi^-$ studies, where the impact of the $\rho^0 - \omega$ interference is not necessarily the same as for the processes studied in this analysis. However, with the current knowledge, this is assumed to be the best available description of the $\rho^0 - \omega$ interference. Nevertheless, to improve the $m(\pi^+\pi^-)$ line-shape description in the fit, the parameters of the model defined in Eq. 7.12 are allowed to differ from the $e^+e^- \rightarrow \pi^+\pi^-$ determinations.

The height of the interference drop in the $m(\pi^+\pi^-)$ distribution is larger in data than in the fit model, as shown in Fig. 7.20. The height of the drop is mostly determined by the parameters L_{eff} and δ, however, the shape variation of $\pm 1\sigma$ defined by $L_{\text{eff}} = 0$ and $L_{\text{eff}} = 2$ used in the initial fit seems to be insufficient to describe data, and the observed discrepancy is therefore assumed to be due to an underestimation of the parameter δ. Thus, L_{eff} is fixed to the optimal value found in the initial fit, and instead, δ is allowed to vary.
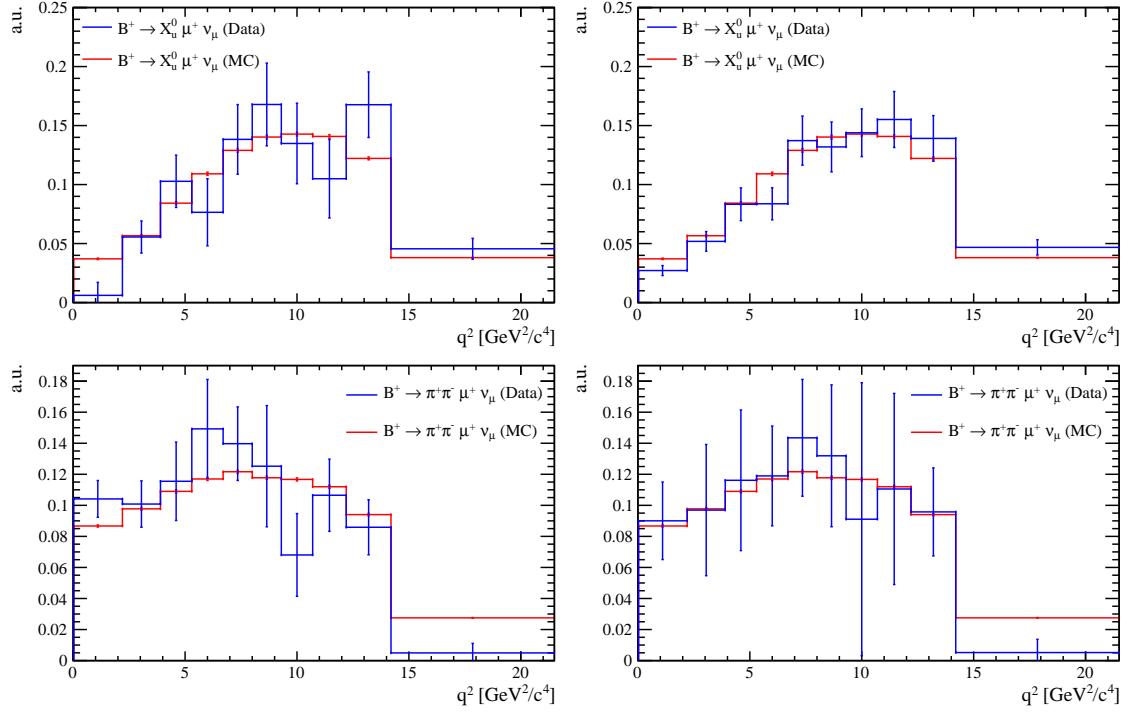
Figure 7.23 – q^2 distributions of the inclusive $B^+ \to X_u^0 \mu^+ \nu_\mu$ (top) and nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ (bottom) decays obtained in the fit (blue) and in simulation (red) before (left) and after (right) constraining their contributions in the fit.

For the signal template, the initial fit in the full q^2 region finds the optimal value to be $L_{\text{eff}} = 1.44 \pm 0.08$, which is consistent with the results in the binned fits. However, the optimal L_{eff} obtained for the incl. $|V_{ub}|^{B^+/B^0}$ templates are associated with large uncertainties within which they agree with the optimal value found for the signal template. Therefore, the new $m(\pi^+\pi^-)$ line-shape correction is redefined with $L_{\text{eff}} = 1.44$, and the new shape variation is defined by δ with a large $\pm 1\sigma$ shape uncertainty corresponding to $\delta = 0$ and $\delta = 0.02$, while $\delta = 0.01$ is the nominal correction. This is the final $m(\pi^+\pi^-)$ line-shape correction denoted $\text{GS}_{\rho^0-\omega}^{L_{\text{eff}}=1.44}$. The corrected signal templates are illustrated in Fig. 7.17.

Applying this to the fit significantly improves the agreement in the region under the ρ^0-meson peak between the model and data, as illustrated in Fig. 7.24. The shape variation of the signal defined by δ in the fit to the full q^2 region is $\sigma_{\text{shape}} = -(0.62 \pm 0.05)\sigma$, which is consistent with the observed shape variations in the binned fits. Thus, the fit results in $\delta = 0.0038$, which is a factor 2.4 larger than the determination from Ref. [189]. The systematic uncertainty associated with the $m(\pi^+\pi^-)$ line-shape is estimated in Sec. 7.5.3.

Table 7.7 – Summary of the template constraints and shape uncertainties included in the nominal signal fit. The relative normalisation of template T with respect to the reference template T_{ref} is denoted $R_{T/T_{ref}}$ and is defined in Eq. 6.6. The yields of the misIDμ and misIDπ backgrounds are denoted $N_{misID\mu}$ and $N_{misID\pi}$. The magnitudes of the constraints are only valid for the fit of the full q^2 region.

Template (T)	Constraint	Shape uncertainty
Signal		$GS_{\rho^0-\omega}^{L_{eff}=1.44}$, $\delta_{min=0.00}^{max=0.02}$
$|V_{cb}|^{B^+}$ $\bar{D}^0 \to K^0\pi^+\pi^-$	$R_{T/T_{ref}} = 1$ (reference template)	
$B^+ \to \eta'\mu^+\nu_\mu$	$R_{T/T_{ref}} = (3.45 \pm 1.23) \times 10^{-2}$	
$B^+ \to f_0(500)\mu^+\nu_\mu$	$R_{T/T_{ref}} = (8.12 \pm 5.72) \times 10^{-2}$	
$B^+ \to f_2(1270)\mu^+\nu_\mu$	$R_{T/T_{ref}} = (1.80 \pm 1.08) \times 10^{-2}$	
$B_s^0 \to D_s^+\pi^+\pi^-\pi^-$	$R_{T/T_{ref}} = (4.32 \pm 0.78) \times 10^{-2}$	
$B^0 \to X_u^-\mu^+\nu_\mu$	$R_{T/T_{ref}} = (3.36 \pm 0.65) \times 10^{-1}$	$GS_{\rho^0-\omega}^{L_{eff}=1.44}$, , $\delta_{min=0.00}^{max=0.02}$
$B^+ \to X_u^0\mu^+\nu_\mu$	$R_{T/T_{ref}} = (1.26 \pm 0.25)$	$GS_{\rho^0-\omega}^{L_{eff}=1.44}$, $\delta_{min=0.00}^{max=0.02}$
$B^+ \to \pi^+\pi^-\mu^+\nu_\mu$	$R_{T/T_{ref}} = (0.47 \pm 1.58) \times 10^{-1}$	
$B^0 \to J/\psi\rho^0$	$R_{T/T_{ref}} = (1.28 \pm 0.13) \times 10^{-2}$	
$B^0 \to X_{c\bar{c}}Y$	$R_{T/T_{ref}} = (1.92 \pm 0.53) \times 10^{-2}$	
$B^+ \to X_{c\bar{c}}Y$	$R_{T/T_{ref}} = (6.41 \pm 1.03) \times 10^{-3}$	
$B_s^0 \to X_{c\bar{c}}Y$	$R_{T/T_{ref}} = (9.56 \pm 0.83) \times 10^{-3}$	
$|V_{cb}|^{B^0}$ $D^- \to \pi^+\pi^-\pi^-$	$R_{T/T_{ref}} = (1.08 \pm 0.11) \times 10^{-2}$	
$|V_{cb}|^{B^+}$ $\bar{D}^0 \to K^0\pi^+\pi^-\pi^0$		$m(\pi^+\pi^-)$ data + CLN
$|V_{cb}|^{B^+}$ $\bar{D}^0 \to K^0\pi^+\pi^-$		CLN
$|V_{cb}|^{B^+}$ $\bar{D}^0 \to \pi^+\pi^-\pi^0$		CLN
$|V_{cb}|^{B^+}$ $\bar{D}^0 \to \pi^+\pi^-\pi^0\pi^0$		CLN
$|V_{cb}|^{B^+}$ $\bar{D}^0 \to K^+\pi^+\pi^-\pi^0$		CLN
MisIDμ	$N_{misID\mu} = 1555 \pm 39$	
MisIDπ	$N_{misID\pi} = 2675 \pm 52$	

7.3.2 Nominal fit in the full q^2 region

The nominal fit in the full q^2 region is performed with the constraints and shape uncertainties defined in Table 7.7. The result of the fit projected in $m_{corr}(B^+)$ and $m(\pi^+\pi^-)$ is illustrated in Fig. 7.24. It should be noted that the normalisation of the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ template does not reflect the numerical fit result due to a bad behaviour of RooFit, and consequently, its contribution displayed in the plot is about twice as large as what it really is. The extracted signal yield is 33093 ± 558 corresponding to a relative uncertainty of 1.7%. Overall, the fit describes data well, however, a discrepancy is observed in the last $m_{corr}(B^+)$ bin. The variation of the template shapes and constraints are mostly within $\pm 1\sigma$. However, the relative contribution of the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ template is $+1.5\sigma$ and for the $B^+ \to f_2(1270)\mu^+\nu_\mu$ template it is -1.7σ, while the shape variation of the incl. $|V_{ub}|^{B^0}$ template is -1.4σ.
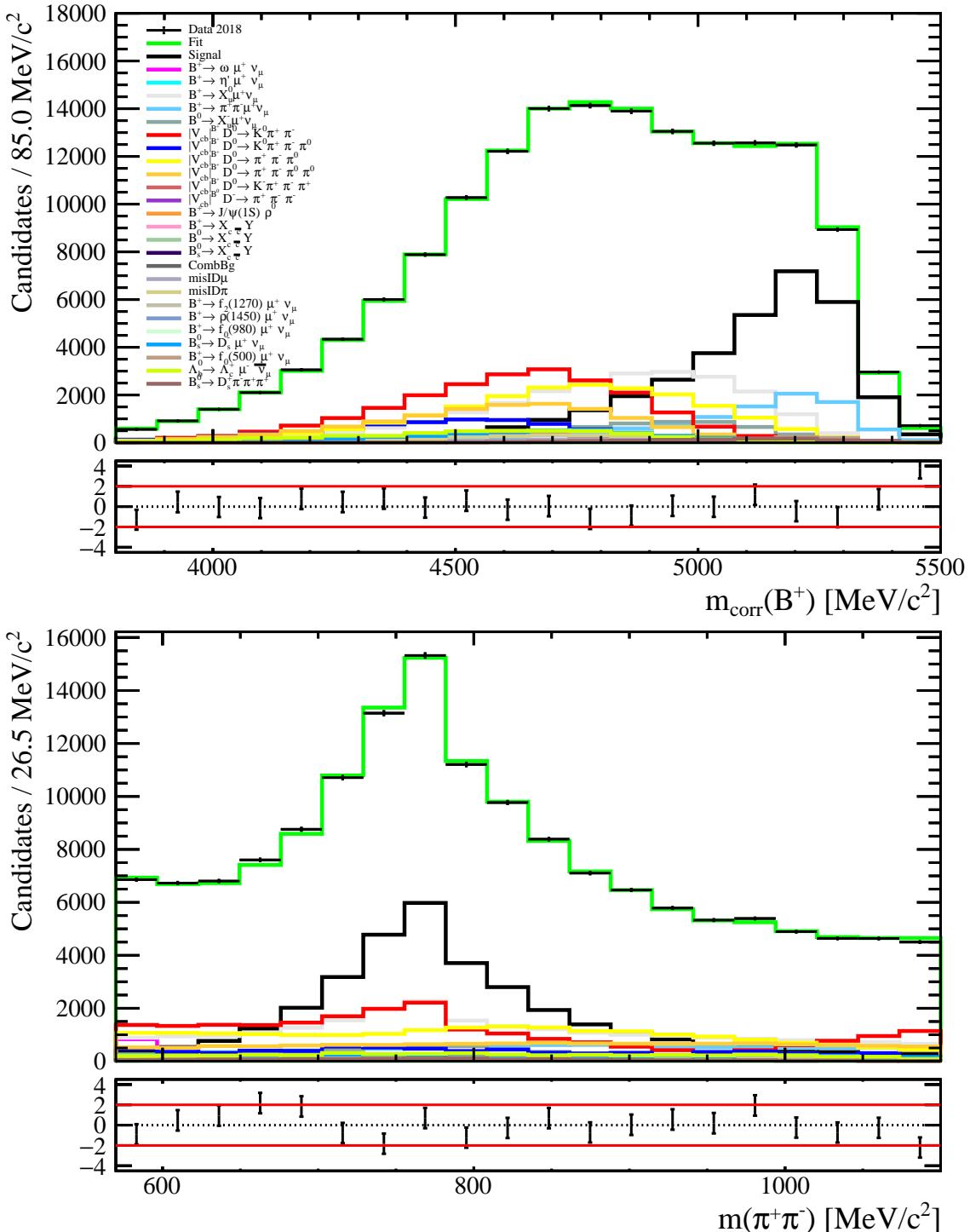
Figure 7.24 – Nominal fit (green) of the signal (black) and backgrounds (other colours) in the full q^2 region projected in $m_{\mathrm{corr}}(B^+)$ (top) and $m(\pi^+\pi^-)$ (bottom).

7.3.3 Nominal fits in bins of q^2

The nominal fits in the ten bins of q^2 defined in Table 5.2 are performed with the same shape uncertainties and constrained templates as in the fit of the full q^2 region. However, the magnitude of the constraints are re-computed for each q^2 bin to account for the change in selection efficiencies. The results of the fits are projected in $m_{\mathrm{corr}}(B^+)$ and $m(\pi^+\pi^-)$ as illustrated in Figs. 7.25 and 7.26. As for the signal fit in the full q^2 region, it should be noted that the normalisations of the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ templates in the different q^2 bins do not in general reflect the numerical fit result due to a bad behaviour of ROOFIT. The extracted signal yields are listed in Table 7.8. A good agreement is found between the fit and data in all bins of q^2 and the relative uncertainties of the signal yields are between 4.0% and 5.2%. In addition to this, the sum of the extracted signal yields in the binned fits is found to agree with the signal yield obtained in the fit of the full q^2 region. In general the variation of template shapes and constraints are found to be within $\pm 1\sigma$, however, similarly to the fit of the full q^2 region, the relative contribution of the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ template is found to be between $+1$ to $+2\sigma$ larger than expected in all bins of q^2 except in the last bin.

7.4 Unfolding the q^2 distribution

The extracted yields listed in Table 7.8 correspond to the reconstructed q^2_{reg} distribution of the signal in data. However, before this can be used to obtain a measurement of the differential branching fraction of the signal as a function of q^2, the extracted yields in bins of q^2_{reg} have to be unfolded to obtain them in bins of the true q^2 variable, q^2_{true}.

To do this, ROOUNFOLD, presented in Sec. 6.2.3, is used to build a 10×10 response matrix relating q^2_{true} and q^2_{reg} in the simulated signal sample. In order to choose the best unfolding algorithm, the performance of the SVD (see Sec. 6.5.2) and Bayes (see Sec. 6.2.3) unfolding algorithms are compared by following the procedure outlined in Sec. 6.5.2. A series of $N_{\mathrm{runs}} = 5000$ pseudo-experiments are carried out for each regularisation strength. For each run, the q^2_{reg} distribution scaled to the extracted signal yield in data is fluctuated according to a Gaussian distribution where σ is defined by the uncertainty of the fitted signal yields. The observables defined in Eqs. 6.25–6.27 that quantify the bin-averaged and run-averaged bias, standard deviation and coverage are plotted as functions of the regularisation strength in Appendix I. For Bayes unfolding method, the best performance is obtained with $r = 3$ where $\sqrt{\Delta^2} = 168$, $R_\sigma = 1.15$ and $p_{\mathrm{cov}} = 71\%$. For the SVD unfolding method, the best performance is obtained with $r = 7$ where $\sqrt{\Delta^2} = 146$, $R_\sigma = 1.01$ and $p_{\mathrm{cov}} = 72\%$. For both methods, the bias between the unfolded and true yields are found to be below 3% of the uncertainty of the unfolded yields in all bins. Thus, the best unfolding performance is obtained with the SVD method, and it is therefore used as the nominal unfolding method. The systematic uncertainty related to the choice of unfolding method is assessed in Sec. 7.5.5.
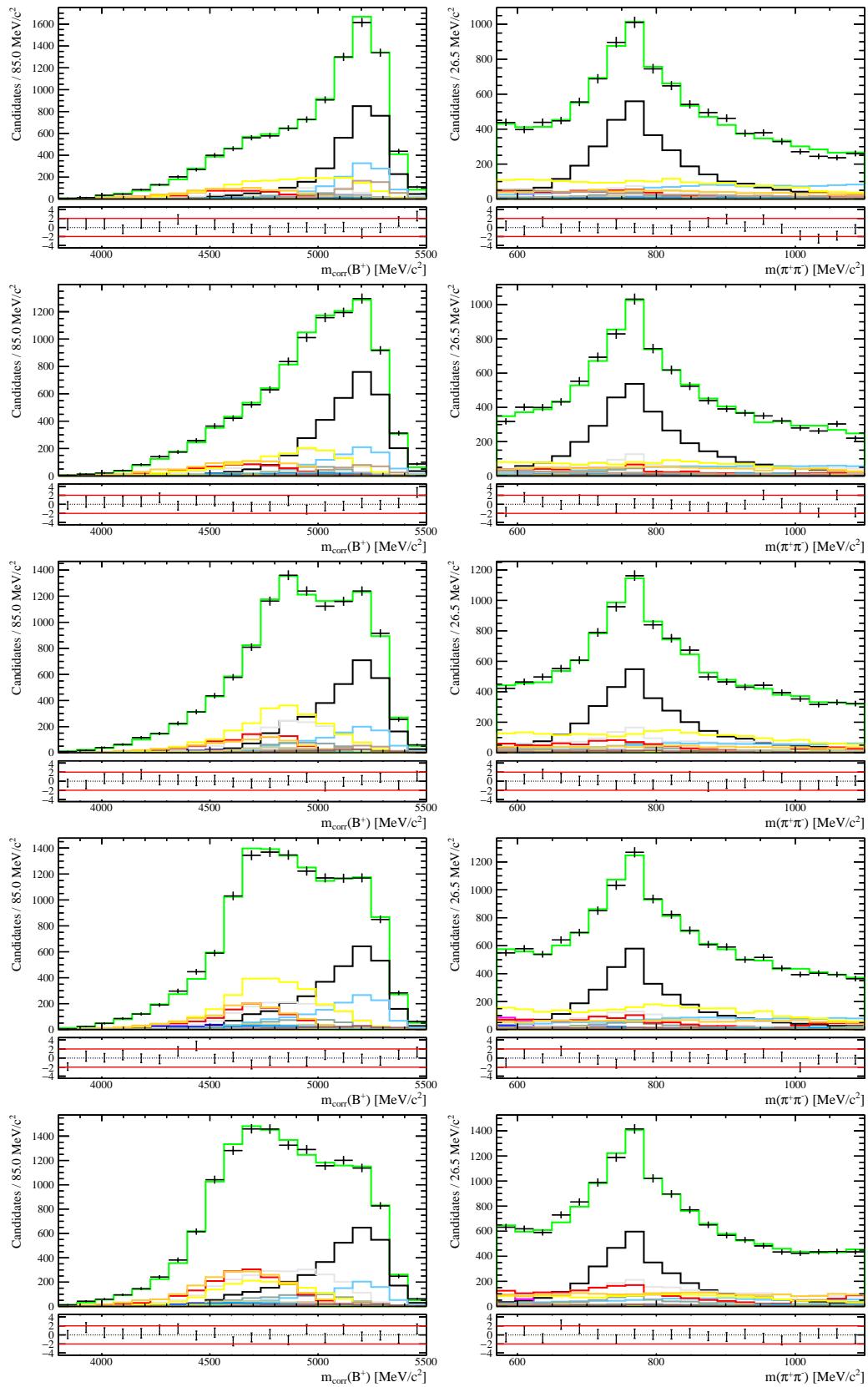
Figure 7.25 – Nominal signal fits projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right) for q^2 bins 1–5 (corresponding to row 1–5). A legend is shown in Fig. 7.24.
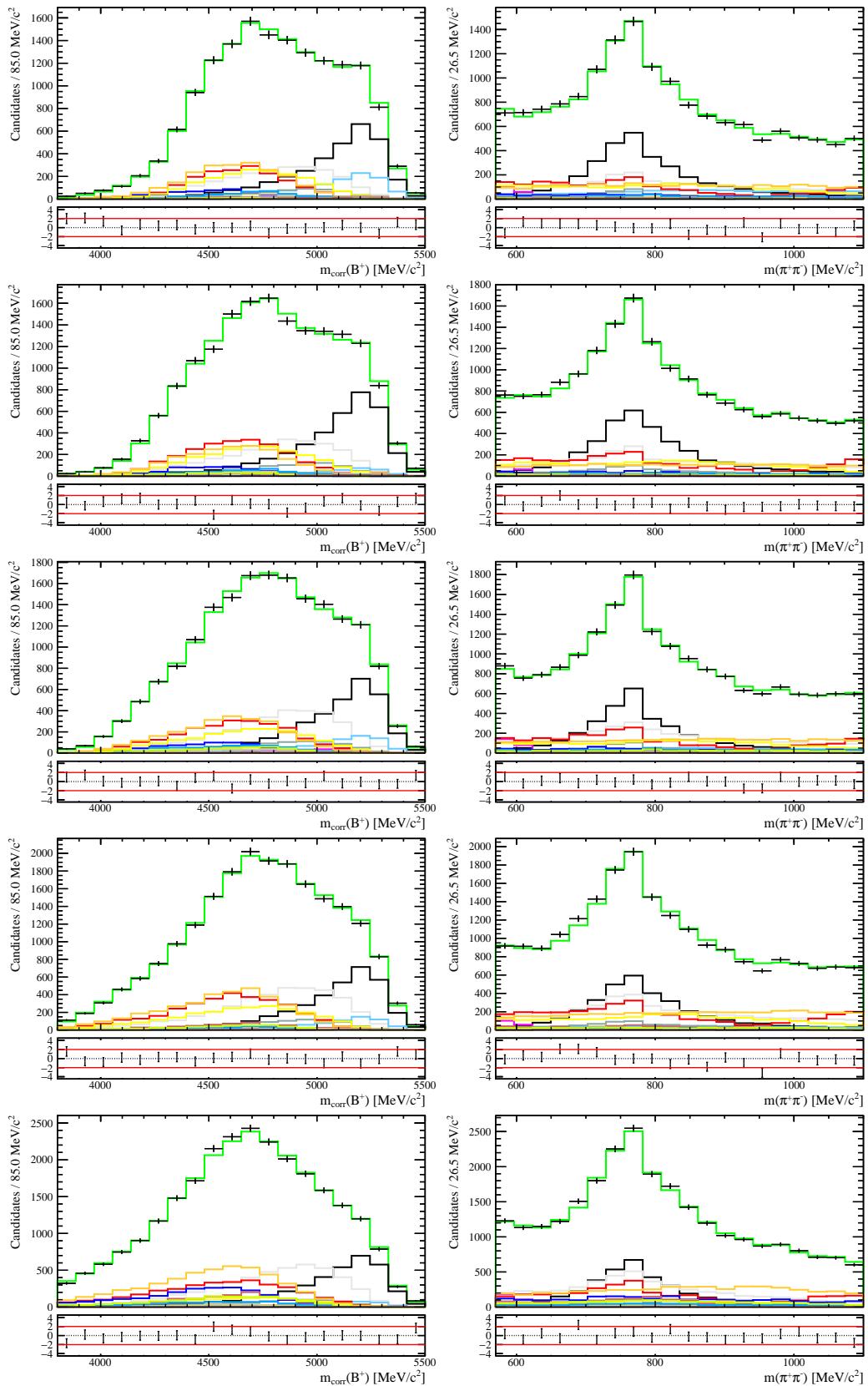
Figure 7.26 – Nominal signal fits projected in $m_{\mathrm{corr}}(B^+)$ (left) and $m(\pi^+\pi^-)$ (right) for q^2 bins 6–10 (corresponding to row 1–5). A legend is shown in Fig. 7.24.
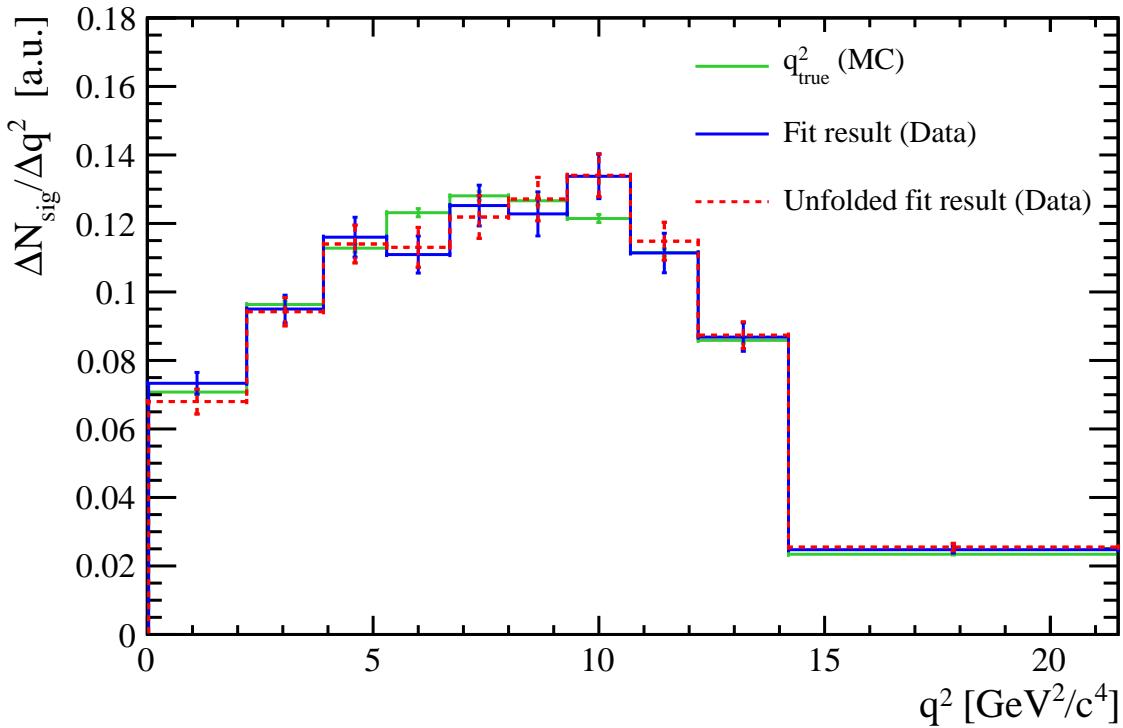
Figure 7.27 – Normalised q^2 distributions for the signal channel, *i.e.* the unfolded q^2_{true} distribution (dashed red) from the measured q^2_{reg} distribution (blue) compared to the expected q^2_{true} distribution (green) based on the BCL parameterisation.

The unfolded q^2 distribution in data is illustrated in Fig. 7.27 and summarised in Table 7.8. The statistical correlation matrix for the unfolded signal yields is given in Table 7.9. In general, a good agreement is found between the unfolded and expected true q^2 distributions with a χ^2 of 12.5 for 9 degrees of freedom corresponding to a p-value of 18.6%. Discrepancies between 1σ and 2σ are observed in bin 4, 7 and 10. The systematic uncertainty of using the simulated signal produced with the BCL form factors to unfold q^2 in data is assessed in Sec. 7.5.6.

Table 7.8 – Fitted signal yields, $N_{\rm sig}^{\rm fit}$, in bins of q^2 with bin boundaries $q_{\rm min}^2$ and $q_{\rm max}^2$. For comparison, the sum of these yields and the signal yield from the fit to the full q^2 region are given. Finally, the unfolded signal yields, $N_{\rm sig}^{\rm unf}$, in bins of the true q^2 are given.

q^2 bin	$q_{\rm min}^2$ [GeV$^2/c^4$]	$q_{\rm max}^2$ [GeV$^2/c^4$]	$N_{\rm sig}^{\rm fit}$	$N_{\rm sig}^{\rm unf}$
1	0.0	2.2	3188 ± 138	2950 ± 157
2	2.2	3.9	3192 ± 136	3160 ± 139
3	3.9	5.3	3209 ± 161	3147 ± 152
4	5.3	6.7	3069 ± 149	3121 ± 161
5	6.7	8.0	3217 ± 153	3125 ± 160
6	8.0	9.3	3155 ± 165	3259 ± 163
7	9.3	10.7	3701 ± 181	3701 ± 171
8	10.7	12.2	3301 ± 171	3396 ± 164
9	12.2	14.2	3431 ± 164	3447 ± 153
10	14.2	21.5	3570 ± 144	3671 ± 159
Sum	0.0	21.5	33034 ± 495	32977 ± 500
Full q^2	0.0	21.5	33093 ± 558	33093 ± 558

Table 7.9 – Statistical correlation matrix in percent for the unfolded signal yields.

q^2 bin	1	2	3	4	5	6	7	8	9	10
1	100.0	5.9	−27.9	−12.0	0.9	1.9	0.3	−0.1	−0.1	0.0
2	5.9	100.0	33.3	−19.3	−14.9	−2.1	1.2	0.4	−0.1	−0.0
3	−27.9	33.3	100.0	35.8	−21.3	−17.2	−2.9	1.3	0.5	−0.1
4	−12.0	−19.3	35.8	100.0	29.7	−21.1	−16.7	−2.9	1.0	0.3
5	0.9	−14.9	−21.3	29.7	100.0	35.6	−18.1	−16.2	−2.7	1.1
6	1.9	−2.1	−17.2	−21.1	35.6	100.0	38.7	−16.0	−14.0	−0.3
7	0.3	1.2	−2.9	−16.7	−18.1	38.7	100.0	37.3	−14.8	−8.5
8	−0.1	0.4	1.3	−2.9	−16.2	−16.0	37.3	100.0	32.2	−16.7
9	−0.1	−0.1	0.5	1.0	−2.7	−14.0	−14.8	32.2	100.0	12.9
10	0.0	−0.0	−0.1	0.3	1.1	−0.3	−8.5	−16.7	12.9	100.0

7.5 Fit and unfolding systematic uncertainties

The different sources of systematic uncertainties associated with the fit (fit syst.) and unfolding procedure (unfolding syst.) that could affect the measured q^2 distribution of the signal are identified as:

- Constraining the $B^+ \to f_0(500)\mu^+\nu_\mu$ and $B^+ \to f_2(1270)\mu^+\nu_\mu$ templates (fit syst.).

- Fixing the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ contribution in the last q^2 bin (fit syst.).

- Correcting the $m(\pi^+\pi^-)$ line-shape of the signal and incl. $|V_{ub}|$ templates (fit syst.).

- Using `PID` particle regions to estimate the misID background contribution (fit syst.).

- Performing the unfolding of q^2 with the SVD method (unfolding syst.).

- Unfolding q^2 with signal MC based on the BCL form factors (unfolding syst.).

