université PARIS-SACLAY

Search for $B^+ \to K^+\tau^+\tau^-$ decays with Belle and Belle II experiments

Recherche de la désintégration $B^+ \to K^+\tau^+\tau^-$ dans les expériences Belle and Belle II

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576 : Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)
Spécialité de doctorat : physique des particules
Graduate School : Physique. Référent : Faculté des sciences d'Orsay

Thèse préparée dans la unité de recherche **IJCLab (Université Paris-Saclay, CNRS)**, sous la direction de **Karim TRABELSI**, directeur de recherche.

Thèse soutenue à Paris-Saclay, le 03 Octobre 2023, par

Vidya Sagar VOBBILISETTI

Composition du jury

Membres du jury avec voix délibérative

Patrick ROBBE Directeur de recherche, IJCLab, Université Paris-Saclay	Président
Olivier SCHNEIDER Professeur, Ecole Polytechnique Fédérale de Lausanne	Rapporteur & Examinateur
Kai-Feng CHEN Professeur, National Taiwan University	Rapporteur & Examinateur
James LIBBY Professeur, Indian Instititute of Technology Madras	Examinateur
Francesco POLCI Chargé de recherche, LPNHE, Université Pierre et Marie Curie	Examinateur

THESE DE DOCTORAT
NNT : 2020UPASA001

université PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)

Titre : Recherche de la désintégration $B^+ \to K^+\tau^+\tau^-$ dans les expériences Belle and Belle II
Mots clés : Universalité de la saveur leptonique meson B, nouvelle physique, Belle, Belle II, lepton tau

Résumé : Ces dernières années, des déviations par rapport au modèle standard (MS) ont été observées dans les désintégrations semi leptoniques des mésons B, notamment dans l'universalité de la saveur leptonique dans les transitions $b \to s\ell\ell$ et $b \to c\tau\nu$. Cela a rapidement suscité un intérêt considérable dans la communauté de la physique des saveurs : cet ensemble d'anomalies peut s'expliquer par des scénarios de la nouvelle physique (NP) qui favorisent la troisième génération de leptons par rapport aux deux premières, comme dans la transition $b \to s\tau\tau$, en particulier lorsque le couplage NP est proportionnel au carré de la masse du lepton. En raison des défis expérimentaux dans la reconstruction des leptons tau qui produisent des neutrinos indétectables, la seule limite supérieure existante, pour le mode de désintégration $B^+ \to K^+\tau^+\tau^-$ par l'expérience Babar, est très loin de la prédiction du MS et une vaste région reste à explorer. Dans ce travail, la recherche du mode $B^+ \to K^+\tau^+\tau^-$ est effectuée dans le contexte des expériences Belle et Belle II.

En raison de la présence de quatre neutrinos (énergie manquante), le signal $B^+ \to K^+\tau^+\tau^-$ ne peut pas être entièrement reconstruit. Dans une usine à B, les collisions e^+e^- produisent des $\Upsilon(4S)$ qui se désintègrent principalement en une paire de mésons B, avec un état initial bien défini. En reconstruisant entièrement l'un des mésons B, nous pouvons déduire les propriétés de l'autre méson B. Cette technique, appelée étiquetage du B, permet de supprimer efficacement les grandes contributions du bruit de fond dans les recherches impliquant l'énergie manquante. La reconstruction complète des mésons B dans les modes hadroniques est effectuée à l'aide d'un algorithme de Belle II, qui utilise un réseau d'arbres de décision boostés entraînés avec le Monte Carlo (MC). On observe un écart important dans les performances de l'étiquetage des B entre le MC et les données, ce qui constitue non seulement une source importante d'incertitude systématique, mais indique également un entraînement sous-optimal. Pour y remédier, une nouvelle procédure de contrôle, utilisant les désintégrations $B \to D\pi$ sans reconstruire explicitement les mésons D, est introduite pour calibrer l'étiquetage des B hadroniques. Cela permet d'obtenir une large statistique, tout en obtenant une grande pureté (contrairement à la procédure traditionnelle $B \to X\ell\nu$), et permet également d'étudier comment la modélisation de la désintégration du B affecte les performances de l'étiquetage du B. En appliquant des corrections aux modèles de désintégration, l'accord est considérablement amélioré. En outre, un nouvel entraînement de l'algorithme d'étiquetage des B à l'aide des simulations MC mises à jour permet d'améliorer le rejet du bruit de fond. Il est important de noter que ces efforts et d'autres améliorations du modèle de désintégration en mesurant les désintégrations hadroniques moins connues des mésons B conduiront à une meilleure performance de l'étiquetage hadronique du B, et ainsi profitera à toutes les recherches futures des modes avec énergie manquante.

Après avoir reconstruit le B étiqueté et les traces chargées du signal, les événements qui contiennent un $B^+ \to K^+\tau^+\tau^-$ ne laissent pas d'énergie supplémentaire dans le calorimètre électromagnétique ($E_{\rm ECL}$). La contamination par les photons du bruit de fond du faisceau et les amas neutres provenant des processus hadroniques dégradent la résolution du $E_{\rm ECL}$. Pour atténuer ce problème et optimiser la séparation avec le bruit de fond composé principalement de désintégrations B semi leptoniques, une étude du $E_{\rm ECL}$ est menée en utilisant l'échantillon de contrôle des désintégrations $B^+ \to J/\psi K^+$. Enfin, en l'absence de signal, une limite supérieure pour le rapport d'embranchement des désintégrations $B^+ \to K^+\tau^+\tau^-$ est estimée en utilisant les données combinées des simulations MC de Belle et Belle II.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)

Title : Search for $B^+ \to K^+\tau^+\tau^-$ with Belle and Belle II experiments

Keywords : Lepton flavour universality, B meson, new physics, Belle, Belle II, tau lepton

Abstract : In recent years, some deviations from the Standard Model (SM) have been observed in semileptonic B meson decays, such as departures from lepton flavour universality in the $b \to s\ell\ell$ and $b \to c\tau\nu$ transitions quantified by the $R_{K^{(*)}}$ and $R_{D^{(*)}}$ ratios. This generated a considerable interest in the flavour physics community : the pattern of anomalies could be explained by New Physics (NP) scenarios predicting large enhancements in the third generation of leptons compared to the first two, as in the $b \to s\tau\tau$ transition, especially when the NP coupling is proportional to the square of the lepton mass. The current limits for all processes mediated by $b \to s\tau\tau$ are far from the SM predictions by four orders of magnitude, due to the experimental challenges of reconstructing τ leptons that decay quickly and produce undetectable neutrinos. Consequently, the possibility of large NP-induced enhancements remains unchallenged. An upper limit for the $B^+ \to K^+\tau^+\tau^-$ decay mode has been set by the BaBar experiment, but it has never been searched for by any other experiment. In this work, the search for $B^+ \to K^+\tau^+\tau^-$ is performed in the context of the Belle and Belle II experiments.

Due to the presence of up to four neutrinos (missing energy), the signal $B^+ \to K^+\tau^+\tau^-$ cannot be fully reconstructed. At a B factory, e^+e^- collisions produce $\Upsilon(4S)$ which mostly decay into a pair of B mesons, with a well-known initial state. By fully reconstructing a charged B mesons, we can infer the four momentum of the other B meson, which may decay into $K^+\tau^+\tau^-$. This technique, called B tagging, effectively mitigates large background contributions in searches involving missing energy. The full reconstruction of hadronic B decays is done using Belle II's exclusive tagging algorithm, employing a network of boosted decision trees trained on Monte Carlo (MC). A significant discrepancy in B-tagging performance between MC and data has been observed, not only being a significant source of systematic uncertainty but also indicating suboptimal training. To address this, a new control procedure, using $B \to D\pi$ decays without explicitly reconstructing D mesons, is introduced to calibrate the hadronic B tagging. This provides large statistics, while having a high purity (in contrast to the traditional $B \to X\ell\nu$), also enabling the study of how B decay modelling affects B-tagging performance. By implementing the corresponding corrections to the decay model, the agreement is significantly improved. Furthermore, retraining the B-tagging algorithm with the updated MC simulations increases the background rejection. It is important to note that these efforts and further improvements to the decay model obtained by measuring poorly known hadronic B decays will lead to better performance of hadronic B tagging, benefiting all future searches involving missing energy.

After reconstructing the tagged B and the charged tracks from the signal side, events that contain a $B^+ \to K^+\tau^+\tau^-$ are expected to leave no extra energy in the electromagnetic calorimeter (E_{ECL}). Contamination from beam background photons and neutral clusters arising from hadronic split-off processes degrades the E_{ECL} resolution. On the other hand, the background mostly composed of semileptonic decays also has low E_{ECL}. To mitigate the resolution issue and optimise the separation from the background composed mostly of semileptonic B decays, a study of E_{ECL} is conducted using the control sample of $B^+ \to J/\psi K^+$ decays. Finally, in the absence of a signal, an upper limit for the branching ratio of the $B^+ \to K^+\tau^+\tau^-$ is derived using combined data sets of Belle and Belle II MC simulations.

Contents

Introduction

The Standard Model of particle physics (SM) has been confirmed by many experiments, but it is known to be an incomplete theory. Many New Physics (NP) models have been proposed to explain open questions such as the description of the gravitational interaction, the origin of the observed matter-antimatter asymmetry, and the existence of 'dark' (neither baryonic nor luminous) matter. Recently, significant deviations from the predictions of SM have been observed in semileptonic B meson decays, such as deviations from lepton flavour universality in the $b \to s\ell\ell$ and $b \to c\tau\nu$ transitions, quantified by the ratios $R_{K^{(*)}}$ and $R_{D^{(*)}}$. This attracted considerable interest in the flavour physics community: the pattern of anomalies could be explained by NP scenarios predicting large enhancements in the third generation of leptons compared to the first two, as in the $b \to s\tau\tau$ transition, in particular when the NP coupling is proportional to the square of the lepton mass. The current limits on all processes mediated by $b \to s\tau\tau$ are four orders of magnitude away from the predictions of SM, due to the experimental challenges in reconstructing τ leptons that decay rapidly and produce undetectable neutrinos. Consequently, the possibility of large NP-induced enhancements remains undisputed. An upper limit for the $B^+ \to K^+\tau^+\tau^-$ decay mode was established by the *BABAR* experiment [1], but this mode has never been searched for by any other experiment. The work presented here aims to search for $B^+ \to K^+\tau^+\tau^-$ decays using the combined data collected by the Belle and Belle II experiments, corresponding to a total of 1.18×10^9 $B\overline{B}$ pairs.

Due to the presence of up to four neutrinos (missing energy), the signal B candidate cannot be completely reconstructed. In such cases, the B-factories use the B-tagging technique to effectively mitigate large background contributions. In our search, hadronic B-tagging is performed using a network of boosted decision trees trained with Monte Carlo (MC). A significant discrepancy has been observed between the performance of MC and the data, which is not only a significant source of systematic uncertainty, but also an indication of suboptimal training. Therefore, considerable effort is being made to investigate the sources of these discrepancies, as any improvements in this area would benefit all future missing energy searches.

The manuscript is organized as follows. Chapter 1 briefly introduces the basic principles and nomenclature of SM, focusing on aspects related to flavour and B anomalies. Relevant previous measurements are summarized, along with some NP models predicting an enhanced branching fraction for $B^+ \to K^+\tau^+\tau^-$ decays.

Chapter 2 describes the (Super)KEKB machines and the corresponding experiments Belle (II), where the data sets used are collected. Key elements about the performance and upgrades are presented. Special attention is given to the Belle II High Level Trigger, as the IJCLab group has put effort into its operation, monitoring and optimization since the early stages of data collection.

Chapter 3 presents the package used to perform hadronic B-tagging in Belle (II) experiments. Calibration procedures using two independent samples and the appropriate combination that is currently the official method are described. Chapter 4 describes the extensive B decay modelling study that has led to improved hadronic B-tagging performance.

In Chapter 5, we report the sensitivity of the search for $B^+ \to K^+\tau^+\tau^-$ decays using hadronic B-tagging in the combined Belle+Belle II data samples corresponding to $1073\,\text{fb}^{-1}$. The analysis is performed on MC, and an upper limit on the branching fraction is derived in the absence of a signal.

Finally, Chapter 6 summarizes the results and discusses some ways to improve hadronic B-tagging and achieve a higher sensitivity for $B^+ \to K^+\tau^+\tau^-$ search.

Chapter 1

Flavour in particle physics

In the Standard Model (SM) there are four different types of fundamental fermions, each in three generations. The term flavour denotes the species of these fermions, each possessing a specific set of quantum numbers[1]. Flavour physics explores how quarks and leptons can change their flavour through the weak force, one of the four fundamental forces of nature in SM. This led to the predictions of the charm quark and the third generation of fermions before they were discovered. Also, the measurement of neutrino flavour transitions, which showed that neutrinos have mass [2].

In recent years, some deviations from the SM predictions have been observed, especially in the semileptonic decays of B mesons, which are summarized in Section 1.2. New Physics (NP) models explaining these anomalies predict an enhanced branching fraction for the decay $B^+ \to K^+\tau^+\tau^-$, which is described in Section 1.3 along with the current experimental status. Observation of this enhancement would indicate the presence of processes beyond SM (BSM).

1.1 The Standard Model

Decades of measurements, predictions and discoveries have led to the Standard Model of particle physics, which coherently describes the interactions of elementary particles through the weak, strong and electromagnetic forces. It is a relativistic quantum field theory with three categories of particles or quantum fields: force carriers called gauge bosons, matter constituents called fermions, and scalar particles.

Gauge bosons are excitations of the field that correspond to spin-1 particles and are responsible for a specific type of interaction: the gluons (g) mediating strong interactions, the photon (γ) for the electromagnetic ones, and the Z and W, particles mediating weak interactions according to the gauge invariance under the group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$[2]. The Standard Model unifies the electroweak interactions, which have the symmetry group $SU(2)_L \otimes U(1)_Y$ with associated coupling constants, and quantum chromodynamics (QCD), characterized by the quantum number colour associated with the group $SU(3)_c$, whose coupling constants are kept independent [3].

The fermions describe excitations corresponding to spin-1/2 particles and are organized into three families or generations. Each family contains four types of particles, two quarks and two leptons, with different quantum numbers as shown in Fig. 1.1. Ordinary matter consists essentially of particles of the first family. The particles of the second and third families appear to be identical copies, except for their heavier masses, which rapidly decay into lighter particles through weak interactions. The masses of the quarks and leptons are the result of a Yukawa interaction that connects the fermion fields to the Higgs field, the only scalar field in SM. The existence of three

[1] electric charge, weak isospin, baryon and lepton number.

[2] Where c denotes the colour, L the chiral (weak isospin) component and Y the hypercharge.

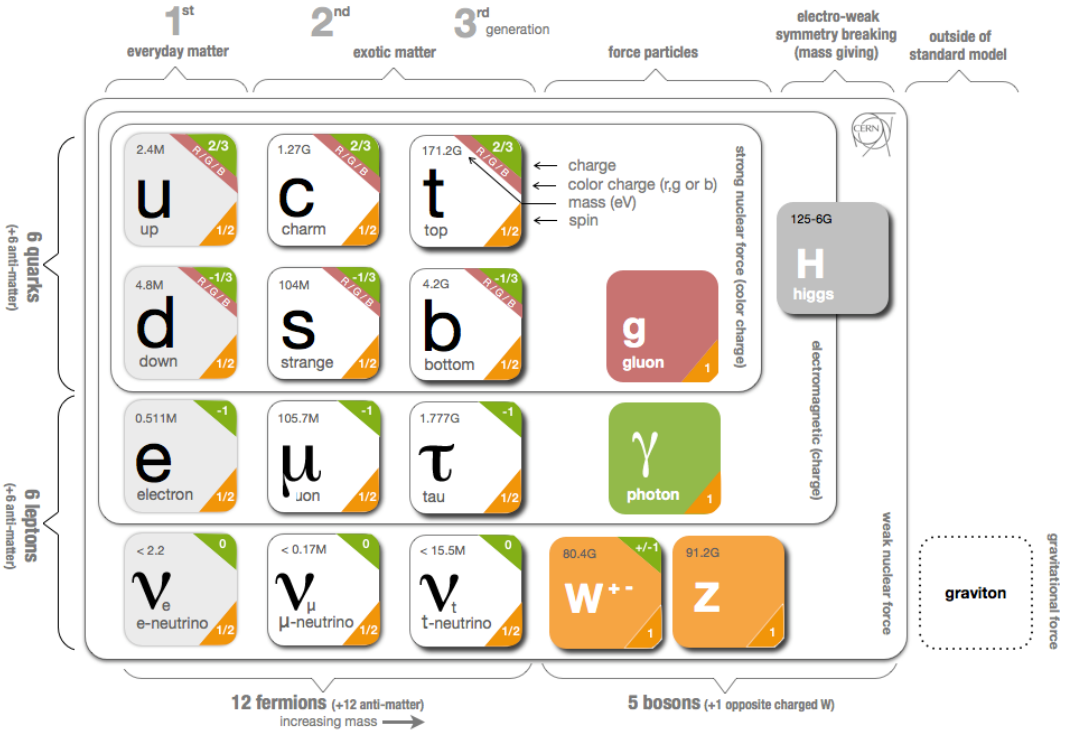


FIGURE 1.1: Infographic of the fundamental particles in the Standard Model (SM). Credits: CERN.

replicas of quarks and leptons and the origin of their different masses is one of the most important open questions in physics. Quarks carry colour, weak, and electromagnetic charges, so they interact with each other through all three forces. Leptons participate in the weak interactions, and additionally charged leptons participate in the electromagnetic processes.

The fermion fields of the SM can be written in terms of Weyl spinors in a $(c, L)_Y$ notation. For example, the left-handed quark doublets can be represented by $Q_i(3,2)_{1/6}$ where $i = 1, 2, 3$ denotes the generation index. It is a triplet under $SU(3)$ colour group and a doublet under $SU(2)_L$ weak isospin doublet, whose hyper-charge is $1/3$. The weak hyper-charge is defined by the Gell-Mann Nishijima formula as $Y = 2 \cdot (Q - T_3)$, where Q is the electric charge and T_3 is the third component of the weak isospin. The right-handed quarks, on the other hand, form singlets $U_i(3,1)_{2/3}$ and $D_i(3,1)_{-1/3}$, which correspond to the up-type and down-type, respectively.