In practice, the fit syst. uncertainties are estimated by varying the considered systematic source in the fit and comparing the unfolded result with the nominal result. The unfolding syst. uncertainties are estimated by varying the considered systematic source in the unfolding procedure and comparing this to the result of applying the nominal unfolding procedure. In both cases, the unfolded yields resulting from the systematic study are denoted N_{syst} and they are compared to the unfolded yields N_{nom} produced by the nominal fit and nominal unfolding procedure. The systematic uncertainty is defined as $\Delta N = N_{\mathrm{syst}} - N_{\mathrm{nom}}$.

7.5.1 Constraining exclusive $|V_{ub}|$ decays

The excl. $|V_{ub}|$ decays $B^+ \to f_0(500)\mu^+\nu_\mu$ and $B^+ \to f_2(1270)\mu^+\nu_\mu$ are constrained in the fit, as described in Sec. 7.3.1. However, since their branching fractions have not been precisely measured, a systematic uncertainty is assigned to constraining them. This is done by performing the fit with both templates unconstrained. The unfolded yields resulting from this are listed and compared to the nominal fit result in Table 7.10. Although, the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ decay has not been precisely measured, its contribution to the fit when being unconstrained is significantly overestimated, which is not the case for the $B^+ \to f_0(500)\mu^+\nu_\mu$ and $B^+ \to f_2(1270)\mu^+\nu_\mu$ decays, and thus, including the nonresonant decay in this estimation would lead to a inaccurate systematic uncertainty.

7.5.2 Nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ contribution

The nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ contribution that is only loosely constrained in the fit due to its large branching fraction uncertainty is assumed to be underestimated in the last q^2 bin based on the comparison displayed in Fig. 7.23. For this reason, the nonresonant

Table 7.10 – Systematic uncertainties associated with constraining the $B^+ \to f_0(500)\mu^+\nu_\mu$ and $B^+ \to f_2(1270)\mu^+\nu_\mu$ templates in the fit.

q^2 bin	N_{syst}	ΔN	$\Delta N/N_{\text{nom}}$ [%]
1	2916 ± 157	-35	-1.2
2	3113 ± 142	-47	-1.5
3	3104 ± 155	-43	-1.4
4	3095 ± 165	-26	-0.8
5	3081 ± 164	-44	-1.4
6	3222 ± 167	-36	-1.1
7	3683 ± 174	-18	-0.5
8	3379 ± 166	-17	-0.5
9	3440 ± 159	-7	-0.2
10	3689 ± 160	$+19$	$+0.5$

contribution is fixed in this bin according to the predicted q^2 distribution based on the DFN model. In order to assign a systematic uncertainty to this, the same procedure is performed with the more recent BLNP heavy-quark-effective-theory-based model [195]. This model provides a more systematic treatment of shape function effects, contains power corrections and applies an improved perturbative treatment. The BLNP model is used by EvtGen to produce the nonresonant decay, and the resulting q^2 distribution is compared to the one produced with the DFN model, as illustrated in Fig. 7.28. The two models produce slightly different q^2 distributions, while the effect on the $m_{\text{corr}}(B^+)$ and $m(\pi^+\pi^-)$ distributions are negligible. Assuming the reconstruction efficiency of the nonresonant decay does not depend on the model, the two distributions can be used to reweight the reconstructed q^2 distribution based on the DFN model to obtain the distribution based on the BLNP model. Using Eq. 7.13, the nonresonant yield in bin 10 is predicted by the BLNP model to be $N_{\text{bin }10} = 439$. The fit is performed with the nonresonant contribution fixed to this value, and the unfolded yields are listed and compared to the nominal fit result in Table 7.11.

7.5.3 Line-shape model for the dipion mass

As described in Secs. 7.2.1 and 7.3.1, the $m(\pi^+\pi^-)$ line-shape is corrected to a model where the ρ^0 line-shape is described with a GS function and the $\rho^0 - \omega$ interference is taken into account. In the final $m(\pi^+\pi^-)$ line-shape correction, the value of the relative angular momentum L_{eff} is fixed to the optimal value for the signal template found by the initial fit in the full q^2 region, *i.e.* $L_{\text{eff}} = 1.44 \pm 0.08$, while the line-shape variation is defined by the mixing parameter δ. In order to assign a systematic uncertainty to the choice of L_{eff}, the $m(\pi^+\pi^-)$ line-shape is varied within the uncertainty of L_{eff}. Thus, the fit is performed with L_{eff} fixed to $L_{\text{eff}}^{\text{min}} = 1.36$ and $L_{\text{eff}}^{\text{max}} = 1.53$, respectively, and each time, the $\pm 1\sigma$ shape variation is defined by $\delta = 0.00$ and $\delta = 0.02$. The unfolded yields are listed and compared to the nominal fit result in Table 7.12.
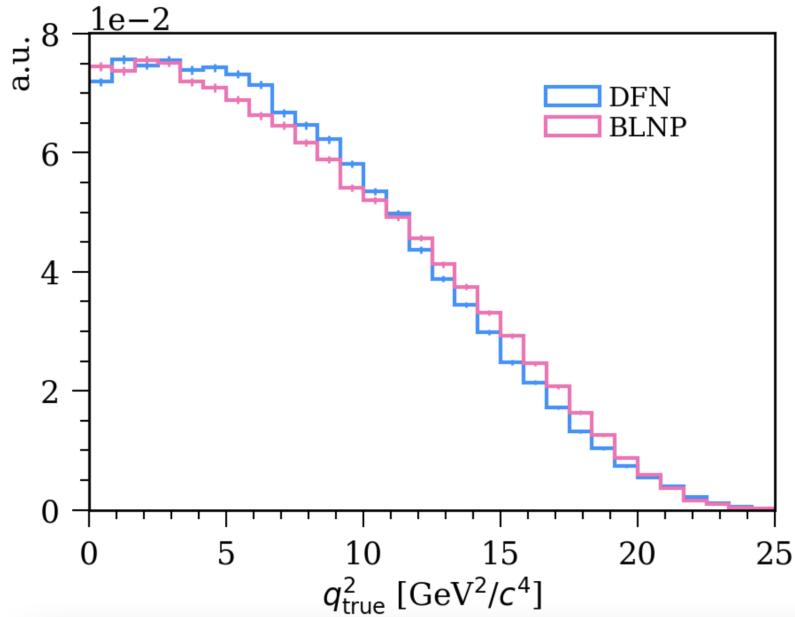
Figure 7.28 – The true q^2 distribution of the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ process based on the DFN (blue) and BLNP (pink) models.

7.5.4 MisID contribution

To estimate the contribution from misIDμ and the misIDπ background made of the misIDπ^+ and misIDπ^- backgrounds the particle regions defined in Tables 6.7, 7.4 and 7.5 are used. However, in the reference samples used to estimate the misID backgrounds, $\sim 9\%, \sim 5\%$ and $\sim 12\%$ of the μ^+, π^+ and π^- candidates, respectively, fall outside the defined particle regions, and they are therefore not included in the estimation of the misID backgrounds. The maximal effect of this is estimated Appendix H. For the misIDμ and misIDπ backgrounds, the maximal increases are found to be $\sim 6\%$ and $\sim 8\%$, respectively. To evaluate the effect of this, the fit is performed with the misID templates constrained to their maximal contributions. The unfolded yields are listed and compared to the nominal fit result in Table 7.13.

7.5.5 Unfolding method

The nominal unfolding is performed with the SVD unfolding algorithm with a regularisation strength of $r = 7$. In order to estimate the systematic uncertainty associated with the method, it is compared to the Bayes unfolding method. As described in Sec. 7.4 and illustrated in in Appendix I, the optimal regularisation strength of the Bayes unfolding method is found to be $r = 3$. This is used to unfold the nominal fit, and the resulting yields are listed and compared to the nominal method in Table 7.14.

Table 7.11 – Systematic uncertainties associated with fixing the nonresonant $B^+ \rightarrow \pi^+\pi^-\mu^+\nu_\mu$ contribution in the last q^2 bin.

q^2 bin	N_{syst}	ΔN	$\Delta N/N_{\text{nom}}$ [%]
1	2950 ± 157	-0	-0.0
2	3160 ± 139	-0	-0.0
3	3147 ± 152	$+0$	$+0.0$
4	3122 ± 161	$+0$	$+0.0$
5	3125 ± 160	-0	-0.0
6	3259 ± 163	$+0$	$+0.0$
7	3703 ± 171	$+2$	$+0.1$
8	3402 ± 164	$+5$	$+0.2$
9	3442 ± 153	-5	-0.1
10	3620 ± 159	-51	-1.4

Table 7.12 – Systematic uncertainties associated with fixing L_{eff} in the $m(\pi^+\pi^-)$ line-shape model. The unfolded yields $N_{\text{syst,min}}$ and $N_{\text{syst,max}}$ are based on the minimum and maximum shape variations.

q^2 bin	$N_{\text{syst,min}}$	$N_{\text{syst,max}}$	ΔN_{min}	ΔN_{max}	$\Delta N_{\text{min}}/N_{\text{nom}}$ [%]	$\Delta N_{\text{max}}/N_{\text{nom}}$ [%]
1	2956 ± 157	2944 ± 157	$+5$	-6	$+0.2$	-0.2
2	3169 ± 141	3152 ± 138	$+9$	-8	$+0.3$	-0.2
3	3157 ± 153	3137 ± 151	$+10$	-11	$+0.3$	-0.4
4	3134 ± 161	3109 ± 160	$+14$	-12	$+0.4$	-0.4
5	3138 ± 161	3108 ± 158	$+13$	-16	$+0.4$	-0.5
6	3274 ± 164	3231 ± 158	$+15$	-28	$+0.4$	-0.9
7	3718 ± 172	3679 ± 169	$+17$	-22	$+0.5$	-0.6
8	3410 ± 164	3382 ± 163	$+14$	-14	$+0.4$	-0.4
9	3453 ± 154	3440 ± 153	$+6$	-6	$+0.2$	-0.2
10	3669 ± 160	3672 ± 159	-2	$+1$	-0.0	$+0.0$

7.5.6 Signal form factors

As described in Sec. 7.4, the response matrix relating the reconstructed and true q^2 is based on the simulated signal produced with the BCL form factors. Since the shape of the q^2 distribution depends on the form factors, the choice of the form factor model gives rise to some degree of model dependence in the unfolding procedure. In order to estimate the systematic uncertainty of this, the BCL form factor parameters are varied within their uncertainties and the resulting simulation is used to perform the unfolding.

The uncertainties of the form factor parameters and their correlations are taken from Ref. [64], where the parameters $\alpha_1^{A_0}$, $\alpha_2^{A_0}$, $\alpha_0^{A_1}$, $\alpha_1^{A_1}$, $\alpha_2^{A_1}$, $\alpha_0^{A_{12}}$, $\alpha_1^{A_{12}}$, $\alpha_2^{A_{12}}$, α_0^{V}, α_1^{V} and α_2^{V}, their uncertainties and correlation matrix are used to generate new sets of parameters according to a multi-dimensional Gaussian distribution. These are used by EVTGEN to produce ten different simulations of the signal that each contains $\sim 270 \times 10^3$ events.

Table 7.13 – Systematic uncertainties associated with defining particle regions to estimate the misID contributions.

q^2 bin	N_syst	ΔN	$\Delta N/N_\text{nom}$ [%]
1	2947 ± 157	-4	-0.1
2	3159 ± 139	-1	-0.0
3	3147 ± 152	-0	-0.0
4	3123 ± 161	$+2$	$+0.1$
5	3125 ± 160	$+1$	$+0.0$
6	3258 ± 163	-1	-0.0
7	3701 ± 171	-0	-0.0
8	3398 ± 164	$+1$	$+0.0$
9	3447 ± 153	$+0$	$+0.0$
10	3671 ± 159	-0	-0.0

Table 7.14 – Systematic uncertainties associated with the unfolding method.

q^2 bin	N_syst	ΔN	$\Delta N/N_\text{nom}$ [%]
1	2956 ± 158	$+5$	$+0.2$
2	3126 ± 154	-33	-1.1
3	3235 ± 181	$+87$	$+2.8$
4	3084 ± 166	-37	-1.2
5	3193 ± 172	$+68$	$+2.2$
6	3168 ± 185	-91	-2.8
7	3820 ± 212	$+119$	$+3.2$
8	3340 ± 202	-56	-1.7
9	3420 ± 197	-26	-0.8
10	3692 ± 170	$+21$	$+0.6$

The q^2 distributions resulting from this procedure are illustrated in Fig. 7.29 together with the nominal q^2 distribution. The nominal q^2 distribution is re-weighted according to each of the q^2 distributions resulting from varying the BCL form factors, and each time, a new response matrix relating the reconstructed and true q^2 is obtained and used to perform the unfolding. The yields that are most different from the yields obtained with the nominal unfolding procedure are used to estimate the systematic uncertainty and are given in Table 7.15. It should be noted that the estimate neglects the correlation between the systematic uncertainties in different q^2 bins and it is therefore considered to be a conservative estimate for the systematic uncertainty related to using the BCL form factors in the unfolding procedure.
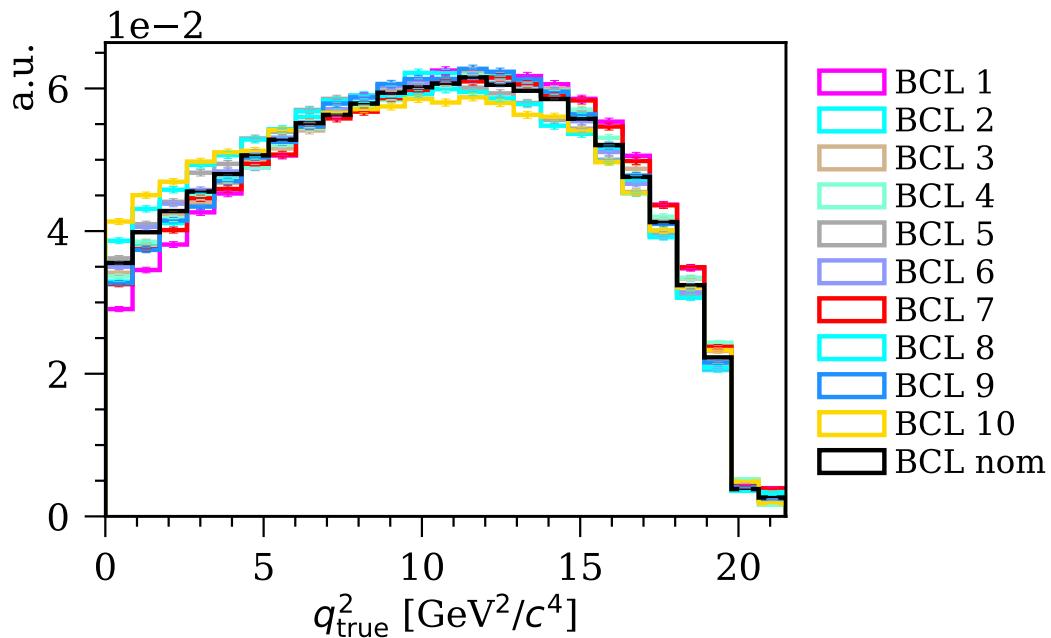
Figure 7.29 – q^2 distributions of the signal decay simulated by EVTGEN with the nominal (nom) BCL form factors (black) and BCL form factors varied within their uncertainties (other colours) denoted BCL $i = 1, 2, .., 10$.

Table 7.15 – Systematic uncertainties associated with using the BCL form factors in the unfolding procedure.

q^2 bin	N_{syst}	ΔN	$\Delta N/N_{\text{nom}}$ [%]
1	2935 ± 152	-15	-0.5
2	3198 ± 134	$+38$	$+1.2$
3	3116 ± 150	-32	-1.0
4	3137 ± 162	$+16$	$+0.5$
5	3146 ± 161	$+21$	$+0.7$
6	3232 ± 161	-27	-0.8
7	3723 ± 172	$+22$	$+0.6$
8	3377 ± 163	-20	-0.6
9	3468 ± 154	$+22$	$+0.6$
10	3678 ± 161	$+7$	$+0.2$

Table 7.16 – Summary of the overall systematic uncertainty related to the fit and unfolding procedure, σ_{sys}, in each bin of q^2 relative to the unfolded signal yield, N_{sig}.

q^2 bin	σ_{sys}	$\sigma_{\mathrm{sys}}/N_{\mathrm{sig}}$ [%]
1	39	1.3
2	70	2.2
3	103	3.3
4	51	1.6
5	86	2.8
6	106	3.3
7	126	3.4
8	65	1.9
9	36	1.0
10	59	1.6

7.5.7 Total systematic uncertainty

The overall systematic uncertainty related to the fit and unfolding procedure in each q^2 bin is obtained by adding in quadrature the systematic uncertainties described in Secs. 7.5.1–7.5.6. These are summarised in Table 7.16, where the largest contribution is coming from the choice of unfolding method.

8 Measuring the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction

This chapter describes the main result of the thesis corresponding to the measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction with respect to q^2. In order to perform this measurement, the selection efficiencies for the signal and normalisation channels are estimated, as described in Sec. 8.1, and the systematic uncertainties related to the computation of the efficiencies are evaluated, as described in Sec. 8.2. The differential branching fraction measurement is presented in Sec. 8.4. Since this analysis is undergoing an internal LHCb review at the time of writing, the final result is blinded, however, the precision will be given unblinded and compared to previous measurements.

8.1 Computing efficiencies

In order to obtain the differential branching fraction of the signal, the measured yields of the signal and normalisation decays must be corrected according to their total selection efficiencies. For the signal channel, the measured yield, $N_{\mathrm{sig},i}$, and the total selection efficiency, $\epsilon_{\mathrm{sig},i}^{\mathrm{tot}}$, are calculated in bins i of the true q^2, while for the normalisation channel, the measured yield, N_{norm}, and the total selection efficiency, $\epsilon_{\mathrm{norm}}^{\mathrm{tot}}$, are calculated for the integrated q^2 region. Mathematically, this is summarised as:

$$\left(\frac{\Delta \mathcal{B}_{\mathrm{sig}}}{\Delta q^2}\right)_i = \frac{N_{\mathrm{sig},i}}{q_{\mathrm{max},i}^2 - q_{\mathrm{min},i}^2} \times \frac{\mathcal{B}(B^+ \to \bar{D}^0 \mu^+ \nu_\mu) \times \mathcal{B}(\bar{D}^0 \to \pi^+ \pi^-)}{N_{\mathrm{norm}}} \times R_{\epsilon,i}, \quad (8.1)$$

where $q_{\mathrm{max},i}^2 - q_{\mathrm{min},i}^2$ is the bin width of the ith q^2 bin and $R_{\epsilon,i}$ is the efficiency ratio:

$$R_{\epsilon,i} = \frac{\epsilon_{\mathrm{norm}}^{\mathrm{tot}}}{\epsilon_{\mathrm{sig},i}^{\mathrm{tot}}}. \quad (8.2)$$

The total selection efficiency of each decay channel is obtained by considering the efficiencies of each selection step of the analysis corresponding to:

- The LHCb acceptance and MC generator-level cuts, ϵ^{GLC}.

- The reconstruction and stripping efficiencies, excluding `PID` requirements, ϵ^{strip}.

- The combined trigger efficiency of the L0, HLT1 and HLT2 trigger selections, ϵ^{trig}.

- The `PID` efficiency of the stripping and preselection, ϵ^{PID}.

- The preselection and MVA selection efficiency, excluding `PID` requirements, $\epsilon^{\text{pre+MVA}}$.

Thus, the total selection efficiency of each channel is:

$$\epsilon^{\text{tot}} = \epsilon^{\text{GLC}} \times \epsilon^{\text{strip}} \times \epsilon^{\text{trig}} \times \epsilon^{\text{PID}} \times \epsilon^{\text{pre+MVA}} . \tag{8.3}$$

For the signal channel all efficiencies are calculated in bins of the true q^2, while for the normalisation channel, the efficiencies are calculated for the integrated q^2 region.

In general, simulation is used to estimate the different efficiencies, however, as the `PID` response in simulation is known to differ from that in data, the efficiencies of the `PID` requirements in the stripping and preselection are estimated separately using data driven methods provided by the PIDCALIB package. The track reconstruction and trigger efficiency in simulation can also differ from that in data and data-driven methods are therefore used to correct for this. As mentioned in Sec. 4.3, the `B2XuMuNuBu2Rho_NoPIDPiLine` stripping line cannot be used to study `PID` efficiencies as it accidentally applies the requirement `PIDK`$(\pi) < -2$. However, for the MC samples based on the 2017 data-taking configuration, which is known to be similar to that of 2018, these requirements are not present in neither the `B2XuMuNuBu2Rho_NoPIDPiLine` stripping line nor in the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line. For this reason, simulation based on 2017 is used to estimate all efficiencies described in Secs. 8.1.1–8.1.4. Only to estimate the efficiency of the preselection and MVA selection, described in Sec. 8.1.5, simulations based on both 2017 and 2018 are used in order to reduce the statistical uncertainty of that efficiency.

To correct for MC/data differences, the simulated events are weighted. The simulated signal is weighted to the optimal $m(\pi^+\pi^-)$ line-shape determined by the fit, described in Sec. 7.3.1. Similarly, the simulated normalisation channel is weighted to the (true) q^2 distribution obtained with the CLN form factors, as explained in Sec. 6.2.1. In addition to this, both samples are corrected with the data/MC weights $w_{\text{nTracks}} \times w_{\text{mva}}$ computed in Sec. 4.6.3. However, since these weights were derived with the trigger selection already applied, they are not applied to samples without this selection.

In the following Secs. 8.1.1–8.1.5, all efficiencies are reported with their statistical uncertainties only. In cases where the efficiency is computed as $\epsilon = n_s/(n_s + n_r)$, where n_s (n_r) is the number of candidates accepted (rejected) by the selection, the calculation of the statistical uncertainty is based on the *the Normal Approximation* [196]. In this approach, n_s and n_r are assumed to be binomially distributed such that ϵ becomes a binomially distributed observation. The distribution of uncertainty about ϵ is approximated with a normal distribution based on the central limit theorem [197]. For small sample sizes or for

ϵ close to 0 or 1 this approximation is not reliable, however, for the efficiencies calculated in this chapter, the approximation holds. As the simulated events used to calculate the efficiencies are in general weighted, a generalised version of the Normal Approximation method [196] is used. The efficiency and its uncertainty are given as:

$$\epsilon = \frac{W_s}{W_s + W_r}\,, \qquad \sigma_\epsilon = \sqrt{\frac{W_s^2 \sigma_{W_r}^2 + W_r^2 \sigma_{W_s}^2}{(W_s + W_r)^4}}\,, \tag{8.4}$$

where

$$W_s = \sum_{i=1}^{n_s} w_i\,, \quad W_r = \sum_{i=1}^{n_s + n_r} w_i - W_s \tag{8.5}$$

$$\sigma_{W_s}^2 = \sum_{i=1}^{n_s} w_i^2\,, \quad \sigma_{W_r}^2 = \sum_{i=1}^{n_s + n_r} w_i^2 - \sigma_{W_s}^2 \tag{8.6}$$

$$\tag{8.7}$$

with w_i being the weight of the simulated event i.

In the following sections the selection efficiencies of the signal and normalisation channels are reported as the efficiency ratio R_ϵ in each true q^2 bin. Since the efficiencies of the signal and normalisation channels are, in general, determined from statistically independent simulations, their statistical uncertainties are assumed to be uncorrelated, and thus, the statistical uncertainty of the ratio is obtained by standard error propagation. In the cases where the samples are not independent it will be notified explicitly.

8.1.1 Acceptance and generator level cut efficiency

Decay candidates can only be reconstructed if their visible decay products are within the LHCb detector acceptance. To calculate the efficiency of this criterion, one must rely on simulation where events are removed if their decay products are far outside the LHCb acceptance. For charged particles the acceptance requirement corresponds to 10–400 mrad with respect to the z axis. While this is the only requirement imposed on the simulated normalisation decays, the simulated signal decays have to pass the acceptance requirement corresponding to 5–400 mrad, and $1.9 < \eta < 5.1$ as well as kinematic requirements on the p_T and p of the muon and pions. This selection results in the so-called *generator level cut (GLC) efficiency*, which is defined as:

$$\epsilon^{\mathrm{GLC}} = \frac{\#\ \mathrm{GLC}}{\#\ \mathrm{Total\ production\ in\ } 4\pi}\,, \tag{8.8}$$

where # denotes the number of events that passed the specified selection.
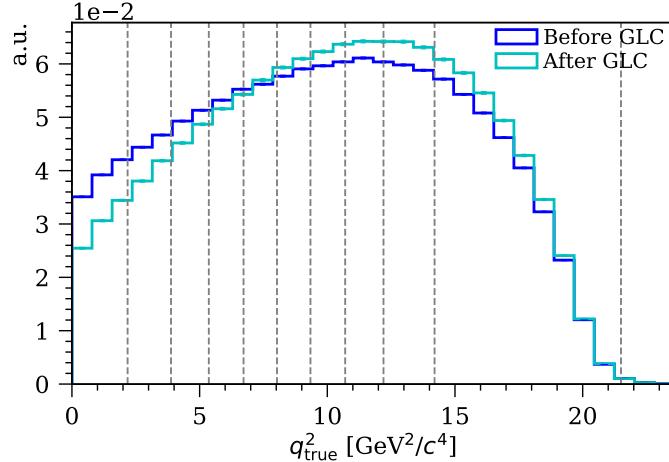
Figure 8.1 – The true q^2 distributions of the simulated signal before (blue) and after (cyan) applying generator level cuts (GLC). The vertical lines illustrate the bin boundaries of the non-uniform q^2 binning scheme defined in Table 5.2.

For the signal channel, the GLC efficiency, averaged over the magnet polarities in the full q^2 region, is $\epsilon^{\mathrm{GLC}} = (10.15 \pm 0.02)\%$ where the number of simulated signal candidates before and after applying GLC are $N_{\mathrm{tot}} \sim 214 \times 10^6$ and $N_{\mathrm{GLC}} \sim 22 \times 10^6$, respectively. However, to obtain ϵ^{GLC} in bins of q^2, EVTGEN is used to simulate a smaller sample consisting of $N_{\mathrm{tot}}^{\mathrm{small}} \sim 2.2 \times 10^6$ signal candidates without GLC applied, and another sample of $N_{\mathrm{GLC}}^{\mathrm{small}} \sim 1.0 \times 10^6$ signal candidates with GLC applied. With these samples, both produced with the BCL form factors, the true q^2 distributions of the signal before and after applying GLC are obtained, as illustrated in Fig. 8.1. The two samples are weighted such that their total yields correspond to N_{tot} and N_{GLC}, respectively, and finally, the GLC efficiency and the statistical uncertainty in each bin of q^2 is obtained with Eq. 8.4.

For the normalisation channel, the GLC efficiency cannot be directly obtained from the simulation defined in Table 6.3, since multiple decay processes are simulated together, and thus, the GLC efficiency represents the combined efficiency of all the simulated processes. Therefore, the decay $B^+ \to \bar{D}^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ is simulated alone resulting in $\epsilon^{\mathrm{GLC}} = (17.60 \pm 0.04)\%$ based on $N_{\mathrm{tot}} \sim 457 \times 10^3$ and $N_{\mathrm{GLC}} \sim 80 \times 10^3$ decay candidates before and after GLC, respectively. In order to compute the stripping efficiency, it is necessary to determine the number of normalisation decays that passes GLC in the simulation defined in Table 6.3 based on 2017. Knowing the total number of decays produced in the MC cocktail of $N_{\mathrm{tot}}^{\mathrm{cocktail}} \sim 152 \times 10^6$ and the fraction of simulated $B^+ \to \bar{D}^0(\to \pi^+\pi^-)\mu^+\nu_\mu$ decays of $f_{\mathrm{norm}} = 0.211$ combined with $\epsilon_{\mathrm{norm}}^{\mathrm{GLC}}$, the number of decays that pass GLC is found to be $N_{\mathrm{GLC}} = 6 \times 10^6$.

The resulting GLC efficiency ratios between the normalisation and signal channels in bins of q^2 are listed in Table 8.1. The GLC efficiency ratios are between 1.6 and 2.2, which reflects that the GLC applied to the signal channel is stricter than that applied to the normalisation channel. Furthermore, the GLC efficiency of the signal increases with q^2.

Table 8.1 – GLC efficiency ratio $R_\epsilon^{\mathrm{GLC}} = \epsilon_{\mathrm{norm}}^{\mathrm{GLC}}/\epsilon_{\mathrm{sig}}^{\mathrm{GLC}}$ between the normalisation and signal channels in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	$R_\epsilon^{\mathrm{GLC}}$
1	$2.241 \pm 0.011\,(0.5\%)$
2	$1.993 \pm 0.010\,(0.5\%)$
3	$1.860 \pm 0.009\,(0.5\%)$
4	$1.786 \pm 0.009\,(0.5\%)$
5	$1.723 \pm 0.008\,(0.5\%)$
6	$1.678 \pm 0.008\,(0.5\%)$
7	$1.662 \pm 0.007\,(0.4\%)$
8	$1.648 \pm 0.007\,(0.4\%)$
9	$1.616 \pm 0.006\,(0.4\%)$
10	$1.632 \pm 0.005\,(0.3\%)$

This can be explained by the kinematic requirements that are particularly strict on the muons ($p_{\mathrm{T}} > 1.2\,\mathrm{GeV}/c$ and $p > 5\,\mathrm{GeV}/c$), and since muons at low q^2 are characterised by softer p_{T} and p distributions, they are less likely to pass the GLC requirement than the muons at higher q^2.

8.1.2 Reconstruction and stripping efficiency

In order to reconstruct signal candidates passing GLC, their visible final state particles must produce three long tracks that form a vertex displaced from the primary vertex. Additionally, to pass the stripping line, they must satisfy different kinematic, geometric and `PID` requirements, as defined in Table 4.3. Since `PID` efficiencies are evaluated separately in Sec. 8.1.4, the combined efficiency of the reconstruction and stripping selection is computed without including the `PID` selection specified in Table 8.6. In the following, the stripping efficiency is understood to include the reconstruction and exclude the `PID` selection, and is therefore defined as:

$$\epsilon^{\mathrm{strip}} = \frac{\text{\# GLC \& stripping}}{\text{\# GLC}}. \tag{8.9}$$

Note that the requirements `nSPDHits` < 450 and `IsMuon`$(\mu) = 1$ are included in the stripping efficiency.

To estimate the numerator of Eq. 8.9, simulated signal and normalisation decays selected with the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line are used. However, since this line imposes `PIDmu` < 2 on both pions, it is necessary to estimate the efficiency, $\epsilon_{\mathtt{PIDmu}(\pi)}$, of this requirement. To do this the `B2XuMuNuBu2Rho_NoPIDPiLine` stripping line is used. Assuming the `PID` selection efficiencies of the pions and muons factorise, $\epsilon_{\mathtt{PIDmu}(\pi)}$ is the ratio of simulated decay candidates with and without the `PIDmu`$(\pi^{\pm}) < 2$ requirement. For the signal channel, $\epsilon_{\mathtt{PIDmu}(\pi)}$ is obtained in each bin of q^2, while for the normalisation channel

Table 8.2 – Stripping efficiency ratio $R_\epsilon^{\mathrm{strip}} = \epsilon_{\mathrm{norm}}^{\mathrm{strip}}/\epsilon_{\mathrm{sig}}^{\mathrm{strip}}$ between the normalisation and signal channels in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	$R_\epsilon^{\mathrm{strip}}$
1	$0.277 \pm 0.002\,(0.6\%)$
2	$0.257 \pm 0.002\,(0.6\%)$
3	$0.259 \pm 0.002\,(0.6\%)$
4	$0.262 \pm 0.002\,(0.6\%)$
5	$0.276 \pm 0.002\,(0.6\%)$
6	$0.292 \pm 0.002\,(0.6\%)$
7	$0.317 \pm 0.002\,(0.6\%)$
8	$0.355 \pm 0.002\,(0.6\%)$
9	$0.435 \pm 0.002\,(0.5\%)$
10	$0.797 \pm 0.004\,(0.5\%)$

it is obtained in the full q^2 region. The relative statistical uncertainties of these efficiencies are $< 0.1\%$ and therefore negligible. The yields of the signal and normalisation decays that pass the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line are weighted according to $\epsilon_{\mathrm{PIDmu}(\pi)}$ to get the yields after stripping without any `PID` requirements. Finally, combining this with the numbers of decays that pass GLC, described in Sec. 8.1.1, the stripping efficiencies $\epsilon_{\mathrm{norm}}^{\mathrm{strip}}$ and $\epsilon_{\mathrm{sig}}^{\mathrm{strip}}$ and their statistical uncertainties are obtained with Eq. 8.4.