1.1.1 The electroweak interaction

The weak charged bosons, $W^{\pm}$, couple only to fermions of left chirality. The Glashow-Weinberg-Salam (GWS) [4, 5] theory describes the electroweak interactions as spontaneously broken symmetry groups

$$SU(2)_L \otimes U(1)_Y \overset{SSB}{\rightarrow} U(1)_{EM}.$$

This spontaneous symmetry breaking (SSB) is iniated by the Higgs mechanism and causes the four primordial massless gauge bosons to mix to form the massive weak mediators $W^{\pm}$, Z^0 and the massless γ and is responsible for the masses of the fermions [6].

In the quark sector, the $W^{\pm}$ bosons couple to different flavours, which also allows for coupling across generations as the mass eigenstates differ from the weak eigenstates.

The relation between them is described by a rotation with the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{weak}} = \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{mass}} ,$$

where the flavour/weak eigenstates of down-type quarks are donated by q', a superposition of the mass eigenstates (q). This principle was first introduced by Cabibbo [7] for two generations of quarks to explain the universality of the weak interaction. The elements V_{ij} describe the coupling strength of the charged current between the quarks of flavours i and j. The complex-conjugated element V_{ij}^* accounts for the coupling of the corresponding Charge Parity(CP)-conjugated processes. In the SM, the complex phase leads to $V_{ij} \neq V_{ij}^*$ and provides the only known mechanism that leads to CP violation. Kobayashi and Maskawa showed that while it is possible to have a fully real 2 x 2 matrix by a suitable definition of the quark phases, this is not possible any more for a 3 x 3 matrix [8]. This led to a prediction of the third generation of quarks, long before the discovery of the b and t quarks[3].

The CKM matrix is also unitary (i.e., $V_{CKM}^{-1} = V_{CKM}^{\dagger}$) which gives the individual unitary conditions:

$$\sum_i V_{ij} V_{ik}^* = \delta_{jk} \quad \text{and} \quad \sum_j V_{ij} V_{ik}^* = \delta_{jk}.$$

These relations can be visualized as triangles in the complex plane. Out of the six, a particularly interesting one is

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = \delta_{db} = 0$$

because they can be extracted from measurements of B decays. To form the so-called Unitary Triangle, it can be dividing by the experimentally best known term $V_{cd} V_{cb}^*$:

$$1 + \frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} + \frac{V_{td} V_{tb}^*}{V_{cd} V_{cb}^*} = \delta_{db} = 0.$$

Measuring the parameters of this triangle is one of the primary goals of flavour physics.

These conditions leave four free parameters to describe the entire V_{CKM}: three mixing angles and a complex phase in the standard parametrization. The latest measurements of these elements correspond to [9]:

$$V_{\text{CKM}} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.0096^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix} .$$

The diagonal elements correspond to the flavour transition within the same generation (the strength is close to one). But the off-diagonal elements are small, for example the quark transition $b \to c$ corresponds to $|V_{cb}| \sim 4 \times 10^{-2}$. This suppression by the CKM mechanism leads to the relatively long lifetime of B mesons.

In the lepton sector, the coupling strength to the $W^{\pm}$ bosons is identical for all the lepton doublets. This property is called lepton-flavour universality (LFU). Coupling

[3] In 2008, Makoto Kobayashi and Toshihide Maskawa were awarded the Nobel Prize in Physics, where the Belle and BaBar experiments were explicitly mentioned for confirming their predictions.

across generations is not possible with leptons[4]. Dominant semileptonic decays of B meson would be from the $b \to c\ell\overline{\nu}_\ell$ transition, as represented in Fig. 1.2.

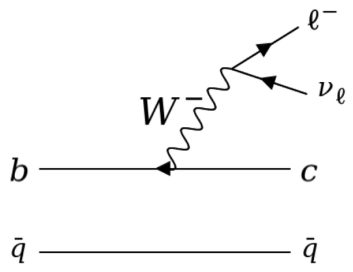


FIGURE 1.2: Feynman diagram representing dominant semileptonic transitions of B mesons. Here, $\bar{q}$ stands for the spectator quark.

1.1.2 Flavour Changing Neutral Currents

In the SM, unlike the charged weak interactions, the neutral weak interactions mediated by the Z boson cannot couple across quark generations. Therefore, transitions such as $b \to s$ or $s \to d$ are forbidden at the tree-level. These Flavour Changing Neutral Currents (FCNC) can only occur via higher order penguin or box diagrams, as shown in Fig. 1.3. Such loops are highly suppressed in SM, since the number of

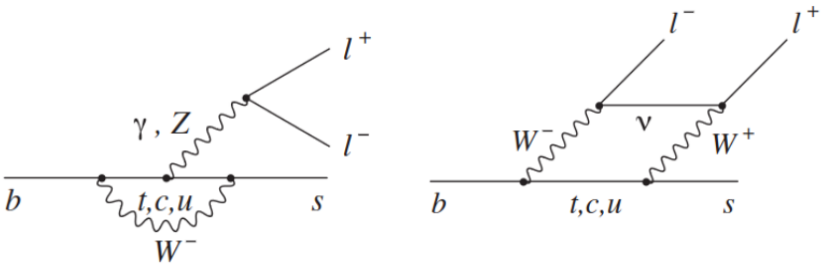


FIGURE 1.3: Lowest-order Feynman diagrams of the $b \to s$ transition, producing a pair of charged leptons. The diagrams are of the penguin/loop type (left), or of the box type (right).

electroweak vertices is larger and heavier virtual particles are involved. However, NP particles can be present in the loops, enhancing the amplitude significantly, making them particularly interesting. In this sense, probing the SM by studying the rare decays (like $b \to s\ell^+\ell^-$ transitions), can be seen as complementary to the searching for heavier new physics particles directly among the collision products.

1.2 Recent summary of B anomalies

The LFU is an accidental symmetry of the SM, which states that the three generations of charged leptons have the same coupling with the W and Z bosons, except for their different masses. Tests of LFU in the last decade, in particular at the $b \to c\ell\overline{\nu}_\ell$ and $b \to s\ell\ell$ transitions by the LHCb, Belle (II), and BaBar experiments, show some deviations from the SM predictions, commonly referred to as 'B-anomalies'. These

[4]Except that neutrinos can oscillate in flavour, described by the Pontecorvo–Maki–Nakagawa–Sakata matrix [10]

tests usually measure the ratios of branching fractions corresponding to the different leptons in the final state.

In the $b \to c\ell\overline{\nu}_\ell$ (charged current) transitions, the discrepancies are found in the ratios

$$R^{\tau/\ell}_{D^{(*)}} = \frac{\Gamma\left(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau\right)}{\Gamma\left(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell\right)} \qquad (\ell = e, \mu)\,, \tag{1.1}$$

The latest measurements come from various experiments, such as BaBar, Belle, Belle II, and LHCb using different reconstruction techniques. The current world averages of these ratios (as summarized in Fig. 1.4) are:

- $R_D = 0.357 \pm 0.029$
- $R_{D^*} = 0.284 \pm 0.012$

which exceed the SM predictions by 2.0σ and 2.2σ respectively. Considering the correlation of -0.37 between R_D-R_{D^*}, the combined fit differs with the SM predictions at about 3.3σ.

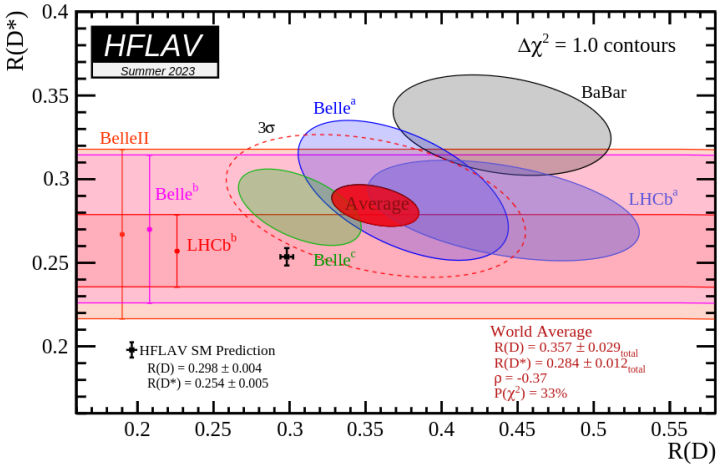


FIGURE 1.4: LFU tests of R_D-R_{D^*} compared with the SM predictions. Taken from HFLAV [11].

A similar 2σ deviation is observed [12] in

$$R^{\tau/\mu}_{J/\psi} = \frac{\Gamma\left(\bar{B}_c \to J/\psi\tau^-\bar{\nu}_\tau\right)}{\Gamma\left(\bar{B}_c \to J/\psi\ell^-\bar{\nu}_\ell\right)} \qquad (\ell = e, \mu).$$

These discrepancies point towards an enhancement for $b \to c\tau^-\bar{\nu}_\tau$ via BSM mediator. This could be due to NP contributions like a tree-level exchange of a W' boson (the partner of Z' in the vector-boson model), a leptoquark or a charged Higgs. In which case, large tree-level contribution is needed corresponding to $\mathcal{O}(10\%)$ at amplitude level. Or it can just be a statistical fluctuation, or a systematic effect that is not well understood. More data and analyses are needed to explore its possible implications for physics beyond the SM.

The neutral currents $b \to s\ell\ell$ transitions are studied via the decays $B \to H\ell\ell$, where H is a hadron with a strange quark. The individual branching fraction of those decays are difficult to compute because of the strong forces among the quarks composing the hadrons, but are identical between different ℓ species. This makes the

prediction on the ratio

$$R_H^{e/\mu} = \frac{\Gamma\left(B \to He^+e^-\right)}{\Gamma\left(B \to H\mu^+\mu^-\right)}\Big|_{q^2\in(q^2_{\min},q^2_{\max})} \tag{1.2}$$

very precise and allows to directly test the LFU. In fact, the ratio is predicted to be $1\pm\mathcal{O}(1\%)$ [13], the uncertainty being due to electromagnetic corrections and effects related to the mass of the e, μ. In these transitions, there are contributions from the $B \to X_{q\bar{q}}K$ resonant states, for example the $B \to KJ/\psi(\to \ell^+\ell^-)$ cross section is much larger because of the favoured $b \to c\bar{c}s$ transition. To exclude this, the $q^2 = m^2_{\ell\ell}$ is chosen in certain ranges: $0.1 < q^2 < 1.1$ GeV$^2/c^4$ called low-q^2 and $1.1 < q^2 < 6$ GeV$^2/c^4$ called central-q^2. While large deviations (around 3σ) were reported by the LHCb experiment in recent past [14–16], the latest measurements [17, 18] with better understanding of the background show that R_K and R_{K^*} have a good overall agreement with SM at the level of 0.2σ (see Fig. 1.5). Belle [19, 20] and BaBar

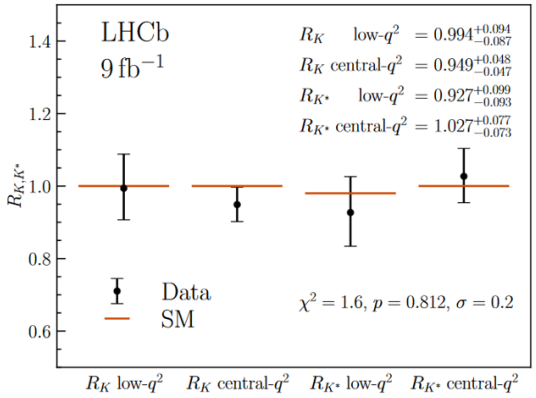


FIGURE 1.5: Comparison of the latest R_K and R_{K^*} measurements from LHCb with the SM prediction is different regions of q^2.

have also reported measurements of R_H compatible with the SM, though with larger statistical uncertainties than LHCb. Some discrepancies are still seen in the angular distributions of $b \to s\ell\ell$ decays, including the P_5' variable [21–24].

Given the hints of enhancement for the third generation leptons (τ) in $b \to c\ell\overline{\nu}_\ell$ transitions, it seems natural to wonder if such an effect can be seen in neutral channel $b \to s\ell\ell$ transition also, which will be the focus from here on.

1.3 Motivation for searching $B^+ \to K^+\tau^+\tau^-$ decays

For the $b \to s\ell\ell$ transitions, replacing one of the ℓ by τ would lead to decays such as $B \to K^{(*)}\tau\ell$. This violates lepton number conservation and is therefore not allowed in SM. Its observation would be a direct proof of new physics, and is studied extensively in the Belle [25], BaBar [26] and LHCb [27, 28] experiments.

Replacing the two ℓ with τ leptons would lead to $B \to K^{(*)}\tau\tau$ decays. Since this is an FCNC decay, this occurs through the penguin diagram or the box diagram at the lowest order (as shown in Fig. 1.3), which are strongly suppressed. However, the decay amplitude would be enhanced in models of new physics models that have larger coupling with particles of third generation (two τ's in this case) or with heavier mass ($m_\tau \sim 20m_\mu$), as described in Sec. 1.3.1. The presence of up to four neutrinos in the final state makes the experimental search particularly challenging, so that only limited measurements are available for such decays,which are summarized in Sec. 1.3.2.

1.3.1 Theoretical predictions

The one-loop $b \to s$ transition can be described by the effective Hamiltonian using the operator product expansion as[5]:

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_{i=1} \mathcal{C}_i(\mu)\mathcal{O}_i(\mu) + h.c.,$$

where $V_{qq'}$ is the corresponding CKM matrix element. The local operators represent effective point-like vertices, in particular: $\mathcal{O}_1$ and $\mathcal{O}_2$ are current-current operators, $\mathcal{O}_{3-6}$ are QCD penguin operators, $\mathcal{O}_7$ and $\mathcal{O}_8$ are electromagnetic and chromomagnetic operators, $\mathcal{O}_{(10)9}$ are the (axial)vector components of the electroweak penguin operator, and $\mathcal{O}_{(P)S}$ are the (pseudo)scalar operators respectively. The $\mathcal{C}_i(\mu)$ functions are called Wilson coefficients (WCs) and are scale-dependent couplings to the corresponding operators (μ denotes the renormalization scale), describing short-distance physics.

For the $b \to s\ell^+\ell^-$ transition, only the operators[6]

$$\mathcal{O}_7^{(\prime)} = \frac{e}{(4\pi)^2} m_b(\bar{s}\sigma_{\mu\nu}P_{R/L}b)F^{\mu\nu}$$

$$\mathcal{O}_9^{(\prime)} = \frac{\alpha}{4\pi}(\bar{s}\gamma_\mu P_{L/R}b)(\bar{\ell}\gamma^\mu\ell), \qquad \mathcal{O}_{10}^{(\prime)} = \frac{\alpha}{4\pi}(\bar{s}\gamma_\mu P_{L/R}b)(\bar{\ell}\gamma^\mu\gamma^5\ell),$$
$$\mathcal{O}_S^{(\prime)} = \frac{\alpha}{4\pi}(\bar{s}P_{R/L}b)(\bar{\ell}\ell), \qquad \mathcal{O}_P^{(\prime)} = \frac{\alpha}{4\pi}(\bar{s}P_{R/L}b)(\bar{\ell}\gamma^5\ell),$$

contribute in leading order (visualized in Fig. 1.6). Here, s, b and ℓ are the strange quark, bottom quark, and lepton fields respectively, and $F_{\mu\nu}$ is the electromagnetic field tensor. For the decay rates of interest, $B \to K\tau^+\tau^-$, the ℓ in the effective

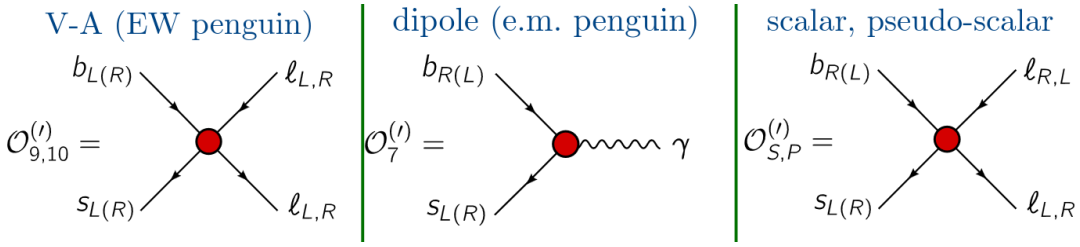


FIGURE 1.6: Feynman diagrams for the underlying process of the effective operators contributing to the $b \to s\ell\ell$ transitions in the leading order. Source: LHCb presentation.

Hamiltonian can be replaced by τ and the corresponding branching fraction can be calculated.

Within the SM, the most precise prediction [29] using lattice QCD form factors is

$$\mathcal{B}_{SM}(B \to K\tau^+\tau^-) = (1.44 \pm 0.15) \times 10^{-7}.$$

Here, the branching fraction is calculated by integrating in the $q^2 > 14.18\ \text{GeV}^2/c^4$ region, to exclude contributions from $B \to K\psi(2S)[\to \tau^+\tau^-]$ (which has an effective branching fraction of 1.8×10^{-6}), as shown in Fig. 1.7. The error contributions from

[5] G_F is the Fermi constant ($1.166\,378\,7 \times 10^{-5}\,\text{GeV}^{-2}$). And h.c. stands for Hermitian conjugate

[6] $\alpha = e^2/4\pi$ and $P_{L,R}$ are the left-handed (LH) and right-handed (RH) projection operators. m_b is the running b-quark mass.