The resulting stripping efficiency ratios between the normalisation and signal channels in bins of q^2 are listed in Table 8.2. The stripping efficiency ratios are in the range 0.3–0.8. The small stripping efficiency of the normalisation channel with respect to that of the signal channel can be explained by the strict $\chi_{\mathrm{vtx}}^2(B^+)$ requirement that only selects D^0 mesons with short flight distances leading to a similar topology between the signal and normalisation channel. Furthermore, the stripping efficiency of the signal decreases with q^2, which can be explained by the stricter kinematic requirements of the pions where at least one pion must satisfy $p_\mathrm{T} > 900\,\mathrm{MeV}/c$ and $p > 5\,\mathrm{GeV}/c$, and since pions at high q^2 are characterised by softer p_T and p distributions, they are less likely to pass the stripping requirement than the pions at lower q^2.

Tracking correction

The track reconstruction efficiency in simulation can differ from that in data. In order to take this possible data/MC difference into account, a data-driven tag-and-probe method [198] based on the $J/\psi \to \mu^+\mu^-$ decay is used to obtain a correction factor, $f_{\mathrm{corr}}^{\epsilon,\mathrm{track}}$, to the track reconstruction efficiency determined from simulation, $\epsilon^{\mathrm{track}}$, *i.e.*:

$$\epsilon_{\mathrm{corr}}^{\mathrm{track}} = f_{\mathrm{corr}}^{\epsilon,\mathrm{track}} \times \epsilon^{\mathrm{track}}\,, \qquad \text{where} \qquad f_{\mathrm{corr}}^{\epsilon,\mathrm{track}} = \frac{\epsilon_{\mathrm{data}}^{\texttt{tag-and-probe}}}{\epsilon_{\mathrm{MC}}^{\texttt{tag-and-probe}}}\,, \tag{8.10}$$

where $\epsilon_{\text{data}}^{\texttt{tag-and-probe}}$ and $\epsilon_{\text{MC}}^{\texttt{tag-and-probe}}$ are the track reconstruction efficiencies determined by the tag-and-probe method for data and simulation, respectively. Note that the track reconstruction efficiency is included in the stripping efficiency.

In the tag-and-probe method, one of the two muons is the *tag* track, *i.e.* it is a fully reconstructed long track that is required to pass a tight selection to ensure that it originates from a $J/\psi \rightarrow \mu^+\mu^-$ decay, while the second muon is the *probe* track, *i.e.* it is only partially reconstructed with one of three dedicated algorithms designed to probe one or two of the three LHCb tracking detectors: the VELO, the T stations and the TT, defined in Sec. 3.3. The probe tracks reconstructed with one of the algorithms are used to calculate the track reconstruction efficiency as the fraction of all probe tracks that can be matched with a long track. Finally, the efficiencies resulting from the three algorithms are combined into one single reconstruction efficiency for long tracks.

The tracking efficiency correction is performed by the TRACKCALIB2 tool [199]. Since the track reconstruction efficiency depends on the track momentum and pseudo rapidity, it is determined in five bins in p and two bins in η. In each bin, the true number of probe tracks that can be matched with a long track and those that fail to do so are determined by fitting their invariant mass with the sum of two Crystal Ball functions [167] corresponding to the $J/\psi \rightarrow \mu^+\mu^-$ component, while the background component is modelled with an exponential, as illustrated in Fig. 8.2.

This procedure is performed for 2018 data and simulation of the $B \rightarrow J/\psi(\rightarrow \mu^+\mu^-)X$ processes where X can be any particle, and the resulting binned tag-and-probe efficiencies are illustrated in Fig. 8.2 together with their ratio corresponding to the correction factor in Eq. 8.10. The tracks of the simulated signal and normalisation decays are assigned a correction factor depending on which p and η bin that they fall into, and their product is used to assign an overall correction factor to the decay candidate. To get the correction factor to the track efficiency ratio, the average correction factor of the normalisation channel is divided with that of the signal channel. A smearing is applied by generating 1000 different versions of the 2D correction histogram, where each bin is varied with a Gaussian around its mean with a sigma given by the uncertainty determined from the tag-and-probe procedure. This results in a distribution of correction factors, where the mean is used as the final correction to the track efficiency ratio, $f_{\text{corr}}^{R,\text{track}}$, while the standard deviation, $\sigma_{\text{corr}}^{R,\text{track}}$, is used to estimate the systematic uncertainty in Sec. 8.2.1. This procedure is repeated for each q^2 bin, and the result is presented in Table 8.3. The corrections to the tracking efficiency ratios are all consistent with one.

8.1.3 Trigger efficiency

The signal and normalisation decays that pass the reconstruction and stripping selection are required to fire each of the three trigger lines `muplus_L0MuonDecision_TOS`, `Bplus_Hlt1TrackMVADecision_TOS` and `Bplus_Hlt2TopoMu3BodyDecision_TOS`, described in Sec. 4.2. Thus, the trigger efficiency is defined as:
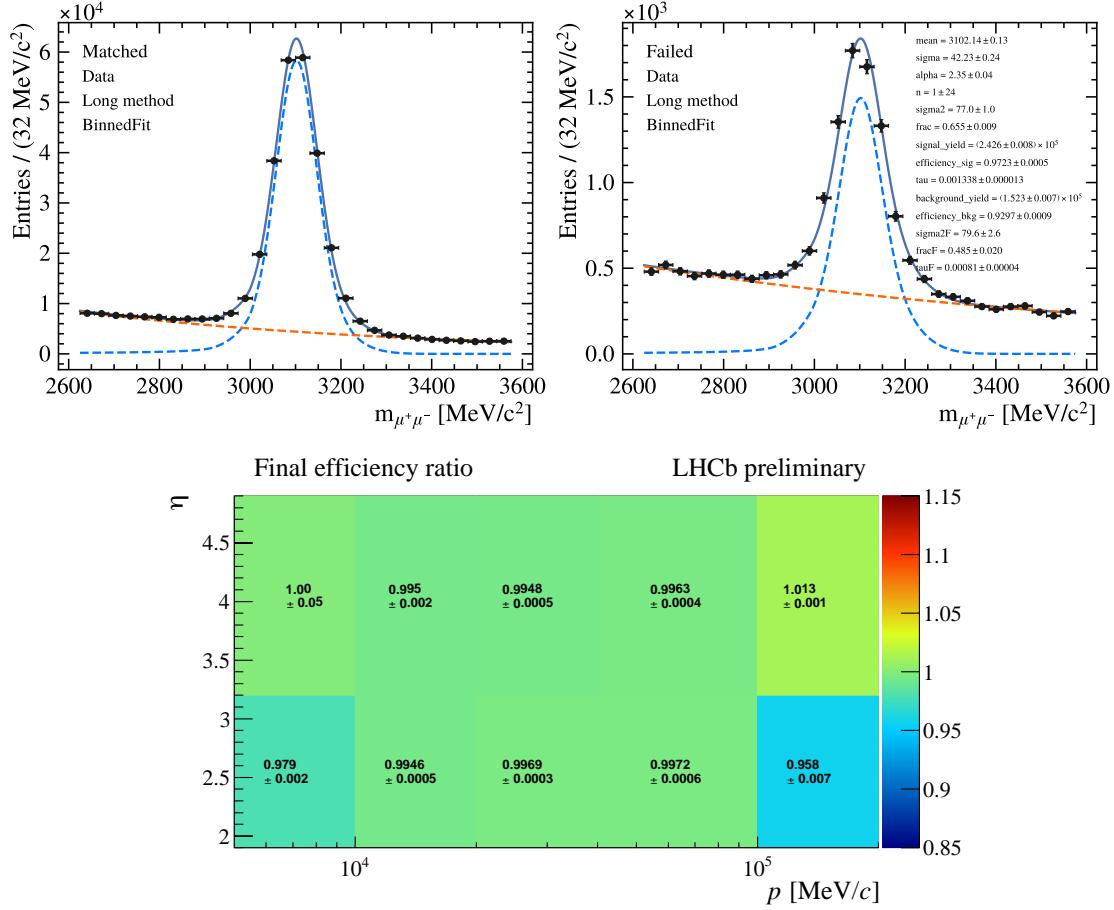
Figure 8.2 – Example of fit to the invariant mass of matched (top left) and failed (top right) probe tracks from the $J/\psi \rightarrow \mu^+ \mu^-$ decay reconstructed with one of the three algorithms called *the long method* using 2018 data. The resulting ratio of the tag-and-probe efficiencies between data and MC in bins of p and η is shown (bottom).

$$\epsilon^{\text{trig}} = \frac{\# \text{ stripping \& trigger}}{\# \text{ stripping}} . \tag{8.11}$$

To estimate the numerator of Eq. 8.11, simulated signal and normalisation decays selected with the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line are used. Again, it is necessary to correct for the `PIDmu`$(\pi^\pm) < 2$ requirement applied by the stripping line. Thus, to obtain the yields of the signal and normalisation decays after the trigger selection without any `PID` requirements, the same approach as explained in Sec. 8.1.2 is used. Finally, combining this with the numbers of decays passing the reconstruction and stripping selection, described in Sec. 8.1.2, the trigger efficiencies $\epsilon^{\text{trig}}_{\text{norm}}$ and $\epsilon^{\text{trig}}_{\text{sig}}$ and their statistical uncertainties are obtained with Eq. 8.4.

The resulting trigger efficiency ratios between the normalisation and signal channels in

Table 8.3 – Correction factors $f_{\text{corr}}^{R,\text{track}}$ in each q^2 bin used to correct the track reconstruction efficiency ratios (included in the stripping efficiency ratio) based on simulation with the relative uncertainty $\sigma_{\text{corr}}^{R,\text{track}}/f_{\text{corr}}^{R,\text{track}}$ given in parenthesis.

q^2 bin	$f_{\text{corr}}^{R,\text{track}} \pm \sigma_{\text{corr}}^{R,\text{track}}$
1	$1.000 \pm 0.001 \,(0.1\%)$
2	$1.000 \pm 0.001 \,(0.1\%)$
3	$1.000 \pm 0.001 \,(0.1\%)$
4	$1.001 \pm 0.001 \,(0.1\%)$
5	$1.001 \pm 0.001 \,(0.1\%)$
6	$1.001 \pm 0.001 \,(0.1\%)$
7	$1.001 \pm 0.001 \,(0.1\%)$
8	$1.001 \pm 0.002 \,(0.2\%)$
9	$1.001 \pm 0.002 \,(0.2\%)$
10	$1.002 \pm 0.004 \,(0.4\%)$

Table 8.4 – Trigger efficiency ratio $R_\epsilon^{\text{trig}} = \epsilon_{\text{norm}}^{\text{trig}}/\epsilon_{\text{sig}}^{\text{trig}}$ between the normalisation and signal channels in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	R_ϵ^{trig}
1	$0.985 \pm 0.002 \,(0.2\%)$
2	$0.974 \pm 0.002 \,(0.2\%)$
3	$0.974 \pm 0.002 \,(0.2\%)$
4	$0.966 \pm 0.002 \,(0.2\%)$
5	$0.959 \pm 0.002 \,(0.2\%)$
6	$0.949 \pm 0.002 \,(0.2\%)$
7	$0.941 \pm 0.001 \,(0.2\%)$
8	$0.930 \pm 0.001 \,(0.2\%)$
9	$0.919 \pm 0.001 \,(0.1\%)$
10	$0.902 \pm 0.001 \,(0.1\%)$

bins of q^2 are listed in Table 8.4. The trigger efficiency ratios are between 0.9 and 1.0. Since the signal and normalisation decays have the same final state and topology after the stripping, their trigger efficiencies are expected to be rather similar. However, the trigger efficiency of the signal increases with q^2. The reason for this is that muons at high q^2 are characterised by higher momenta, and since the q^2 distribution of the normalisation channel has its endpoint at a smaller value of q^2, the momenta of the signal muons in this region of q^2 are higher compared to that of the muons from the normalisation channel, and thus, these signal decays are more likely to be selected by the L0 trigger line.

Trigger correction

The trigger response in simulation can differ from that in data. In order to take possible data/MC differences into account, the so-called `TISTOS` method [200] is applied on a sample

of $B^0 \to J/\psi K^{*0}$ decays in data and simulation. This produces a correction factor to the efficiency ratio between the normalisation and signal channels as well as an estimate of the systematic uncertainty related to the trigger efficiency. Ideally this procedure would be performed on the control channel $B^0 \to J/\psi \rho^0$ used in Sec. 4.6. However, due to its small branching fraction and the resulting small yield in data, the statistical uncertainty would become too big. The decay $B^0 \to J/\psi K^{*0}(\to K^\pm \pi^\mp)$ with a ~ 50 times larger branching fraction has the same topology and angular structure as the $B^0 \to J/\psi \rho^0(\to \pi^+ \pi^-)$ decay when replacing a π with a K, and the trigger selections used in this analysis are not sensitive to this exchange.

The `TISTOS` method is used to determine the trigger efficiency in data. Here the trigger efficiency is defined as the number of signal candidates that pass both the stripping and trigger selection versus those that only pass the stripping selection [200]:

$$\epsilon^{\text{trig}|\text{strip}} = \frac{N_{\text{trig}|\text{strip}}}{N_{\text{strip}}} \, . \tag{8.12}$$

However, it is not possible to determine N_{strip} directly from data. For this reason, the `TIS` and `TOS` categories defined in Sec. 4.2 are used to obtain the trigger efficiency in data. Under the assumption that the `TIS` and `TOS` efficiencies are uncorrelated, the trigger efficiency can be obtained by the `TISTOS` efficiency:

$$\epsilon^{\text{TISTOS}} = \frac{N_{\text{TISTOS}|\text{strip}}}{N_{\text{TIS}|\text{strip}}} \, , \tag{8.13}$$

where $N_{\text{TISTOS}|\text{strip}}$ is the number of events where both the signal candidate and the rest of the event are sufficient to trigger the event, and N_{TIS} is the number of events where the rest of the event is sufficient to trigger the event. In both cases, only events passing the stripping selection are considered.

However, in reality the `TIS` and `TOS` efficiencies are not completely uncorrelated due to the other b hadron of the event, which is likely to trigger the event in the `TIS` configuration. Thus, ϵ^{TISTOS} in Eq. 8.13 is biased, and it can therefore not be directly used as the trigger efficiency. However, the bias cancels to a very large degree in the ratio of the `TISTOS` efficiency between the data, $\epsilon^{\text{TISTOS}}_{\text{data}}$, and simulation, $\epsilon^{\text{TISTOS}}_{\text{MC}}$, and it is therefore used as a correction factor to the trigger efficiency obtained with MC:

$$\epsilon^{\text{trig}}_{\text{corr}} = f^{\epsilon, \text{trig}}_{\text{corr}} \times \epsilon^{\text{trig}} \, , \qquad \text{where} \qquad f^{\epsilon, \text{trig}}_{\text{corr}} = \frac{\epsilon^{\text{TISTOS}}_{\text{data}}}{\epsilon^{\text{TISTOS}}_{\text{MC}}} \, . \tag{8.14}$$

The decay $B^0 \to J/\psi K^{*0}$ is used to mimic the signal and normalisation channels. Thus, a K^{*0} meson is combined with one muon, forming a B meson, which is then subsequently combined with a muon of the opposite charge. The dimuon is obtained with the stripping line `StrippingFullDSTDiMuonJpsi2MuMuDetachedLine`. The selection

of the B meson resembles the stripping selection of the $\pi^+\pi^-\mu^+$ combination for the signal and normalisation channels without the `PID` requirements defined in Table 8.6. The aforementioned dimuon stripping line applies a `PID` requirement corresponding to `PIDmu`$(\mu^\pm) > 0$, however, its effect on the `TISTOS` efficiency is assumed to be negligible.

Since the trigger selection depends mainly on the transverse momentum and the impact parameter, the $B^0 \to J/\psi K^{*0}$ samples based on the data and MC are split in four bins of $\ln(\sum p_\mathrm{T})$ and $\ln(\chi^2_{\mathrm{IP}\mu})$, respectively. The $\ln(\sum p_\mathrm{T})$ variable is chosen for the binning as it is used in the `Hlt2Mu3BodyTopo` [147] trigger line. In each bin, the yield of $B^0 \to J/\psi K^{*0}$ decays is obtained by fitting the invariant mass of the B^0 meson with a Crystal Ball function [167], introduced in Sec. 4.6.2, while the background is modelled with a Chebyshev polynomial, as illustrated in Fig. 8.3. To obtain the yields used to calculate the efficiency in Eq. 8.13, the aforementioned fitting procedure is performed with the following `TIS` and `TOS` requirements applied:

$$\mathtt{TIS} \quad = \quad \mathtt{L0\_TIS=1}\,\&\,\mathtt{Hlt1\_TIS=1}\,\&\,\mathtt{Hlt2\_TIS=1}\,, \tag{8.15}$$

$$\mathtt{TOS} \quad = \quad \mathtt{L0Muon=1}\,\&\,\mathtt{Hlt1TrackMVA=1}\,\&\,\mathtt{Hlt2Mu3BodyTopo=1}\,, \tag{8.16}$$

$$\epsilon^{\mathtt{TISTOS}} \quad = \quad \frac{\#(\mathtt{TIS}\,\&\,\mathtt{TOS})}{\#(\mathtt{TIS})}\,. \tag{8.17}$$

The binned `TISTOS` efficiencies for the data and simulation are illustrated in Fig. 8.3 together with their ratio corresponding to the correction factor in Eq. 8.14. The simulated signal and normalisation decays are assigned a correction factor depending on which $\ln(\sum p_\mathrm{T})$ and $\ln(\chi^2_{\mathrm{IP}\mu})$ bin that they fall into. To get the overall correction factor to the trigger efficiency ratio, the correction factor of the normalisation channel averaged over $\ln(\sum p_\mathrm{T})$ and $\ln(\chi^2_{\mathrm{IP}\mu})$ is divided with that of the signal channel. A smearing is applied by generating 1000 different versions of the 2D correction histogram, where each bin is varied with a Gaussian around its mean with a sigma given by the uncertainty determined from the $B^0 \to J/\psi K^{*0}$ decay. This results in a distribution of correction factors, where the mean is used as the final correction to the trigger efficiency ratio, $f_{\mathrm{corr}}^{R,\mathrm{trig}}$, while the standard deviation, $\sigma_{\mathrm{corr}}^{R,\mathrm{trig}}$, is used to estimate the systematic uncertainty in Sec. 8.2.2. This procedure is repeated for each q^2 bin, and the result is presented in Table 8.5. The corrections to the trigger efficiency ratios are all very close to one.

8.1.4 PID efficiency

Since the `PID` response in simulation is known to differ from that in data, the efficiencies of the `PID` requirements present in the stripping and preselection are estimated with the PIDCALIB package described in Sec. 6.2.3. Each visible final state particle is required to pass a specific set of `PID` requirements denoted `PID`$_{\mu^+}$, `PID`$_{\pi^+}$ and `PID`$_{\pi^-}$ and summarised
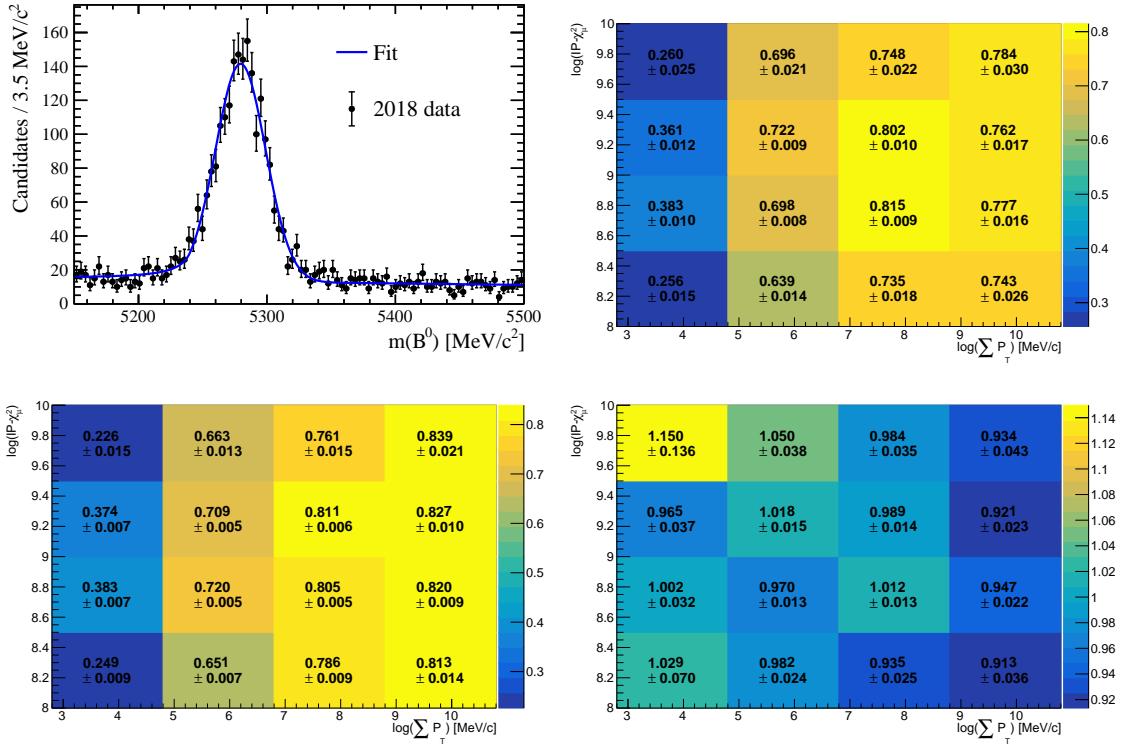
Figure 8.3 – Example of `TIS` fit of the $B^0 \rightarrow J/\psi K^{*0}$ decay in 2018 data (top left) for the determination of the `TISTOS` efficiency in one specific bin. The `TISTOS` efficiency in bins of $\ln(\sum p_{\mathrm{T}})$ and $\ln(\chi^2_{\mathrm{IP}\mu})$ for the $B^0 \rightarrow J/\psi K^{*0}$ decay for 2018 data (top right), 2018 MC (bottom left) and the ratio between the two (bottom right).

in Table 8.6. The total `PID` efficiency resulting from these requirements is defined as:

$$\epsilon^{\mathtt{PID}} = \frac{\# \text{ trigger \& } \mathtt{PID}}{\# \text{ trigger}}, \tag{8.18}$$

With PIDCALIB, the `PID` efficiency is binned in the track p, η and `nSPDHits` in data calibration samples containing particles of only one known species. Thus PIDCALIB is used to obtain the efficiencies of $\mathtt{PID}_{\mu^+}$ for true muons and of $\mathtt{PID}_{\pi^+}$ and $\mathtt{PID}_{\pi^-}$ for true pions. Subsequently, these `PID` efficiencies are assigned to the relevant tracks of the reference sample according to their p, η and `nSPDHits` bins.

The reference samples consist of simulated signal and normalisation decay candidates that are selected with the `B2XuMuNuBu2Rho_NoPIDMuLine` and `B2XuMuNuBu2Rho_NoPIDPiLine` stripping lines. As the requirements `nSPDHits` < 450 and `IsMuon`$(\mu^+) = 1$ as well as the trigger selection have already been included in the previous steps, these requirements are applied to the reference samples before assigning the `PID` efficiencies.

The two binning schemes defined in Table 8.7 are used to compute the $\mathtt{PID}_{\mu^+}$ and $\mathtt{PID}_{\pi^\pm}$ efficiencies. Using these two binning schemes, all bins of the calibration samples are

Table 8.5 – Correction factors $f_{\text{corr}}^{R,\text{trig}}$ in each q^2 bin used to correct the trigger efficiency ratios based on simulation with the relative uncertainty $\sigma_{\text{corr}}^{R,\text{trig}}/f_{\text{corr}}^{R,\text{trig}}$ given in parenthesis.

q^2 bin	$f_{\text{corr}}^{R,\text{trig}} \pm \sigma_{\text{corr}}^{R,\text{trig}}$
1	$0.999 \pm 0.004\,(0.4\%)$
2	$0.999 \pm 0.004\,(0.4\%)$
3	$1.000 \pm 0.003\,(0.3\%)$
4	$1.000 \pm 0.003\,(0.3\%)$
5	$1.001 \pm 0.003\,(0.3\%)$
6	$1.001 \pm 0.003\,(0.3\%)$
7	$1.002 \pm 0.003\,(0.3\%)$
8	$1.002 \pm 0.003\,(0.3\%)$
9	$1.002 \pm 0.003\,(0.3\%)$
10	$1.001 \pm 0.005\,(0.5\%)$

Table 8.6 – `PID` requirements applied to final state particles i, denoted PID_i.

Particle (i)	PID_i
μ^+	`PIDmu > 3.0`
π^+	`PIDmu < 2.0, PIDK < −2.0, ProbNNp < 0.7`
π^-	`IsMuon ≠ 1.0, PIDmu < 0.0, PIDK < −2.0, ProbNNp < 0.7`

populated with tracks, and the `PID` efficiency can therefore be computed in all bins. Moreover, the binning ensures that all of the tracks in the reference samples are within the bin limits, and they are therefore assigned `PID` efficiencies.

For the signal channel, the average PID_{μ^+}, PID_{π^+} and PID_{π^-} efficiencies are obtained for each true q^2 bin, while for the normalisation channel, they are obtained in the full q^2 region. As the PID_{μ^+}, PID_{π^+} and PID_{π^-} efficiencies are assumed to factorise for each channel, the total `PID` efficiencies $\epsilon_{\text{norm}}^{\text{PID}}$ and $\epsilon_{\text{sig}}^{\text{PID}}$ are computed as the product of the individual `PID` efficiencies.

The statistical uncertainties of the efficiencies arise from the limited size of the calibration and reference samples. However, their calculation becomes complicated when the `PID` selection is applied to more than one track of the event leading to correlations between events if they share the same efficiency bin. In general, the statistical uncertainties are expected to be small due to the large calibration samples used by PIDCALIB, in particular, compared to the systematic uncertainties. This can be illustrated by the following estimate. Assuming that the statistical uncertainties of the efficiencies assigned by PIDCALIB to the tracks associated with either muons or pions are fully correlated, the total statistical uncertainty is obtained by adding all uncertainties and dividing by the number of tracks in the considered region of the reference sample. The uncertainty of the combined efficiency of muons and pions is obtained with standard error propagation where their statistical uncertainties are assumed to be uncorrelated, since they are determined

Table 8.7 – Binning of the variables p, η and `nSPDHits` used to compute `PID` efficiencies.

Variable	PID_{μ^+} efficiency binning
p	[6.0, 21.7, 37.3, 53.0, 68.7, 84.3, 100.0, 3000.0] GeV/c
η	[1.8, 2.5, 3.0, 3.4, 3.8, 5.0]
`nSPDHits`	[5, 95, 184, 272, 361, 450]
Variable	$\text{PID}_{\pi^\pm}$ efficiency binning
p	[3.0 , 19.2, 35.3, 51.5, 67.7 , 83.8, 100.0, 560.0] GeV/c
η	[1.8, 2.4, 2.8, 3.3, 3.7, 4.2, 5.2]
`nSPDHits`	[5, 95, 184, 272, 361, 450]

from different calibration samples. Finally, to estimate the uncertainty of the efficiency ratio, the statistical uncertainties of the normalisation and signal channels are first assumed to be fully correlated, as they use the same calibration samples, however, to obtain a conservative estimate, they are also assumed to be completely uncorrelated. This results in two estimations for the statistical uncertainties denoted $\sigma_{\min}^{\text{PID}}$ and $\sigma_{\max}^{\text{PID}}$, respectively. The maximal relative $\sigma_{\min}^{\text{PID}}$ is $\sim 0.02\%$ and for $\sigma_{\max}^{\text{PID}}$ it is $\sim 0.2\%$.

The resulting `PID` efficiency ratios as well as $\sigma_{\min}^{\text{PID}}$ and $\sigma_{\max}^{\text{PID}}$ are listed in Table 8.8 for each q^2 bin. The true statistical uncertainty is expected to be somewhere between $\sigma_{\min}^{\text{PID}}$ and $\sigma_{\max}^{\text{PID}}$, however, $\sigma_{\max}^{\text{PID}}$ is used as a conservative estimation for the statistical uncertainty in the following calculations. The `PID` efficiency ratios are between 1.0 and 0.9. Due to the different coverages of the q^2 distributions of the normalisation and signal channels, they become kinematically more different at high q^2. Thus, for the signal channel the muons at high q^2 are characterised by higher momenta compared to the muons of the normalisation channel, while the opposite is true for the pions. The `PID` efficiency of the muons and pions are both found to increase with q^2. For the pions, the decreasing momenta makes it easier for the `PID` system to distinguish them from muons and kaons.

8.1.5 Preselection and MVA selection efficiency

After the `PID` selection, the signal and normalisation decays are required to pass the preselection and MVA selection that are not identical for the two decay modes. All preselection requirements imposed on the signal decays are defined in Tables 4.4. For the normalisation channel, the same requirements are applied, except for the one involving $\sigma_{m_{\text{corr}}(B^+)}$. This requirement is loosened, as defined in Table 6.1. The MVA selection for the signal and normalisation channels are different and are defined in Tables 7.1 and 6.1, respectively. The combined efficiency of the preselection and MVA selection without including the `PID` requirements is defined as:

$$\epsilon^{\text{pre+MVA}} = \frac{\# \; \text{PID \& pre + MVA}}{\# \; \text{PID}}, \tag{8.19}$$

Table 8.8 – `PID` efficiency ratio $R_\epsilon^{\mathrm{PID}} = \epsilon_{\mathrm{norm}}^{\mathrm{PID}}/\epsilon_{\mathrm{sig}}^{\mathrm{PID}}$ between the normalisation and signal channels in each q^2 bin with the relative uncertainties given in parentheses. The estimated minimum and maximum statistical uncertainties are denoted $\sigma_{\mathrm{min}}^{\mathrm{PID}}$ and $\sigma_{\mathrm{max}}^{\mathrm{PID}}$, where $\sigma_{\mathrm{min}}^{\mathrm{PID}}$ is found to be negligible.

q^2 bin	$R_\epsilon^{\mathrm{PID}}$	$\sigma_{\mathrm{min}}^{\mathrm{PID}}$	$\sigma_{\mathrm{max}}^{\mathrm{PID}}$
1	1.007	0.000 (0.0%)	0.002 (0.2%)
2	0.995	0.000 (0.0%)	0.002 (0.2%)
3	0.986	0.000 (0.0%)	0.002 (0.2%)
4	0.974	0.000 (0.0%)	0.002 (0.2%)
5	0.966	0.000 (0.0%)	0.002 (0.2%)
6	0.957	0.000 (0.0%)	0.002 (0.2%)
7	0.949	0.000 (0.0%)	0.002 (0.2%)
8	0.940	0.000 (0.0%)	0.002 (0.2%)
9	0.931	0.000 (0.0%)	0.002 (0.2%)
10	0.917	0.000 (0.0%)	0.002 (0.2%)

To estimate the numerator, simulated signal and normalisation decays based on the 2017 and 2018 data-taking configurations are used. The two samples are combined to reduce the statistical uncertainty of the efficiencies. In particular, it is done to reduce the statistical uncertainty of the signal efficiency, which reaches $\sim 1.2\%$ in certain bins of q^2 when using the 2017 simulation alone. In both simulations, decays are selected with the `B2XuMuNuBu2Rho_NoPIDMuLine` stripping line, and after applying the `PID` requirements in Sec. 8.1.4, the decays of different data-taking years are ensured to have undergone the same selection. Finally, with the numbers of signal and normalisation decays that pass the combined preselection and MVA selection versus those that pass the `PID` selection, the efficiencies $\epsilon_{\mathrm{norm}}^{\mathrm{pre+MVA}}$ and $\epsilon_{\mathrm{sig}}^{\mathrm{pre+MVA}}$ and their statistical uncertainties are obtained with Eq. 8.4.