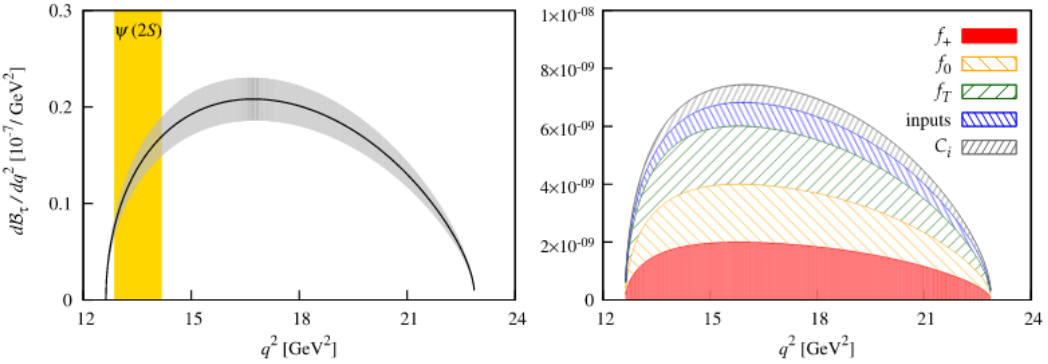


FIGURE 1.7: Left: Predicted SM $d\mathcal{B}/dq^2$ for $B \to K\tau^+\tau^-$ decays. Right: Error contributions for SM $d\mathcal{B}/dq^2$ from form factors, input parameters, and Wilson coefficients ($\mathcal{C}_i$). Taken from Ref. [29].

form factors, input parameters, and Wilson coefficients ($\mathcal{C}_i$) are also shown in the figure, and it can be seen that the uncertainty in the form factors dominates.

The effect of NP can be included here with modified Wilson coefficients, $\mathcal{C}_i \to \mathcal{C}_i^{SM} + \Delta_i^{NP}$. For the relevant operators, a fit is performed to all the measurements that together form the B anomalies to find the best values of the Wilson coefficients. For a given BSM hypothesis, several (constrained) Wilson coefficients may be allowed to vary simultaneously in such a fit to explain the anomalies.

One particular BSM hypothesis, the leptoquark U_1 (represented by $(3,1)_{2/3}$ in Weyl notation) model, is currently the preferred one [30], as it can explain both sets of anomalies[7], the low-energy constraints and the direct searches at the LHC. Leptoquarks (LQs) are coloured states [32] that can mediate interactions between quarks and leptons. In general, a LQ can be a scalar or a vector field, denoted by its quantum number with respect to the SM gauge group $(SU(3)_c, SU(2)_L)_Y$, where the electric charge can take the values $Q = \pm 5/3, \pm 4/3, \pm 2/3, \pm 1/3$. The LQs enter at the loop level in interactions where no deviation from the SM has been found (into four quarks or four leptons), while they contribute to the semileptonic transitions at the tree level, where they show significant anomalies. The fact that they have not yet been directly observed is compatible with the predicted mass of $\sim\mathcal{O}(\text{TeV})$.

To analyse the compatibility of the U_1 LQ model with the recent charged current B anomalies, the effective Hamiltonian is first transformed into the notation introduced in Ref. [30] and the two Wilson coefficients ($\mathcal{C}^c_{LL}$ and $\mathcal{C}^c_{LR}$) are fitted. Three sets of measurements are used:

- $b \to c\tau\overline{\nu}_\tau$: the LFU ratios R_D, R_{D^*} and R_{Λ_c} [33];
- $b \to u\tau\overline{\nu}_\tau$: $\mathcal{B}(B^- \to \tau^-\overline{\nu}_\tau)$ [34, 35];
- Drell-Yan production $pp \to \tau^+\tau^-$ with the main contribution due to $b\bar{b} \to \tau^+\tau^-$ [36, 37].

The first two are collectively referred to as low-energy observables and the latter as high-energy observables. The most recent fit [38] to low-energy observables is shown in Fig. 1.8. Two scenarios of U_1 LQs are considered: a purely left-handed interaction ($\mathcal{C}^c_{LR} = 0$), and a scenario with right-handed currents of equal magnitude ($\mathcal{C}^c_{LL}$ and $\mathcal{C}^c_{LR}$). Both scenarios are found to be equally compatible (with the pull at the 3σ level) with the $b \to c$ observables as seen in Fig. 1.8. Examination of the model on

[7] Due to the absence of a tree-level constraint from $\mathcal{B}(B \to K\nu\bar{\nu})$ [31].

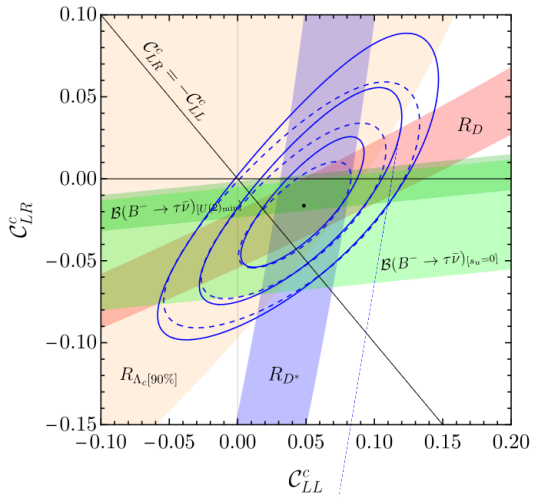


FIGURE 1.8: 2D fits of the Wilson coefficients to the low-energy 2023 observables. The blue ellipses denote the 1, 2 and 3σ contours, which fit only $b \to c$ observables. The black dot indicates the best fit point of $(0.05, -0.02)$. The dotted lines are obtained including the $\mathcal{B}(B^- \to \tau^-\overline{\nu}_\tau)$. Taken from Ref. [38].

the $pp \to \tau^+\tau^-$ constraints shows that low and high energy data are currently well compatible.

Using the preferred parameter space from the low-energy fit for the U_1 model, the predictions for $\mathcal{B}_{NP}(B^+ \to K^+\tau^+\tau^-)$ are also reported in Ref. [38]. Fig. 1.9 shows this prediction in both scenarios of the U_1 LQ model as a function of the relative deviation in R_{D^*} from the SM prediction. Compared to the SM prediction, which is

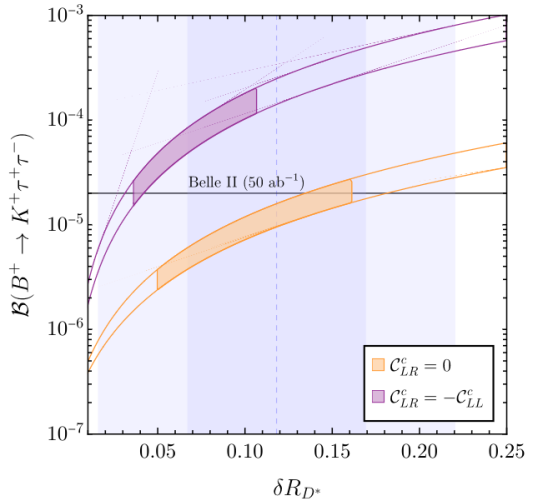


FIGURE 1.9: Predicted $\mathcal{B}_{NP}(B^+ \to K^+\tau^+\tau^-)$ by the U_1 model, as a function of $\delta R_{D^*} = R_{D^*}/R^{SM}_{D^*} - 1$. The filled orange and purple regions correspond to the 90% CL preferred regions from the low-energy charged current fit. The blue vertical bands denote the current experimental 1σ and 2σ regions for δR_{D^*}. Taken from [38].

in $\mathcal{O}(10^{-7})$, the NP predictions are about two orders of magnitude larger, allowing to probe the presence of the LQs indirectly. The black line in the figure, showing the estimated sensitivity of the Belle II experiment with 50 ab^{-1}, is based on Ref. [39].

In summary, extensions to the SM (such as U_1 LQs) that explain B anomalies point to the presence of new couplings that favour τ leptons and increase the decay

rates of the $b \to s\tau^+\tau^-$ transitions. However, the experimental search for these modes is challenging, as discussed in the next section.

1.3.2 Experimental status

The τ leptons decay rapidly and produce neutrinos, making their detection/reconstruction a challenge. Decays containing a τ (or a ν in general) in the final state cannot be fully reconstructed and are referred to as missing energy modes. To study such decays, some kind of tagging is required. When there is only one missing particle in the decay, hadronic B-tagging in the B factories allows the signal yield to be obtained from the recoil mass distribution, which peaks at the mass of the missing particle, as described in Sec. 3.2.2. This technique was used by the Belle experiment to obtain the most stringent upper limits for $B^+ \to K^+\tau\ell$ [25], and an earlier upper limit by the BaBar experiment [26]. A different strategy was used at LHCb [27] to study the same mode. The B^+ candidates are tagged by searching for $B_{s2}^{*0} \to B^+K^-$ decays, which allows to calculate the direction of the B^+ and its energy up to a quadratic ambiguity and provides the missing mass.

The presence of two τ's in $B \to K\tau^+\tau^-$ means that the recoil/missing mass method is not applicable. Instead, an alternative observable is required. Moreover, the large background from semileptonic B decays and the low momentum of the final-state charged tracks make the search for these modes challenging. For this reason, only two measurements have been performed so far.

In 2016, the BaBar experiment searched for $B^+ \to K^+\tau^+\tau^-$ using its full data set of 471×10^6 $B\overline{B}$ pairs [1]. One of the two B mesons is reconstructed exclusively by hadronic B-tagging, and the remaining tracks, clusters, and missing energy are assigned to the B meson on the signal side. Only leptonic τ decays are considered, resulting in three signal topologies with a $K^\pm$ and either e^+e^-, $\mu^+\mu^-$, or $e^\pm\mu^\mp$ in the final state. A neural network with seven input variables is trained to suppress the background from semileptonic decays and extract the signal yield. An important variable used here is the extra energy in the calorimeter, defined as the energy sum of all neutral clusters associated with neither the tag B nor the signal B candidates. The $B^+ \to K^+\tau^+\tau^-$ events would peak at zero in this variable. By calculating the signal efficiency and the expected background yield in the defined signal region in the neural network output, the branching fraction can be calculated. No signal was observed, and an upper limit of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 2.25 \times 10^{-3}$ was set at 90% confidence level is set. While the yields in the e^+e^- and $\mu^+\mu^-$ channels are consistent with the expected background estimate, an excess of 3.7σ over the background expectation is reported for the yield in the $e^\pm\mu^\mp$ channel. Unfortunately, there are no other measurements of this decay to confirm these results. Using a much larger data set from the Belle and Belle II experiments, totaling 1.18×10^9 $B\overline{B}$ pairs, we aim to search for this decay.

Recently, the Belle experiment reported $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 3.1 \times 10^{-3}$ at 90% confidence level using a data set of 771×10^6 $B\overline{B}$ pairs [40]. This uses an older and less efficient hadronic B-tagging package called Full Reconstruction (see Chapter 3), which makes a fit directly to the extra energy in the calorimeter information to obtain signal and background yields. The latest hadronic B-tagging package, which is further improved in this thesis, is used for our search and extracts the signal from the same observable.

Chapter 2

The Belle and Belle II experiments

In this chapter, we describe the main features of the SuperKEKB accelerator and the Belle II experiment, which started taking data in March 2019. It is the successor to the Belle experiment at the KEKB accelerator, which operated from December 1998 to June 2010, with the goal of collecting a 50 times larger data set to make a significant contribution to flavour physics. Some components of the detector have remained unchanged, while other components have been significantly improved to withstand a harsher background due to increased instantaneous luminosity.

For the analyses discussed in the next chapters, data from both experiments will be used, not only to take advantage of the largest possible statistics, but also to allow independent comparisons. The key elements of both experiments are outlined, with improvements highlighted, in preparation for a discussion of the prospects of our analysis.

2.1 The KEKB and SuperKEKB accelerators

KEKB [41] and SuperKEKB [42] accelerators are e^+e^- colliders located at KEK, Tsukuba, Japan. The SuperKEKB consists of two rings: a low-energy ring (LER) for the 4 GeV-positrons, and a high-energy ring (HER) for the 7 GeV-electrons (compared to 3.5 GeV and 8 GeV at KEKB respectively), and an injector linear accelerator (Linac), as sketched in Fig. 2.1. The two beams collide at the interaction point (IP) around which the Belle (II) detectors are installed.

On the 15th June 2020, SuperKEKB reached an instantaneous luminosity of $2.22 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ setting a record for the highest luminosity at a particle collider, which was held by KEKB for almost ten years since June 2009. Since then, the luminosity kept increasing until the latest record on 22 June 2022: $4.71 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The target luminosity for SuperKEKB is $6 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which will be 30 times larger than the peak luminosity at KEKB [43].

The luminosity of a collider is given by:

$$L = \frac{\gamma_\pm}{2er_e}\left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)\left(\frac{I_\pm \xi_{y\pm}}{\beta_y^*}\right)\left(\frac{R_L}{R_{\xi_{y\pm}}}\right), \qquad \xi_{y\pm} \propto \frac{N_\mp \beta_y^*}{\sigma_x^* \sigma_y^*} \tag{2.1}$$

where the subscript $(+)$ is for positrons and $(-)$ for electrons and r_e, e and γ are respectively the classical electron radius, its charge and the Lorentz factor. The main parameters entering the expression are briefly described as follows (more details are given in Ref. [43]). $\sigma^*_{x(y)}$ denote the beam sizes at the IP in the horizontal (vertical) plane, and I the total beam current. The vertical beam-beam parameter (ξ_y) describes the focusing force exerted on a bunch by the electromagnetic field of the opposite bunch and depends on $N_\mp$, the number of particles (e^-/e^+) in a bunch. β_y^* is the vertical beta function at the IP; $\beta(s)$ measuring the beam cross-section, which depends

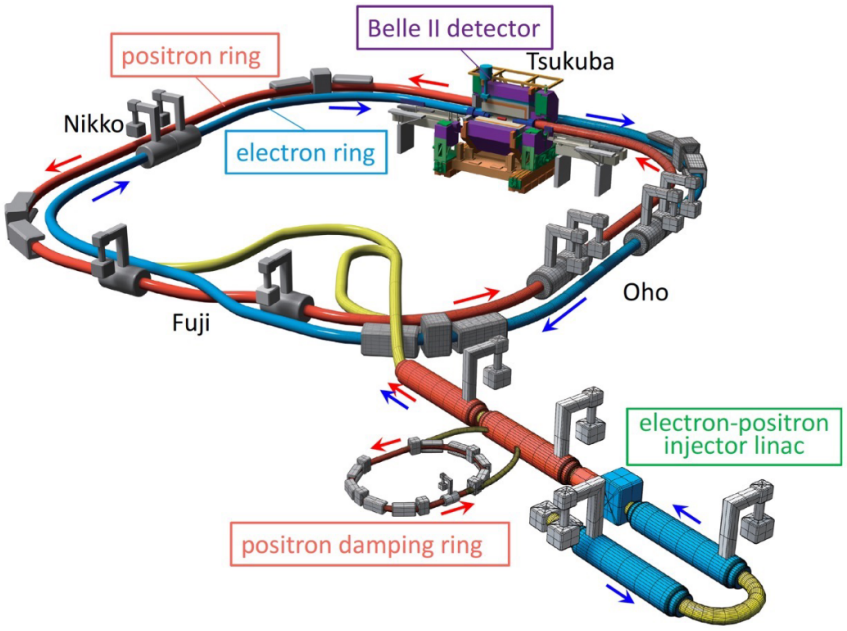


FIGURE 2.1: Pictorial representation of the SuperKEKB facility complex, mainly composed of a Linac and the two main storage rings.

on the beam focusing and varies with the position s along the ring due to the betatron oscillations [44]. Finally, R_L and $R_{\xi_{y\pm}}$ are respectively the geometrical reduction factor for the luminosity and the beam-beam parameter.

From Eq. 2.1 one can see that the luminosity is directly proportional to the beam-beam parameter ξ_y and the beam current I, and inversely to the vertical β function at the IP. The luminosity can be increased by rising the beam currents and lowering the beam sizes. At SuperKEKB, the design LER/HER currents are 3.6/2.6 A (compared to 1.6/1.2 A for KEKB), while the vertical size of the bunch at IP, σ_y^*, is reduced by a factor 20 – from 1.1 μm to $\sim$ 60 nm (*nano-beam* scheme). However, these manipulations lead to instabilities due to the increase of the non-linear beam-beam interactions. Besides, the transverse beam size varies along the bunch length: the minimum is at the collision focal point and this value is kept only over a short distance, of the order of β_y^*, then it rapidly grows towards the tails. Since the bunch length (σ_z, of the order of few mm) is longer than β_y^*, the actual luminosity is usually smaller than the nominal one; this effect is known as the *hourglass effect*. To overcome this, the nanobeam scheme works with a large half crossing angle ($\phi_x \approx 40$ mrad) and a small bunch width in the x direction ($\sigma_x = 10$ μm). By doing so, even if the actual length σ_z of the bunches in the z direction is of the order of 10 mm, only a small part of each bunch is overlapping at any given time, resulting in an effective bunch length of $\sigma_z^{\text{eff}} = \sigma_x / \sin\phi_x \approx 0.25$ mm $< \sigma_z$.

At KEKB, the peak luminosity achieved was twice the design value, although the LER beam current was lower than the design value (but the HER beam current was higher). The daily integrated luminosity was also twice the design value due to the continuous injection mode and the use of two bunches per radio-frequency (RF) pulse at the Linac [45]. In addition, chromatic corrections of the IP β functions were applied using skew-sextupole magnets installed in the winter of 2009 and contributed to the great performance summarized in Tab. 2.1, and are compared to the expected values for SuperKEKB.

As shown in Table. 2.1, the latest β_y^* is a factor $\sim$5 higher than the design value

	KEKB Achieved	**SuperKEKB Latest**	**SuperKEKB Design**
Energy (GeV)	3.5/8.0	4.0/7.0	4.0/7.0
ξ_y	0.129/0.090	0.040/0.028	0.088/0.081
β_y^*(mm)	5.9/5.9	1.0/1.0	0.27/0.30
σ_y^*(nm)	940/940	177/177	59/59
I (A)	1.64/1.19	1.32/1.10	3.60/2.62
Luminosity ($10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	2.11	4.71	60

TABLE 2.1: Fundamental design parameters (LER/HER) and the latest achieved parameters of SuperKEKB compared to the values achieved by KEKB.

in SuperKEKB and the current is a factor ∼2 lower, which could lead to the target luminosity in near future.