The resulting preselection and MVA selection efficiency ratios of the normalisation and signal channels in each q^2 bin are listed in Table 8.9. The efficiency ratios are between 2.7 and 4.5. The low selection efficiency of the signal channel with respect to that of the normalisation channel reflects the stricter selection applied to the signal. Furthermore, the selection efficiency of the signal decreases with q^2 because the distribution of the output variable of the DNN becomes more populated at low value, *i.e.* less consistent with the signal hypothesis, and the signal decays at high q^2 are therefore less likely to pass the MVA selection.

8.2 Systematic uncertainties of the efficiencies

The different sources of systematic uncertainties related to the estimation of the efficiencies in Secs. 8.1.1–8.1.5 are identified as:

Table 8.9 – Preselection and MVA selection efficiency ratio $R_\epsilon^{\text{pre+MVA}} = \epsilon_{\text{norm}}^{\text{pre+MVA}}/\epsilon_{\text{sig}}^{\text{pre+MVA}}$ between the normalisation and signal channels in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	$R_\epsilon^{\text{pre+MVA}}$
1	$2.698 \pm 0.020\,(0.8\%)$
2	$2.930 \pm 0.022\,(0.8\%)$
3	$2.999 \pm 0.022\,(0.7\%)$
4	$3.060 \pm 0.023\,(0.7\%)$
5	$3.168 \pm 0.024\,(0.7\%)$
6	$3.211 \pm 0.024\,(0.7\%)$
7	$3.263 \pm 0.024\,(0.7\%)$
8	$3.363 \pm 0.025\,(0.7\%)$
9	$3.647 \pm 0.027\,(0.7\%)$
10	$4.456 \pm 0.033\,(0.8\%)$

- Determining the track reconstruction efficiency with simulations.

- Determining the trigger efficiency with simulations.

- Calculating PID efficiencies with PIDCALIB.

Since a good agreement is found for the variables used in the preselection and MVA selection between simulation and data, in particular, after applying the $w_{\text{nTracks}} \times w_{\text{mva}}$ weights, the preselection and MVA selection efficiency based on simulation is considered to be accurate. For the remaining efficiencies that rely on simulation only, *i.e.* the stripping efficiency (without the contribution from the track reconstruction efficiency) and the GLC efficiency, the simulation is assumed to provide a reasonably accurate description, in particular, when forming the ratio between the normalisation and the signal channels.

The systematic uncertainty associated with a specific selection step is estimated by varying the considered systematic source, applying the selection requirement (req), determining the resulting efficiency ratio, $R_{\epsilon,\text{syst}}^{\text{req}}$, and finally, comparing it to the nominal efficiency ratio, R_ϵ^{req}. In practice, the systematic uncertainty is calculated as $\Delta R_{\epsilon,\text{syst}}^{\text{req}} = |R_{\epsilon,\text{syst}}^{\text{req}} - R_\epsilon^{\text{req}}|$.

8.2.1 Tracking systematics

As described in Sec. 8.1.2, the track reconstruction efficiency that is included in the stripping efficiency obtained from simulation is corrected to account for possible data/MC differences. The uncertainties, $\sigma_{\text{corr}}^{R,\text{track}}$, of the final correction factors $f_{\text{corr}}^{R,\text{track}}$ are used to estimate the systematic uncertainties of the stripping efficiency ratio:

$$\Delta R_{\epsilon,\text{syst}}^{\text{strip}} = \sigma_{\text{corr}}^{R,\text{track}} \times R_\epsilon^{\text{strip}}, \tag{8.20}$$

Table 8.10 – Systematic uncertainty $\Delta R_{\epsilon,\text{syst}}^{\text{strip}}$ in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	$\Delta R_{\epsilon,\text{syst}}^{\text{strip}}$
1	$0.000\,(0.1\%)$
2	$0.000\,(0.1\%)$
3	$0.000\,(0.1\%)$
4	$0.000\,(0.1\%)$
5	$0.000\,(0.1\%)$
6	$0.000\,(0.1\%)$
7	$0.000\,(0.1\%)$
8	$0.001\,(0.2\%)$
9	$0.001\,(0.2\%)$
10	$0.004\,(0.4\%)$

where $R_{\epsilon}^{\text{strip}}$ is the stripping efficiency ratio based on simulation listed in Table 8.2. The resulting systematic uncertainties are summarised in Table 8.10 with small relative uncertainties between 0.1% and 0.4%.

8.2.2 Trigger systematics

As described in Sec. 8.1.2, the trigger efficiency obtained from simulation is corrected to account for possible data/MC differences. The uncertainties, $\sigma_{\text{corr}}^{R,\text{trig}}$, of the final correction factors $f_{\text{corr}}^{R,\text{trig}}$ are used to estimate the systematic uncertainties of the trigger efficiency ratio:

$$\Delta R_{\epsilon,\text{syst}}^{\text{trig}} = \sigma_{\text{corr}}^{R,\text{trig}} \times R_{\epsilon}^{\text{trig}}, \tag{8.21}$$

where $R_{\epsilon}^{\text{trig}}$ is the trigger efficiency ratio based on simulation listed in Table 8.4. The resulting systematic uncertainties are summarised in Table 8.11 with small relative uncertainties between 0.3% and 0.5%.

8.2.3 PID systematics

To assign a systematic uncertainty to the PID efficiencies estimated with PIDCalib, a different tool called PIDCorr which is also part of the PIDCalib package [153] is used.

The PIDCorr tool corrects the PID response in simulation by transforming a given PID variable in simulation, PID_{MC}, in such a way that it is distributed as in data. The corrected PID variable is computed as [129]

$$\text{PID}_{\text{corr}} = P_{\text{exp}}^{-1}(P_{\text{MC}}(\text{PID}_{\text{MC}}|p_{\text{T}},\eta,\texttt{nTracks})|p_{\text{T}},\eta,\texttt{nTracks}), \tag{8.22}$$

Table 8.11 – Systematic uncertainty $\Delta R_{\epsilon,\text{syst}}^{\text{trig}}$ in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	$\Delta R_{\epsilon,\text{syst}}^{\text{trig}}$
1	$0.004\,(0.4\%)$
2	$0.004\,(0.4\%)$
3	$0.003\,(0.3\%)$
4	$0.003\,(0.3\%)$
5	$0.003\,(0.3\%)$
6	$0.003\,(0.3\%)$
7	$0.002\,(0.3\%)$
8	$0.003\,(0.3\%)$
9	$0.003\,(0.3\%)$
10	$0.004\,(0.5\%)$

where $P_{\text{MC}}(\text{PID}_{\text{MC}}|p_{\text{T}}, \eta, \texttt{nTracks})$ is the cumulative distribution function (CDF) of PID_{MC} and $P_{\text{exp}}^{-1}(x|p_{\text{T}}, \eta, \texttt{nTracks})$ is the inverse CDF of the $\texttt{PID}$ variable in the calibration samples based on data. These functions are computed using a kernel density estimation procedure, also used for the PIDGEN tool described in Sec. 4.4.4. Finally, the PIDCORR tool computes a PID_{corr} distribution, which in contrast to the PIDGEN method, preserves the correlation with the simulation, and consequently, the correlations between different $\texttt{PID}$ variables related to the same track are mostly reproduced.

With PIDCORR the simulated distributions of the $\texttt{PID}$ variables in Table 8.6 are corrected, *i.e.* the $\texttt{PID}$ variables used in PID_{μ^+} are corrected for the simulated decays selected with the $\texttt{B2XuMuNuBu2Rho\_NoPIDMuLine}$ stripping line, while the $\texttt{PID}$ variables of the pions used in PID_{π^+} and PID_{π^-} are corrected for the simulated decays selected with the $\texttt{B2XuMuNuBu2Rho\_NoPIDPiLine}$ stripping line. The $\texttt{PID}$ requirements PID_{μ^+}, PID_{π^+} and PID_{π^-} are then applied to the samples using the corrected $\texttt{PID}$ variables only, and the product of the resulting efficiencies are used to obtain the total $\texttt{PID}$ efficiency. The efficiency ratios $R_{\epsilon,\text{syst}}^{\text{PID}}$ resulting from using PIDCORR and the estimated systematic uncertainties are listed in Table 8.12 with relative contributions between 0.4% and 1.3%.

8.3   Total efficiency ratio

The total efficiency ratios as well as their statistical and systematic uncertainties are given in Table 8.13. Here the total efficiency ratios in bins of q^2 defined in Eq. 8.2 are obtained by multiplying the efficiency ratios and correction factors of each selection step given in Tables 8.1–8.9, *i.e.*

$$R_\epsilon = R_\epsilon^{\text{GLC}} \times f_{\text{corr}}^{R,\text{track}} \times R_\epsilon^{\text{strip}} \times f_{\text{corr}}^{R,\text{trig}} \times R_\epsilon^{\text{trig}} \times R_\epsilon^{\text{PID}} \times R_\epsilon^{\text{pre+MVA}} . \qquad (8.23)$$

Since the efficiencies of each selection step are multiplied, the statistical uncertainties cancel

Table 8.12 – Systematic uncertainty $\Delta R_{\epsilon,\mathrm{syst}}^{\mathtt{PID}}$ in each q^2 bin with the relative uncertainty given in parenthesis.

q^2 bin	$R_{\epsilon}^{\mathrm{syst}}$	$\Delta R_{\epsilon,\mathrm{syst}}^{\mathtt{PID}}$
1	1.003	0.004 (0.4%)
2	0.988	0.007 (0.7%)
3	0.973	0.013 (1.3%)
4	0.965	0.009 (0.9%)
5	0.955	0.010 (1.1%)
6	0.945	0.012 (1.2%)
7	0.937	0.012 (1.2%)
8	0.928	0.012 (1.3%)
9	0.919	0.012 (1.3%)
10	0.910	0.007 (0.7%)

when the numerator and denominator of two efficiencies are the same. This is the case between ϵ^{GLC} and $\epsilon^{\mathrm{strip}}$, and between $\epsilon^{\mathrm{strip}}$ and ϵ^{trig}, where the number of events passing GLC and the stripping selection cancel. However, this is not the case for $\epsilon^{\mathtt{PID}}$, which is determined with PIDCALIB, and $\epsilon^{\mathrm{pre+MVA}}$, which is determined from the combined 2017 and 2018 simulations. Thus, to estimate the overall statistical uncertainty of R_{ϵ} the statistical uncertainties of the following $R_{\epsilon}^{\mathrm{req}}$ are propagated with standard error propagation assuming no correlation:

$$R_{\epsilon} = R_{\epsilon}^{\mathrm{trig/GLC}} \times R_{\epsilon}^{\mathtt{PID}} \times R_{\epsilon}^{\mathrm{pre+MVA}} , \tag{8.24}$$

Here $R_{\epsilon}^{\mathrm{trig/GLC}} = \epsilon_{\mathrm{norm}}^{\mathrm{trig/GLC}} / \epsilon_{\mathrm{sig}}^{\mathrm{trig/GLC}}$ where $\epsilon^{\mathrm{trig/GLC}}$ is the number of events that pass the trigger selection versus the total number of events produced before GLC. Since the different factors of Eq. 8.24 are to some degree positively correlated, which reduces the combined statistical uncertainty, this provides a conservative estimate for the overall statistical uncertainty of R_{ϵ}. The overall systematic uncertainty of the efficiency ratio in each q^2 bin is obtained by propagating the systematic uncertainties described in Sec. 8.2.1–8.2.3 with standard error propagation assuming no correlation.

As summarised in Table 8.13, the relative statistical and systematic uncertainties in bins of q^2 are in the ranges 0.9%–1.0% and 0.6%–1.4%, respectively. In general, the smallness of these uncertainties are due to the similarity of the signal and normalisation channels leading to a cancellation of uncertainties in the ratio between them.

The efficiency corrected yields of the signal and normalisation channels, *i.e.* $N_{\mathrm{sig},i}^{\mathrm{corr}} = N_{\mathrm{sig},i}/\epsilon_{\mathrm{sig},i}^{\mathrm{tot}}$ and $N_{\mathrm{norm}}^{\mathrm{corr}} = N_{\mathrm{norm}}/\epsilon_{\mathrm{norm}}^{\mathrm{tot}}$, are found in Appendix J.

Table 8.13 – Total efficiency ratio $R_\epsilon = \epsilon^{\text{tot}}_{\text{norm}}/\epsilon^{\text{tot}}_{\text{sig}}$ between the normalisation and signal channels in each q^2 bin. The total statistical and systematic uncertainties are denoted σ^{stat}_R and σ^{syst}_R, and their relative contributions to R_ϵ are given in parentheses.

q^2 bin	$R_\epsilon \pm \sigma^{\text{stat}}_R \pm \sigma^{\text{syst}}_R$
1	$1.659 \pm 0.016\,(1.0\%) \pm 0.010\,(0.6\%)$
2	$1.452 \pm 0.014\,(0.9\%) \pm 0.011\,(0.8\%)$
3	$1.387 \pm 0.013\,(0.9\%) \pm 0.019\,(1.4\%)$
4	$1.348 \pm 0.013\,(0.9\%) \pm 0.013\,(1.0\%)$
5	$1.396 \pm 0.013\,(0.9\%) \pm 0.016\,(1.1\%)$
6	$1.432 \pm 0.013\,(0.9\%) \pm 0.018\,(1.3\%)$
7	$1.539 \pm 0.014\,(0.9\%) \pm 0.019\,(1.2\%)$
8	$1.726 \pm 0.016\,(0.9\%) \pm 0.022\,(1.3\%)$
9	$2.202 \pm 0.020\,(0.9\%) \pm 0.030\,(1.3\%)$
10	$4.811 \pm 0.043\,(0.9\%) \pm 0.048\,(1.0\%)$

8.4 Final results

The measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction as a function of q^2 can now be performed. To do this, the measured yields of the signal decays in bins of q^2 and the normalisation decays in the full q^2 region are combined with the total efficiency ratios. While the final result of the differential branching fraction is blinded its uncertainties are given and compared to previous measurements.

8.4.1 Measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction

The differential branching fraction $\Delta\mathcal{B}_{\text{sig}}/\Delta q^2$ is computed with Eq. 8.1, where the unfolded signal yields in data, $N_{\text{sig},i}$, the measured yield of the normalisation channel, N_{norm}, and the total efficiency ratios, $R_{\epsilon,i}$, are taken from Tables 7.8, 6.11 and 8.13, respectively, while the branching fractions $\mathcal{B}(B^+ \to \bar{D}^0 \mu^+ \nu_\mu) = (2.30 \pm 0.09)\%$ and $\mathcal{B}(\bar{D}^0 \to \pi^+ \pi^-) = (1.454 \pm 0.024) \times 10^{-3}$ [15] are based on the external determinations reported in Ref. [15].

In each q^2 bin, the different uncertainties related to $N_{\text{sig},i}$, N_{norm}, $R_{\epsilon,i}$, $\mathcal{B}(\bar{D}^0 \to \pi^+ \pi^-)$ and $\mathcal{B}(B^+ \to \bar{D}^0 \mu^+ \nu_\mu)$ are all assumed to be uncorrelated. Thus, the statistical, systematic and external uncertainties of the differential branching fraction in each q^2 bin are obtained with standard error propagation and the total uncertainty is obtained by adding them in quadrature. The statistical and systematic uncertainties of $N_{\text{sig},i}$ are given in Table 7.8 and 7.16, respectively, and for $R_{\epsilon,i}$ both uncertainties are given in Table 8.13. The statistical uncertainty of N_{norm} is given in Table 6.11, while the uncertainty of $\mathcal{B}(\bar{D}^0 \to \pi^+ \pi^-)$ and $\mathcal{B}(B^+ \to \bar{D}^0 \mu^+ \nu_\mu)$ are taken as external uncertainties. The measurement of the differential branching fraction and its absolute uncertainties as a function of q^2 are blinded by multiplying their values in each bin with a common random number uniformly distributed

Table 8.14 – Blinded result of the differential branching fraction $\Delta\mathcal{B}_\mathrm{sig}/\Delta q^2$ of the $B^+ \to \rho^0\mu^+\nu_\mu$ decay with respect to q^2 using 2018 LHCb data. In each q^2 bin the value of $\Delta\mathcal{B}_\mathrm{sig}/\Delta q^2$ has been multiplied with a random number between zero and two. The statistical, systematic, external and total uncertainties are denoted σ^stat, σ^syst, σ^ext and σ^tot, respectively, and the relative uncertainties are given in parentheses. All quantities related to $\Delta\mathcal{B}_\mathrm{sig}/\Delta q^2$ are given in units of $10^{-6}\,\mathrm{GeV}^{-2}c^4$, while the q^2 range is given in units of GeV^2/c^4.

q^2 range	$\Delta\mathcal{B}_\mathrm{sig}/\Delta q^2$	σ^stat	σ^syst	σ^ext	σ^tot
0.0– 2.2	8.466	0.536 (6.3%)	0.124 (1.5%)	0.360 (4.2%)	0.657 (7.8%)
2.2– 3.9	10.087	0.562 (5.6%)	0.235 (2.3%)	0.428 (4.2%)	0.745 (7.4%)
3.9– 5.3	0.658	0.039 (5.9%)	0.023 (3.5%)	0.028 (4.2%)	0.053 (8.1%)
5.3– 6.7	1.036	0.064 (6.2%)	0.020 (1.9%)	0.044 (4.2%)	0.080 (7.7%)
6.7– 8.0	10.971	0.676 (6.2%)	0.327 (3.0%)	0.466 (4.2%)	0.883 (8.1%)
8.0– 9.3	10.345	0.626 (6.1%)	0.362 (3.5%)	0.439 (4.2%)	0.847 (8.2%)
9.3–10.7	11.064	0.638 (5.8%)	0.399 (3.6%)	0.470 (4.2%)	0.887 (8.0%)
10.7–12.2	4.751	0.281 (5.9%)	0.110 (2.3%)	0.202 (4.2%)	0.363 (7.6%)
12.2–14.2	9.068	0.509 (5.6%)	0.154 (1.7%)	0.385 (4.2%)	0.657 (7.2%)
14.2–21.5	5.791	0.320 (5.5%)	0.109 (1.9%)	0.246 (4.2%)	0.418 (7.2%)

between zero and two, different for each bin. The result of this procedure is summarised in Table 8.14, where the relative uncertainties represent the non-blinded precision of the $\Delta\mathcal{B}_\mathrm{sig}/\Delta q^2$ measurement. The relative statistical, systematic and external uncertainties in bins of q^2 are in the ranges 5.5%–6.3%, 1.5%–3.6% and 4.2%, respectively. Thus the uncertainty of the measurement is dominated by the statistical uncertainty which is mostly coming from the signal fit, while the largest systematic uncertainty is, in general, coming from the choice of the unfolding method. The individual relative systematic uncertainties related to the fit, unfolding procedure and selection efficiencies are summarised in Table 8.15. The relative total uncertainties in bins of q^2 are between 7.2% and 8.2%.

8.4.2 Comparison with previous measurements

The $B^+ \to \rho^0\mu^+\nu_\mu$ differential branching fraction has previously been measured by the BaBar [63] and Belle [62] experiments, as described in Sec. 2.2. The q^2 binning schemes used in the two measurements are different from that used in the measurement presented in this thesis. In the BaBar measurement, a binning scheme with three q^2 bins corresponding to $q^2 \in [0.0, 20.3]\,\mathrm{GeV}^2/c^4$ is used, while in the Belle measurement a uniform binning scheme with eleven q^2 bins corresponding to $q^2 \in [0.0, 22.0]\,\mathrm{GeV}^2/c^4$ is used. Thus, a comparison between the bin-by-bin precision of this measurement and that of the BaBar and Belle measurements cannot be made directly. However, a rough comparison can be performed by merging the bins of the different measurements in such a way that their bin boundaries are approximately matching.

Table 8.15 – Relative systematic uncertainties in % of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction in bins of q^2.

q^2 range [GeV^2/c^4]	0.0–2.2	2.2–3.9	3.9–5.3	5.3–6.7	6.7–8.0	8.0–9.3	9.3–10.7	10.7–12.2	12.2–14.2	14.2–21.5
Constraining exclusive $|V_{ub}|$ decays in Sec. 7.5.1.	1.2	1.5	1.4	0.8	1.4	1.1	0.5	0.5	0.2	0.5
Nonresonant $B^+ \to \pi^+ \pi^- \mu^+ \nu_\mu$ contribution in Sec. 7.5.2.	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.1	1.4
Line-shape model for the dipion mass (min) in Sec. 7.5.3.	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.4	0.2	0.0
Line-shape model for the dipion mass (max) in Sec. 7.5.3.	0.2	0.2	0.4	0.4	0.5	0.9	0.6	0.4	0.2	0.0
MisID contribution in Sec. 7.5.4.	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Unfolding method in Sec. 7.5.5.	0.2	1.1	2.8	1.2	2.2	2.8	3.2	1.7	0.8	0.6
Signal form factors in Sec. 7.5.6.	0.5	1.2	1.0	0.5	0.7	0.8	0.6	0.6	0.6	0.2
Tracking efficiency in Sec. 8.2.1.	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.4
Trigger efficiency in Sec. 8.2.2.	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.5
PID efficiency in Sec. 8.2.3.	0.4	0.7	1.3	0.9	1.1	1.2	1.2	1.3	1.3	0.7
Total systematic uncertainty.	1.5	2.3	3.5	1.9	3.0	3.5	3.6	2.3	1.7	1.9

Table 8.16 – Comparison between the relative total uncertainties (precision) of the $\Delta\mathcal{B}_{\mathrm{sig}}/\Delta q^2$ measurements performed at LHCb and Belle [62] in q^2 intervals with approximately matching boundaries.

Belle q^2 range $[\,\mathrm{GeV}^2/c^4\,]$	LHCb q^2 range $[\,\mathrm{GeV}^2/c^4\,]$	Belle precision	LHCb precision
0.0– 2.0	0.0– 2.2	27.4%	7.8%
2.0– 4.0	2.2– 3.9	20.3%	7.4%
4.0– 8.0	3.9– 8.0	11.5%	6.0%
8.0–12.0	8.0–12.2	12.5%	6.0%
12.0–20.0	12.2–21.5	8.3%	6.2%

Since the measurement performed by Belle is more precise than that of BaBar, the Belle measurement is used for the comparison.[1]

For the measurement presented in this thesis, referred to as the LHCb measurement, bins 3–4, bins 6–8 and bins 9–10 are merged, while bin 1 and bin 2 are unchanged. For the Belle measurement, bin 1 and bin 2 are also unchanged, while bins 3–4, bins 5–6 and bins 7–11 are merged. The resulting bin boundaries are shown in Table 8.16. For the Belle measurement, the total uncertainties of the partial branching fractions in bins of q^2 as well as their correlations are taken from Ref. [62] and used with standard error propagation to obtain the total uncertainties in the merged bins. For the LHCb measurement, the correlation between the signal yields of different q^2 bins are obtained from the unfolding procedure and used to propagate the statistical uncertainty of the signal yields in the merged bins. The possible correlation between the other statistical and systematic uncertainties of the bins making up the merged q^2 bins are not considered. The resulting precision in the five approximately matching q^2 intervals for the Belle and LHCb measurements are presented in Table 8.16.

In all q^2 intervals, the measurement of $\Delta\mathcal{B}_{\mathrm{sig}}/\Delta q^2$ at LHCb is found to be more precise than that at Belle, and in particular, for $q^2 < 12\,\mathrm{GeV}^2/c^4$ the relative uncertainty of the LHCb measurement is less than half of that of the Belle measurement. The precision of the LHCb measurement is expected to further improve when adding data collected during 2017. Based on the precision of the $\Delta\mathcal{B}_{\mathrm{sig}}/\Delta q^2$ measurement presented in this thesis, the prospects of obtaining a precise measurement of the $|V_{ub}|$ matrix element and the $B^+ \to \rho^0$ form factor parameters based on the $B^+ \to \rho^0\mu^+\nu_\mu$ channel at LHCb are promising.

---

[1]The BaBar [63] and Belle [62] measurements are combined into one measurement of the differential branching fraction in Ref. [64] to improve the precision of the $B^+ \to \rho^0\mu^+\nu_\mu$ differential branching fraction, however, since the results of the two measurements disagree strongly, this result is not used for the comparison.

9 Summary and outlook

The decay $B^+ \to \rho^0 \mu^+ \nu_\mu$ provides an exclusive semileptonic determination of the CKM matrix element $|V_{ub}|$ valuable for understanding the long-standing tension between determinations based on either inclusive and exclusive semileptonic decays. Moreover, the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay is associated with a rich helicity structure expressed by the full differential branching fraction that provides more variables to test the Standard Model and to probe new physics than measuring $|V_{ub}|$ alone.

This thesis presents the first measurement of the exclusive semileptonic decay $B^+ \to \rho^0 \mu^+ \nu_\mu$ at the LHCb experiment, using data collected during 2018. A measurement of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction as a function of q^2 is performed relative to the normalisation channel $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ with $\bar{D}^0 \to \pi^+ \pi^-$, where the calculation of the ratio between the differential branching fraction of the signal and the integrated branching fraction of the normalisation channel leads to the cancellation of many uncertainties.

The current precision of the result is given in the thesis, with the central values still kept blind to avoid accidental biases before the final publication. The relative statistical, systematic and external uncertainties in bins of q^2 are in the ranges 5.5%–6.3%, 1.5%–3.6% and 4.2%, respectively. A rough comparison shows that the uncertainties of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ differential branching fraction measured at LHCb are significantly smaller than for those of previous measurements. This is expected to further improve when adding data collected in 2017.

Based on the measurement presented in this thesis, the prospects of obtaining a precise measurement of the $|V_{ub}|$ matrix element from the $B^+ \to \rho^0 \mu^+ \nu_\mu$ channel at LHCb are promising. Moreover, this work paves the way for measuring the full differential branching fraction of the $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay at LHCb.

A The corrected B mass

The decay $B^+ \to \rho^0 \mu^+ \nu_\mu$ with an unmeasured neutrino is illustrated in Fig. A.1. The direction of the B^+ meson momentum is constrained to that defined by the PV and SV.

The coordinate system is defined with one axis parallel to the B flight direction, $\hat{p}_\parallel$, and the other perpendicular to it, $\hat{p}_\perp$. Thus, the momenta of all particles, $\vec{p}(x_i)$, can be split up in a component parallel, $p_\parallel(x_i)$, and perpendicular, $p_\perp(x_i)$, to the B decay flight direction.
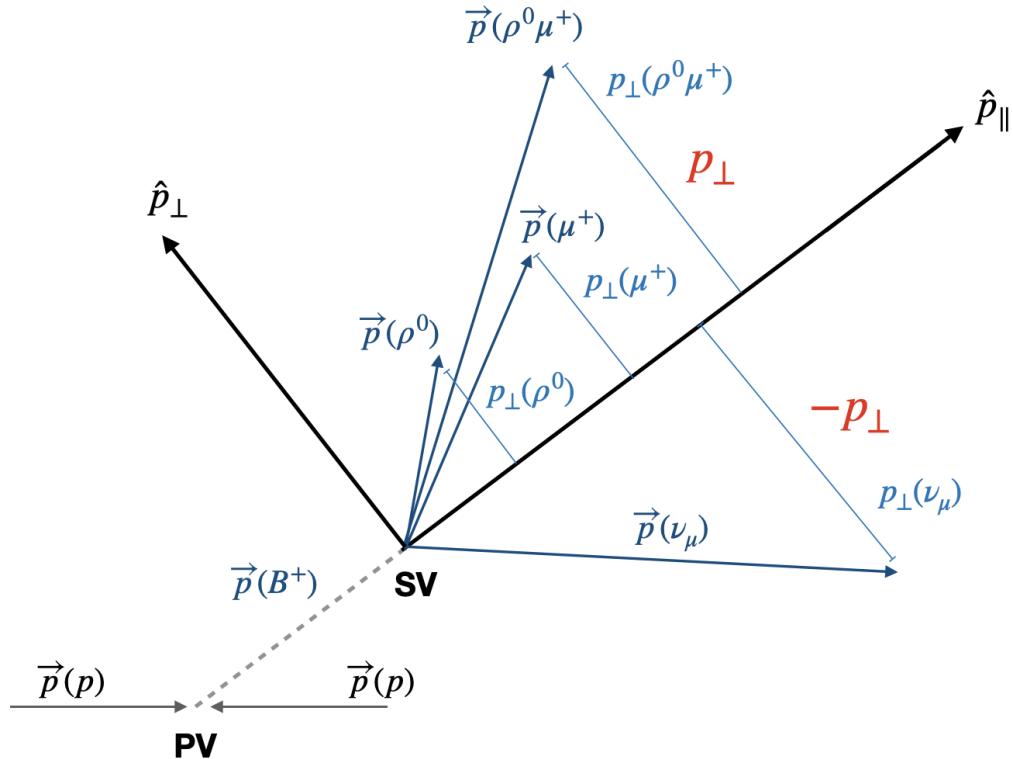

Figure A.1 – The $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay seen in the laboratory frame in a coordinate system with one axis parallel to the B flight direction, $\hat{p}_\parallel$, and the other perpendicular to it, $\hat{p}_\perp$ leading to $\vec{p}_\perp(\mu^+) + \vec{p}_\perp(\rho^0) = -\vec{p}_\perp(\nu_\mu) \equiv \vec{p}_\perp$.

Appendix A. The corrected B mass

Under the assumption of energy and momentum conservation, the four-vectors of the particles B, ρ^0, μ^+ and ν_μ satisfy:

$$P_{B^+} = P_{\rho^0} + P_{\mu^+} + P_{\nu_\mu} \ . \tag{A.1}$$

In the rest frame of the B^+ meson, assuming the neutrino mass to be negligible, *i.e.* $E_{\nu_\mu} = |\vec{p}(\nu_\mu)| \equiv p_{\nu_\mu}$, this can be written as:

$$\begin{pmatrix} E_{B^+} \\ \vec{0} \end{pmatrix} = \begin{pmatrix} E_{\rho^0} + E_{\mu^+} + p_{\nu_\mu} \\ \vec{0} \end{pmatrix} \ . \tag{A.2}$$

After squaring both sides of the equation, we get the invariant mass of the B^+ meson:

$$m_{B^+}^2 = (E_{\rho^0} + E_{\mu^+})^2 + p_{\nu_\mu}^2 + 2(E_{\rho^0} + E_{\mu^+})p_{\nu_\mu} \ , \tag{A.3}$$

and since $\vec{p}_{\rho^0} + \vec{p}_{\mu^+} = -\vec{p}_{\nu_\mu}$, the invariant mass of the $\rho^0\mu^+$ system can be expressed as:

$$m_{\rho^0\mu^+}^2 = (E_{\rho^0} + E_{\mu^+})^2 - p_{\nu_\mu}^2 \ , \tag{A.4}$$

where $m_{\rho^0\mu^+} \equiv m_{\mathrm{vis}}$ corresponds to the mass of the visible final state.

Combining Eq. A.3 and A.4 as well as rewriting $\vec{p}_{\nu_\mu}$ in terms of its parallel and perpendicular components, *i.e.* $p_{\nu_\mu}^2 = p_{\nu_\mu\|}^2 + p_{\nu_\mu\perp}^2 \equiv p_\|^2 + p_\perp^2$, we arrive at the expression:

$$m_{B^+}^2 = m_{\mathrm{vis}}^2 + 2(p_\|^2 + p_\perp^2) + 2(E_{\rho^0} + E_{\mu^+})\sqrt{p_\|^2 + p_\perp^2} \tag{A.5}$$

$$= m_{\mathrm{vis}}^2 + 2(p_\|^2 + p_\perp^2) + 2\sqrt{m_{\mathrm{vis}}^2 + p_\|^2 + p_\perp^2}\sqrt{p_\|^2 + p_\perp^2} \ . \tag{A.6}$$

As $p_\|$ cannot be measured, it is neglected in Eq. A.6, and we arrive at the expression for the corrected B^+ mass:

$$m_{\mathrm{corr}}(B^+)^2 = m_{\mathrm{vis}}^2 + 2p_\perp^2 + 2p_\perp\sqrt{m_{\mathrm{vis}}^2 + p_\perp^2} \tag{A.7}$$

$$m_{\mathrm{corr}}(B^+) = \sqrt{m_{\mathrm{vis}}^2 + p_\perp^2} + p_\perp \ . \tag{A.8}$$

Since the boost is along the B^+ flight direction, $p_\perp$ is Lorentz invariant, *i.e.* its value is the same in the lab frame as in the center of mass frame. Thus, in Eq. A.8 the information on the transverse momentum of the neutrino is used to apply a kinematic correction to the invariant mass of the visible final state particles resulting in a better discriminating variable.

B Charge isolation MVA with VELO tracks

The potential of performing charge isolation with VELO tracks, where no momentum information is available, is assessed.

The MVA using VELO tracks is a NN identical to the charge isolation MVA based on long and upstream tracks described in Sect. 4.5.3. The same simulation of the decay $B^0 \rightarrow D^{*-}(\rightarrow \bar{D}^0(\rightarrow K^+ 2\pi^- \pi^+)\pi^-)\mu^+\nu_\mu$ (see Table 4.1) is used to train the NN.

The signal and background input samples consist of $\sim$ 2k and $\sim$ 134k VELO tracks obtained from the `StdAllNoPIDsVeloPions` [148] particle container. The variables found to be the most powerful in separating non-isolated and isolated tracks are $\min(\chi^2_{\rm IP})$ and $\Delta\chi^2_{\rm vtx}$. The distributions of these input variables for signal and background are illustrated in Fig. B.1. The performance of the charge isolation MVA trained on VELO tracks is illustrated by the ROC curve with an AUC $= 97\%$ and the MVA output distribution in Fig. B.2.

However, the number of non-isolated VELO tracks present in the training sample is 8–10 times smaller than the number of non-isolated upstream and long tracks. One would therefore expect the background with non-isolated VELO tracks to be small in data compared to the background with non-isolated upstream and long tracks. This is confirmed
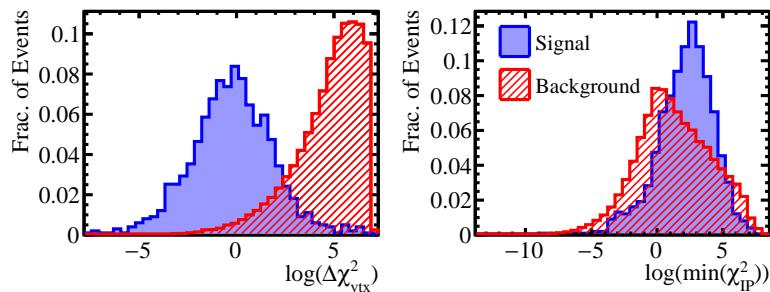

Figure B.1 – The input variables for signal (blue) and background (red) used by the charge isolation MVA based on VELO tracks.

Figure B.2 – Performance of the charge isolation MVA trained on VELO tracks: ROC curve (left) with an AUC = 97% and MVA output (right) for signal (blue) and background (red) for the training (markers) and testing (filled histograms) samples.

when using the response variable of the MVA based on VELO tracks in the main MVA described in Sec. 4.5.4, as it does not improve the MVA performance. Therefore, the charge isolation MVA based on VELO tracks is not used in the final MVA selection.

C Neutral isolation MVA

In addition to charge isolation described in Sec. 4.5.3, the potential of neutral isolation is assessed. Two neutral isolation MVAs are developed to evaluate the isolatedness of signal candidates with respect to photons (γ) and neutral pions (π^0), respectively. Both MVAs are NNs with the same architecture and training as the charge isolation MVA described in Sec. 4.5.3.

A MC cocktail of $B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ decays with $\bar{D}^0 \to \pi^+\pi^- X$ (see Table 4.1) is used to train a NN. This sample contains typical background processes involving neutral particles, *e.g.*

$$B^+ \to \bar{D}^0\mu^+\nu_\mu \text{ where } \bar{D}^0 \to K_S^0\pi^+\pi^- \text{ with } K_S^0 \to \pi^0\pi^0 \;,$$

$$B^+ \to \bar{D}^{*0}\mu^+\nu_\mu \text{ where } \bar{D}^{*0} \to \bar{D}^0\pi^0 \text{ or } \bar{D}^{*0} \to \bar{D}^0\gamma \;.$$

As for the charge isolation, the signal and background training samples consist of isolated and non-isolated neutral particles corresponding to either γ or π^0.

Since neutral particles do not create tracks, the variables $\min(\chi_{IP}^2)$ and $\Delta\chi_{vtx}^2$ used to train the charge isolation MVA are not available. The variables p_T, p, $p_\perp$ and ΔR are therefore used for the MVA training. For both MVAs the best performance is obtained using p_T, p and $p_\perp$. The distributions of these variables for signal and background are illustrated in Fig. C.1 for γ and π^0 samples, respectively.

Neutral isolation MVA based on γ

The signal and background input samples consist of ~ 26k and ~ 745k γ obtained from the `StdLoosePhotons` [148] particle container where the minimum requirement on $p_T(\gamma)$ is removed. The performance of the neutral isolation MVA based on γ is illustrated by the ROC curve with AUC $= 86\%$ and the MVA output distribution in Fig. C.2. The MVA output distributions show a limited separation power as both signal and background peak at zero.

Appendix C. Neutral isolation MVA
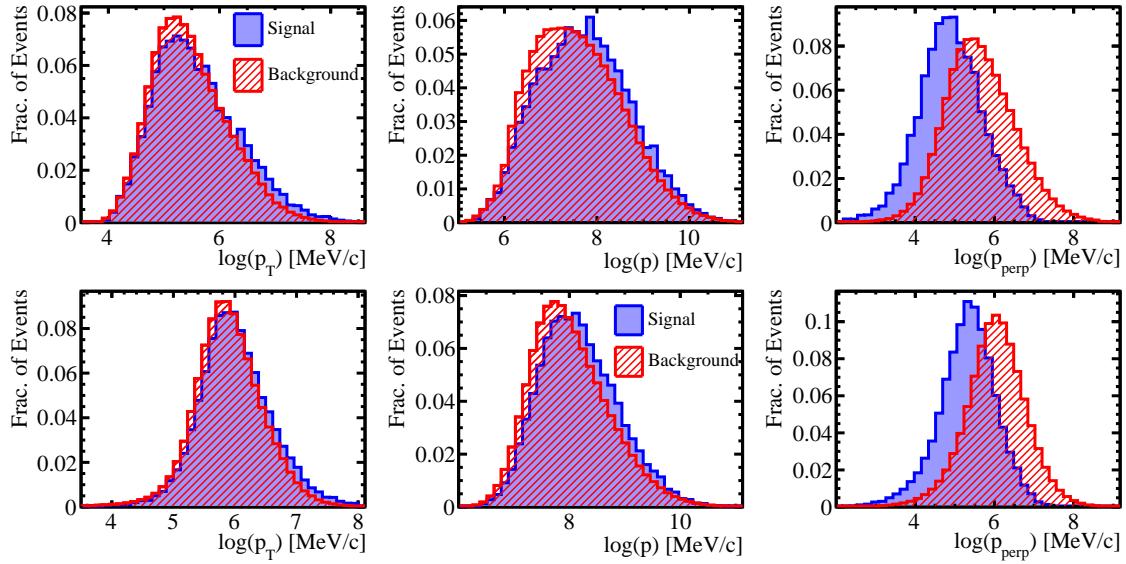
Figure C.1 – The input variables for signal (blue) and background (red) used by the neutral isolation MVAs trained with γ (top) and π^0 (bottom) samples, respectively.

Neutral isolation MVA based on π^0

The signal and background input samples consist of $\sim 61\text{k}$ and $\sim 2\text{M}$ π^0 mesons obtained from the `StdLooseResolvedPi0` and `StdLooseMergedPi0` [148] particle containers where the minimum requirement on $p_\text{T}(\pi^0)$ is removed. The performance of the neutral isolation MVA based on π^0 is illustrated by the ROC curve with AUC = 85% and the MVA output distribution in Fig. C.3. The MVA output distributions show a limited separation power as both signal and background peak at zero.

DNN based on neutral isolation variables

The response variables of the two MVAs, *i.e.* $\text{mva}_\text{out}(\gamma)$ and $\text{mva}_\text{out}(\pi^0)$, are used together with *neutral cone variables* to train a DNN to reject backgrounds with additional neutral particles in their final states.

The computation of neutral cone variables uses the cone around the B^+ flight direction defined as $\Delta R = \sqrt{((\Delta\eta)^2 + (\Delta\phi)^2)}$ with $\Delta\eta$ and $\Delta\phi$ being the difference in pseudorapidity and azimuthal angle between the neutral particle and the B^+ flight direction. All γ candidates from `StdLooseAllPhotons` (without a minimum requirement on the $p_\text{T}(\gamma)$) inside the cone are used to compute different variables. The neutral cone variables resulting in the best performance of the DNN are:

- $\text{asym}_{\Delta R = 0.6}(p)$: p asymmetry between signal and the γ candidates inside $\Delta R = 0.6$.

- $\text{asym}_{\Delta R = 0.6}(p_T)$: p_T asymmetry between the signal and the γ candidates inside $\Delta R = 0.6$.
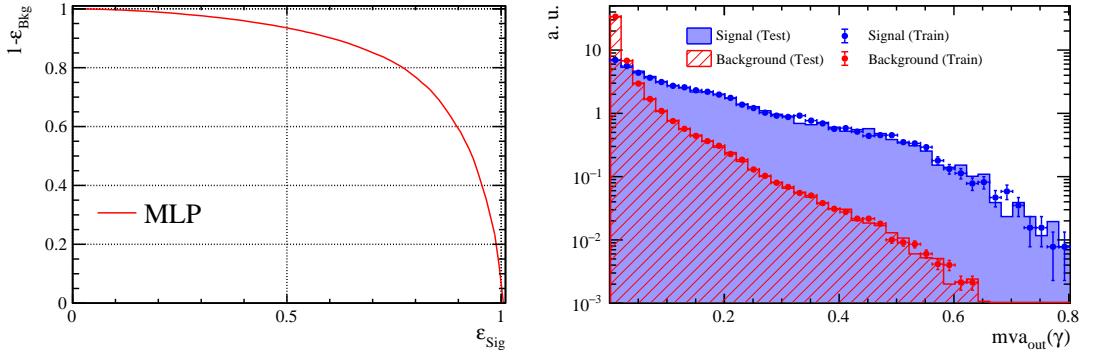
Figure C.2 – Performance of the neutral isolation MVA trained to distinguish between isolated and nonisolated γ in $B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ decays with $\bar{D}^0 \to \pi^+\pi^- X$: ROC curve (left) with an AUC = 86% and MVA output (right) for signal (blue) and background (red) for the training (markers) and testing (filled histograms) samples.
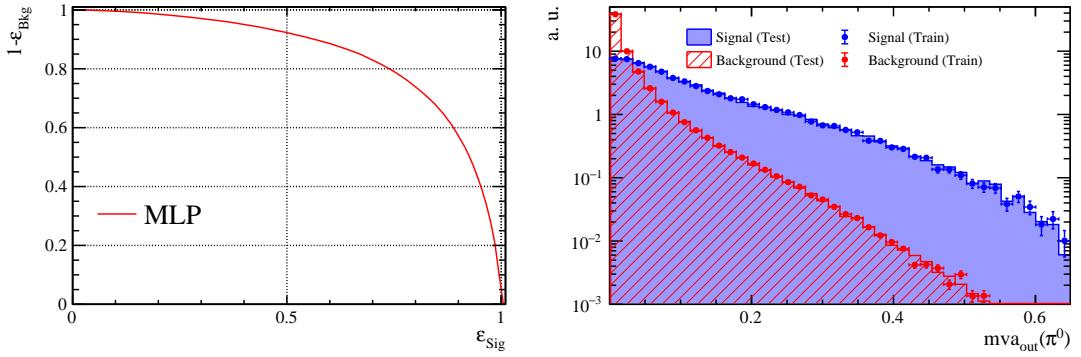

Figure C.3 – Performance of the neutral isolation MVA trained to distinguish between isolated and nonisolated π^0 in $B^+ \to D^{(*,**)}\mu^+\nu_\mu(Y)$ decays with $\bar{D}^0 \to \pi^+\pi^- X$: ROC curve (left) with an AUC = 85% and MVA output (right) for signal (blue) and background (red) for the training (markers) and testing (filled histograms) samples.

- $\max_{\Delta R=0.6}(p_T)$: maximum $p_\text{T}(\gamma)$ inside $\Delta R = 0.6$.

- $\text{avg}_{\Delta R=0.6}(\text{CL})$: average confidence level of the γ candidates inside $\Delta R = 0.6$.

- $\max_{\Delta R=0.6}(\text{CL})$: maximum confidence level of the γ candidates inside $\Delta R = 0.6$.

The asymmetry variables are given by:

$$\text{asym}_{\Delta R=0.6}(p_{(T)}) = \frac{|\sum_{i=1}^{N_{\text{sig},\Delta R}} p_{(T)}(s_i)| - |\sum_{j=1}^{N_{\gamma,\Delta R}} p_{(T)}(\gamma_j)|}{|\sum_{i=1}^{N_{\text{sig},\Delta R}} p_{(T)}(s_i)| + |\sum_{j=1}^{N_{\gamma,\Delta R}} p_{(T)}(\gamma_j)|} \tag{C.1}$$

where $p_{(T)}(s_i)$ is the (transverse) momentum of the signal particle s_i inside the cone ΔR, while the total number of signal particles inside the cone is $N_{\text{sig},\Delta\text{R}}$. The same notation is used for the photons γ_j inside the cone.
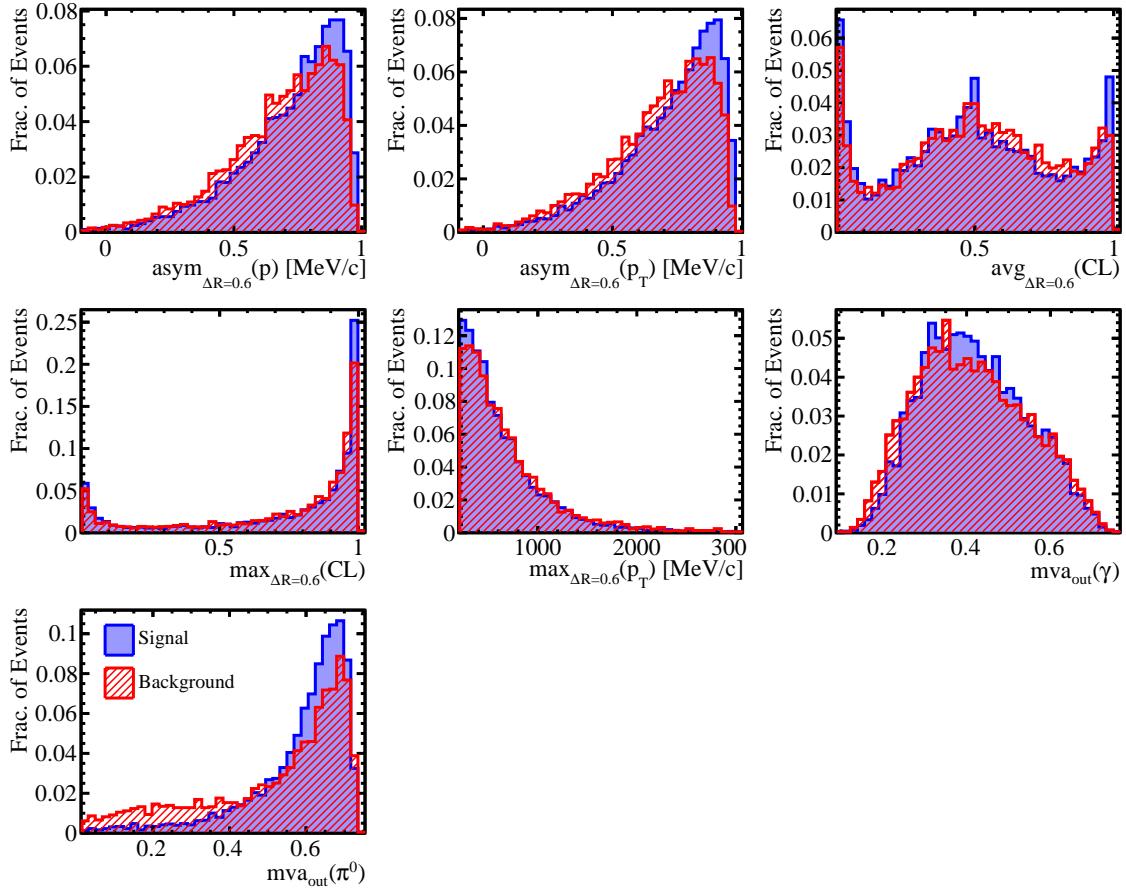
Figure C.4 – The neutral isolation DNN input variables for signal (blue) and background (red).

The confidence level of a γ candidate is calculated from its PID information, described in Sec. 3.4.4, to indicate the confidence of the γ hypothesis. More specifically, it is calculated as the ratio of the likelihood of the γ hypothesis with respect to the sum of the likelihoods of all particle hypotheses.

As the DNN based on charge isolation, kinematic and geometric variables, described in Sec. 4.5.4, has proven powerful in rejecting backgrounds in data, it is applied to the input samples before training the DNN based on neutral isolation variables. An non-optimised requirement of $dnn_{out} > 0.7$ is imposed.

As background the WS-MS data sample defined in Sec. 4.3 is used. The distributions of the variables used to train the DNN for signal and background are illustrated in Fig. C.4.

The signal and background input samples consist of $\sim$ 6k and $\sim$ 9k candidates. The DNN has the same architecture and training as the DNN described in Sec. 4.5.4. The only difference is the number of epochs, which is 60. Although overtraining is avoided, the training is found to be unstable. The performance of the DNN based on neutral isolation
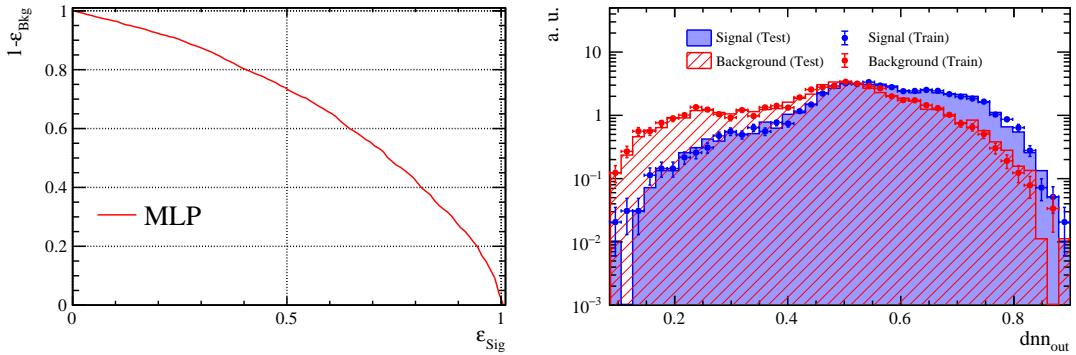
Figure C.5 – Performance of the DNN based on neutral isolation variables: ROC curve (left) with AUC = 67% and MVA output (right) for signal (blue) and background (red) for the training (markers) and testing (filled histograms) samples.

is illustrated by the ROC curve in Fig. C.5, where the AUC reaches 67%. The distribution of the output value for signal and background from the training and testing sample are illustrated in log scale in Fig. C.5. This shows a poor separation power between signal and background.

Different neutral cone variables were tested, in particular, those computed with $\Delta R = 0.4$ and $\Delta R = 0.5$. In addition to this, different architectures and model parameters were tested as well as different MVA methods, *e.g.* BDTs and MLPs, however, in all cases a poor performance was obtained. Some improvement might be achieved by computing the cone variables with π^0 (not only γ), however, it is doubtful that it would result in a significant performance of the neutral isolation MVA, and it is therefore not used in this analysis.

D Additional plots for the control studies

Consistency in specific variables between the simulated signal and the signal in data has to be ensured. These are the variables used to train the main MVA described in Sec. 4.5.4, the variables used in the preselection described in Sec. 4.4, except PID variables that are treated in Chapter 8, and the variable `nTracks`.

The comparison of variables between the signal and control channels in simulation can be found in Sec. D.1, while the comparison of variables between the simulated control channel and the control channel reconstructed in data can be found in Sec. D.2. All plots shown in this appendix are made with 2018 data and MC samples.

D.1 Comparing signal MC and control MC variables

Initially, only the stripping and trigger selections are applied to the two samples, and the resulting distributions of preselection and DNN input variables are shown in Figs. D.1 and D.2, respectively. The control channel has the same coverage as the signal channel, except for the variable $p_\perp(\rho^0)$ where the distribution for the control channel is narrower than the one for the signal. This is due to the fact that for the control channel $p_\perp(\rho^0)$ peaks when $m(\mu^+\mu^-) \sim m(J/\psi)$, however, for the signal, the mass of $\mu^+\nu_\mu$ is not fixed, and therefore $p_\perp(\rho^0)$ does not peak. In addition to this, discrepancies between the signal and control channel are observed in several variables, in particular, in $m_{\text{corr}}(B^+)$, $p_{B\text{rest}}(\mu^+)$, $p_\perp(\rho^0)$, $\chi^2_{\text{IP}}(B^+)$, $m(B^+ + 1^{\text{st}} \text{ track})$, $\text{mva}_{\text{out}}(1^{\text{st}} \text{ track} + \text{vtx}_{B+})$ and $\text{mva}_{\text{out}}(1^{\text{st}} \text{ track} + \text{vtx}_{\rho^0})$.

To understand the origin of the observed discrepancies, the difference in the leptonic system of the two modes is compensated with the three requirements described in Sec. 4.6.1. After applying the three requirements, a good overall agreement is found between signal and control MC variables, as illustrated in Figs. D.3 and D.4. The small disagreement between in $\chi^2_{\text{IP}}(B^+)$ is explained in Sec. 4.6.1. The discrepancy in $p_\perp(\rho^0)$ is significantly reduced by restricting the q^2 region of signal MC, and can be completely eliminated by restricting it to be even closer around the $q^2 = (m(J/\psi))^2$ peak of the control channel.
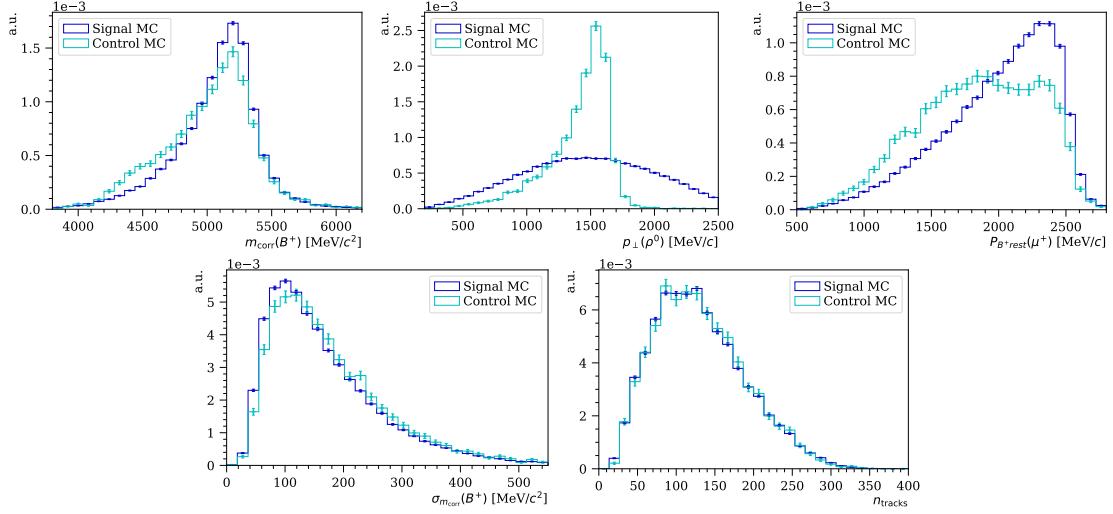
Figure D.1 – Preselection variables for signal MC (blue) and control MC (cyan) after applying trigger and stripping selection.
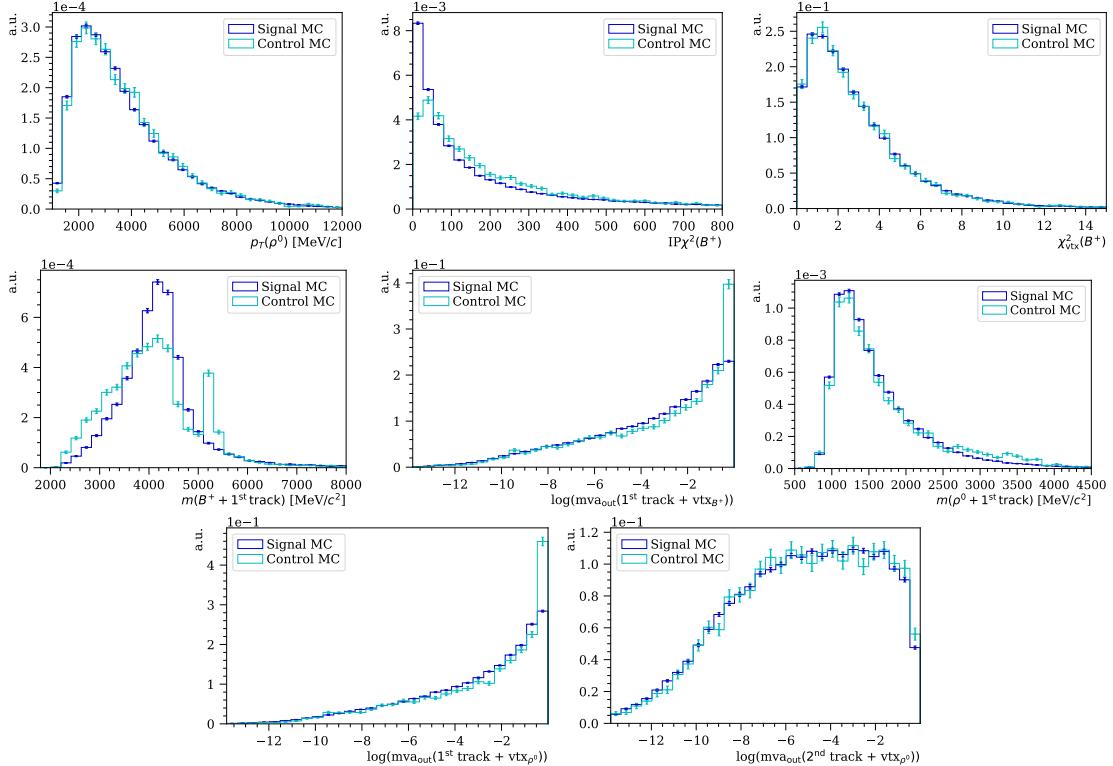
Figure D.2 – DNN input variables for signal MC (blue) and control MC (cyan) after applying trigger and stripping selections. The peak at $m(B^0 + 1^{\text{st}} \text{ track}) = m(B^+)$ in control MC corresponds to fully reconstructed the $B^0 \to J/\psi(\to \mu^+\mu^-)\rho^0(\to \pi^+\pi^-)$ decays. For these candidates the true missing muon was not identified by the charge isolation MVA and therefore not excluded from the computation of the isolation variables.
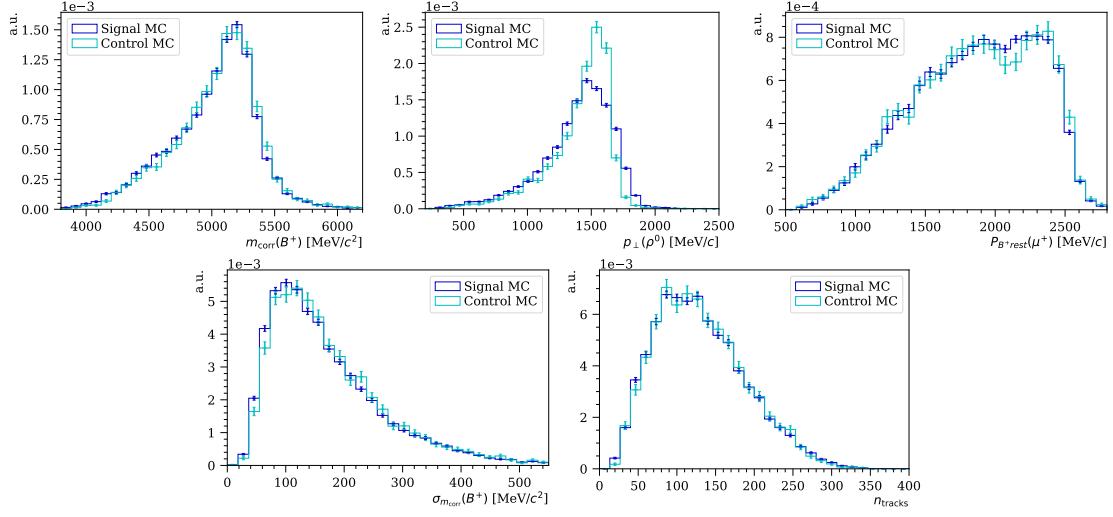
Figure D.3 – Preselection variables for signal MC (blue) and control MC (cyan) after applying requirements to compensate for the difference in the leptonic system of the two modes.
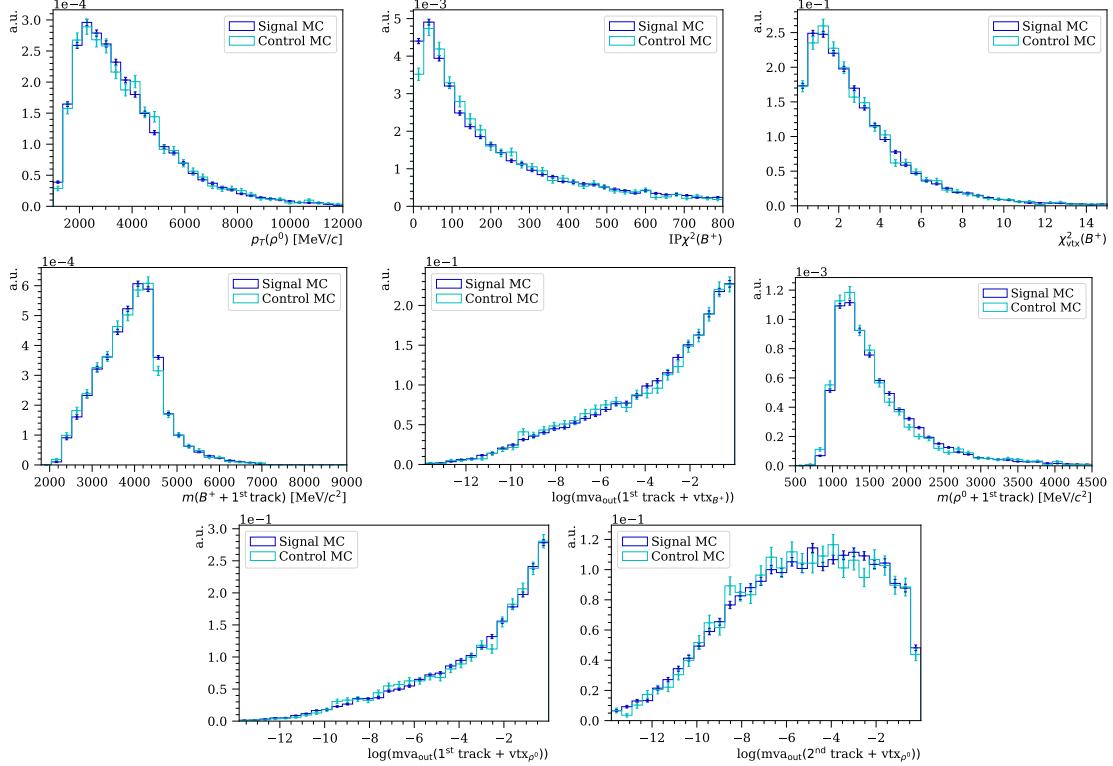

Figure D.4 – DNN input variables for signal MC (blue) and control MC (cyan) after applying requirements to compensate for the difference in the leptonic system of the two modes.