In order to produce B mesons at an electron-positron collider, the center-of-mass energy has to be greater than the beauty mesons' threshold $2m_B \simeq 10.56\,\mathrm{GeV}$. The $\Upsilon(4S)$, the lightest strong resonance - the third among the radial excitations of the Υ[1] with a mass of $m_{\Upsilon(4S)} \simeq 10.58\,\mathrm{GeV}$, decays mostly to b-flavoured mesons ($B^0\overline{B}^0$ and B^+B^- pairs); this also explains the fact that this resonance has larger width than those of lower mass (see Fig. 2.2).

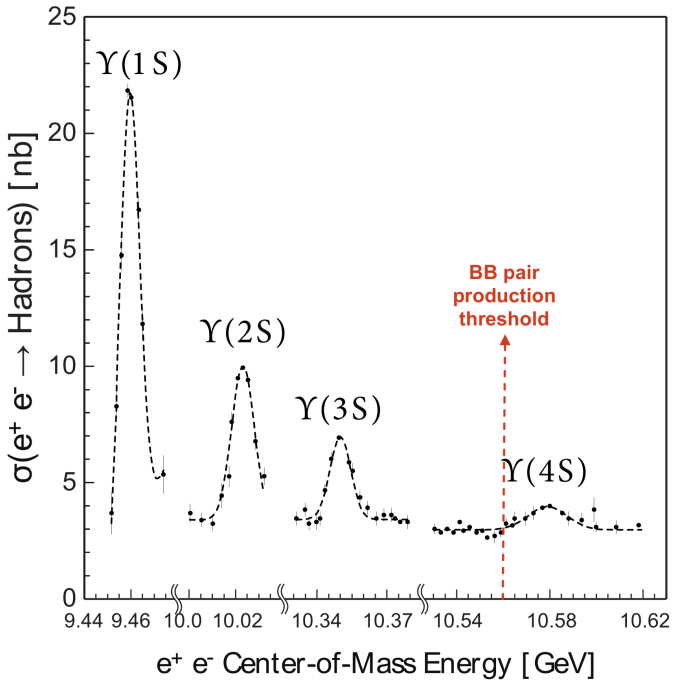


FIGURE 2.2: Hadronic cross section as a function of the center-of-mass energy in the region of the first four S-wave Υ resonances.

The KEKB and SuperKEKB mostly operate by tuning the center-of-mass energy to the $\Upsilon(4S)$ resonance. However, as shown with the continuum background in Fig. 2.2, e^+e^- collisions at that energy produce also a large amount of $q\bar{q}$ pairs, QED processes (Bhabha, di-gamma events), di-muon, and di-tau events. The cross-sections of the main processes are listed in Tab. 2.2.

[1] $b\bar{b}$ states with $J^{\mathrm{PC}} = 1^{--}$.

$e^+e^- \to$	Cross Section (nb)
$\Upsilon(4S)$	1.1
$u\bar{u}$ (γ)	1.6
$d\bar{d}$ (γ)	0.4
$s\bar{s}$ (γ)	0.4
$c\bar{c}$ (γ)	1.3
$e^+e^-(\theta_{\rm lab} \geq 17°)$	$44^\dagger$
$\gamma\gamma(\theta_{\rm lab} \geq 17°)$	$2.4^\dagger$
$\mu^+\mu^-$ (γ)	1.1
$\tau^+\tau^-$ (γ)	0.9

TABLE 2.2: Cross sections of various physics processes generated from e^+e^- collisions at the $\Upsilon(4S)$ resonance. The superscript † indicates that the values are prescaled by a factor 1/100.

The initial goal of B-factories (Belle and BaBar experiments) was to measure precisely the decay-time-dependent CP violation parameters. For that, along with measurements of lifetimes and mixing, the pair of B mesons are produced with a significant relativistic boost $\beta\gamma$ along the beam axis. Hence, the different energies for the two storage rings. The boost vector is defined as

$$\boldsymbol{\beta} = \frac{1}{E_H + E_L}(\mathbf{p}_H + \mathbf{p}_L), \qquad |\boldsymbol{\beta}| \simeq \frac{E_H - E_L}{E_H + E_L}$$

while the CM energy can be computed from the energy of the two beams, neglecting the crossing angle between them, $2\phi_x$.

$$\begin{aligned}\sqrt{s} &= \sqrt{p_H^\mu p_{\mu,L}} = \sqrt{(E_H + E_L)^2 - (\mathbf{p}_H + \mathbf{p}_L)^2} \\ &= \sqrt{2E_H E_L(1 + \cos(2\phi_x))} \simeq 2\sqrt{E_H E_L} \simeq 10.58\,\text{GeV}\,.\end{aligned}$$

The resulting boost, considering KEKB parameters

$$\beta\gamma = \frac{|\mathbf{p}_H + \mathbf{p}_L|\,c}{\sqrt{s}} \simeq \frac{E_H - E_L}{2\sqrt{E_H E_L}} \simeq 0.42$$

leads to a separation between the two B mesons of about $200\,\mu$m. At SuperKEKB, the boost factor has decreased to 0.28 and hence the separation is only $\sim 130\,\mu$m.

The total recorded samples at both the experiments are listed in Tab. 2.3. KEKB not only allowed to collect the largest $\Upsilon(4S)$ data sample to date, but also significantly large samples at other $b\bar{b}$ resonances, like at the $\Upsilon(2S)$ and $\Upsilon(5S)$ resonances. Though SuperKEKB didn't tune to take data at other Υ resonances yet, it collected scan data of $\sim$20 fb^{-1} around the $\Upsilon(5S)$ mass for quarkonium studies.

Resonance	On-resonance Luminosity (fb^{-1})	Off-resonance Luminosity (fb^{-1})	Υ number (10^6)
$\Upsilon(5S)$	121.4	1.7	7
$\Upsilon(4S)$	711.0	73.8	772
$\Upsilon(3S)$	2.9	0.2	11
$\Upsilon(2S)$	24.9	1.7	158
$\Upsilon(1S)$	5.7	1.8	102
$\Upsilon(4S)$	361.6	42.3	393

TABLE 2.3: Samples collected during Belle (top, from Ref. [46]) and Belle II operation at different center-of-mass energies.

2.2 The Belle and Belle II detectors

Detectors in B-factories must provide wide acceptance and high efficiency in reconstructing B decays, good momentum and energy resolution to separate the small signal from the background, very good vertex position resolution, and efficient and robust particle identification capabilities for hadrons and leptons. Since the average momentum of charged particles is below $1\,\mathrm{GeV}/c$, minimizing the amount of material that causes multiple scattering is critical. The same need applies for the amount of uninstrumented material in front of the electromagnetic calorimeter, which must detect showers with energies as low as 20 MeV. The detector has an asymmetric polar geometry that reflects the boost, and includes: silicon vertex detectors near the beam pipe for secondary vertex measurement and to provide additional tracking points, a drift chamber with helium-based gas for charged particle tracking and particle identification through ionization and energy loss, Cherenkov detectors for particle identification, and a CsI(Tl) crystal calorimeter for the neutral particle measurement. The high light yield and small Molière radius of the CsI crystals provide excellent energy and angular resolutions; in addition, the high yield allows the use of silicon photodiodes, which can work in magnetic fields; all of the above detectors are inside a 1.5 T cryogenic superconducting solenoid. The magnetic flux return yokes are used to absorb hadrons and contain scintillator bars and resistive plate chambers (RPCs) to perform muon and neutral hadron detectors. The Belle II detector, as shown in Fig. 2.3 in comparison to the Belle detector, reuses the structure, solenoid, CsI(Tl) crystals, and some of the barrel RPCs from Belle, while most of the other components of the sub-detectors are new.

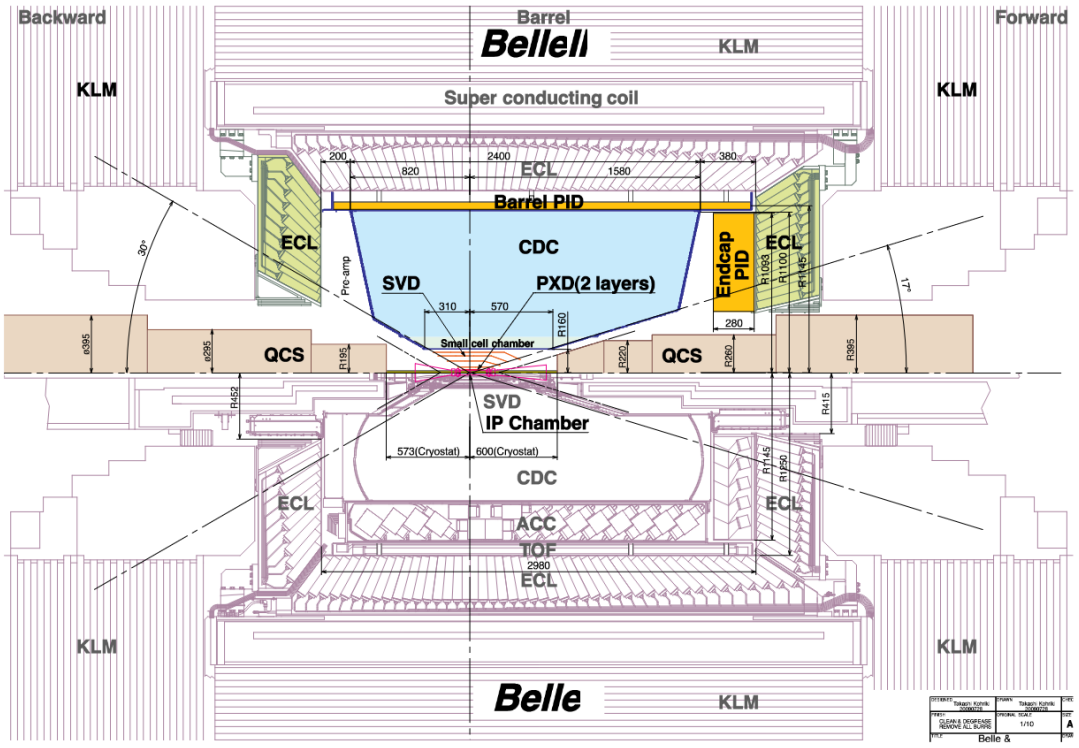


FIGURE 2.3: Schematic view of the Belle II detector (top half) in comparison to the previous Belle detector (bottom half). Taken from Ref. [47].

2.2.1 Coordinate system

The standard coordinate system of the Belle (II) is as follows:

(i) x: lies in the horizontal plane and points towards the outside part of the ring tunnel,

(ii) y: vertical and points upwards,

(iii) z: coincides with the Belle (II) solenoid axis[2] and the bisector of the two beams. Given the small crossing-angle between them, it almost coincides with the direction of the electron beam (Nikko→Oho, see Fig. 2.1).

Due to the cylindrical structure, it is convenient to use polar coordinates:

(i) $\rho = \sqrt{x^2 + y^2}$: radius in the $x - y$ plane,

(ii) ϕ: (azimuthal) angle $\in [0°, 360°]$ with x axis,

(iii) θ: (zenith) angle $\in [-180°, 180°]$ with z axis, where $\theta = 0$ corresponds to $(x, y, z) = (0, 0, 1)$.

We refer to *longitudinal* for the direction parallel to the z axis, while *transverse* means lying in the x-y (or r-ϕ) plane. Also, *forward* as the direction in positive z coordinates, *backward*, otherwise.

[2]The magnetic field lines go from $-z$ to z.

2.2.2 Tracking system

The Belle (II) tracking system relies on two components: a central drift chamber (CDC) and a vertex detector. A superconducting solenoid provides the magnetic field of 1.5 T in a cylindrical volume of 3.4 m in diameter and 4.4 m in length. The coil is surrounded by a multilayer structure consisting of iron plates and calorimeters, which is integrated into a magnetic return circuit. The iron structure of the Belle (II) detectors serve as the return path of magnetic flux and an absorber material for KLM. It also provides the overall support for the sub-detectors.

Central Drift Chamber

The central drift chamber (CDC) is the most important detector for charged particle tracking. It also plays a crucial role in particle identification (PID), especially for low-momentum tracks. The CDC at Belle contains 8,400 sense wires in 50 layers inside a mixture of low-Z gas (50% helium, 50% ethane) to reduce multiple Coulomb scattering contributions to momentum resolution. The sense wires in each layer are in either an axial or stereo orientation. Axial wires are aligned with the detector's z-axis, and thus with the solenoidal magnetic field inside the detector. In contrast, stereo wires are skewed with respect to the detector axis (imagine holding a bunch of uncooked spaghetti in both hands and then slightly twisting your hands in opposite directions: this is effectively the geometry of stereo wires).

As a charged particle moves through the CDC, it ionizes atoms and molecules in the gas mixture. The ionization electrons drift toward the sense wires, since there is a voltage difference between the sense and field wires, and this permits the detection of the original charged particle. The charge, positive or negative, is determined by the curvature of the track. While both axial and stereo wires give information about the particle's r-ϕ position (with a typical hit resolution of 100μm), the measurement angle of the stereo wires allows us to extract the particle's z position as well. Combining this information with the drift time measurement gives the momentum vector of the particle. The relative resolution of the CDC on the transverse momentum p_T is given by

$$\sigma(p_T) = (0.28 p_T \oplus \frac{0.35}{\beta})\%, \tag{2.2}$$

Particles of different masses will exhibit different amounts of ionization energy loss as they travel through the CDC gas. These characteristic curves, called dE/dx (i.e., energy loss per distance) curves, are used for PID within the CDC. While essentially all charged particles will have dE/dx information (and thus PID information) from the CDC, the helical track shape of low-momentum particles (due to the Lorentz force from the solenoidal field) means they may spiral in on themselves and never reach the dedicated PID system in the TOP or ARICH. Additionally, the TOP and ARICH do not cover the backward endcap. In these cases, it is particularly critical to have high-quality PID information from the CDC.

Upgraded components in Belle II: CDC

The CDC at Belle II contains 14,336 sense wires in 56 layers. Smaller cells and longer level arm are the key ingredients to improve the p_T resolution performance:

$$\sigma(p_T) = (0.18 p_T \oplus \frac{0.32}{\beta})\%. \tag{2.3}$$

Silicon Vertex Detector

The vertex detector is the innermost detector component outside the beryllium beam pipe and is crucial to the vertex resolution (and hence, a large part of the physics program) of a B-factory. The Silicon Vertex Detector (SVD) plays a central role in measurements of time dependent CP violation, as it provides a precise measurement of the B meson decay vertex position. Besides a good spatial resolution, it is also important for this sub-detector that the amount of material placed inside the detector acceptance is kept sufficiently low (to reduce the multiple scattering of particles).

At that time, the most natural choice that meets these criteria is the use of double-sided silicon strip detectors (DSSDs).

In the SVD as initially implemented at Belle, the DSSD units were arranged in three (later, in 2003, four) cylindrical layers around the beam pipe, as shown in Fig. 2.4. The modules in each layer are of different lengths, to cover the polar angle $17° < \theta < 150°$, corresponding to the full angular acceptance, since 2003. Each DSSD module consists of 1280 sensitive strips. On the side that is oriented perpendicularly to the beam direction (to measure z position of vertex) they are separated by 42 μm, while on the other side, with strips in the beam direction (to measure r-ϕ position), the separation is 25 μm. On the z side of the strips only every second channel is read (to reduce the number of readout channels) making the effective separation of strips 84 μm.

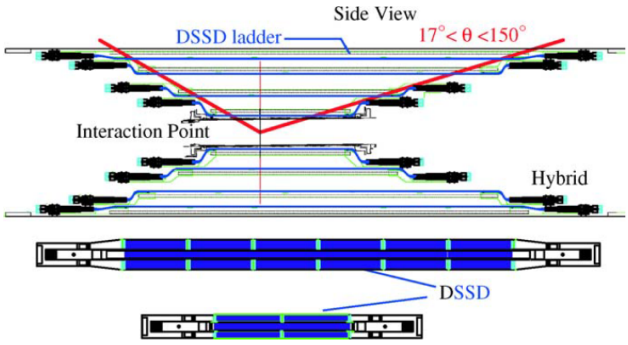


FIGURE 2.4: The longitudinal cross section of Belle's SVD. The layer 1 and layer 4 ladders are also depicted. Taken from Ref. [48].

The performance of the SVD can be characterized by its resolution on charged tracks impact parameters, which is measured to be

$$\sigma_{xy} = 21.9 \oplus \frac{32.5}{p\beta \sin^{3/2}\theta}\mu\text{m}, \ \sigma_z = 27.8 \oplus \frac{31.9}{p\beta \sin^{5/2}\theta}\mu\text{m}, \tag{2.4}$$

where p is the particle's momentum in GeV and θ its polar angle.

Upgraded components in Belle II: PXD and SVD

In Belle II, the vertexing system is made of six total layers and two sub-detectors: the inner silicon pixel vertex detector (PXD) and the outer SVD. The PXD is two layers, and the remaining four layers are the SVD, made up of double-sided silicon strip sensors. Fig. 2.5 shows a schematic of the beam crossing; the beam pipe radius in the collision region is only 10 mm. While this is good for vertex reconstruction, the small radius increases the background from the collision region. The silicon strip sensors of the SVD would not be able to handle this larger background, hence the inclusion of the inner PXD layers. The PXD sensors can withstand a 20 Mrad radiation dose

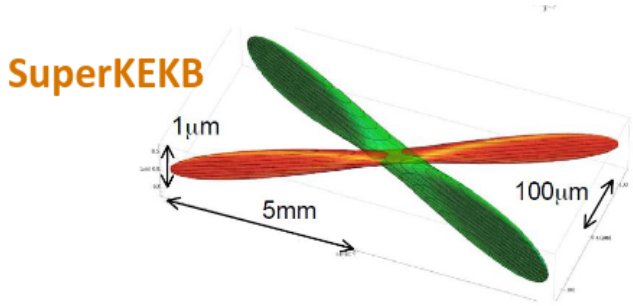


FIGURE 2.5: Schematic of the beam crossing at SuperKEKB with the nanobeam scheme.

and have a ∼0.2% radiation length per layer, while maintaining an average spatial resolution of approximately 15 μm and a hit efficiency of 98% after 4 years of data taking [49]. This system provides, $\sigma_{d0} = 14.2 \pm 0.1$ μm, about 25% better resolution on charged tracks' impact parameters than Belle. The readout time per pixel row is 100 ns, while four rows are treated in parallel, which makes the PXD's readout time 20 μs per module [49]. This is about a factor 5 higher than the other subdetectors, therefore treated separately in event building (see Sec. 2.2.6).