In conclusion, discrepancies between the signal and control channel are well understood, and the control channel covers the range of all relevant variables in signal, except for $p_\perp(\rho^0)$, which is expected from the difference in the leptonic system of the two channels.

D.2   Comparing control MC and control data variables

The distributions of variables are extracted in data for the control channel by applying the *sWeights*, and a comparison between variable distributions in data and MC are illustrated in Figs. D.5 and D.6. Discrepancies are observed in several variables, in particular, for $\mathrm{mva_{out}}(1^{\mathrm{st}}\ \mathrm{track} + \mathrm{vtx}_{B+})$, $\mathrm{mva_{out}}(1^{\mathrm{st}}\ \mathrm{track} + \mathrm{vtx}_{\rho^0})$ and `nTracks`.

To correct for the observed discrepancies, the two variables `nTracks` and $\mathrm{mva_{out}}(1^{\mathrm{st}}\ \mathrm{track} + \mathrm{vtx}_{B+})$ are used to compute weights, as described in Sec. 4.6.3. The two variables are uncorrelated, thus the product of the weights $w_{\mathrm{tot}} = w_{\mathtt{nTracks}} \times w_{\mathrm{mva}}$ is used to reweight MC. After reweighting, a better agreement is obtained between MC and data for the isolation variables $\mathrm{mva_{out}}(1^{\mathrm{st}}\ \mathrm{track} + \mathrm{vtx}_{\rho^0})$, $\mathrm{mva_{out}}(2^{\mathrm{nd}}\ \mathrm{track} + \mathrm{vtx}_{\rho^0})$, $\mathrm{mva_{out}}(1^{\mathrm{st}}$ and `nTracks`, as illustrated in Figs. D.7 and D.8. A small unexplained discrepancy remains for high $p_{B\mathrm{rest}}(\mu^+)$. However, for the signal decay this variable is highly correlated with the form factors, and it can therefore not be assumed that the observed discrepancy in the control channel between MC and data is the same for the signal channel. All other variables used in the preselection and DNN selection agree reasonably well between MC and data.
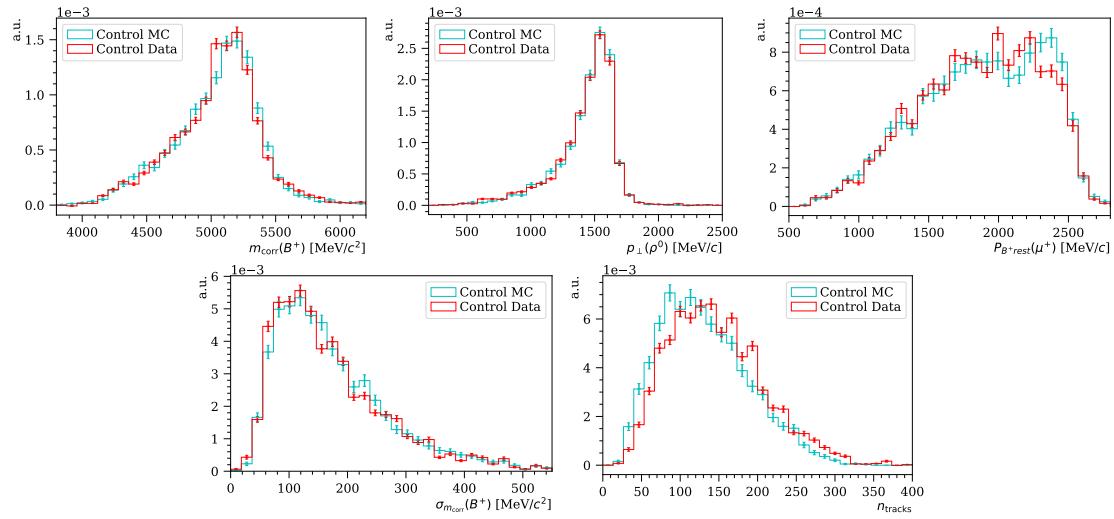
Figure D.5 – Preselection variables for control MC (cyan) and control data (red), where *sWeights* are applied to data.

Figure D.6 – DNN input variables for control MC (cyan) and control data (red), where *sWeights* are applied to data.
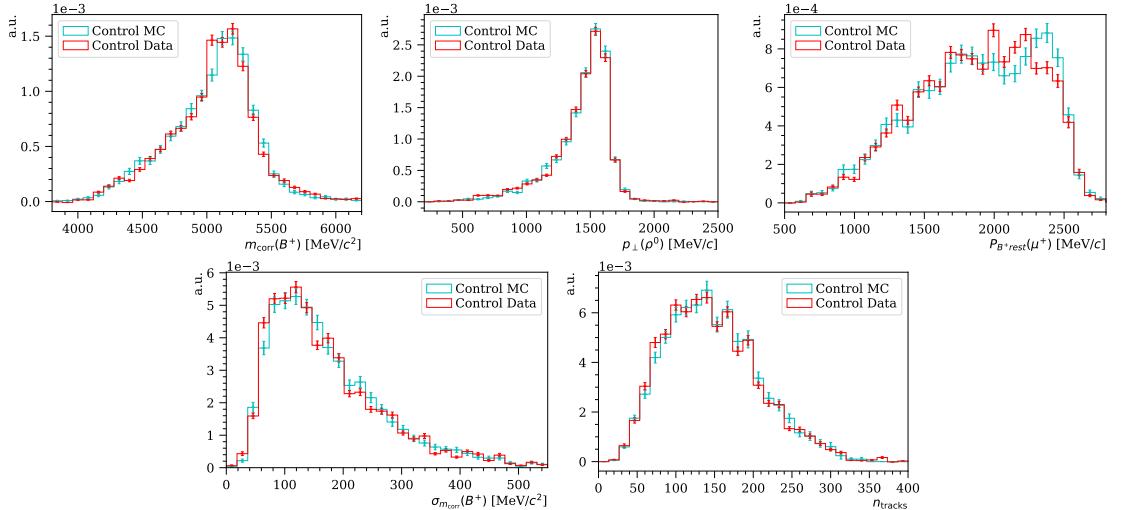

Figure D.7 – Preselection variables for control MC (cyan) and control data (red), where *sWeights* are applied to data, and control MC is reweighted with $w_{\texttt{nTracks}} \times w_{\text{mva}}$.
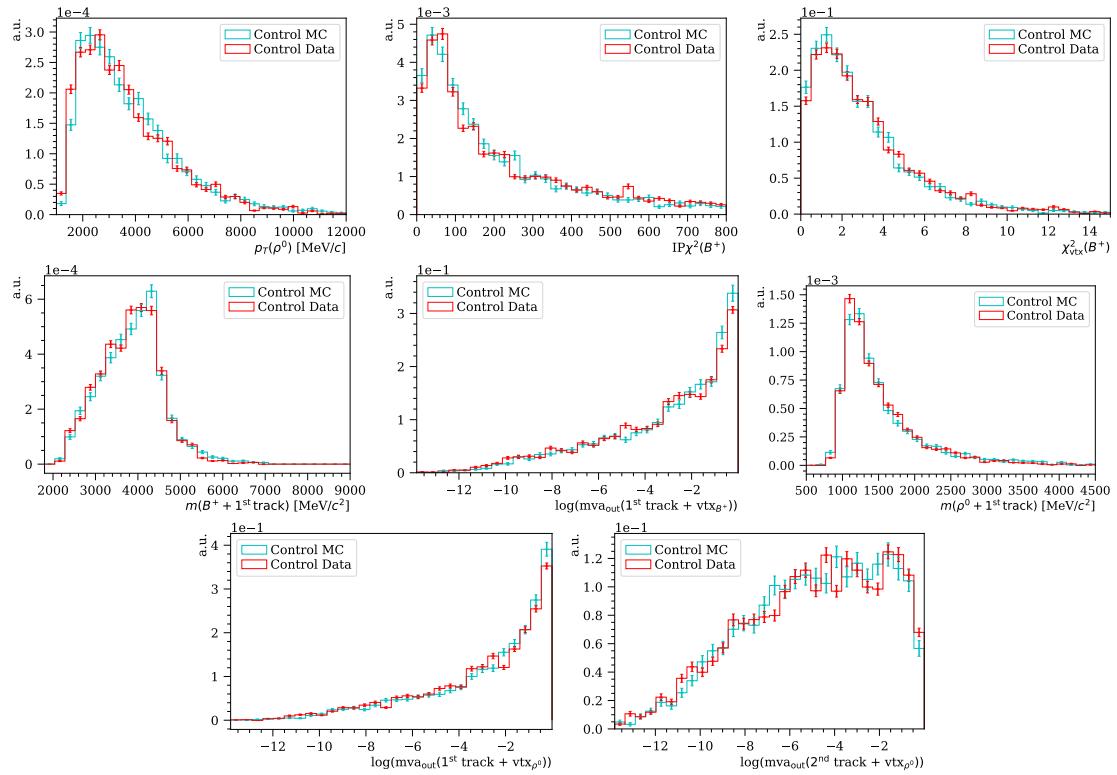
Figure D.8 – DNN input variables for control MC (cyan) and control data (red), where *sWeights* are applied to data and control MC is reweighted with $w_{\texttt{nTracks}} \times w_{\text{mva}}$.

E Selection efficiencies for signal fit processes

This appendix contains the selection efficiencies of all the physics processes listed in Table 7.2. These efficiencies are used to compute Gaussian constraints in the signal channel fit. For the fit to the full q^2 region these constraints are given in Table 7.7.

All efficiencies are estimated from simulation as the ratio of the number of decays passing all selection requirements including that of the q^2 bin boundaries (see the non-uniform binning scheme in Table 5.2) and the total number of decays produced in 4π. The resulting efficiencies are listed in Table E.1. Note that the selection efficiency of the signal decay is only given for comparison and is not used directly to obtain the differential branching fraction. The proper estimation of the signal efficiency that accounts for possible data/MC differences is performed in Chapter 8.

Table E.1 – Summary of the estimated selection efficiencies of the fitting templates with the IDs T_i defined in Table 7.2. The total efficiency in bin i is denoted ϵ_i, while that in the full q^2 region is denoted ϵ_{tot}.

ID	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6	ϵ_7	ϵ_8	ϵ_9	ϵ_{10}	ϵ_{tot}
T_1	3.99×10^{-5}	4.00×10^{-5}	4.00×10^{-5}	3.99×10^{-5}	4.00×10^{-5}	3.99×10^{-5}	3.94×10^{-5}	3.98×10^{-5}	4.12×10^{-5}	3.94×10^{-5}	4.00×10^{-4}
T_2	1.70×10^{-7}	1.85×10^{-7}	2.80×10^{-7}	3.82×10^{-7}	5.05×10^{-7}	6.10×10^{-7}	6.32×10^{-7}	7.00×10^{-7}	8.34×10^{-7}	9.03×10^{-7}	5.20×10^{-6}
T_3	9.24×10^{-9}	1.67×10^{-8}	2.79×10^{-8}	5.68×10^{-8}	8.24×10^{-8}	1.18×10^{-7}	1.69×10^{-7}	2.40×10^{-7}	3.81×10^{-7}	7.23×10^{-7}	1.82×10^{-6}
T_4	7.00×10^{-7}	6.27×10^{-7}	7.44×10^{-7}	8.39×10^{-7}	9.05×10^{-7}	9.38×10^{-7}	8.78×10^{-7}	8.64×10^{-7}	8.75×10^{-7}	8.11×10^{-7}	8.18×10^{-6}
T_5	3.27×10^{-9}	6.28×10^{-9}	8.52×10^{-9}	1.42×10^{-8}	2.13×10^{-8}	3.54×10^{-8}	3.62×10^{-8}	4.36×10^{-8}	6.28×10^{-8}	8.64×10^{-8}	3.18×10^{-7}
T_6	3.55×10^{-9}	6.80×10^{-9}	9.23×10^{-9}	1.54×10^{-8}	2.31×10^{-8}	3.84×10^{-8}	3.92×10^{-8}	4.72×10^{-8}	6.80×10^{-8}	9.36×10^{-8}	3.44×10^{-7}
T_7	2.24×10^{-7}	1.99×10^{-7}	2.06×10^{-7}	2.45×10^{-7}	2.35×10^{-7}	2.39×10^{-7}	2.27×10^{-7}	1.91×10^{-7}	1.91×10^{-7}	1.73×10^{-7}	2.13×10^{-6}
T_8	5.78×10^{-8}	6.09×10^{-8}	7.27×10^{-8}	9.34×10^{-8}	1.04×10^{-7}	1.21×10^{-7}	1.39×10^{-7}	1.34×10^{-7}	1.45×10^{-7}	1.72×10^{-7}	1.10×10^{-6}
T_9	1.50×10^{-9}	2.64×10^{-9}	5.06×10^{-9}	9.63×10^{-9}	1.35×10^{-8}	1.70×10^{-8}	2.95×10^{-8}	4.47×10^{-8}	8.50×10^{-8}	1.76×10^{-7}	3.85×10^{-7}
T_{10}	2.13×10^{-7}	3.13×10^{-7}	6.26×10^{-7}	9.04×10^{-7}	1.28×10^{-6}	1.66×10^{-6}	1.95×10^{-6}	2.26×10^{-6}	2.90×10^{-6}	3.94×10^{-6}	1.60×10^{-5}
T_{11}	6.20×10^{-6}	6.31×10^{-6}	6.64×10^{-6}	6.31×10^{-6}	6.45×10^{-6}	7.17×10^{-6}	6.10×10^{-6}	6.89×10^{-6}	6.06×10^{-6}	5.99×10^{-6}	6.41×10^{-5}
T_{12}	4.14×10^{-6}	3.98×10^{-6}	3.70×10^{-6}	3.97×10^{-6}	3.73×10^{-6}	3.64×10^{-6}	3.05×10^{-6}	3.02×10^{-6}	2.89×10^{-6}	1.85×10^{-6}	3.40×10^{-5}
T_{13}	5.32×10^{-5}	3.63×10^{-5}	2.68×10^{-5}	2.23×10^{-5}	1.87×10^{-5}	1.64×10^{-5}	1.31×10^{-5}	1.10×10^{-5}	1.02×10^{-5}	7.25×10^{-6}	2.15×10^{-4}
T_{14}	9.43×10^{-5}	6.74×10^{-5}	5.44×10^{-5}	4.42×10^{-5}	3.93×10^{-5}	3.30×10^{-5}	2.79×10^{-5}	2.42×10^{-5}	2.08×10^{-5}	1.03×10^{-5}	4.16×10^{-4}
T_{15}	5.78×10^{-6}	4.68×10^{-6}	3.78×10^{-6}	3.48×10^{-6}	3.03×10^{-6}	2.70×10^{-6}	2.19×10^{-6}	1.91×10^{-6}	1.80×10^{-6}	1.22×10^{-6}	3.06×10^{-5}
T_{16}	1.79×10^{-6}	2.08×10^{-6}	2.69×10^{-6}	3.22×10^{-6}	3.63×10^{-6}	3.96×10^{-6}	4.18×10^{-6}	4.53×10^{-6}	5.31×10^{-6}	6.04×10^{-6}	3.74×10^{-5}
T_{17}	2.00×10^{-5}	1.79×10^{-5}	1.77×10^{-5}	1.77×10^{-5}	1.76×10^{-5}	1.71×10^{-5}	1.72×10^{-5}	1.78×10^{-5}	1.88×10^{-5}	1.87×10^{-5}	1.81×10^{-4}
T_{18}	5.79×10^{-7}	6.91×10^{-7}	9.39×10^{-7}	1.07×10^{-6}	1.18×10^{-6}	1.22×10^{-6}	1.29×10^{-6}	1.29×10^{-6}	1.35×10^{-6}	1.24×10^{-6}	1.08×10^{-5}
T_{19}	9.57×10^{-7}	1.91×10^{-7}	4.78×10^{-6}	8.61×10^{-6}	1.53×10^{-5}	6.12×10^{-5}	1.71×10^{-4}	2.01×10^{-5}	5.74×10^{-6}	9.57×10^{-7}	2.91×10^{-4}
T_{20}	2.36×10^{-9}	3.53×10^{-9}	9.42×10^{-9}	1.94×10^{-8}	2.53×10^{-8}	3.95×10^{-8}	5.42×10^{-8}	8.84×10^{-8}	2.07×10^{-8}	4.55×10^{-8}	9.05×10^{-7}
T_{21}	6.25×10^{-6}	7.14×10^{-6}	2.50×10^{-8}	4.46×10^{-8}	9.11×10^{-8}	1.37×10^{-7}	2.01×10^{-7}	3.70×10^{-7}	1.02×10^{-6}	2.57×10^{-7}	4.47×10^{-6}
T_{22}	8.01×10^{-8}	1.95×10^{-7}	3.17×10^{-7}	5.16×10^{-7}	8.54×10^{-7}	4.33×10^{-6}	4.43×10^{-6}	4.49×10^{-6}	3.62×10^{-6}	2.29×10^{-6}	2.11×10^{-5}
T_{23}	4.30×10^{-6}	1.34×10^{-5}	3.93×10^{-5}	4.12×10^{-5}	2.55×10^{-5}	1.78×10^{-5}	1.34×10^{-5}	1.08×10^{-5}	7.00×10^{-6}	4.14×10^{-5}	1.77×10^{-4}

F Estimating misID in signal fit

This appendix contains details regarding the estimation of the misIDμ and misIDπ backgrounds for the signal channel in the full q^2 region. This include the efficiency matrices (see Eq. 6.9) computed with PIDCALIB, the binning schemes in p, η and `nSPDHits` used to parameterise the PID efficiency, the number of measured particle species and the unfolded number of true particle species in the reference sample.

Estimating misIDμ

Using the particle regions defined in Table 6.7 together with the binning scheme in Table F.1, the resulting efficiency matrix computed with PIDCALIB is:

$$E^{\mathrm{misID}\mu} = \begin{pmatrix} 0.9907 & 0.0027 & 0.0060 & 0.0006 \\ 0.0053 & 0.9095 & 0.0716 & 0.0136 \\ 0.0010 & 0.1017 & 0.8646 & 0.0327 \\ 0.0005 & 0.0247 & 0.0376 & 0.9372 \end{pmatrix}. \tag{F.1}$$

Here only the efficiency matrix based on the MagDown calibration samples is shown, however, the efficiency matrix based on the MagUp calibration samples is very similar.

Table F.1 – Binning of the variables p, η and `nSPDHits` used to compute PID efficiencies.

Variable	Standard binning scheme
p	[6000, 29500, 53000, 76500, 100000, 500000] MeV/c
η	[1.6, 2.4, 3.3, 4.1, 4.9]
`nSPDHits`	[25, 131, 238, 344, 450]
Variable	Small binning scheme
p	[6000, 48000, 90000, 500000] MeV/c
η	[1.6 , 3.3, 4.9]
`nSPDHits`	[25 , 238, 450]

Appendix F. Estimating misID in signal fit

The number of measured μ, π, K and p candidates in the reference sample is:

$$\vec{N}_{\text{meas}}^{\text{misID}\mu} = \begin{pmatrix} 3626 \\ 5799 \\ 3520 \\ 596 \end{pmatrix}. \tag{F.2}$$

As for the normalisation channel, the Baysian unfolding algorithm in the RooUnfold package is used to find the true particle composition of the reference sample:

$$\vec{N}_{\text{true}}^{\text{misID}\mu} = \begin{pmatrix} 3614 \\ 6211 \\ 3257 \\ 458 \end{pmatrix}. \tag{F.3}$$

Estimating misIDπ

For the signal candidates $B^+ \to \rho^0(\to \pi^+\pi^-)\mu^+\nu_\mu$, a stricter PID selection is applied to π^- than to π^+, as described in Sec. 4.4.4. The misIDπ^+ and misIDπ^- backgrounds are therefore estimated individually. Using the particle regions defined in Table 7.4 and Table 7.5 together with the binning scheme in Table F.2, the resulting efficiency matrices computed with PIDCalib are:

$$E^{\text{misID}\pi^+} = \begin{pmatrix} 0.8585 & 0.0964 & 0.0108 & 0.0343 \\ 0.0039 & 0.9859 & 0.0088 & 0.0015 \\ 0.0439 & 0.0114 & 0.8579 & 0.0869 \\ 0.0169 & 0.0025 & 0.0665 & 0.9141 \end{pmatrix}, \tag{F.4}$$

$$E^{\text{misID}\pi^-} = \begin{pmatrix} 0.9366 & 0.0184 & 0.0117 & 0.0333 \\ 0.0051 & 0.9811 & 0.0113 & 0.0025 \\ 0.0522 & 0.0132 & 0.8507 & 0.0839 \\ 0.0178 & 0.0030 & 0.0666 & 0.9126 \end{pmatrix}. \tag{F.5}$$

Here only the efficiency matrices based on the MagDown calibration samples are shown, however, the efficiency matrices based on the MagUp calibration samples are very similar.

The numbers of measured π, μ, K and p candidates in the two reference samples are:

$$\vec{N}_{\text{meas}}^{\text{misID}\pi^+} = \begin{pmatrix} 12725 \\ 254 \\ 7847 \\ 679 \end{pmatrix}, \tag{F.6}$$

Table F.2 – Binning of the variables p, η and `nSPDHits` used to compute PID efficiencies.

Variable	Standard binning scheme
p	[3000, 27250, 51500, 75750, 100000, 500000] MeV/c
η	[1.7, 2.5, 3.4, 4.2, 5.0]
`nSPDHits`	[25, 131, 238, 344, 450]
Variable	Small binning scheme
p	[3000, 41500, 80000, 500000] MeV/c
η	[1.7 , 3.4, 5.0]
`nSPDHits`	[25 , 238, 450]

$$\vec{N}_{\text{meas}}^{\text{misID}\pi^-} = \begin{pmatrix} 11948 \\ 3068 \\ 991 \\ 492 \end{pmatrix}. \tag{F.7}$$

Using again the Bayesian unfolding algorithm of the RooUnfold package the true particle compositions of the two reference samples are found to be:

$$\vec{N}_{\text{true}}^{\text{misID}\pi^+} = \begin{pmatrix} 13536 \\ 247 \\ 7297 \\ 425 \end{pmatrix}, \tag{F.8}$$

$$\vec{N}_{\text{true}}^{\text{misID}\pi^-} = \begin{pmatrix} 12535 \\ 3087 \\ 486 \\ 391 \end{pmatrix}. \tag{F.9}$$

G Estimating the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ contribution

The nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ decay has not been measured, and its contribution to the signal fit is therefore difficult to constrain. However, the combined resonant and nonresonant branching fraction of $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ decays has been measured by the Belle experiment in Ref. [185]. A rough estimate of the nonresonant contribution can therefore be obtained by fitting the $m(\pi^+\pi^-)$ spectrum reported in this publication. Under the assumption that the $m(\pi^+\pi^-)$ shapes of the various components are not too different between data recorded by the Belle and LHCb detector, the excl. $|V_{ub}|$ templates used in the signal fit can also be used to fit the Belle spectrum. Thus, in HISTFACTORY a fit model is build with the templates:

- $B^+ \to \rho^0\mu^+\nu_\mu$ (SIGNAL)

- $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ (NON-RES)

- $B^+ \to \rho(1450)\mu^+\nu_\mu$ (RHO(1450))

- $B^+ \to f_0(980)\mu^+\nu_\mu$ (F0(980))

- $B^+ \to f_2(1270)\mu^+\nu_\mu$ (F2(1270))

- $B^+ \to f_0(500)\mu^+\nu_\mu$ (F0(500))

This results in the fit illustrated in Fig. G.1 with the extracted yields listed in Table G.1.

## Appendix G. Estimating the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ contribution
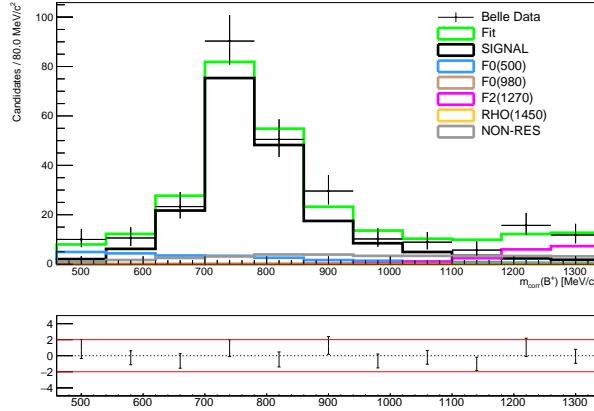
Figure G.1 – Fit of $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ spectrum.

Table G.1 – Yields obtained from fitting $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ spectrum.

Template	Fit results
N_{SIGNAL}	228.7 ± 24.6
$N_{\text{NON-RES}}$	8.7 ± 29.0
$\text{corr}(N_{\text{SIGNAL}}, N_{\text{NON-RES}})$	-0.01
$N_{\text{F0(500)}}$	22.0 ± 9.0
$N_{\text{F0(980)}}$	0.0 ± 8.8
$N_{\text{F2(1270)}}$	25.9 ± 10.7
$N_{\text{RHO(1450)}}$	0.0 ± 21.7

Based on this result, the relative contribution of the nonresonant component with respect to the signal is found to be approximately $4\% \pm 12\%$. Using the fit results, the branching fraction of the signal $\mathcal{B}_{\text{SIGNAL}} = (1.58 \pm 0.11) \times 10^{-4}$ [15] and their selection efficiencies ϵ estimated from simulations, the nonresonant branching fraction is estimated:

$$\mathcal{B}_{\text{NON-RES}} = \frac{N_{\text{NON-RES}}}{N_{\text{SIGNAL}}} \times \frac{\mathcal{B}_{\text{SIGNAL}} \times \epsilon_{\text{SIGNAL}}}{\epsilon_{\text{NON-RES}}} = (0.9 \pm 3.1) \times 10^{-5} \qquad \text{(G.1)}$$

Clearly this is a rough estimate, however, it can be used to constrain the nonresonant $B^+ \to \pi^+\pi^-\mu^+\nu_\mu$ contribution in the fit of the signal channel. The effect of implicitly assuming the branching fraction of the signal in this estimation will be addressed in the evaluation of the systematic uncertainties.

H MisID regions

This appendix contains details of the particle regions used to estimate the misIDμ and misIDπ backgrounds for the signal channel. In particular, the nature of the candidates that fall outside the defined particle regions are studied and their effect on the misID contributions are estimated.

The misIDμ regions

The particle regions used to estimate the misIDμ background are defined in Table 6.7. Of all the μ^+ candidates in the reference sample, $\sim 9\%$ fall outside the regions. The `PID` information of these candidates is illustrated in Fig. H.1. The single largest contribution is assumed to be true μ^+ based on `ProbNNmu` for `IsMuon = 1` corresponding to approximately 30% of the candidates. Only for `IsMuon = 0` smaller contributions of true π^+, K^+ and p are visible in `ProbNNpi`, `ProbNNK` and `ProbNNp`. Thus, the $\sim 70\%$ of candidates with `IsMuon = 0` are assumed to be a more or less equal contributions of π^+, true K^+, p^+ and μ^+.

The maximal effect of not including the μ candidates that fall outside the particle regions in the calculation of the misIDμ contribution is estimated in the following way. The μ^+ candidates with `IsMuon = 0` and either `ProbNNpi > 0.2`, `ProbNNK > 0.2` or `ProbNNp > 0.2` are assumed to be true π^+, K^+ or p^+ corresponding to $\sim 60\%$ of the candidates that fall outside the regions. Given that these $\sim 60\%$ consist of equal proportions of π^+, K^+ or p^+, the numbers of π^+, K^+ or p^+ in the true particle vector in Eq. F.3 are each increased with 221 particles. Assuming that the efficiencies in the response matrix in Eq. F.1 are unchanged, the new misID fraction is estimated to be 1.07% instead of the nominal fraction of 1.01% corresponding to an increase of $\sim 6\%$. Thus, the maximal contribution from the misIDμ background to the fitted data is found to be $N^{\max}_{\mathrm{misID}\mu} = 1649 \pm 41$. However, to finally evaluate the effect and assign a systematic uncertainty, the fit is performed with the maximal misIDμ contribution and its effect on the number of signal candidates is determined, as described in Sec. 7.5.4.
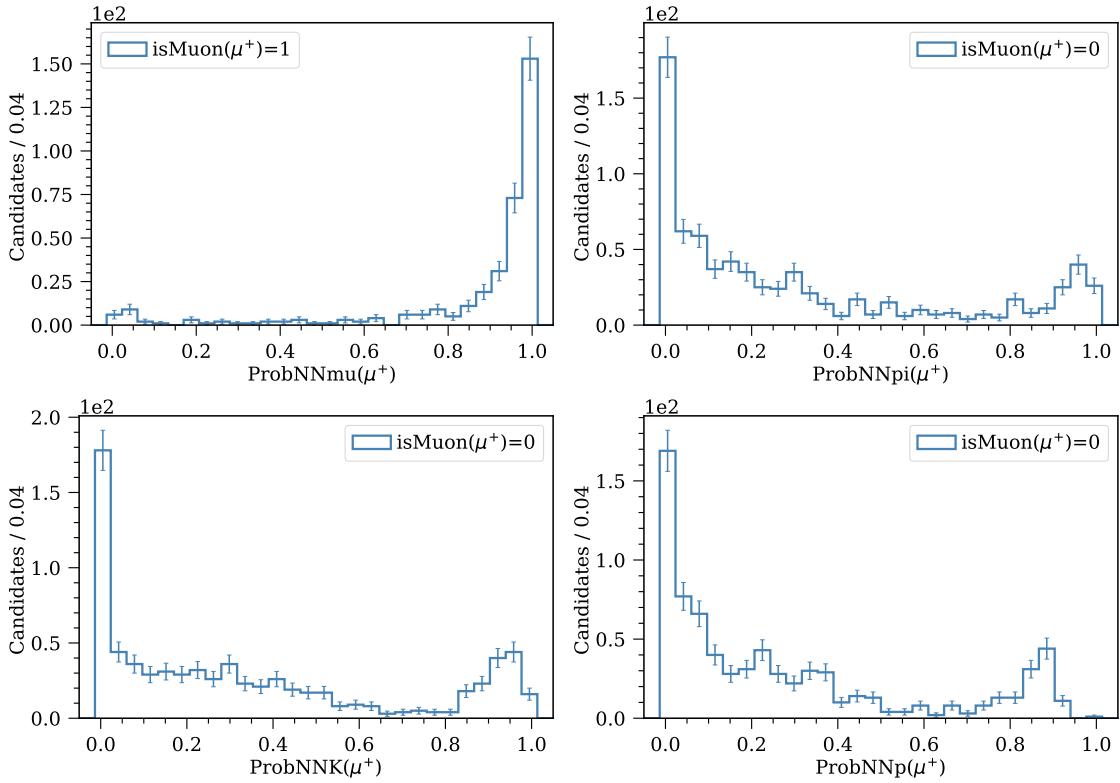
Figure H.1 – `PID` information of μ candidates that fall outside the the misIDμ particle regions, *i.e.* `ProbNNmu` for `IsMuon` $= 1$ (top-left) and `ProbNNpi` (top-right), `ProbNNK` (bottom-right) and `ProbNNp` (bottom-right) for `IsMuon` $= 0$.

The misIDπ^+ regions

The particle regions used to estimate the misIDπ^+ background are defined in Table 7.4. Of all the π^+ candidates in the reference sample, $\sim 5\%$ fall outside the regions. The `PID` information of these candidates are illustrated in Fig. H.2. The largest contribution is assumed to be from true π^+ based on `ProbNNpi` and a smaller contribution is assumed to be from true K^+ based on `ProbNNK`. However, the contribution from μ^+ and p are assumed to be negligible based on `ProbNNmu` with `IsMuon` $= 1$ and `ProbNNp`. If all π^+ candidates that fall outside the defined particle regions are true π^+, the estimation of the misIDπ^+ background is unaffected. However, considering the tail of `ProbNNK`, one can assume some contribution from true K^+.