2.2.3 Charged hadron identification

The Aerogel Cherenkov Counter (ACC) is used to provide separation between charged kaons and pions with high momenta (from 1.2 GeV to 3.5 GeV). A good separation between pions and kaons is crucial for efficient flavour tagging. When a charged particle travels through a medium with a velocity that exceeds the speed of light in that medium, it emits Cherenkov photons. In a medium with the refractive index n a particle of mass m and given momentum p emits Cherenkov photons only if $m < p\sqrt{n^2 - 1}$ is satisfied. By selecting appropriate n one can ensure that pions of given momentum are below that threshold, while kaons are not. On this basis, they are separated in the ACC.

The Time of Flight Counter (TOF) gives particle identification information in order to separate charged pions and kaons with low momenta, below 1.2 GeV. By measuring the time T that the particle needs from the IP to the TOF ($L \sim 1.2$ m), and knowing its momentum p, the particle's mass can be inferred by

$$m = p\sqrt{\frac{c^2 T}{L^2} - 1}. \tag{2.5}$$

Basic building blocks of the TOF are photomultiplier tubes with attached plastic scintillation counters. They provide a timing resolution of 100 ps, allowing the separation between kaons and pions at 1.0 GeV with 3σ significance. The TOF also provides fast timing signals for the data acquisition trigger system. For this purpose, a thin trigger scintillation counter (TSC) is placed directly in front of two basic units to form a TOF module. In total, there are 64 TOF modules placed in the barrel region at the radius of 1.2 m from the IP, as shown in Fig. 2.3. The polar angle covered is $34° < \theta < 120°$.

Upgraded components in Belle II: TOP and ARICH

In the barrel and forward endcap regions, particles that leave the CDC enter either the TOP (barrel) or ARICH (endcap) detectors (Fig. 2.3, top). TOP stands for time-of-propagation and is a special type of Cherenkov detector. Fig. 2.6 gives the conceptual

overview of the TOP. The TOP detector's main purpose is to improve K/π separation,

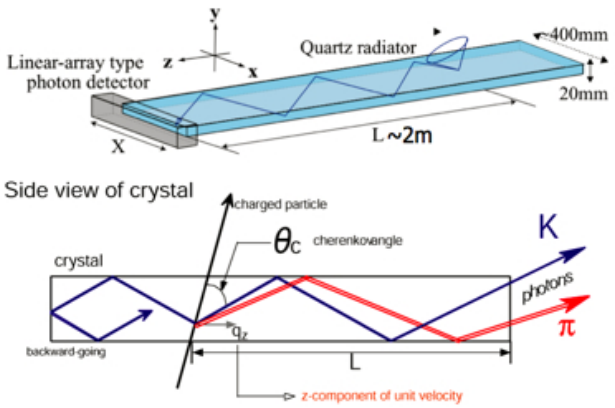


FIGURE 2.6: Conceptual overview of the TOP detector, showing the basic principle of it working when a kaon track passes (blue) vs when a pion track (red) passes.

but it provides information about other charged particles as well.

In the forward endcap, the aerogel ring-imaging Cherenkov detector (ARICH) provides particle identification (PID) information. Fig. 2.7 shows the working principle of ARICH. As the name implies, the ARICH uses an aerogel radiator, while the TOP

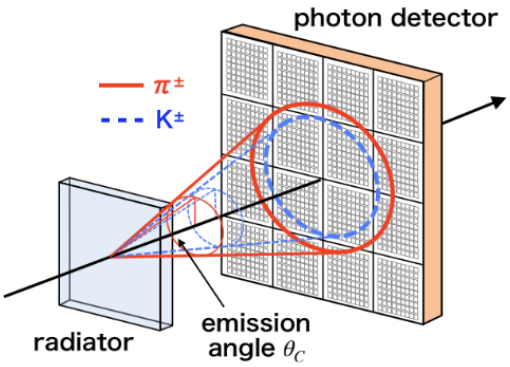


FIGURE 2.7: The principle of the particle identification by the ARICH counter.

uses a quartz radiator. The ARICH is also a more traditional Cherenkov detector, using ring-imaging of the Cherenkov cone to identify particles where the TOP is a time-of-propagation device (though the time-of-propagation is related to the Cherenkov angle).

The reason for the different PID systems in different regions of the detector largely results from geometric constraints. Additionally, there is no PID system in the backward endcap. Due to the asymmetric beam energies, decays in Belle II are forward-boosted, and thus we expect the absence of a dedicated PID system in the backward endcap to have a small effect on the detector's capabilities. Overall, only about 10% of the CDC angular coverage is outside the TOP and ARICH coverage.

Hadron identification

The particle identification (PID) at Belle was performed with likelihood-based selectors. For hadron identification, the likelihood for a candidate particle α against a particle hypothesis β was calculated based on $\mathrm{d}E/\mathrm{d}x$ information from the CDC

($\mathcal{L}_\alpha^{\mathrm{CDC}}$), time of flight from the TOF ($\mathcal{L}_\alpha^{\mathrm{TOF}}$) and the number of photons from the ACC ($\mathcal{L}_\alpha^{\mathrm{ACC}}$), respectively. Then, the ratio [50]:

$$L(\alpha:\beta) = \frac{\mathcal{L}_\alpha^{\mathrm{CDC}}\mathcal{L}_\alpha^{\mathrm{TOF}}\mathcal{L}_\alpha^{\mathrm{ACC}}}{\mathcal{L}_\alpha^{\mathrm{CDC}}\mathcal{L}_\alpha^{\mathrm{TOF}}\mathcal{L}_\alpha^{\mathrm{ACC}} + \mathcal{L}_\beta^{\mathrm{CDC}}\mathcal{L}_\beta^{\mathrm{TOF}}\mathcal{L}_\beta^{\mathrm{ACC}}} \tag{2.6}$$

was calculated and used for the identification by analysts. By construction, these variables returned the value 0.5 in case there was no likelihood available for the given hypothesis and a factor of 0.5 would appear in the product if any of the involved sub-detectors could not provide a likelihood. Therefore, pions (kaons) can be selected by requiring a low (high) value of $L(K:\pi)$:

$$K\mathrm{ID} = L(K:\pi) = \frac{\mathcal{L}_\mathrm{K}}{\mathcal{L}_\pi + \mathcal{L}_K}\,, \pi\mathrm{ID} = L(\pi:K) = 1 - K\mathrm{ID}\,. \tag{2.7}$$

2.2.4 Neutrals: Electromagnetic calorimeter

Given that the signal extraction for this analysis is through the extra energy deposited in the calorimeter after reconstructing all the tracks of signal B, a high-resolution electromagnetic calorimeter is a fundamental component of the detector. Also, around a third of B-decay products are π^0's or other neutral particles, producing photons in a wide energy range (20 MeV to 4 GeV), making a calorimeter essential.

CsI(Tl) was chosen as the scintillation crystal material for the Belle experiment because of its high light output, relatively short radiation length, good mechanical properties and moderate price. The main tasks of the calorimeter are:

- detection of photons with high efficiency,
- precise determination of the photon energy and angular coordinates,
- electron identification (eID),
- generation of the proper signal for trigger,
- on-line and off-line luminosity measurement obtained from Bhabha events, whose signature is a geometrical coincidence of high energy deposits in the forward and backward parts of ECL,
- K_L detection together with the KLM.

The ECL consists of a barrel section and two end-caps of segmented arrays of 8736 CsI(Tl) crystals in total. The former part was 3.0 m long and had an inner radius of 1.25 m. The end caps were located at $z = +2.0\,\mathrm{m}$ and $z = -1.0\,\mathrm{m}$. The scintillation light produced by particles in the crystals was detected with silicon photodiodes.

Each crystal had a tower-like shape and pointed almost to the IP. The crystals were tilted by a small angle in the θ and ϕ directions to prevent photons escaping through the gaps between the crystals. The angular coverage of the ECL was $17.0° < \theta < 150.0°$. Small gaps were left intentionally between the barrel and end cap compartments, providing the necessary space for cables and supporting parts of the inner detector; however, these gaps resulted in a loss of acceptance at the level of 3% and induced a low electron identification performance. The ECL ensured an energy resolution from 4% at 100 MeV to about 1.6% at 8 GeV, while the angular resolution ranged from 3 to 13 mrad. The combination of the two determined a π^0 mass resolution of about $4.5\,\mathrm{MeV}/c^2$.

For electron identification, in addition to CDC ($\mathrm{d}E/\mathrm{d}x$) and ACC (light yield) information, the ECL was used to form the ratio of likelihoods. In fact, ECL provided the matching between the position of the track and the energy cluster, the E/p[3] and the information on the transverse shower shapes.

The variable used for electron identification is defined as

$$e\mathrm{ID} = \frac{\Pi_i \mathcal{L}_i^e}{\Pi_i \mathcal{L}_i^e + \Pi_i \mathcal{L}_i^\pi} ,$$

where all the discriminant variables described above are involved in the products. The electron identification efficiency obtained with the selections eID> 0.5 is shown in the left side plot of Fig. 2.8. It was measured using a sample of radiative Bhabha events, where very forward and backward regions were excluded because of the lower efficiency due to the larger amount of material budget degrading both CDC and ECL performance. Nonetheless, for the barrel region, the efficiency was very high (over 90% for tracks with momenta above $1\,\mathrm{GeV}/c$) and in very good agreement with data. The kaon fake rate at the same cut value reached 0.43(21)% in data (MC) while the pion fake rate (see right side plot of Fig. 2.8), obtained with $K^0_S \to \pi^+\pi^-$ events, was $(0.25 \pm 0.02)\%$ for tracks' momenta between 1.0 and 3.0 GeV/c.

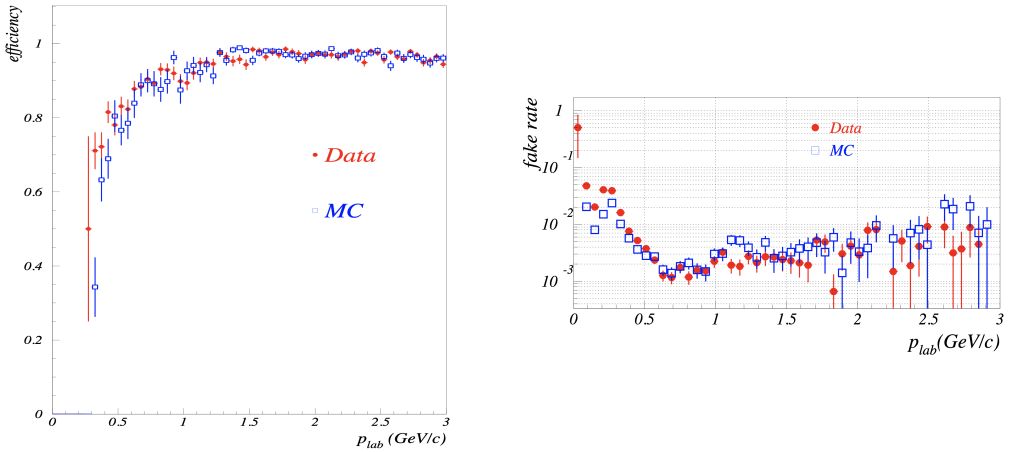


FIGURE 2.8: eID> 0.5 performance. Left: Efficiency in radiative Bhabha events as a function of the tracks' momentum. Left: Fake rate for $\pi^\pm$ as a function of momentum. Data points are shown in red circles and MC in blue squares. Tracks are selected within the barrel ECL acceptance. Taken from Ref. [51].

Upgraded in Belle II

For Belle II, the CsI(Tl) crystals, preamplifiers, and support structures have been reused, while the readout electronics and reconstruction software have been improved. In the absence of background, a very similar performance is obtained at Belle II. However, due to the much higher background, the relatively long decay time of scintillations in CsI(Tl) crystals is expected to increase the overlap of pulses from contiguous background events. To mitigate the resulting pile-up noise, the photo-sensors are equipped with waveform-sampling read-out electronics that allow the storage of ADC samples in FPGA internal buffers and the waveform fitting to discriminate signal from off-timing hits. The latest performance measurements with $eID > 0.9$ selection

[3] An electron is expected to release all its energy in the calorimeter, yielding to an $E/p \sim 1$.

in Belle II data are shown in Fig. 2.9, calculated using many different channels as listed in the legend. The efficiency and fake rate are similar to that of Belle.

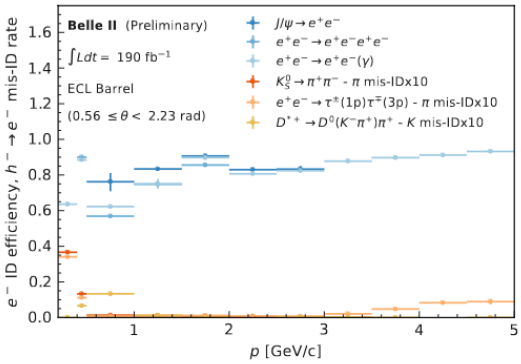


FIGURE 2.9: Electron identification and hadron-electron misidentification rates as a function of momentum in the ECL barrel region, with $eID > 0.9$ selection. Note that the misidentification rate is multiplied by a factor of 10 for illustration purposes. Taken from Ref. [52].

2.2.5 Muon and K_L^0 identification: KLM detector

Electrons mostly lose all of their energy in the calorimeter and stop. Unlike them, muons don't undergo bremsstrahlung radiation and tend to travel further. In order to detect them and K_L mesons, the Belle (II) experiments contain a dedicated detector in the outermost layer after the solenoid. Muons need a certain minimum momentum to reach there, above 600 MeV/c they can be identified with high efficiency.

It consists of two regions: the barrel-shaped one, covering a polar angular range of 45° to 125°, and the end-caps, extending the total acceptance from 20° to 155° (see Fig. 2.3). Due to the absence of a magnetic field, charged particles travel in a straight line through the KLM until they escape ($|p| > 1.5$ GeV/c) or range out due to energy depositions.

The detector consists of alternating layers of 4.7 cm thick iron plates and active detector elements. Through interaction with the absorber material, muons and K_L mesons create a shower. The iron plates provide an interaction lengths of 3.9 on top of the 0.8 X_0 from the ECL for hadrons travelling normal to the detector plane. Resistive Plate Chambers (RPC) with glass electrodes act as the active detector elements in the barrel region of the KLM. But at Belle II, due to the expected larger backgrounds in end-caps and the innermost layers in the barrel region due to neutrons, the RPCs have been replaced by layers of scintillator strips with wavelength shifting fibers, which are tolerant to higher rates. This upgrade also helps to mitigate the long dead time of the RPCs due to the recovery of the electric field after a discharge, impacting the detection efficiency at high rates.

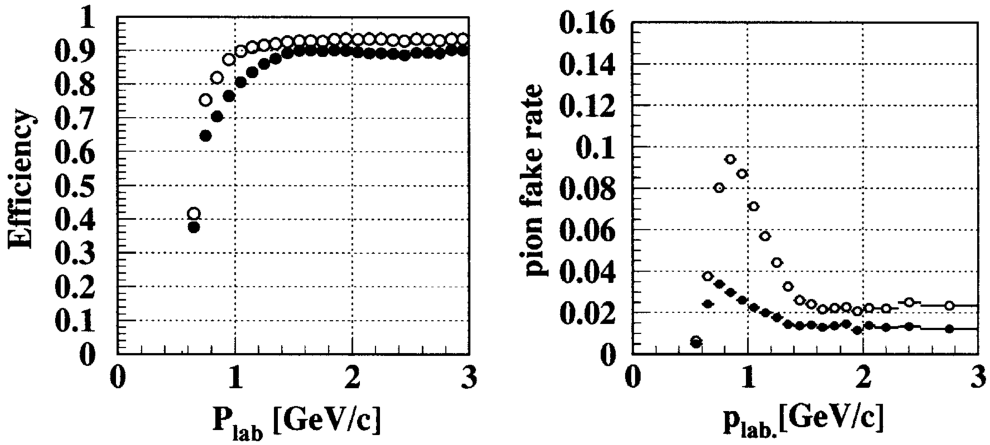
The large fluctuations of the production of hadronic showers does not allow for the reconstruction of the energy of K_L, just the direction (assuming it origins from the IP). The range and transverse deflection of a non-showering charged particle allow to discriminate between muons and hadrons. The double-gap design resulted in a super layer efficiency of over 98%; the hit position was resolved to about 1.1 cm when either one or two adjacent strips fire, resulting in an angular resolution of under 10 mrad from the IP.

To identify muons, the CDC tracks are extrapolated to KLM acceptance and compared to the reconstructed hits in the KLM, using the difference ΔR between

measured and expected range of the track, and the statistic χ^2_r constructed from the transverse deviations of all hits associated to the track, normalized by the number of hits. Likelihoods for the muon, pion and kaon hypotheses were formed based on probability density functions (PDFs) in ΔR and χ^2_r.
The normalized ratio:

$$\mu\text{ID} = \frac{\mathcal{L}_\mu}{\mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K}$$

was then used as a discriminating variable. At Belle, the muon identification efficiency at $\mu\text{ID} > 0.9$ was close to 90% for tracks with momenta above 1 GeV/c while the fake rate for pions reached $(1.35 \pm 0.07)\%$ in the momentum region 1.0-3.0 GeV/c (Fig. 2.10).



(A) (B)

FIGURE 2.10: μID> 0.9 performance measured with two-photon events ($e^+e^- \to e^+e^-\mu^+\mu^-$) at Belle. A: Efficiency as a function of the momentum. B: Fake rate of pions. The filled (open) dots refer to $\mu\text{ID} > 0.9(0.1)$. For both plots, the full polar region ($25° < \theta < 145°$) is considered. Taken from Ref. [53].

At Belle II, to identify muons with momentum less than 600 MeV that do not reach the KLM, machine learning techniques are trained using information from the inner detectors. This means that the muon identification efficiency is non-zero even at low momentum, but has much lower purity (fake rate for pions of $\sim 7\%$) as shown in Fig. 2.11.

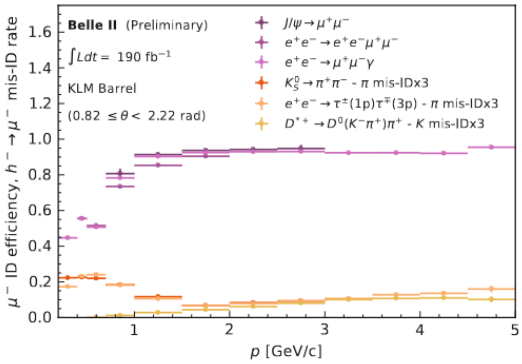


FIGURE 2.11: Muon identification and hadron-muon misidentification rates as a function of momentum, with $\mu ID > 0.9$ selection. Note that the misidentification rate is multiplied by a factor of 3 for illustration purposes. Taken from Ref. [52].

2.2.6 Trigger and Data Acquisition system at Belle II

In addition to the main subdetectors described above, there are several other components of the Belle II detector that are critical to its operation. Two important ones are the trigger system and the data acquisition system (DAQ). The trigger system consists of two levels: the hardware-based low-level trigger (L1) and the software-based high-level trigger (HLT). The function of both triggers is to filter out the beam background and retain only events of interest, and to prescale common interactions like Bhabha scattering. Filtering out hadronic events (events where a meson or baryon is produced) is relatively easy because they look quite different to the detector than noise, beam backgrounds, and other events that are not of interest. This function couples into the DAQ, which would be overwhelmed at high rates without the trigger filtering out these events. The DAQ is responsible for reading data from all Belle II subdetectors (Fig. 2.12). It is designed to handle a 10 kHz maximum event rate from the HLT (reduced for online storage from a 30 kHz trigger rate out of the L1), the value expected at peak luminosity.

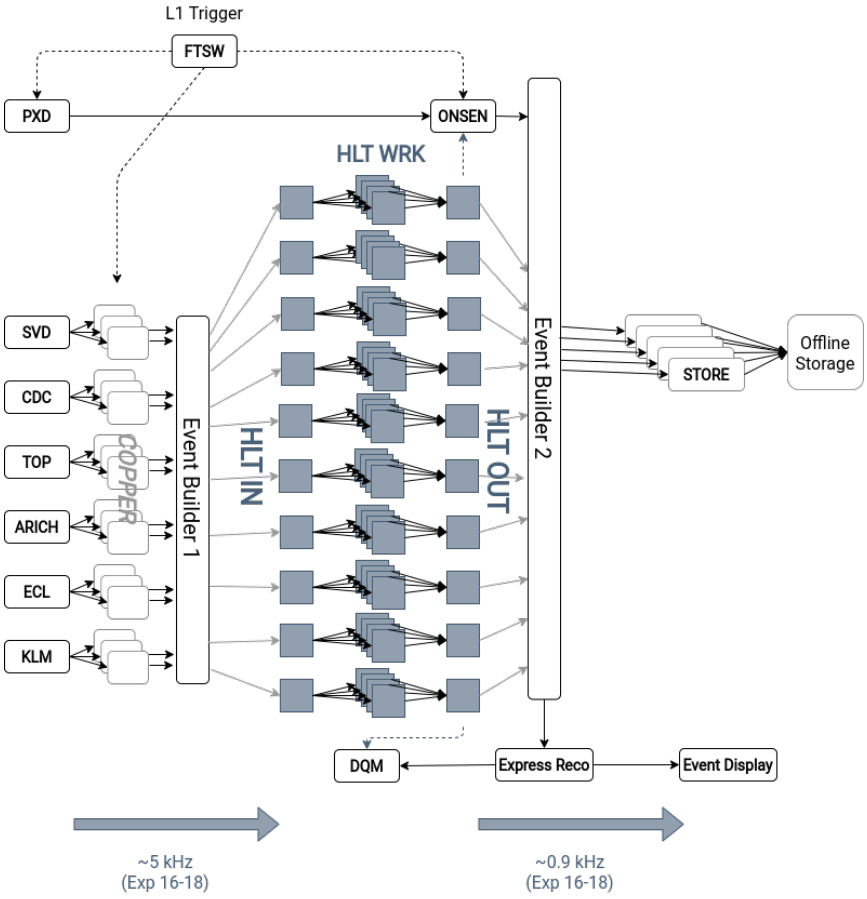


FIGURE 2.12: The Belle II Data Acquisition system. The data from detectors are transferred to the common readout cards (COPPER) via the unified optical link (Belle2link). Event building is performed in multiple steps by the combination of the readout PCs and the event builder switch, and the event data are fed into multiple HLT units. The data from PXD are merged with the HLT output and recorded in storage units.

2.2.7 Hardware trigger L1

Belle II's online trigger (also called Level 1 or L1 trigger) is mainly built upon two independent trigger lanes, the CDC and the ECL trigger. The different subdetectors produce near real-time trigger signals in the order of 16 ns intervals. All subdetector triggers are combined to a global L1 trigger decision in the global decision logic (GDL) after a fixed latency of 5 μs. If the decision is positive, this information is transferred via the Belle2link connection to the front-end readout boards of all subdetectors (except PXD). Those boards were storing the full event data in a ring buffer during the 5 μs. The hardware triggers and the front-end electronics are implemented using field programmable gate arrays (FPGA). The detector data is then delivered to the common pipelined platform for electronics readout (COPPER). A solution involving a PCI-express-based readout module (PCIe40) with high data throughput of up to 100 Gb/s is now adopted to upgrade the Belle II DAQ system.

This platform transfers the data eventually to the event builder computers. The information is collected and sent via an Ethernet connection to the high level trigger (HLT) running the software reconstruction. There are different trigger channels planned for the L1 trigger, and the final decision on the output rate for each physics process is still under investigation. However, it is already planned to have a total output rate of 20 kHz at the anticipated luminosity of $L = 6 \times 10^{35} cm^{-1} s^{-2}$. Including a safety margin, the rate may be as high as 30 kHz. An overview on the current status of the hardware trigger menu is given in Ref. [54].

2.2.8 The High Level Trigger of Belle II

The decision on whether the event is eventually retained after being triggered by the software is based on variables calculated using the event topology and the physical properties of the particles. This requires that the particles be fully reconstructed based on the detector data, which is a very challenging task. This task cannot be implemented only on low-level hardware, but requires software running on a server farm, similar to offline reconstruction. The full detector information without PXD is distributed from the Event Builder computers over Ethernet to the HLT nodes and finally to the HLT workers. About 20 HLT units are planned for the full luminosity run, each hosting 16 HLT workers with 20 CPUs. Since the online trigger software runs on each of the worker CPU, there are about 6400 parallel processes computing the trigger decision. At the expected L1 output rate of 20 kHz (maximum 30 kHz), each process has an average of $\sim$ 300 ms ($\sim$ 200 ms) to calculate the trigger decision. HLT should reduce the event rate to about 10 kHz. In addition to simply making a binary decision whether to store an event, the HLT outputs classification information (called a tag) that can be used to divide the recorded events into different groups according to the underlying physical process. This tag information is later used to quickly skim the recorded data when searching for a specific process used for calibration or for physics analysis. As described above, the integration time of the PXD is very long, resulting in a large latency and many hit pixels in a single readout. Since it is not possible to transfer or store the entire data, a way must be found to reduce the number of hits in a single readout. Therefore, a very important task of online reconstruction is to provide information about where to expect the PXD measurements needed to increase the resolution of the tracks already found. This is done by extrapolating the reconstructed particles into the volume of the PXD. A rectangular region around the extrapolated positions on each sensor plane where a particle is expected to pass through the detector (also called region of interest or ROI) is then sent to the PXD's readout board. The online selection nodes (ONSEN) combine these ROIs with similar

information extracted by a faster but more course FPGA-based ROI extraction (the data concentrator – DATCON) and transfers only the PXD hits in the specified regions to the second event builder. The information from the pixel detector is then combined with the other subdetectors and sent to the storage system. During the recording of the collisions, the data must be monitored to ensure consistent data quality without degradation in neither the detector systems, the triggers, or the online reconstruction. A certain portion of triggered events, along with the PXD information, is sent not only to the storage system, but also to the ExpressReco farm, which performs the standard offline reconstruction and extracts a series of monitoring plots that are shown integrated over multiple events to the control room shifters.

Monitoring

Even at the HLT level, the reconstructed information is used to create monitoring plots for various subdetectors that summarize performance. Monitoring systems consist of: Data Quality Monitoring (DQM), which displays live plots during each run, and Mirabelle, which collects information from the DQM and displays it as a function of each run. Each run typically lasts a few hours, and the collection of runs taken under the same experimental conditions are grouped together as an "experiment" that lasts a few months. Because the HLT is critical to the stable operation of the experiment, we have worked to establish monitoring of the HLT. In addition to monitoring the physics related variables of the HLT filter and skim decisions, we have focused on monitoring the CPU and memory performance of the HLT units. For illustration, the HLT-related Mirabelle plots for some runs are shown in Fig. 2.13. The top plot shows the trigger rate normalized to luminosity, an effective cross section, at the level of the L1 output (in blue) and at the level of the HLT output (in red). The reduction by a factor of 5 is clearly visible. The lower plot in the same figure shows the average time it takes an HLT worker to process an event (in violet), which is about 300 ms. And the budget time (in green) represents the maximum time available to process an event at the current input rate. This effort allowed us to estimate the safe limits of HLT operation and better prepare for stable HLT operation in the immediate future and beyond. Monitoring was also critical in building the confidence to move the HLT into the filtering phase in 2021, where events are actually rejected based on the HLT decision, reducing the data flow rate by a factor of 5.

Software Optimization

As already seen from the monitoring plots, the average processing time per event was already 300 ms, which should be the limit at the expected design luminosity. In the fall of 2020, when the instantaneous luminosity of SuperKEKB was $L = 2 \times 10^{34} cm^{-1} s^{-2}$, the maximum HLT throughput limit was calculated to be 9 kHz. To maintain the HLT performance with the expected increase in luminosity, the CPU processing time per event had to be significantly reduced. The entire software processing chain that each event goes through is shown in Fig. 2.14. All modules in the chain were profiled to determine the slowest components, and the code in some unpackers, DQM, and reconstruction components was optimized. Most of the optimisations relate to efficient handling of ROOT objects and reduction of unnecessary operations. They were performed during the 2020–2021 data taking and implemented during the maintenance days that occur every two weeks, as shown in the left plot of Fig. 2.15. Here, the average processing time is plotted as a function of the date of release of the new software version, illustrating the overall reduction. The name of the module that is modified to obtain each improvement is written in orange. Over a period of several months, we

FIGURE 2.13: HLT related plots on Mirabelle monitoring system for certain runs in Experiment 18. The top plot shows the effective cross section, while the bottom plot shows the average processing time per event.

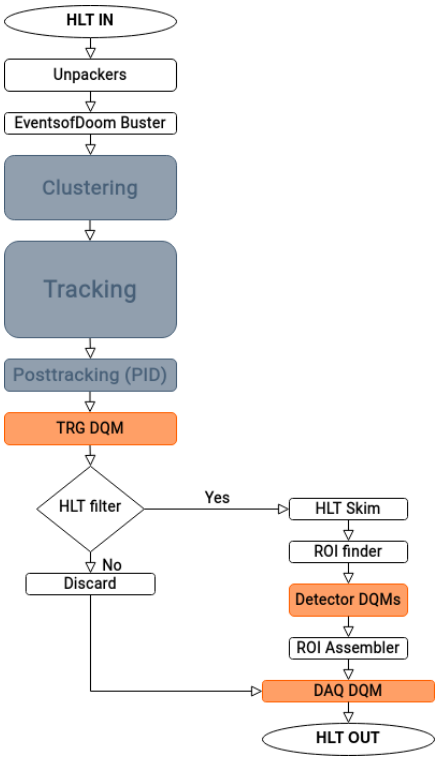


FIGURE 2.14: The software processing chain each event goes through in HLT. The reconstruction related tasks are shown in grey and DQM related tasks are shown in orange.

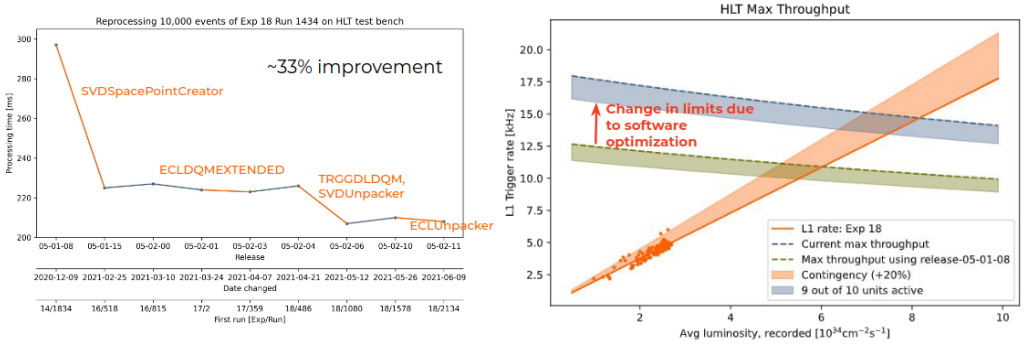


FIGURE 2.15: Left: Change in average processing time per event for different releases, optimized over time, showing the modules that changed. Right: HLT throughput and the L1 trigger rate as function of instantaneous luminosity. The intersection of both curves corresponds to the maximum limit.

were able to improve processing time by a factor of almost 2 and effectively improve the throughput limits of HLT, as shown in the right plot of the same figure. Here, the orange curve represents the increase in the HLT's input rate (the L1 output rate) as a function of instantaneous luminosity. A contingency of 20% is assumed (in the shaded region) because the L1 trigger rate is sensitive to the operation conditions. The grey and green curves represent the maximum throughput of HLT. As the luminosity increases, the background contribution in the events increases, making the processing time longer, which translates into lower HLT throughput. Again, some buffer is assumed (shown in shaded regions) to account for the possibility that one of the HLT units is not functioning. The intersection of the two curves represents the maximum possible HLT throughput, which has shifted from 10 kHz to 15 kHz for the given background conditions as a result of software optimization. These efforts allowed the HLT to operate smoothly until Long Shutdown 1 (LS1), at the end of which we exceeded a maximum L1 trigger rate of 10 kHz at $L = 3.5 \times 10^{34} cm^{-1} s^{-2}$ average luminosity, which is very close to the safe maximum HLT throughput under these experimental conditions.

In April 2021, some HLT units crashed during some runs due to memory leaks. The cause of this memory leak was determined and fixed on the next maintenance day. This incident also strengthened the monitoring systems for memory usage in the HLT units.

Future possibilities

Improving the HLT processing time by a factor 2 to about 150 ms per event already increases the safe HLT throughput limit. Nevertheless, the HLT was operated very close to this limit at the end of June 2022, when we reached the highest peak luminosity, because the trigger rate and the processing time depend on the conditions of the machine and the collimators. Therefore, the processing time needs to be reduced further. Currently, tracking takes up most of the processing time in the reconstruction chain, especially the track fitting task based on the GENFIT package [55]. The track fitting would only slow down with increasing background conditions, so serious efforts are needed to control and optimize it. New HLT units will also be added over time to achieve the design specification.

In addition, a significant improvement is needed to make the HLT system more robust and reliable to all kinds of machine conditions, ready in case we encounter some limitations and do not want to limit the increase of luminosity. One could explore the possibility of a preliminary filter decision that could be computed immediately after the partial reconstruction of the event (e.g., using only SVD+ECL information). If such a faster decision could be computed within 10 ms per event and the input rate could be reduced by at least a factor of 3, this would reduce the total processing time from about 150 ms to bout 55 ms per event. This allows for a greater throughput and more flexibility for a looser L1 trigger criteria, which has implications for low multiplicity physics studies. Like the HLT, this prefilter should initially be implemented in a monitoring phase and could be turned on as we build confidence and reach limitations in data taking.

In the future, the HLT can also be used to perform a cut-based B-tagging algorithm, which can also be integrated with the HLT skim to optimize the physics analysis workflow.

2.3 Reconstruction software: basf2

BASF (Belle AnalysiS Framework) [56] is the software developed by the Belle collaboration, providing the code for data-taking and 'offline' analyses. Major changes were needed to match the upgrades for the Belle II detector, and a completely new software framework called basf2 (Belle II Analysis Software Framework) was written [57, 58]. It combines the modern features of other HEP experiments while preserving the experience and good algorithms of Belle. The basf2 core is written in C++ and provides a Python interface, allowing easy access to different modules and their configuration for all the tasks. It is responsible for the generating MC simulation, performing the reconstruction chain shown in Fig. 2.14 both online and offline, and calculating high-level variables for the physics analyses.

2.3.1 b2bii

In order to exploit the basf2 software on Belle data as well, a framework called b2bii [59] is developed, which converts the Belle mDST (mini-data summary tapes) files from the PANTHER [60] format (used in BASF) into ROOT-objects [61], used in basf2. As the necessary input for physics analyses, only the mDST file format was targeted. DST files, which contain raw data including detector hits, are deeply linked with geometry and calibration information stored elsewhere and would be much harder to convert than their processed output. Thus, reprocessing of raw Belle data to, e.g., profit from improved track reconstruction will not be possible. After the conversion, all the Belle II analysis tools (like hadronic B-tagging described in the next chapter) can be used and the same scripts can be used for studies on data sets from both experiments, hence essential for combined Belle+Belle II measurements like our search for $B^+ \to K^+\tau^+\tau^-$.