The maximal effect of not including the π^+ candidates that fall outside the particle regions is estimated in the same way as for the misIDμ background. Making the conservative assumption that the π^+ candidates with `ProbNNK` > 0.2 are true K^+ corresponding to $\sim 16\%$ of the candidates that fall outside the regions, the number of K^+ in the true particle vector in Eq. F.8 is increased with 165 particles. Given that the efficiencies in the response matrix in Eq. F.4 are unchanged, the new misID fraction is estimated to be 1.21% instead of the nominal fraction of 1.10% corresponding to an increase of $\sim 10\%$.
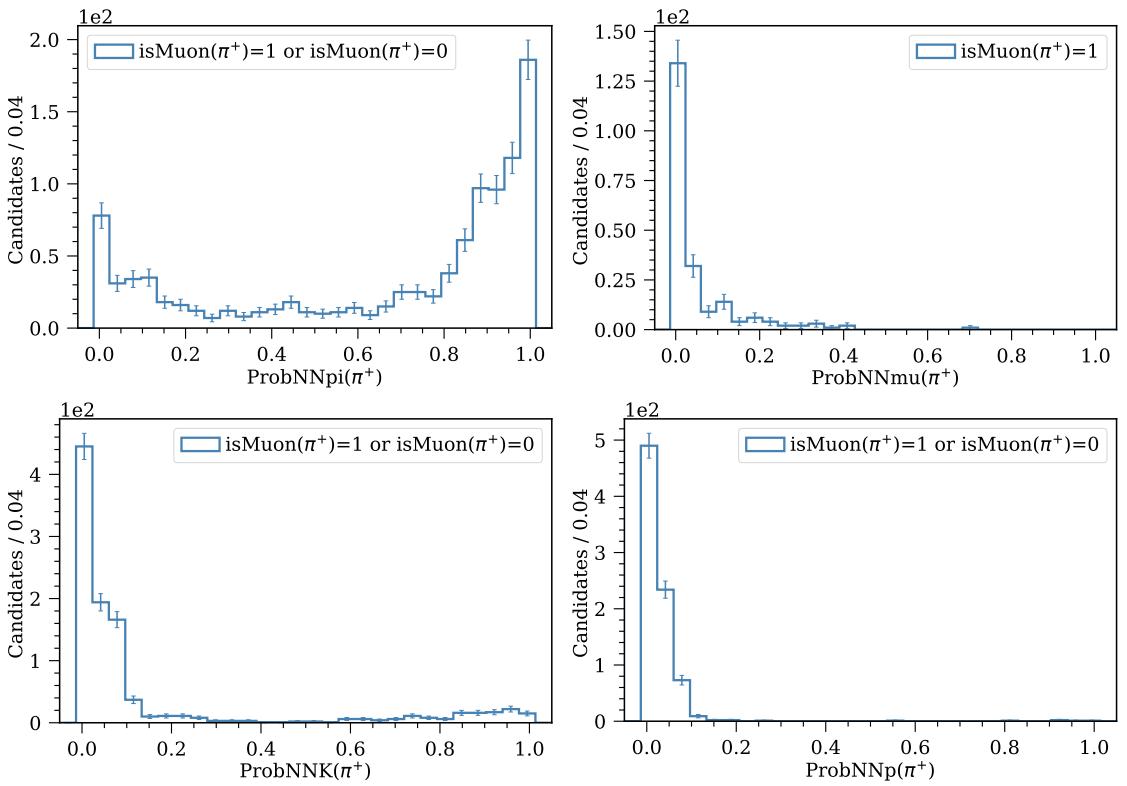
Figure H.2 – `PID` information of π^+ candidates that fall outside the misIDπ^+ particle regions, *i.e.* `ProbNNpi` (top-left), `ProbNNmu` for `IsMuon` $= 1$ (top-right), `ProbNNK` (bottom-right) and `ProbNNp` (bottom-right).

The misIDπ^- regions

The particle regions used to estimate the misIDπ^- background are defined in Table 7.5. Of all the π^- candidates in the reference sample, $\sim 12\%$ fall outside the regions. The `PID` information of these candidates are illustrated in Fig. H.3. The largest contribution is assumed to be from true π^- based on `ProbNNpi` and a smaller contribution is assumed to be from true μ^- based on `ProbNNmu` with `IsMuon` $= 1$. However, the contribution from K^- and p^- are assumed to be negligible based on `ProbNNK` and `ProbNNp`. If all π^- candidates that fall outside the defined particle regions are true π^-, the estimation of the misIDπ^- background is unaffected. However, considering the tail of `ProbNNmu`, one can assume some contribution from true μ^-.

Making the conservative assumption that the π^- candidates with `ProbNNmu` > 0.2 are true μ^- corresponding to $\sim 5\%$ of the candidates that fall outside the regions, the number of μ^- in the true particle vector in Eq. F.9 is increased with 117 particles. Given that the efficiencies in the response matrix in Eq. F.5 are unchanged, the new misID fraction is estimated to be 0.66% instead of the nominal fraction of 0.63% corresponding to an increase of $\sim 4\%$.
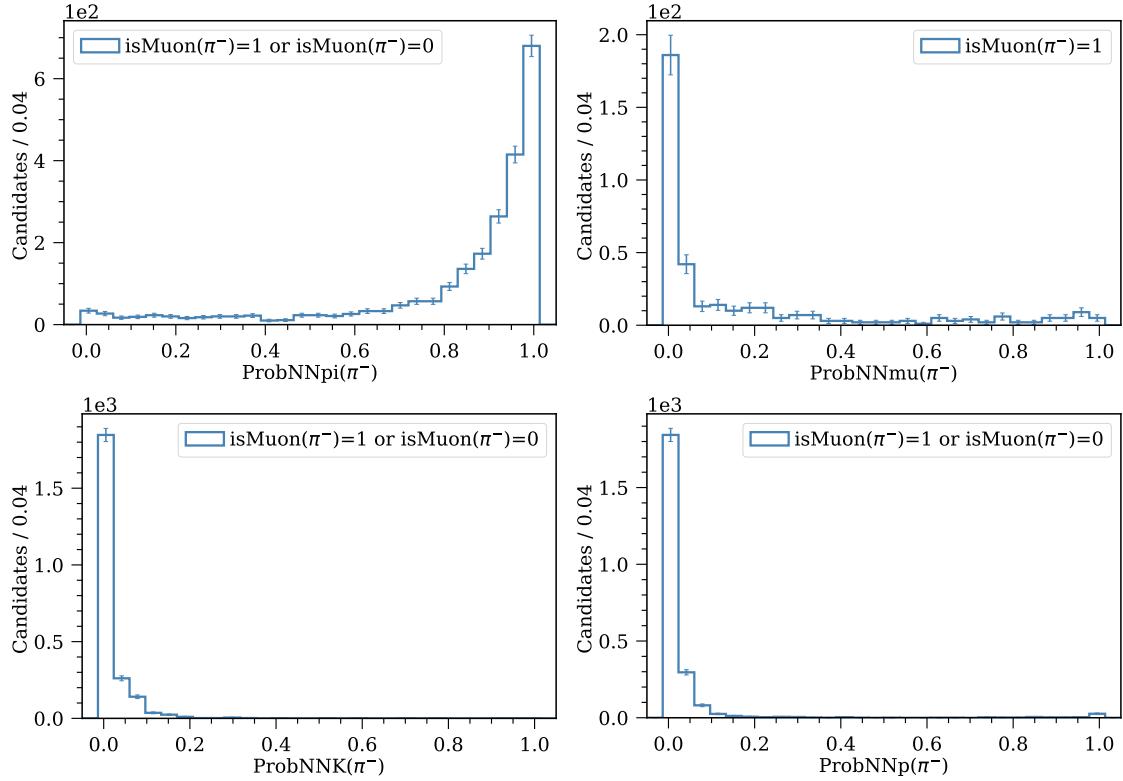
Figure H.3 – `PID` information of π^- candidates that fall outside the misIDπ^- particle regions, *i.e.* `ProbNNpi` (top-left), `ProbNNmu` for `IsMuon` $= 1$ (top-right), `ProbNNK` (bottom-right) and `ProbNNp` (bottom-right).

The combined misIDπ background

Increasing the misIDπ^+ and misIDπ^- contribution with 10% and 4%, respectively, the maximal contribution from the combined misIDπ background to the fitted data is found to be $N_{\text{misID}\pi}^{\text{max}} = 2884 \pm 54$ corresponding to an increase of $\sim 8\%$ with respect to the nominal yield. However, to finally evaluate the effect and assign a systematic uncertainty, the fit is performed with the maximal misIDπ contribution and its effect on the number of signal candidates is determined, as described in Sec. 7.5.4.

I Comparing unfolding methods

This appendix contains information and plots used to evaluate which of the two unfolding algorithms, Bayes or SVD, is more suited for the unfolding of the measured q^2 distributions. To do this, the procedure described in Sec. 6.5.2 is followed, and the observables $\sqrt{\Delta^2}$, R_σ and p_{cov} defined in Eqs. 6.25–6.27 are evaluated for different regularisation strengths. The optimal regularisation strength should minimise $\sqrt{\Delta^2}$, satisfy $R_\sigma \approx 1$ (but not $R_\sigma < 1$), and $p_{\mathrm{cov}} \geq 68.3\%$ within one percent. In addition to this, $\Delta/\sigma_{\mathrm{unf}}$ is considered for each bin. This evaluation is performed in Sec. I.1 for the measured q^2 distributions of the $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ (NORM) and $B^+ \to D^{*0} \mu^+ \nu_\mu$ (Bu2) decays, as described in Sec. 6.5, and in Sec. I.2 for the measured q^2 distribution of the signal, as described in Sec. 7.4.

I.1 Unfolding q^2 of the normalisation channel fit

In this section, the unfolding of the q^2 distributions of the normalisation channel (NORM) and $B^+ \to D^{*0} \mu^+ \nu_\mu$ decay (Bu2) are evaluated.

Optimisation of Bayes unfolding method (NORM)

The performance plots of Bayes unfolding method are given in Fig. I.1. The optimal regularisation strength is determined to be $r = 3$, as it satisfies all the above criteria, *i.e.*:

$$\sqrt{\Delta^2}^{\mathrm{Bayes}} = 310\,, \tag{I.1}$$

$$R_\sigma^{\mathrm{Bayes}} = 1.10\,, \tag{I.2}$$

$$p_{\mathrm{cov}}^{\mathrm{Bayes}} = 68\%\,. \tag{I.3}$$

## Appendix I. Comparing unfolding methods
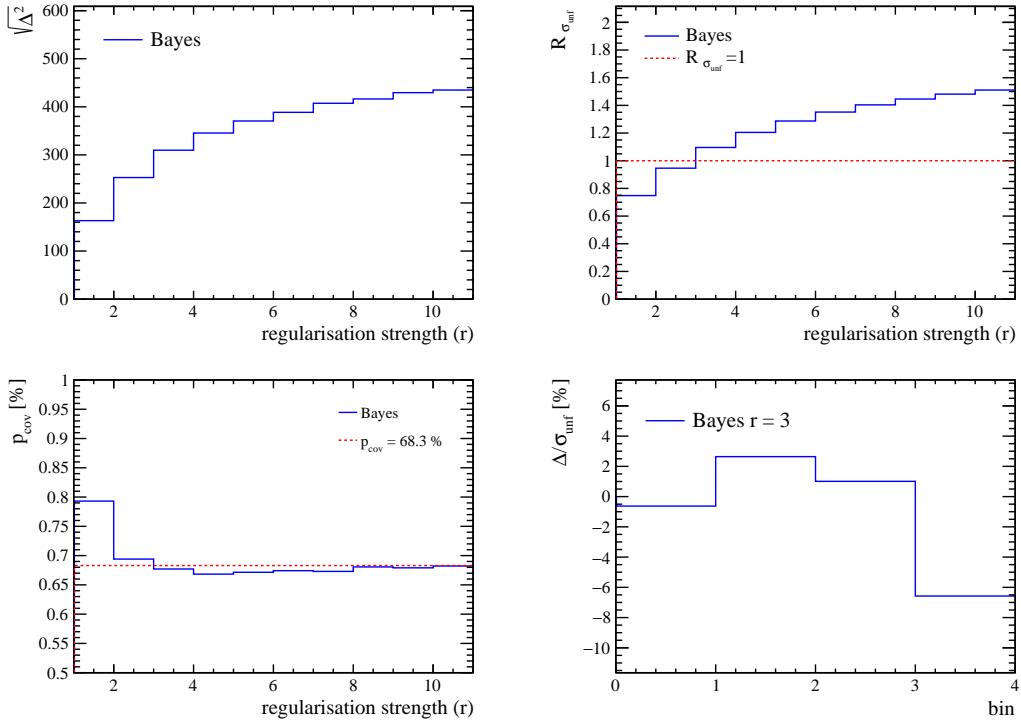
Figure I.1 – Performance of the Bayes method for NORM, *i.e.* $\sqrt{\Delta^2}$, R_σ and p_{cov} as a function of the regularisation r and Δ for $r = 3$ as a function of the bin number.

Optimisation of the SVD unfolding method (NORM)

The performance plots of the SVD unfolding method are given in Fig. I.6. The optimal regularisation strength is determined to be $r = 4$, as it satisfies all the above criteria, *i.e.*:

$$\sqrt{\Delta^2}^{\text{SVD}} = 369\,, \tag{I.4}$$
$$R_\sigma^{\text{SVD}} = 1.30\,, \tag{I.5}$$
$$p_{\text{cov}}^{\text{SVD}} = 68\%\,. \tag{I.6}$$

Comparing Bayes and SVD method (NORM)

The Bayes method shows the best unfolding performance, *i.e.*:

$$\sqrt{\Delta^2}^{\text{Bayes}} < \sqrt{\Delta^2}^{\text{SVD}}\,, \tag{I.7}$$
$$R_\sigma^{\text{Bayes}} < R_\sigma^{\text{SVD}}\,, \tag{I.8}$$
$$p_{\text{cov}}^{\text{Bayes}} = p_{\text{cov}}^{\text{SVD}}\,. \tag{I.9}$$

For both methods $\Delta/\sigma_{\text{unf}}$ is less than 7% in all bins.

Figure I.2 – Performance of the SVD method for NORM, *i.e.* $\sqrt{\Delta^2}$, R_σ and p_{cov} as a function of the regularisation r and Δ for $r = 4$ as a function of the bin number.

Optimisation of Bayes unfolding method (Bu2)

The performance plots of Bayes unfolding method are given in Fig. I.3. The optimal regularisation strength is determined to be $r = 4$, as it satisfies all the above criteria, *i.e.*:

$$\sqrt{\Delta^2}^{\text{Bayes}} = 245 \,, \tag{I.10}$$

$$R_\sigma^{\text{Bayes}} = 1.07 \,, \tag{I.11}$$

$$p_{\text{cov}}^{\text{Bayes}} = 79\% \,. \tag{I.12}$$

Optimisation of the SVD unfolding method (Bu2)

The performance plots of the SVD unfolding method are given in Fig. I.6. The optimal regularisation strength is determined to be $r = 3$, as it satisfies all the above criteria, *i.e.*:

$$\sqrt{\Delta^2}^{\text{SVD}} = 312 \,, \tag{I.13}$$

$$R_\sigma^{\text{SVD}} = 1.27 \,, \tag{I.14}$$

$$p_{\text{cov}}^{\text{SVD}} = 76\% \,. \tag{I.15}$$

Figure I.3 – Performance of the Bayes method for Bu2, *i.e.* $\sqrt{\Delta^2}$, R_σ and p_{cov} as a function of the regularisation r and Δ for $r = 4$ as a function of the bin number.
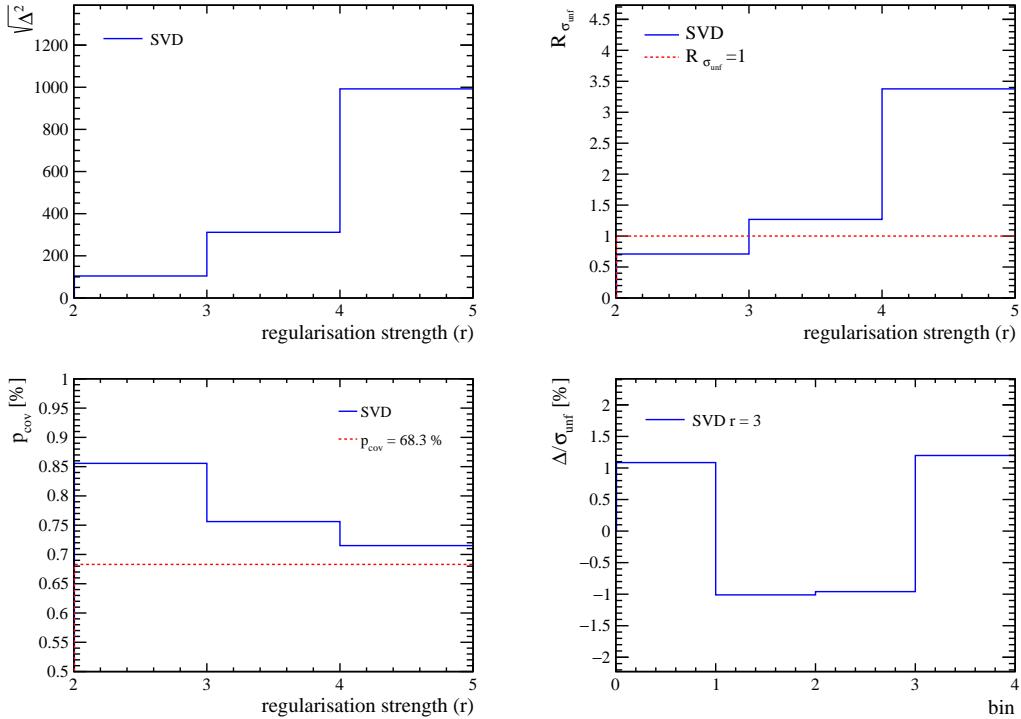
Figure I.4 – Performance of the SVD method for Bu2, *i.e.* $\sqrt{\Delta^2}$, R_σ and p_{cov} as a function of the regularisation r and Δ for $r = 3$ as a function of the bin number.

Comparing Bayes and SVD method (Bu2)

The Bayes method shows the best unfolding performance, *i.e.*:

$$\sqrt{\Delta^2}^{\text{Bayes}} < \sqrt{\Delta^2}^{\text{SVD}}, \tag{I.16}$$

$$R_\sigma^{\text{Bayes}} < R_\sigma^{\text{SVD}}, \tag{I.17}$$

$$p_{\text{cov}}^{\text{Bayes}} > p_{\text{cov}}^{\text{SVD}}. \tag{I.18}$$

For both methods $\Delta/\sigma_{\text{unf}}$ is less than 3% in all bins.

I.2 Unfolding q^2 of the signal channel fit

In this section, the unfolding of the q^2 distributions of the signal channel is evaluated.

Optimisation of Bayes unfolding method

The performance plots of Bayes unfolding method are given in Fig. I.5. The optimal regularisation strength is determined to be $r = 3$, as it satisfies all the above criteria, *i.e.*:

$$\sqrt{\Delta^2}^{\text{Bayes}} = 168, \tag{I.19}$$

$$R_\sigma^{\text{Bayes}} = 1.15, \tag{I.20}$$

$$p_{\text{cov}}^{\text{Bayes}} = 71\%. \tag{I.21}$$

Optimisation of the SVD unfolding method

The performance plots of the SVD unfolding method are given in Fig. I.6. The optimal regularisation strength is determined to be $r = 7$, as it satisfies all the above criteria, *i.e.*:

$$\sqrt{\Delta^2}^{\text{SVD}} = 146, \tag{I.22}$$

$$R_\sigma^{\text{SVD}} = 1.01, \tag{I.23}$$

$$p_{\text{cov}}^{\text{SVD}} = 72\%. \tag{I.24}$$
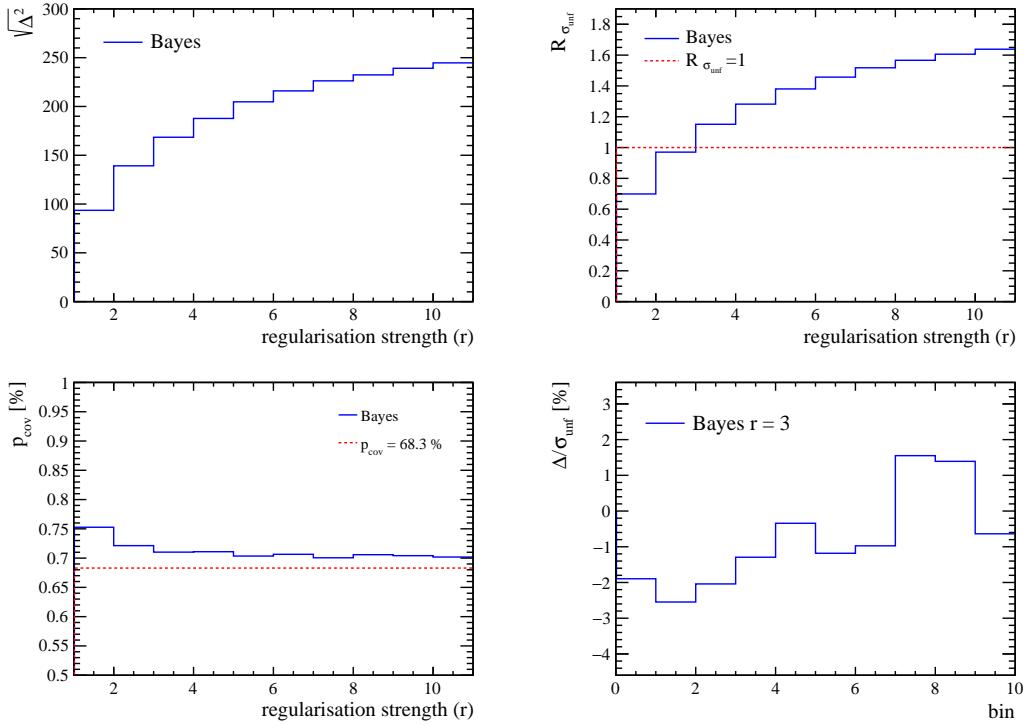
Figure I.5 – Performance of the Bayes method for the signal channel, *i.e.* $\sqrt{\Delta^2}$, R_σ and p_{cov} as a function of the regularisation r and Δ for $r = 3$ as a function of the bin number.
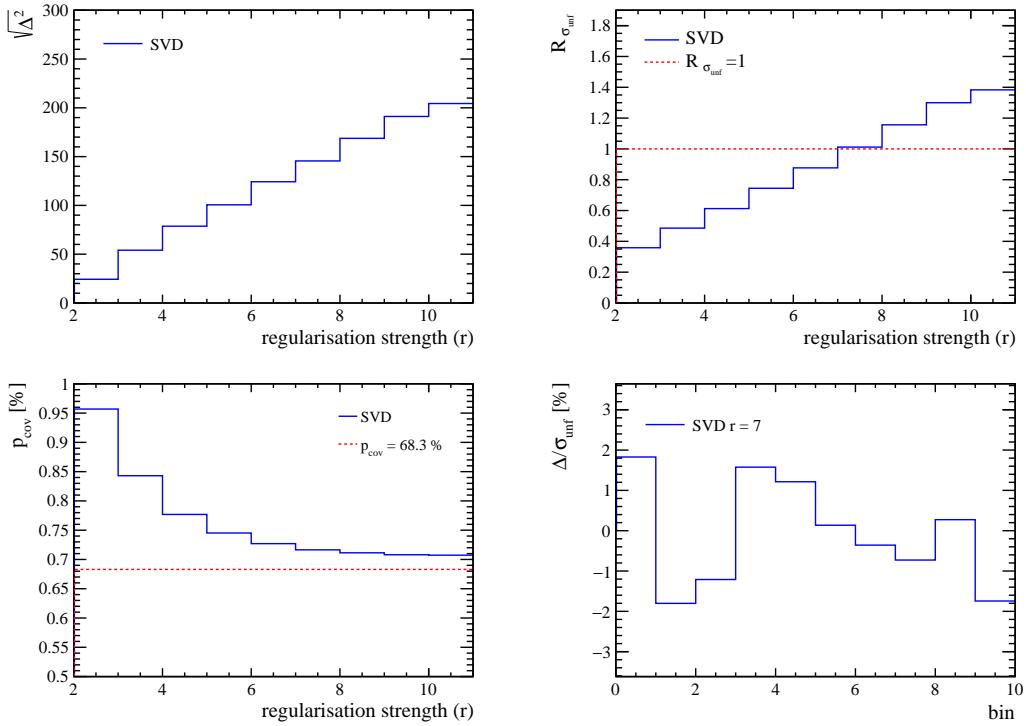
Figure I.6 – Performance of the SVD method for the signal channel, *i.e.* $\sqrt{\Delta^2}$, R_σ and p_{cov} as a function of the regularisation r and Δ for $r = 7$ as a function of the bin number.

Comparing Bayes and SVD method

The SVD method shows the best unfolding performance, *i.e.*:

$$\sqrt{\Delta^2}^{\text{SVD}} < \sqrt{\Delta^2}^{\text{Bayes}}, \tag{I.25}$$

$$R_\sigma^{\text{SVD}} < R_\sigma^{\text{Bayes}}, \tag{I.26}$$

$$p_{\text{cov}}^{\text{SVD}} > p_{\text{cov}}^{\text{Bayes}}. \tag{I.27}$$

For both methods $\Delta/\sigma_{\text{unf}}$ is less than 3% in all bins.

J Efficiency corrected yields

This appendix contains the efficiency corrected yields of the signal and normalisation channels defined as

$$Y_{S_i}^{\mathrm{corr}} = N_{\mathrm{sig},i}/\epsilon_{\mathrm{sig},i}^{\mathrm{tot}} \qquad \text{and} \qquad Y_N^{\mathrm{corr}} = N_{\mathrm{norm}}/\epsilon_{\mathrm{norm}}^{\mathrm{tot}} . \tag{J.1}$$

The yields $N_{\mathrm{sig},i}$ and N_{norm} are obtained from the fits to the signal and normalisation channels (see Tables 7.8 and 6.11), and the total selection efficiencies $\epsilon_{\mathrm{sig},i}^{\mathrm{tot}}$ and $\epsilon_{\mathrm{norm}}^{\mathrm{tot}}$ are calculated in Secs. 8.1.1–8.1.5 and given in Table J.1. The corrections to the tracking reconstruction and trigger efficiencies (see Tables 8.3 and 8.5) are not included, as they are obtained for the efficiency ratios. The resulting efficiency corrected yields and their statistical uncertainties are given in Table J.1. The statistical uncertainties of the efficiencies are estimated according to Sec. 8.3, and they are combined with the statistical uncertainties of the fitted yields as explained in Sec. 8.4.

Table J.1 – Total efficiency, ϵ^{tot}, and efficiency corrected yields, Y^{corr}, for the signal $B^+ \to \rho^0 \mu^+ \nu_\mu$ decay in each q^2 bin with bin boundaries q_{min}^2–q_{max}^2 and for the normalisation $B^+ \to \bar{D}^0 \mu^+ \nu_\mu$ decay with $\bar{D}^0 \to \pi^+ \pi^-$ in the integrated q^2 region. All uncertainties are statistical and the relative uncertainties are given in parentheses.

Decay mode X	q_{min}^2–q_{max}^2 [GeV^2/c^4]	$\epsilon_X^{\mathrm{tot}}$ [10^{-4}]	Y_X^{corr} [10^6]
Signal, bin 1	0.0–2.2	$4.56 \pm 0.04\,(0.9\%)$	$6.47 \pm 0.35\,(5.4\%)$
Signal, bin 2	2.2–3.9	$5.21 \pm 0.04\,(0.8\%)$	$6.06 \pm 0.27\,(4.5\%)$
Signal, bin 3	3.9–5.4	$5.46 \pm 0.05\,(0.8\%)$	$5.76 \pm 0.28\,(4.9\%)$
Signal, bin 4	5.4–6.7	$5.62 \pm 0.05\,(0.8\%)$	$5.55 \pm 0.29\,(5.2\%)$
Signal, bin 5	6.7–8.0	$5.43 \pm 0.05\,(0.8\%)$	$5.75 \pm 0.30\,(5.2\%)$
Signal, bin 6	8.0–9.3	$5.30 \pm 0.04\,(0.8\%)$	$6.15 \pm 0.31\,(5.1\%)$
Signal, bin 7	9.3–10.7	$4.93 \pm 0.04\,(0.8\%)$	$7.50 \pm 0.35\,(4.7\%)$
Signal, bin 8	10.7–12.2	$4.40 \pm 0.04\,(0.8\%)$	$7.72 \pm 0.38\,(4.9\%)$
Signal, bin 9	12.2–14.2	$3.45 \pm 0.03\,(0.8\%)$	$9.99 \pm 0.45\,(4.5\%)$
Signal, bin 10	14.2–21.5	$1.58 \pm 0.01\,(0.8\%)$	$23.25 \pm 1.03\,(4.4\%)$
Normalisation	0.0–21.5	$7.57 \pm 0.03\,(0.4\%)$	$22.40 \pm 0.75\,(3.3\%)$

Bibliography

[1] J. Dalton, *A new system of chemical philosophy*, Manchester: Printed by S. Russel, Deansgate 125, for R. Bickerstaff, Strand, London, 1808.

[2] J. J. Thomson, *Cathode rays*, Phil. Mag. Ser. 5 **44** (1897) 293.

[3] E. Rutherford, *The scattering of alpha and beta particles by matter and the structure of the atom*, Phil. Mag. Ser. 6 **21** (1911) 669.

[4] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214.

[5] CMS collaboration, S. Chatrchyan *et al.*, *A new boson with a mass of 125 GeV observed with the CMS experiment at the Large Hadron Collider*, Science **338** (2012) 1569.

[6] W. N. Cottingham and D. A. Greenwood, *An introduction to the Standard Model of particle physics*, Cambridge University Press, 2007.

[7] D. H. Perkins, *Introduction to High Energy Physics*, Cambridge University Press, 4th ed., 2000. doi: 10.1017/CBO9780511809040.

[8] A. Einstein, *Zur Electrodynamik bewegt Körper*, Annalen der Physik **17** (1905) 891.

[9] W. Buchmuller and C. Ludeling, *Field Theory and Standard Model*, in *2005 European School of High-Energy Physics*, 9, 2006. arXiv:hep-ph/0609174.

[10] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.

[11] E. Fermi, *On the quantization of the monoatomic ideal gas*, arXiv:cond-mat/9912229.

[12] P. A. M. Dirac, *On the theory of quantum mechanics*, Proc. Roy. Soc. Lond. A **112** (1926) 661.

[13] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

Bibliography

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

[15] Particle Data Group, R. L. Workman *et al.*, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[16] Super-Kamiokande collaboration, Y. Ashie *et al.*, *Evidence for an oscillatory signature in atmospheric neutrino oscillation*, Phys. Rev. Lett. **93** (2004) 101801, arXiv:hep-ex/0404034.

[17] SNO collaboration, B. Aharmim *et al.*, *Electron energy spectra, fluxes, and day-night asymmetries of B-8 solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory*, Phys. Rev. C **72** (2005) 055502, arXiv:nucl-ex/0502021.

[18] MINOS collaboration, P. Adamson *et al.*, *Measurement of neutrino oscillations with the MINOS detectors in the NuMI beam*, Phys. Rev. Lett. **101** (2008) 131802, arXiv:0806.2237.

[19] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343.

[20] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973) 1346.

[21] P. A. M. Dirac, *Quantum theory of emission and absorption of radiation*, Proc. Roy. Soc. Lond. A **114** (1927) 243.

[22] D. J. Griffiths, *Introduction to elementary particles*, Physics textbook, 2nd rev. version, Wiley, New York, NY, 2008.

[23] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.

[24] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[25] A. Salam, *Weak and electromagnetic interactions*, Conf. Proc. C **680519** (1968) 367.