Chapter 3

Hadronic B tagging

The B-tagging technique, an essential tool for studying the missing energy modes at B-factories, is presented with a historical context leading up to the current B-tagging package of the Belle II experiment, the so-called Full Event Interpretation (FEI) [62]. The advantages of a hadronic B-tagging over other approaches are presented, but also its limitations: a low efficiency, but also a large discrepancy between the performance in Monte Carlo (MC) and data. Since the FEI package is trained on MC, this discrepancy leads to suboptimal performance. The standard Belle II procedure for calibrating the hadronic B-tagging using the $B \to X\ell\nu$ sample is described. To improve the purity of the calibration sample and to investigate the sources of discrepancies between data and MC, a new sample based on $B \to D\pi$ events is introduced. The performance of the hadronic B-tagging in the Belle and Belle II data sets is evaluated using this $B \to D\pi$ procedure. The combination of calibration factors from the standard and new procedures is explained in Sec. 3.4.3; this is the current official method used in the most recent Belle II measurements.

3.1 B-tagging approach for the missing energy modes

Taking data at the center-of-mass energy of 10.58 GeV, the B-factories produce $\Upsilon(4S)$ mesons with a well known initial state that decay into B^+B^- and $B^0\overline{B}^0$ pairs with a branching ratio of more than 96%. When studying the decays of the B mesons with missing energy in the form of undetected neutrinos in the final state, as $B \to D^{(*,**)}\ell\nu$, $D^{(*)}\tau\nu$, $K^{(*)}\tau\ell$, $K^{(*)}\tau\tau$, $K^{(*)}\nu\nu$, $\pi\ell\nu$, $\tau\ell$, $\tau\tau$, $\tau\nu$, $\ell\nu$ decays (modes covering a large part of the Belle II experiment programme), one can gain information and additional constraints by also considering the other B meson in the event.

Together with the B meson of interest, called B_{sig}, the reconstruction of the other B meson, called B_{tag}, can greatly reduce the combinatorial background and significantly improve the signal purity. This technique is called *tag side reconstruction* or B-tagging. In addition to determining that the event is a $B\overline{B}$ event, it can also provide, partially or completely depending on the type of B-tagging used, the four-momentum and the decay vertex position of the B_{tag}.

The B-tagging approaches can be broadly classified as: *inclusive* and *exclusive*.

3.1.1 Inclusive B-tagging

The inclusive B-tagging procedure starts with the reconstruction of the signal side, followed by the assignment of all remaining particles of the event to the B_{tag} candidate, without checking the correspondence with specific B meson decays. Background suppression for B_{tag} candidates can be improved by additional requirements, e.g., zero net charge, no additional leptons, and zero net proton/antiproton number as was the case for the first observation of $B \to D^{(*)}\tau^+\nu_\tau$ decays in Belle [63, 64].

More recently, for the search for $B^+ \to K^+\nu\bar{\nu}$ in Belle II [65], BDTs were trained on MC using the inclusive properties of the other B meson to suppress the background. By exploiting the distinctive topological and kinematic properties of the B decay on the signal side, this approach has shown a high signal efficiency of up to 4%. The inclusive nature of the algorithm increases the amount of background from unrelated tracks. In addition, correlations between the particles and kinematic constraints in the form of intermediate particles cannot be exploited, resulting in a higher background level or, in other words, low purity. When searching for modes such as $B^+ \to K^+\tau^+\tau^-$, which already have a high background on the signal side originating from the semileptonic B decays, it is better to use a B-tagging with a high purity. The dependence of this inclusive approach on the signal side requires that the method be optimized, validated, and calibrated independently for each study, which prevents this B-tagging approach from becoming a common tool for Belle II analyses.

3.1.2 Exclusive B-tagging

In exclusive B-tagging, the accompanying B_{tag} candidate is reconstructed in several specific decay modes. An illustration of an event with an exclusively reconstructed B_{tag} in which B_{sig} decays to $K\tau^+\tau^-$ is shown in Fig. 3.1. Here, the reconstructed particles are shown as solid lines, while the missing particles (neutrinos) are shown as dashed lines.

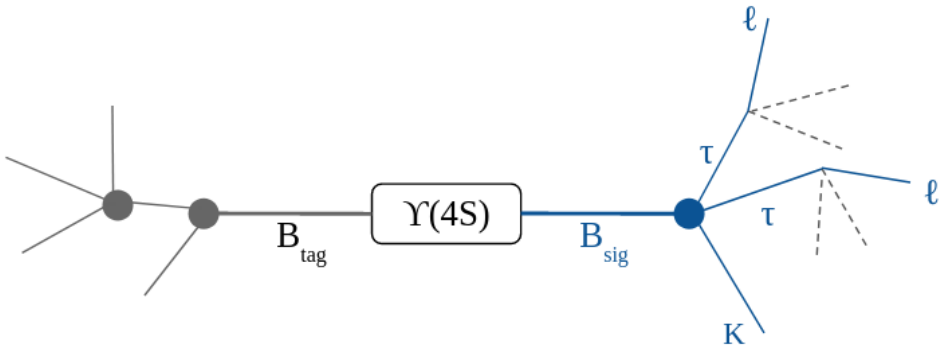


FIGURE 3.1: A schematic of the tag side B and the signal side B. The $\Upsilon(4S)$ decays into an exclusively reconstructed B_{tag} and $B_{\text{sig}} \to K\tau^+\tau^-$, where the missing neutrinos are shown as dashed lines.

The exclusive B-tagging can be further subdivided into the *hadronic* and *semileptonic B-tagging*, depending on whether the decay modes used are hadronic or semileptonic, respectively. The selection of specific decays for B mesons along with the reconstruction of all final-state daughters severely limits the maximum efficiency that can be obtained compared to the inclusive B-tagging. However, exclusive B-tagging provides important additional information, such as the charge/flavour of the B. It allows constraints such as the requirement that no additional charged tracks remain in the event after the tracks of both B candidates have been assigned (i.e., the entire visible $\Upsilon(4S)$ decay chain has been reconstructed). It also allows the estimation of the extra energy in the electromagnetic calorimeter, E_{ECL}, from the neutral clusters not used for the reconstruction of the two B candidates. This observable is used in many searches as a powerful discriminator or even as a variable for signal extraction, as in the search for $B \to \tau\nu$ decays [66,67] and $B \to h^{(*)}\nu\bar{\nu}$ decays [68]. This will also be the case in our search for $B^+ \to K^+\tau^+\tau^-$ decay. It is not possible to calculate the same observable, E_{ECL}, with inclusive B-tagging, since by construction all remaining

clusters are associated with the B_{tag} candidate. It is also important to remember that exclusive B-tagging algorithms, as opposed to inclusive B-tagging, are used as a common tool because they are less dependent on the signal side. This allows for a common calibration of the tool for all analyses and a common skim where the events containing B_{tag} candidates can be preselected from the entire data and MC samples for faster access.

The semileptonic decays account for about a quarter of the B meson decays and are dominated by only a few modes, as in Tab. 3.1. The large branching fractions of the $B \to D^{(*)}\ell\nu$ decays used for the reconstruction of the B_{tag} candidate lead to a relatively high efficiency for the semileptonic B-tagging of up to ∼2% [62]. However, the presence of neutrinos does not allow the reconstruction of the four-momentum of the B_{tag} candidate, resulting in the loss of crucial information.

Mode	B^+	B^0
$D\ell\nu$	$(2.30 \pm 0.09)\%$	$(2.24 \pm 0.09)\%$
$D^*\ell\nu$	$(5.58 \pm 0.22)\%$	$(4.97 \pm 0.12)\%$
$D^{**}\ell\nu$	$(1.79 \pm 0.12)\%$	$(1.67 \pm 0.11)\%$
gap to fill $X_c\ell\nu$	$\sim 1\%$	$\sim 1\%$
$D\tau\nu$	$(0.77 \pm 0.25)\%$	$(1.05 \pm 0.23)\%$
$D^*\tau\nu$	$(1.88 \pm 0.20)\%$	$(1.58 \pm 0.09)\%$
$D^{**}\tau\nu + X_u\tau\nu$	$\sim 0.5\%$	$\sim 0.5\%$
$X_u\ell\nu$	$(0.16 \pm 0.02)\%$	$(0.15 \pm 0.02)\%$
Total	∼25%	∼25%

TABLE 3.1: Branching fractions of semileptonic B decays from Ref. [34]. ℓ indicates an e or a μ mode, not a sum over these modes. The total includes both e and μ modes.

On the other hand, hadronic B decays account for the remaining three-quarters of total B meson decays. But even the largest decay modes among them have a branching fraction of only a few 10^{-3}, so the explicit reconstruction of some of them will cover only a modest branching fraction. Moreover, the reconstruction efficiency decreases rapidly with a higher number of particles in the decay chain. Overall, this leads to a low hadronic B-tagging efficiency on the order of 1% or less. But despite the low efficiency, hadronic B-tagging provides strong background suppression constraints, giving it an advantage over semileptonic B-tagging in terms of purity.

3.2 Hadronic B-tagging

Over time, several packages have been developed to perform hadronic B-tagging. Regardless of the package used, hadronic B-tagging provides important constraints based on the fact that the four-momentum of the B mesons can be reconstructed. These general constraints are described in Sec. 3.2.1, while the more package-dependent selection are described in later sections. The ability to reconstruct the direction of the B_{tag} candidate allows the observation of a missing particle on the signal side using the recoil mass method. This is discussed in Sec. 3.2.2 and is the main idea behind the $B \to D\pi$ calibration sample. A summary of previous hadronic B-tagging packages developed at the Belle experiment is then presented, focusing on their performance.

3.2.1 General selection of the hadronic B_{tag} candidates

If only hadronic B decays are considered for reconstructing B_{tag} candidates, the four-momentum can be obtained, providing the two variables ΔE and M_{bc}. The energy difference ΔE, is defined as

$$\Delta E = E_B^* - E_{\text{beam}}^* \tag{3.1}$$

where E_B^* is the reconstructed energy of the B candidates in the center of mass (CM) frame[1] and E_{beam}^* is the energy equal to half the total energy of the e^+e^- system in the CM frame. Since $\Upsilon(4S) \to B\bar{B}$ is a two-body decay with two daughters of the same mass, correctly reconstructed B mesons should have an energy equal to E_{beam}^*, i.e., ΔE should be centered around zero.

The uncertainty of ΔE is due to the error in the measurement of the energy of the B mesons, $\sigma^2_{E_B^*}$, and the beam energy spread, $\sigma^2_{E_{\text{beam}}^*}$:

$$\sigma^2_{\Delta E} = \sigma^2_{E_B^*} + \sigma^2_{E_{\text{beam}}^*}.$$

The resolution of the beam energy is small, less than $\sim$ 3 MeV. Therefore, the resolution of ΔE is dominated by the uncertainty of the reconstructed B meson energy, so that ΔE is in the range of [-0.150, 0.100] GeV depending on the final state. The range is asymmetric, with a tendency to lower values due to modes containing photons. The left plot in Fig. 3.2 shows the ΔE distributions for some of the hadronic B_{tag} candidates in MC. A clear difference in ΔE resolution can be seen between the $B^+ \to \overline{D}^0\pi^+$ and $B^+ \to \overline{D}^0\pi^+\pi^0\pi^0$ decay modes, which is due to additional π^0's in the final state that result in poorer E_B^* resolution. And ΔE shifts away from zero, for example to negative values, when a daughter is missing in the reconstruction (partially reconstructed events). This effect is demonstrated in the same figure by the $B^+ \to \overline{D}^{*0}\pi^+$ events, which are partially reconstructed as $B^+ \to \overline{D}^0\pi^+$ missing either a π^0 or a γ from the $\overline{D}^{*0}$ decay. Similarly, misidentification of a particle in the decay chain would shift the ΔE distribution. The candidates in the continuum events ($e^+e^- \to q\bar{q}$) should have a flat ΔE distribution. However, the FEI package uses ΔE as a feature when training the BDTs for B_{tag} candidates, as will be described in Sec. 3.3. This distorts the distribution as shown in Fig. 3.2. This bias towards $\Delta E = 0$ limits the use of ΔE as a control variable, and also limits the ability to recover partially reconstructed events from different bins of ΔE, as it will be discussed in Sec. 4.5.2.

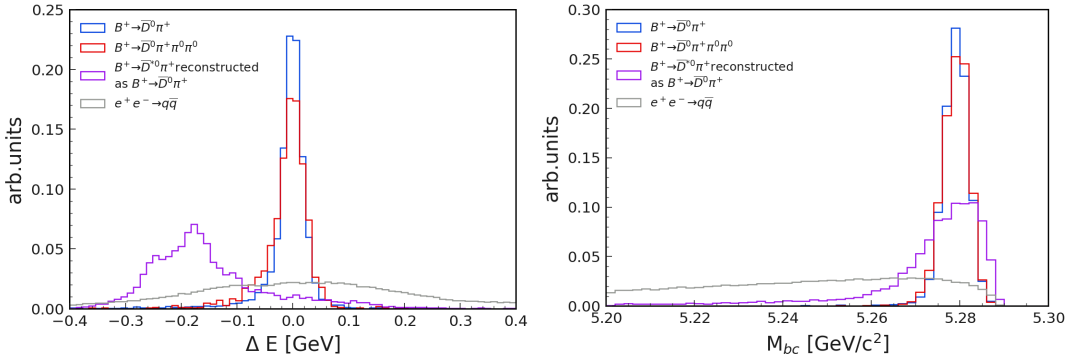


FIGURE 3.2: Left: Comparison of the ΔE (left) and M_{bc} (right) distributions of the B_{tag} candidates reconstructed by the FEI package through different B decay modes in the MC.

[1]From here on, all the observables denoted by an asterisk are measured in the CM frame, unless otherwise mentioned.

The beam-constrained mass defined as

$$M_{\mathrm{bc}} = \sqrt{E^{*2}_{\mathrm{beam}} - |\mathbf{p}^*_B|^2} \tag{3.2}$$

provides additional information. It uses the beam energy E^*_{beam} instead of the reconstructed energy of the B candidate. $\mathbf{p}^*_B$ represents the fully reconstructed three-momentum of the B candidate in the CM frame. Since the energy of the B candidate is replaced by a quantity derived from the beam parameters, the beam-constrained mass no longer depends on the mass hypotheses of the particles used to reconstruct the B meson.

For correctly reconstructed B candidates, M_{bc} takes values around $5.279\,\mathrm{GeV/c^2}$, which corresponds to the mass of the B meson. Since the B mesons are almost at rest in the CM frame, the uncertainty in the B momentum measurement (boosted to the CM frame) becomes small and the resolution of M_{bc} is dominated by the spread in the beam energy, giving $\sigma_{M_{\mathrm{bc}}} \sim 3\ \mathrm{MeV/c^2}$. This is illustrated in the right side plot in Fig. 3.2, which shows that the signal resolution of M_{bc} is much less affected by the uncertainty of the measured B meson four-momentum compared to ΔE. The incorrectly reconstructed candidates tend to have a broader distribution in M_{bc}, as shown in the same figure by the $B^+ \to \overline{D}^{*0}\pi^+$ events, which are partially reconstructed as $B^+ \to \overline{D}^0\pi^+$ and contribute to the peaking background. The M_{bc} distribution for continuum events is also shown. Most of these events have low M_{bc} values and can be suppressed by requiring M_{bc} to be greater than $5.27\ \mathrm{GeV}/c^2$.

Since ΔE and M_{bc} use different components of the four-momentum of B mesons, they are only weakly correlated. They are both very efficient at suppressing the background, as shown in Fig. 3.2. The fact that the ΔE is used as a classifier feature for the B_{tag} reconstruction in the FEI package, makes it an unreliable control variable. In general, the peaking background in M_{bc} and the dependence of ΔE on the B decay mode make those two variables unreliable for measuring the yields of B_{tag} candidates for calibration purposes.

3.2.2 Using the direction of the fully reconstructed B_{tag}

In the cases where the signal decay contains only one missing particle (either a single neutrino or a τ lepton which decays to neutrinos), the hadronic B-tagging also provides the full kinematic information of the missing particle. The four-momentum of the missing particle and the corresponding invariant mass are given respectively by:

$$p_{\mathrm{miss}} = p_{B_{\mathrm{sig}}} - p^{\mathrm{rec}}_{B_{\mathrm{sig}}}$$

$$M_{\mathrm{miss}} = \sqrt{\left(E_{B_{\mathrm{sig}}} - E^{\mathrm{rec}}_{B_{\mathrm{sig}}}\right)^2 - \left(\mathbf{p}_{B_{\mathrm{sig}}} - \mathbf{p}^{\mathrm{rec}}_{B_{\mathrm{sig}}}\right)^2}$$

where the $p_{B_{\mathrm{sig}}}$ represents the entire four-momentum of B_{sig} and $p^{\mathrm{rec}}_{B_{\mathrm{sig}}}$ represents the four-momentum of the reconstructed daughters of B_{sig}, except the missing particle. In the CM frame, where the $\Upsilon(4S)$ resonance is at rest, the two B mesons are back to back, hence:

$$\mathbf{p}^*_{B_{\mathrm{sig}}} = -\mathbf{p}^*_{B_{\mathrm{tag}}};$$

furthermore, the two B's have the same energy, which is half the energy of the $\Upsilon(4S)$:

$$E^*_{B_{\mathrm{sig}}} = E^*_{B_{\mathrm{tag}}} = \frac{\sqrt{s}}{2} = E^*_{\mathrm{beam}}\,.$$

As mentioned earlier, $\sigma^2_{E^*_{\rm beam}} << \sigma^2_{E^*_B}$, hence using $E^*_{\rm beam}$ instead of $E^*_{B_{\rm tag}}$ provides better resolution. With this transformation, the missing mass, also called as recoil mass, can be calculated as:

$$M_{\rm miss} = M_{\rm recoil} = \sqrt{\left(E^*_{\rm beam} - E^{\rm rec*}_{B_{\rm sig}}\right)^2 - \left(-\mathbf{p}^*_{B_{\rm tag}} - \mathbf{p}^{\rm rec*}_{B_{\rm sig}}\right)^2}. \tag{3.3}$$

A fit to this missing/recoil mass is used in searches like $B^+ \to \ell^+\nu_\ell\gamma$ [69], $B^0 \to \tau\ell$ [70], $B^+ \to K^+\tau\ell$ [25]. For example, in the search for $B^+ \to K^+\tau\ell$ where the τ cannot be exclusively reconstructed, the $p^{\rm rec}_{B_{\rm sig}}$ corresponds to $p_K + p_\ell$. The left plot in Fig. 3.3 shows the fit performed on the recoil mass in the τ mass range using the Belle data set for this search. As the right plot in Fig. 3.3 shows, the ability of hadronic B-tagging to reconstruct the four-momentum of $B_{\rm tag}$ leads to much better resolution for $M_{\rm recoil}$, as opposed to semileptonic B-tagging.