[26] M. D'Onofrio and K. Rummukainen, *Standard Model cross-over on the lattice*, Phys. Rev. D **93** (2016) 025003, arXiv:1508.07161.

[27] G. Luders, *Proof of the TCP theorem*, Annals Phys. **2** (1957) 1.

[28] C. S. Wu *et al.*, *Experimental test of parity conservation in β decay*, Phys. Rev. **105** (1957) 1413.

[29] R. L. Garwin, L. M. Lederman, and M. Weinrich, *Observations of the failure of conservation of parity and charge conjugation in meson decays: the magnetic moment of the free muon*, Phys. Rev. **105** (1957) 1415.

[30] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π decay of the K_2^0 meson*, Phys. Rev. Lett. **13** (1964) 138.

[31] Belle collaboration, K. Abe *et al.*, *Observation of large CP violation in the neutral B meson system*, Phys. Rev. Lett. **87** (2001) 091802, `arXiv:hep-ex/0107061`.

[32] BaBar collaboration, J. P. Lees *et al.*, *Observation of time reversal violation in the B^0 meson system*, Phys. Rev. Lett. **109** (2012) 211801, `arXiv:1207.5832`.

[33] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in charm decays*, Phys. Rev. Lett. **122** (2019) 211803, `arXiv:1903.08726`.

[34] B. C. Odom, D. Hanneke, B. D'Urso, and G. Gabrielse, *New measurement of the electron magnetic moment using a one-electron quantum cyclotron*, Phys. Rev. Lett. **97** (2006) 030801.

[35] A. Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Annalen der Physik **49** (1916) .

[36] L. Canetti, M. Drewes, and M. Shaposhnikov, *Matter and antimatter in the Universe*, New J. Phys. **14** (2012) 095012, `arXiv:1204.4186`.

[37] I. S. Sogami and T. Shinohara, *Universal seesaw mechanisms for quark lepton mass spectrum*, Phys. Rev. D **47** (1993) 2905.

[38] V. Trimble, *Existence and nature of dark matter in the Universe*, Ann. Rev. Astron. Astrophys. **25** (1987) 425.

[39] NASA, *Dark energy, dark matter*, https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy.

[40] P. J. E. Peebles and B. Ratra, *The cosmological constant and dark energy*, Rev. Mod. Phys. **75** (2003) 559, `arXiv:astro-ph/0207347`.

[41] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.

[42] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.

[43] M. Neubert, *B decays and CP violation*, Int. J. Mod. Phys. A **11** (1996) 4173, `arXiv:hep-ph/9604412`.

[44] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules*, JHEP **08** (2016) 098, `arXiv:1503.05534`.

[45] F. Knechtli, *Lattice Quantum Chromodynamics*, PoS **CORFU2016** (2017) 020, `arXiv:1706.00282`.

[46] J. D. Richman and P. R. Burchat, *Leptonic and semileptonic decays of charm and bottom hadrons*, Rev. Mod. Phys. **67** (1995) 893, `arXiv:hep-ph/9508250`.

[47] C. Bouchard, L. Cao, and P. Owen, *Summary of the 2018 CKM working group on semileptonic and leptonic b-hadron decays*, in *10th International Workshop on the CKM Unitarity Triangle*, 2019. `arXiv:1902.09412`.

[48] V. G. Luth, *Semileptonic B meson decays*, Ann. Rev. Nucl. Part. Sci. **61** (2011) 119.

[49] Belle collaboration, A. Abdesselam *et al.*, *Precise determination of the CKM matrix element $|V_{cb}|$ with $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, `arXiv:1702.01521`.

[50] BaBar and Belle collaborations, A. J. Bevan *et al.*, *The physics of the B factories*, Eur. Phys. J. C **74** (2014) 3026, `arXiv:1406.6311`.

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

[52] L. Evans and P. Bryant, *LHC machine*, JINST **3** (2008) S08001.

[53] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, `arXiv:1612.05140`, erratum: Phys. Rev. Lett. **119** (2017) 1699012.

[54] C. G. Boyd, B. Grinstein, and R. F. Lebed, *Constraints on form factors for exclusive semileptonic heavy to light meson decays*, Phys. Rev. Lett. **74** (1995) 4603, `arXiv:hep-ph/9412324`.

[55] I. I. Balitsky, V. M. Braun, and A. V. Kolesnichenko, *Radiative decay $\sigma^+ \to p\gamma$ in Quantum Chromodynamics*, Nucl. Phys. B **312** (1989) 509.

[56] S. Weinzierl and O. I. Yakovlev, *New approach for calculating heavy to light form-factors with QCD sum rules on the light cone*, JHEP **01** (2001) 005, `arXiv:hep-ph/0007317`.

[57] HFLAV group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron and tau lepton properties as of 2021*, `arXiv:2206.07501`.

[58] Belle collaboration, R. Glattauer, C. Schwanda *et al.*, *Measurement of the decay $B \to D\ell\nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$*, Phys. Rev. D **93** (2016) 032006, `arXiv:1510.03657`.

[59] D. Bigi, P. Gambino, and S. Schacht, *A fresh look at the determination of $|V_{cb}|$ from $B \to D^*\ell\nu$*, Phys. Lett. B **769** (2017) 441, `arXiv:1703.06124`.

[60] D. Bigi and P. Gambino, *Revisiting $B \to D\ell\nu$*, Phys. Rev. D **94** (2016) 094008, `arXiv:1606.08030`.

[61] F. U. Bernlochner, Z. Ligeti, and S. Turczyk, *A new way to search for right-handed currents in semileptonic $B \to \rho\ell\bar{\nu}$ decay*, Phys. Rev. D **90** (2014) 094003, `arXiv:1408.2516`.

[62] Belle collaboration, A. Sibidanov, K. E. Varvel *et al.*, *Study of exclusive $B \to X_u \ell \nu$ decays and extraction of $|V_{ub}|$ using full reconstruction tagging at the Belle experiment*, Phys. Rev. D **88** (2013) 032005, `arXiv:1306.2781`.

[63] BaBar collaboration, P. del Amo Sanchez *et al.*, *Study of $B \to \pi \ell \nu$ and $B \to \rho \ell \nu$ decays and determination of $|V_{ub}|$*, Phys. Rev. D **83** (2011) 032007, `arXiv:1005.3288`.

[64] F. U. Bernlochner, M. T. Prim, and D. J. Robinson, *$B \to \rho \ell \bar{\nu}$ and $B \to \omega \ell \bar{\nu}$ in and beyond the Standard Model: Improved predictions and $|V_{ub}|$*, Phys. Rev. D **104** (2021) 034032, `arXiv:2104.05739`.

[65] LHCb collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Phys. **11** (2015) 743, `arXiv:1504.01568`.

[66] A. Crivellin, *Effects of right-handed charged currents on the determinations of $|V_{ub}|$ and $|V_{cb}|$*, Phys. Rev. D **81** (2010) 031301, `arXiv:0907.2461`.

[67] D. Becirevic and A. B. Kaidalov, *Comment on the heavy $\to$ light form-factors*, Phys. Lett. B **478** (2000) 417, `arXiv:hep-ph/9904490`.

[68] P. Ball and R. Zwicky, *New results on $B \to \pi, K, \eta$ decay form factors from light-cone sum rules*, Phys. Rev. D **71** (2005) 014015, `arXiv:hep-ph/0406232`.

[69] C. Bourrely, I. Caprini, and L. Lellouch, *Model-independent description of $B \to \pi \ell \nu$ decays and a determination of $|V_{ub}|$*, Phys. Rev. D **79** (2009) 013008, `arXiv:0807.2722`, erratum: Phys. Rev. D **82** (2010) 099902.

[70] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form factors from light-cone sum rules revisited*, Phys. Rev. D **71** (2005) 014029, `arXiv:hep-ph/0412079`.

[71] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: An update*, Phys. Rev. D **52** (1995) 2783, `arXiv:hep-ph/9503486`.

[72] Belle II collaboration, F. Abudinén *et al.*, *Reconstruction of $B \to \rho \ell \nu_\ell$ decays identified using hadronic decays of the recoil B meson in 2019–2021 Belle II data*, `arXiv:2211.15270`.

[73] J. Kahn (on behalf of the Belle II collaboration), *The Belle II experiment*, in *CERN-BINP Workshop for young scientists in e^+e^- colliders*. CERN Proceedings **1 (2017) 45**.

[74] SLD collaboration, K. Abe *et al.*, *A measurement of R_b using a vertex mass tag*, Phys. Rev. Lett. **80** (1998) 660, `arXiv:hep-ex/9708015`.

[75] S. Myers, *The LEP collider, from design to approval and commissioning*, CERN Yellow Report, CERN-91-008 (1991).

[76] P. Mouche, *Overall view of the LHC*, OPEN-PHO-ACCEL-2014-00.

[77] O. Grobner, *Overview of the LHC vacuum system*, Vacuum **60** (2001) 25.

Bibliography

[78] CERN, *Accelerating: Radiofrequency cavities*, https://home.cern/science/engineering/accelerating-radiofrequency-cavities.

[79] R. Bruce, N. Fuster-Martínez, A. Mereghetti, D. Mirarchi, and S. Redaelli, *Review of LHC Run 2 machine configurations*, in *9th LHC operations Evian workshop*. CERN-ACC-2019-059 pp. 187–197.

[80] CERN, *Facts and figures about the LHC*, https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[81] X. C. Vidal and R. C. Manzano, *Momentum, taking a closer look at LHC*, https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.momentum.

[82] L. Rossi, *The LHC superconducting magnets*, Conf. Proc. C **030512** (2003) 141.

[83] X. C. Vidal and R. C. Manzano, *Magnetic multipoles, taking a closer look at LHC*, https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.magnetic_multipoles.

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

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

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

[87] X. C. Vidal and R. C. Manzano, *Proton source*, https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.proton_source.

[88] CERN, *Linear accelerator 2*, https://home.cern/science/accelerators/linear-accelerator-2.

[89] CERN, *The proton synchrotron booster*, https://home.cern/science/accelerators/proton-synchrotron-booster.

[90] CERN, *The proton synchrotron*, https://home.cern/science/accelerators/proton-synchrotron.

[91] CERN, *The super proton synchrotron*, https://home.cern/science/accelerators/super-proton-synchrotron.

[92] M. Zinser, *Search for new heavy charged bosons and measurement of high-mass Drell-Yan production in proton-proton collisions*, PhD thesis, Mainz U., CERN-THESIS-2016-212.

[93] J. Gillies, *Luminosity*, https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.luminosity.

[94] K. Wille, *The physics of particle accelerators: An introduction*, Oxford University Press, 2000. https://inspirehep.net/literature/541588.

[95] J. Gillies, *Luminosity? why don't we just say collision rate?*, https://home.cern/news/opinion/cern/luminosity-why-dont-we-just-say-collision-rate.

[96] B. Salvachua, *Overview of proton-proton physics during Run 2*, in *9th LHC operations Evian workshop*. CERN-ACC-2019-059, pp. 7–14.

[97] R. Alemany-Fernandez, E. Bravin, L. Drosdal, A. Gorzawski, V. Kain, M. Lamont, A. Macpherson, G. Papotti, M. Pojer, and L. Ponce, *Operation and configuration of the LHC in Run 1*, CERN-ACC-NOTE-2013-0041.

[98] CERN news, *End of LHC run 1: First shutdown begins*, https://timeline.web.cern.ch/end-lhc-run-1-first-shutdown-begins.

[99] J. Wenninger, *Operation and configuration of the LHC in Run 2*, CERN-ACC-NOTE-2019-0007.

[100] CERN, *LHC Run 3*, https://home.cern/press/2022/run-3.

[101] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv:1703.05747.

[102] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment*, JHEP **04** (2020) 188, arXiv:1910.12127.

[103] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$*, Phys. Rev. Lett. **115** (2015) 111803, arXiv:1506.08614, erratum: Phys. Rev. Lett. **115** (2015) 159901.

[104] LHCb collaboration, *Measurement of the ratios of branching fractions $\mathcal{R}(D^*)$ and $\mathcal{R}(D^0)$*, arXiv:2302.02886.

[105] G. M. Ciezarek (on behalf of the LHCb collaboration), *First joint measurement of $R(D^*)$ and $R(D)$ at LHCb*, https://indico.cern.ch/event/1187939/2022.

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

[107] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, arXiv:2103.11769.

[108] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, arXiv:2212.09152.

[109] M. Pepe Altarelli and F. Teubert, *B physics at LHCb*, Int. J. Mod. Phys. A **23** (2008) 5117, arXiv:0802.1901.

[110] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. A **30** (2015) 1530022, arXiv:1412.6352.

[111] LHCb collaboration, *Tracking system*, https://lhcb-outreach.web.cern.ch/detector/tracking-system/.

[112] LHCb collaboration, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[113] C. Abellan Beteta, M. Atzeni, V. Battista, A. Bursche, B. Dey, A. Dosil Suarez, C. Elsasser, A. Fernandez Prieto, J. Fu, E. Graverini *et al.* (on behalf of the LHCb Silicon Tracker Group), *Monitoring radiation damage in the LHCb Tracker Turicensis*, JINST **15** (2020) P08016, arXiv:1809.05063.

[114] V. Pugatch, V. M. Aushev, V. Kyva, I. Kolomiets, Y. N. Pavlenko, S. Prystupa, D. Volyanskyy, and Y. Vasilev (on behalf of the LHCb collaboration), *Radiation monitoring system for the LHCb Inner Tracker*, LHCb-2002-067, CERN-LHCb-2002-067.

[115] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, LHCb-TDR-005.

[116] LHCb collaboration, *LHCb Magnet: Technical Design Report*, CERN-LHCC-2000-007, LHCb-TDR-001.

[117] LHCb Silicon Tracker group, J. Luisier, *Performance of LHCb Silicon Tracker Detector in the LHC*, Phys. Procedia **37** (2012) 851.

[118] Stadt Zürich, *History of the city of Zürich*, https://www.stadt-zuerich.ch/portal/en/index/portraet_der_stadt_zuerich/geschichte_der_stadt_zuerich.html.

[119] LHCb collaboration, *LHCb: Inner tracker technical design report*, CERN-LHCC-2002-029, LHCb-TDR-8.

[120] LHCb Outer Tracker group, R. Arink, S. Bachmann, Y. Bagaturia, H. Band, Th. Bauer, A. Berkien, Ch. Färber, A. Bien, J. Blouw, L. Ceelie *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[121] LHCb collaboration, *Tracking strategies used in LHCb*, https://twiki.cern.ch/twiki/pub/LHCb/LHCbTrackingStrategies/tracktypes.svg.

[122] R. Aaij, S. Akar, J. Albrecht, M. Alexander, A. Alfonso Albero, S. Amerio, L. Anderlini, P. d'Argent, A. Baranov, W. Barter *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[123] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, LHCb-PUB-2011-001, CERN-LHCb-PUB-2011-001.

[124] O. Callot and M. Schiller, *PatSeeding: A standalone track reconstruction algorithm*, CERN-LHCB-2008-042, LHCB-2008-042.

[125] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and performance studies*, LHCb-2007-015, CERN-LHCb-2007-015.

[126] M. Needham and J. Van Tilburg, *Performance of the track matching*, LHCb-2007-020, CERN-LHCb-2007-020.

[127] M. Needham, *Performance of the track matching*, CERN-LHCB-2007-129, LHCB-2007-129.

[128] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444.

[129] R. Aaij, L. Anderlini, S. Benson, M. Cattaneo, P. Charpentier, M. Clemencic, A. Falabella, F. Ferrari, M. Fontana, V. Gligorov *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) 1, arXiv:1803.00824.

[130] HyperPhysics hosted by the Dep. of Physics and Astronomy at Georgia State University, *Cerenkov radiation*, http://hyperphysics.phy-astr.gsu.edu/hbase/Relativ/einvel.html.

[131] LHCb calorimeter group, C. Abellán Beteta, A. Alfonso Albero, Y. Amhis, S. Barsuk, C. Beigbeder-Beau, I. Belyaev, R. Bonnefoy, D. Breton, O. Callot, M. Calvo Gomez *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, arXiv:2008.11556.

[132] C. Coca, T. Preda, A. Rosca, I. Ajinenko, A. E. Dorokhov, R. I. Dzhelyadin, A. K. Konoplyannikov, V. Matveev, V. Novikov, O. P. Yushchenko, and Y. Ranyuk, *The hadron calorimeter prototype beam-test results*, LHCb-2000-036, CERN-LHCb-2000-036.

[133] A. Puig Navarro, *First measurements of radiative B decays in LHCb,* PhD thesis, University of Barcelona, 2012, CERN-THESIS-2012-025.

[134] C. Marin Benito (on behalf of the LHCb collaboration), *PID strategy and performance at LHCb in Run 2*, PoS **ICHEP2018** (2019) 687.

[135] LHCb RICH group, M. Adinolfi, G. Aglieri Rinella, E. Albrecht, T. Bellunato, S. Benson, T. Blake, C. Blanks, S. Brisbane, N. H. Brook, M. Calvi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, arXiv:1211.6759.

[136] S. Akar, J.-P. Cachemiche, H. Chanal, J. Cogan, C. Drancourt, R. Lefèvre, R. Le Gac, N. Neufeld, P. Robbe, and M. Vesterinen, *Review document: Low Level Trigger (LLT)*, LHCb-PUB-2014-037, CERN-LHCb-PUB-2014-037.

Bibliography

[137] G. Corti, M. Cattaneo, P. Charpentier, M. Frank, P. Koppenburg, P. Mato, F. Ranjard, S. Roiser, I. Belyaev, and G. Barrand, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.

[138] G. Barrand, I. Belyaev, P. Binko, M. Cattaneo, R. Chytracek, G. Corti, M. Frank, G. Gracia, J. Harvey, E. Van Herwijnen, B. Jost *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[139] M. Clemencic, G. Corti, S. Easo, C. R. Jones, S. Miglioranzi, M. Pappagallo, and P. Robbe (on behalf of the LHCb collaboration), *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[140] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[141] A. Ryd, D. Lange, N. Kuznetsova, S. Versille, M. Rotondo, D. P. Kirkby, F. K. Wuerthwein, and A. Ishikawa, *EvtGen: A Monte Carlo generator for B-physics*, EVTGEN-V00-11-07 (2005).

[142] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comput. Phys. Commun. **199** (2016) 86, arXiv:1011.0937.

[143] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[144] LHCb collaboration, *LHCb computing technical design report*, CERN-LHCC-2005-019, LHCb-TDR-11.

[145] M. Frank, C. Gaspar, E. van Herwijnen, B. Jost, N. Neufeld, and R. Schwemmer, *Optimization of the HLT resource consumption in the LHCb experiment*, J. Phys. Conf. Ser. **396** (2012) 012021.

[146] A. Bharucha, *Two-loop corrections to the $B \to \pi$ form factor from QCD sum rules on the light-cone and $|V_{ub}|$*, JHEP **05** (2012) 092, arXiv:1203.1359.

[147] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, LHCb-PUB-2011-016, CERN-LHCb-PUB-2011-016, LHCb-INT-2011-030.

[148] LHCb collaboration, *LHCb stripping lines*, http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/config/stripping34r0p1/index.html, 2022.

[149] G. Lanfranchi, X. Cid Vidal, S. Furcas, M. Gandelman, J. A. Hernando, J. H. Lopez, E. Polycarpo, and A. Sarti, *The muon identification procedure of the LHCb experiment for the first data*, LHCb-PUB-2009-013, CERN-LHCb-PUB-2009-013.

[150] E. Barberio, *Inclusive semileptonic B decays*, presentation at Flavor Physics and CP Violation Conference, Vancouver, 9–12 April (2006).

[151] T. J. Boettcher, A. Di Canto, M. Dorigo, B. E. Maurin, and D. Tonelli, *Measurement of the decay-width difference between the B_s^0 and B^0 mesons and the D_s and D mesons*, LHCb-ANA-2016-068, CERN-LHCb-ANA-2016-068.

[152] G. Ciezarek, B. Hamilton, A. Jawahery, M. Patel, and F. Redi (on behalf of the LHCb collaboration), *Measurement of B semileptonic decays $B^0 \to D^{*-}\tau^+\nu_\tau$ using the $\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau$ decay*, LHCb-ANA-2014-052, CERN-LHCb-ANA-2014-052.

[153] L. Anderlini, A. Contu, C. R. Jones, S. S. Malde, D. Muller, S. Ogilvy, G. Otalora, M. Juan, A. Pearce, I. Polyakov *et al.*, *PIDCalib package*, LHCb-PUB-2016-021, CERN-LHCb-PUB-2016-021.

[154] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, JINST **10** (2015) P02011, arXiv:1411.5528.

[155] P. Speckmayer, A Höcker1, J. Stelzer, and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, J. Phys. Conf. Ser. **219** (2010) 032057.

[156] A. Höcker, P. Speckmayer, J. Stelzer, F. Tegenfeldt, and H. Voss, *TMVA - Toolkit for multivariate data analysis*, arXiv:physics/0703039.

[157] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognit. Lett. **27** (2006) (861).

[158] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, 2016. http://www.deeplearningbook.org.

[159] C. Peterson and T. Rognvaldsson, *An introduction to artificial neural networks*, CERN School of Computing, LU-TP-91-23, LUTP-91-23 (1991).

[160] A. Nader and D. Azar, *Evolution of activation functions: an empirical investigation*, CoRR **abs/2105.14614** (2021), arXiv:2105.14614.

[161] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees,* Routledge, 1st ed, 1984, https://doi.org/10.1201/9781315139470.

[162] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[163] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin, *TensorFlow: Large-scale machine learning on heterogeneous distributed systems*, arXiv:1603.04467.

[164] F. Chollet *et al.*, *Keras*, https://keras.io.

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

Bibliography

[166] F. James, *MINUIT function minimization and error analysis: reference manual version 94.1*, CERNLIB-D506 (1994).

[167] J. E. Gaiser, *Charmonium spectroscopy from radiative decays of the J/ψ and ψ'*, SLAC-0255, UMI-83-14449-MC, SLAC-R-0255, SLAC-R-25 (1982).

[168] M. Pivk and F. R. Le Diberder, *SPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A **555** (2005) 356, arXiv:physics/0402083.

[169] G. Ciezarek, A. Lupato, M. Rotondo, and M. Vesterinen, *Reconstruction of semileptonically decaying beauty hadrons produced in high energy pp collisions*, JHEP **02** (2017) 021, arXiv:1611.08522.

[170] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, Wiley Series in Probability and Statistics, 2$^{\mathrm{rd}}$ ed., 2003. https://doi.org/10.1002/9780471722199.refs.

[171] I. Caprini, L. Lellouch, and M. Neubert, *Dispersive bounds on the shape of $\bar{B} \to D^{(*)}\ell\bar{\nu}_\mu$ form factors*, Nucl. Phys. B **530** (1998) 153, arXiv:hep-ph/9712417.

[172] J. L. Goity and W. Roberts, *Soft pion emission in semileptonic B meson decays*, Phys. Rev. D **51** (1995) 3459, arXiv:hep-ph/9406236.

[173] H.-Y. Cheng, C.-Y. Cheung, W. Dimm, G.-L. Lin, Y. C. Lin, T.-M. Yan, and H.-L. Yu, *Heavy quark and chiral symmetry predictions for semileptonic decays $\bar{B} \to D(D^*)\pi\ell\bar{\nu}_\mu$*, Phys. Rev. D **48** (1993) 3204.

[174] BaBar collaboration, B. Aubert *et al.*, *Measurement of the branching fractions of $\bar{B} \to D^{**}\ell^-\bar{\nu}_\mu$ decays in events tagged by a fully reconstructed B meson*, Phys. Rev. Lett. **101** (2008) 261802, arXiv:0808.0528.

[175] H. B. Prosper and L. Lyons, *Proceedings, PHYSTAT 2011 workshop on statistical issues related to discovery claims in search experiments and unfolding, CERN, Geneva, Switzerland, 17-20 January 2011*, CERN Yellow Reports, CERN-2011-006.

[176] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Meth. A **362** (1995) 487.

[177] G. D'Agostini, *Improved iterative Bayesian unfolding*, in *Alliance Workshop on unfolding and data correction*, 2010. arXiv:1010.0632.

[178] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016.

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

[180] A. Hocker and V. Kartvelishvili, *SVD approach to data unfolding*, Nucl. Instrum. Meth. A **372** (1996) 469, arXiv:hep-ph/9509307.

[181] LHCb collaboration, R. Aaij *et al.*, *Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B_s^0 decay branching fractions*, Phys. Rev. D **104** (2021) 032005, `arXiv:2103.06810`.

[182] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron fractions in 13 TeV pp collisions*, Phys. Rev. D **100** (2019) 031102, `arXiv:1902.06794`.

[183] M. Pervin, W. Roberts and S. Capstick, *Semileptonic decays of heavy lambda baryons in a quark model*, Phys. Rev. C **72** (2005) 035201, `arXiv:nucl-th/0503030`.

[184] C. A. Beleño de la Barrera, *Measurement of $B \to \pi\pi\ell\nu$ with full hadronic reconstruction at Belle*, PhD thesis, Gottingen U., 2018, http://dx.doi.org/10.53846/goediss-7172.

[185] Belle collaboration, C. Beleño, A. Frey *et al.*, *Measurement of the branching fraction of the decay $B^+ \to \pi^+\pi^-\ell^+\nu_\ell$ in fully reconstructed events at Belle*, Phys. Rev. D **103** (2021) 112001, `arXiv:2005.07766`.

[186] M. T. Prim, F. U. Bernlochner, and D. J. Robinson, *Precision predictions for $B \to \rho\tau\nu_\tau$ and $B \to \omega\tau\nu_\tau$ in the SM and beyond*, PoS **EPS-HEP2019** (2020) 250, `arXiv:2001.06170`.

[187] F. De Fazio and M. Neubert, *$B \to X_u\ell\bar{\nu}_\ell$ decay distributions to order α_s*, JHEP **06** (1999) 017, `arXiv:hep-ph/9905351`.

[188] H. B. O'Connell, B. C. Pearce, A. W. Thomas, and A. G. Williams, *$\rho - \omega$ mixing, vector meson dominance and the pion form-factor*, Prog. Part. Nucl. Phys. **39** (1997) 201, `arXiv:hep-ph/9501251`.

[189] CMD-2 collaboration, R. R. Akhmetshin *et al.*, *Measurement of $e^+e^- \to \pi^+\pi^-$ cross section with CMD-2 around ρ-meson*, Phys. Lett. B **527** (2002) 161, `arXiv:hep-ex/0112031`.

[190] BaBar collaboration, J. P. Lees *et al.*, *Precise measurement of the $e^+e^- \to \pi^+\pi^-(\gamma)$ cross section with the initial-state radiation method at BaBar*, Phys. Rev. D **86** (2012) 032013, `arXiv:1205.2228`.

[191] G. J. Gounaris and J. J. Sakurai, *Finite width corrections to the vector meson dominance prediction for $\rho \to e^+e^-$*, Phys. Rev. Lett. **21** (1968) 244.

[192] LHCb collaboration, R. Aaij *et al.*, *Amplitude analysis of the $B^+ \to \pi^+\pi^+\pi^-$ decay*, Phys. Rev. D **101** (2020) 012006, `arXiv:1909.05212`.

[193] J. M. Blatt and V. F. Weisskopf, *Theoretical nuclear physics*, Springer, New York, 1952. doi: 10.1007/978-1-4612-9959-2.

[194] F. Von Hippel and C. Quigg, *Centrifugal-barrier effects in resonance partial decay widths, shapes, and production amplitudes*, Phys. Rev. D **5** (1972) 624.

[195] B. O. Lange, M. Neubert, and G. Paz, *Theory of charmless inclusive B decays and the extraction of V_{ub}*, Phys. Rev. D **72** (2005) 073006, arXiv:hep-ph/0504071.

[196] R. Waldi, *Efficiency estimation and errors*, https://gitlab.cern.ch/lhcb/statistics-guidelines/-/blob/master/resources/rw1902.pdf.

[197] G. Cowan, *Statistical data analysis*, Claredon press, Oxford, 1998. https://inspirehep.net/literature/483505.

[198] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[199] F. Archilli, M. Kolpin, R. Kopecna, R. Skuza, *TrackCalib2*, https://gitlab.cern.ch/farchill/trackcalib2.

[200] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039, CERN-LHCb-PUB-2014-039.

Veronica Sølund Kirsebom
Physicist

Nationality:
Danish

Email address:
`veronica.kirsebom@`
`gmail.com`

Phone:
00 41 76 244 82 05

Address:
Rue Des Crêtes 16
1018 Lausanne
Switzerland

About Me I am currently finalising my PhD thesis in experimental particle physics. During the past years, I have analysed data collected by the LHCb experiment at CERN to obtain more precise measurements of fundamental properties of nature. As part of the PhD program, I have taught university students in different topics of physics, undertaken a managing role within the experiment, presented my work at collaboration meetings and developed my skills in data analysis, statistics and machine learning by following PhD courses and by participating in summer schools.

Education

2018–2023, EPFL, Switzerland
PhD in experimental high energy particle physics

2015–2018, Niels Bohr Institute, University of Copenhagen, Denmark
M.Sc. in Physics with a specialization in Quantum Physics

2015, University of Canterbury, New Zealand
Exchange Semester as part of my M.Sc. degree

2012–2015, Niels Bohr Institute, University of Copenhagen, Denmark
B.Sc. in Physics

2008–2011, Aarhus Katedralskole, Denmark
High school diploma with a specialization in Physics and Mathematics

Work experience

Laboratory of High Energy Physics, EPFL, 4/2018-4/2023
PhD student:

- I worked on an analysis that allowed for the first measurement of the decay process $B^+ \rightarrow \rho^0 \mu^+ \nu_\mu$ at the LHCb detector valuable for determining fundamental parameters of the Standard Model of Particle Physics.
- I was appointed the liaison of my analysis working group to the LHCb software development group.
- I supervised the lab work of 3rd year bachelor students.
- Teaching assistant in "Nuclear and particle physics".
- Teaching assistant in "Selected topics in nuclear and particle physics".

Niels Bohr Institute, Copenhagen University, 9/2016-1/2018
M.Sc. Thesis:

I developed a model, which simulated the resonant production of sterile neutrinos and feedback in the supernova environment.

CERN summer student program, 6/2017-8/2017
I participated in the CERN summer student program where I attended lectures in various topics within particle physics and worked in the LHCb experiment for nine weeks.

Center of Quantum Devices, Copenhagen University, 1/2015-6/2015
B.Sc. Thesis:

I designed and fabricated hybrid semiconductor/superconductor devices for the study of topological quantum computing.

ActionAid Denmark, Kundapur, India, 11/2011-1/2012
Volunteer:

I worked as a volunteer in a school for mentally handicapped children over a period of three months. During that time, I helped out with everyday activities in the school as well as developing different games and tools to make the teaching more fun and motivating for the students.

Selected talks

Swiss and Austrian Physical Society meeting, Innsbruck, Austria, 31/8/2021.
Title: "Measuring the decay $B \to \rho^0 \mu^+ \nu_\mu$ at LHCb"

LHCb collaboration week, CERN, 9/12/2021
Title: "Measuring $dBR(B \to \rho^0 \mu^+ \nu_\mu)/dq^2$ and extracting $|V_{ub}|$ at LHCb"

Programming and language skills

- Programming: Python, C++/C, bash.
- Data analysis : Tensorflow, Pandas, Jupyter Notebook, ROOT.
- Other: Git, LaTex.
- Languages: Danish (mother tongue), English (fluent) and French (intermediate).

Awards

2017, *Moritz Karbach Summer Student Prize*, CERN
The prize is awarded to outstanding LHCb summer students for the contribution, progress, motivation and clarity of their presentation.

Interests

Hiking
I have done plenty of hiking in the Alps including the sentier de grande randonnée 5 (GR5), tour du Mont Blanc, Alpenpässe-Weg and several others. I have also been hiking in Norway, Canada and New Zealand.

The small things
I am a passionate podcast listener, I like drawing and painting and I enjoy struggling with learning French and drinking beers with my friends.