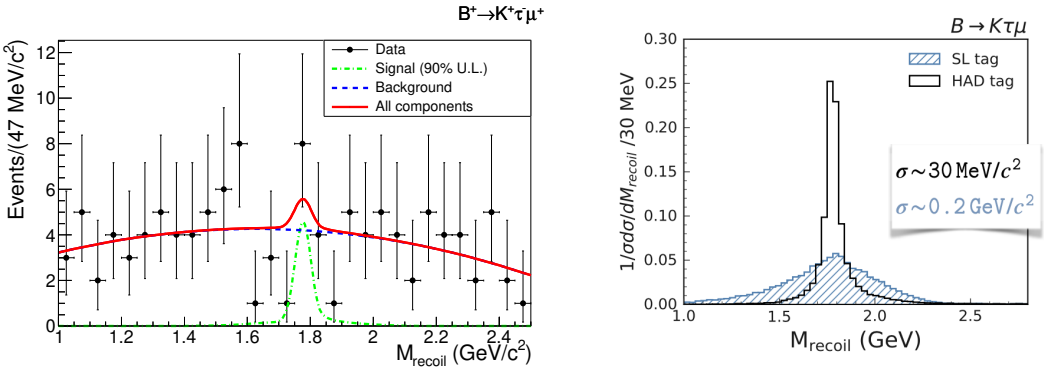


FIGURE 3.3: Left: Fit in data of the recoil mass in τ mass range in the search for $B^+ \to K^+\tau\ell$ decays. Taken from Ref. [25]. Right: Comparison of the resolution obtained with hadronic B-tagging (black) in contrast to the semileptonic B-tagging (blue).

This recoil mass method can also be used to reconstruct events without missing particles. Instead of reconstructing all the daughters of a B meson, one of them could be observed with the $M_{\rm recoil}$ of the $B^{\rm rec}_{\rm sig}$ and $B_{\rm tag}$ system. In the case of $B \to D\pi$, for example, instead of exclusively reconstructing the D meson in one of its decay modes, it can be observed through the recoil mass of the prompt π and the hadronic $B_{\rm tag}$. This provides a complementary sample to the exclusively reconstructed $B \to D\pi$ and simultaneously allows the observation of the $B \to D^*\pi$ and $B \to D^{**}\pi$ modes. This method is not limited to two-body decays of B, but can be extended to multi-body decays as long as exactly one daughter is not exclusively reconstructed. The method is often used to obtain a control sample for studies based on the recoil mass method. For example, the analysis procedure and the acceptance calculations in the search for $B^0 \to \tau\ell$ at the Belle experiment [70] are checked using a $B^0 \to D^{(*)-}\pi^+$ sample, as shown in the left plot of Fig. 3.4. The same method is used to measure the absolute branching fraction of the $B \to D^{**}\pi$ by *BABAR* [71]. Here, the $B^- \to D^{**0}\pi^-$ yield is measured by fitting the background-subtracted $M_{\rm miss}$ distribution in the mass range 2.2 - 2.8 GeV/c^2 as shown in the right plot of Fig. 3.4.

In our Belle and Belle II analysis, the $B \to D\pi$ signal reconstructed with a recoil mass method was introduced to evaluate the performance of the hadronic B-tagging, as explained in Sec. 3.4.2. The purity of this sample allows us to examine the kinematics of the $B_{\rm tag}$ modes used in the reconstruction, compare the data with MC, and identify the sources of disagreement, as described in Chap. 4.

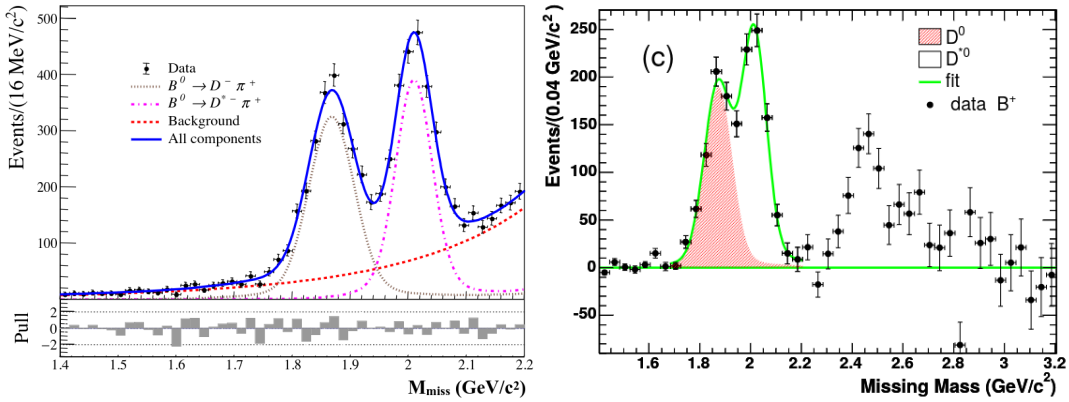


FIGURE 3.4: Left: The fit to $M_{\rm miss}$ distribution of $B^0 \to D^{(*)-}\pi^+$ control sample in Ref. [70]. Right: The fit of the background-subtracted $M_{\rm miss}$ distribution in Ref. [71].

In the search for $B^+ \to K^+\tau^+\tau^-$, the recoil mass method cannot be applied directly because two particles are missing on the signal side. Instead, only the invariant mass of the $\tau^+\tau^-$ system can be evaluated. However, hadronic B-tagging provides the best possible purity on the tag side using variables such as ΔE and $M_{\rm bc}$ described in Sec. 3.2.1. Given that this is the first search for $B^+ \to K^+\tau^+\tau^-$ in the Belle and Belle II experiments, hadronic B-tagging is a preferred option. Before describing the hadronic B-tagging algorithm used for the search, a brief summary of previous algorithms developed at the Belle experiment follows.

3.2.3 Hadronic B-tagging algorithms at the Belle experiment

In the Belle experiment, the first B-tagging algorithm, the initial version of the *Full Reconstruction*, followed a cut based approach [72]. For hadronic B^+-tagging[2], it exclusively reconstructed candidates through the modes $B^+ \to \overline{D}^{(*)0}(\pi, \rho, a_1, D_s^{(*)})^+$, with $\overline{D}^{*0}$ mesons reconstructed by two different decay modes and $\overline{D}^0$ reconstructed in seven modes with a rather loose selection (e.g., 4σ to 5σ mass window).

The reconstruction efficiency of each exclusive B-tagging tool is essentially the sum of the efficiencies multiplied by the branching fractions of the hadronic decays involved in the reconstruction of the $B_{\rm tag}$ candidates:

$$\varepsilon_{\rm tag} = \sum_i^N \varepsilon_i \mathcal{B}_i \,. \tag{3.4}$$

where the ε_i terms take into account geometrical acceptance (particles escaping the detector), tracking efficiency, particle-ID requirements, photon reconstruction efficiency, etc. Although certain modes with high combinatorics, such as those with many π^0's in the final state, must be avoided, the cut-based Full Reconstruction algorithm had a B-tagging efficiency of 0.14% for correctly reconstructing B^+ candidates.

In the next version of the *Full Reconstruction* (FR), neural networks trained with Monte Carlo (MC) were used instead of rectangular cuts [73, 74]. The particle candidates are reconstructed hierarchically, starting the particle from the final state, which are combined to form heavier particles until B candidates are formed. At each stage, a neural network output is computed for the candidate and propagated to form heavier particles. The neural networks allow the usage of additional information in the

[2]From here on, the focus is on the charged B-tagging.

selection process, such as the shower shape of the photon candidates, the particle identification of the charged particles in the final state, the invariant mass, and the angular information of the daughters. Including more information in the training can lead to better separation between correctly and incorrectly reconstructed candidates. By increasing the number of B^+ decay modes (to fifteen) along with D decay modes, i.e., by increasing the branching fraction covered, the neural network-based FR achieves an efficiency of 0.28% for B^+-tagging, while significantly improving purity. The availability of a single scalar neural network output, o_{NB}, which tends to zero for background-like candidates and to one for signal-like candidates and can be interpreted as the signal probability, makes it possible to increase the purity of B_{tag} candidates at the expense of efficiency. This ability to be interpreted as Bayesian probability is provided by the NeuroBayes [75], the neural network package used, given there is no overtraining. The output o_{NB} for B candidates in the final stage also allows ranking and selection of the best candidate.

As described in Eq. 3.4, the efficiency of B-tagging depends on the branching fractions ($\mathcal{B}_i$) of the decay modes involved. In MC, the branching fractions are determined by interpreting the measurements performed so far [34, 76]. Due to limitations in the measurements and the resulting poor modelling of MC, it is expected that the performance in the data will be different from MC, which can have a significant impact on the physics results. By combining the B_{tag} with a B meson reconstructed in a channel with a large and already accurately measured branching fraction (like the semileptonic decays $B \to X_c \ell \nu$), the reconstruction efficiency can be measured for each hadronic decay mode. Averaging over all decay modes, one obtains a correction factor $\epsilon_{\mathrm{Data}}/\epsilon_{\mathrm{MC}} \approx 0.75$; for many modes, the estimated reconstruction efficiency is significantly larger in MC [77].

3.3 Full Event Interpretation

For Belle II, the exclusive B-tagging algorithm is rewritten with modern tools (Python interface instead of `C++`; FastBDT [78] instead of NeuroBayes) under the name *Full Event Interpretation* (FEI) to further increase the reconstruction efficiency, automate the training and to make it more modular, i.e., simplify the configuration of the decay channels and the input variables for the classifiers [79]. To increase the covered branching fraction of B decays, it is only necessary to specify additional channels that are automatically converted into reconstruction tasks, such as selecting or combining particles, fitting the decay vertex, or training a multivariate classifier. These tasks are performed simultaneously, but some of them depend on the results of others. Therefore, FEI training automatically resolves the dependencies and orders the tasks for parallel execution.

Like the neural network-based Full Reconstruction algorithm, the FEI package builds the B candidates in a hierarchical fashion, as shown in Fig. 3.5. The training starts with the preparation of final state particles $\left(e^{\pm}, \mu^{\pm}, K^{\pm}, \pi^{\pm}, p^{\pm}, K_L^0, \gamma\right)$ directly from the tracks, ECL and KLM clusters. Before feeding them into the associated multivariate classifiers, the combinatorics are reduced to a reasonable level through a series of pre-cuts without significantly affecting efficiency. For charged tracks, the pre-cuts are on the track impact parameters, the radial and z distances from the beam line to the closest track point, in order to reject secondary particles produced from the interaction with the detector. For photons, different energy thresholds are used in different regions of the detector, as mentioned in Tab. 3.2. Tighter cuts are used in the endcap regions compared to the barrel to suppress the higher beam

backgrounds. The charged particle and photon candidates are ranked by particle identification probability and energy, respectively. A cut is applied to this rank to limit the number of candidates per event used to train the classifier. This ensures that combinatorial tracks and photons do not bias the classifier training.

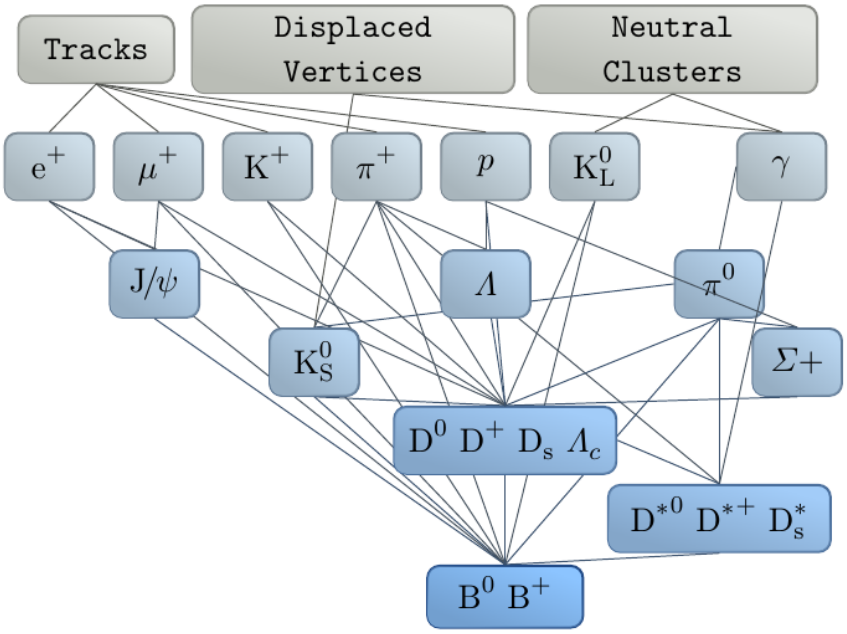


FIGURE 3.5: Hierarchical reconstruction of B candidates in the FEI algorithm. Taken from Ref. [62].

For each final state particle, individual BDT classifiers are trained based on FastBDT, with the target variable indicating whether the reconstructed particle matches the generated one. It is ensured that the total number of candidates in the training is appropriately limited so that both classes (signal and background) are balanced. The charged tracks are trained using the following features: impact parameters, PID likelihoods, and other variables that affect the performance of the PID such as momentum and track quality (χ^2 of the track fit). Photon classifiers use detector region information to remove the beam background (which tends to be more abundant in the endcap regions), as well as cluster timing, energy, direction, and shower shape information. In both cases, the rank computed as part of the pre-cut is also used as a feature for the classifiers. After training, a cut is applied to the classifier output ($\mathcal{P}_{\text{FEI}}$), which embeds information from all classifier input variables and their correlations to improve selection. Depending on the particle, only certain top candidates ranked by $\mathcal{P}_{\text{FEI}}$ are saved in order to reduce combinatorics in forming combined particles from the final state particles. The exact number of candidates saved per event is shown in the FEI post-cuts column in Tab. 3.2.

After all final state particles have been trained, the combined particles of the next stage are reconstructed as shown in Fig. 3.5. π^0 candidates are built up from photon pairs. After a loose pre-cut to the invariant mass to remove the obvious background, the classifiers are trained using the following features: the invariant mass, the angle between the photons, the energy, the direction and the classifier outputs of the daughters. After applying a cut to its $\mathcal{P}_{\text{FEI}}$ and selecting the first ten candidates, the π^0 candidates are propagated to form heavier combined particles.

Particle List	FEI pre-cuts	FEI post-cuts
$e^{\pm}$, $\mu^{\pm}$	$dr < 2\,\text{cm}, \lvert dz\rvert < 4\,\text{cm}$, 10 hi ℓID	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 5 hi $\mathcal{P}_{\text{FEI}}$
$K^{\pm}$, $\pi^{\pm}$, $p^{\pm}$	$dr < 2\,\text{cm}, \lvert dz\rvert < 4\,\text{cm}$, 20 hi pID	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 10 hi $\mathcal{P}_{\text{FEI}}$
γ in Belle	$E > \{0.10, 0.05, 0.15\}$ GeV in {forward, barrel, backward}, `clusterBelleQuality`=0, 40 hi E	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 20 hi $\mathcal{P}_{\text{FEI}}$
γ in Belle II	$E > \{0.10, 0.09, 0.16\}$ GeV in {forward, barrel, backward} regions, 40 hi E	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 20 hi $\mathcal{P}_{\text{FEI}}$
π^0	$0.08 < M < 0.18$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 10 hi $\mathcal{P}_{\text{FEI}}$
J/ψ	$2.6 < M < 3.7$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-3}$, 10 hi $\mathcal{P}_{\text{FEI}}$
Λ	$0.9 < M < 1.3$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 10 hi $\mathcal{P}_{\text{FEI}}$
K^0_S	$0.4 < M < 0.6$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 10 hi $\mathcal{P}_{\text{FEI}}$
Σ^+	$1.0 < M < 1.4$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-2}$, 10 hi $\mathcal{P}_{\text{FEI}}$
$D^0/\bar{D}^0, D^{\pm}$	$1.7 < M < 1.95$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-3}$, 10 hi $\mathcal{P}_{\text{FEI}}$
$D^{\pm}_s$	$1.68 < M < 2.1$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-3}$, 10 hi $\mathcal{P}_{\text{FEI}}$
Λ_c	$2.2 < M < 2.4$, 20 lo $\lvert dM\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-3}$, 10 hi $\mathcal{P}_{\text{FEI}}$
$D^{*0}/\bar{D}^{*0}, D^{*\pm}_{(s)}$	$0.0 < Q < 0.3$, 20 lo $\lvert dQ\rvert$	$\mathcal{P}_{\text{FEI}} > 10^{-3}$, 10 hi $\mathcal{P}_{\text{FEI}}$
$B^{\pm}$	$\lvert\Delta E\rvert < 0.5\,\text{GeV}$, $M_{\text{bc}} > 5.2\,\text{GeV}/c^2$ 20 hi $\Pi_{daughters}\ \mathcal{P}_{\text{FEI}}$	20 hi $\mathcal{P}_{\text{FEI}}$

TABLE 3.2: FEI selection: pre-cuts and post-cuts applied to each particle lists before and after dedicated BDT trainings from Ref. [62]. ℓID stands for electron-ID or muon-ID for electrons and muons, respectively. pID stands for binary ratio of kaon vs. pion ($K^{\pm}$), pion vs. kaon ($\pi^{\pm}$), proton vs. kaon ($p^{\pm}$) (see Eq. 2.7). 'hi' stands for 'highest' and 'lo' stands for lowest'.

For other combined particles, individual classifiers are trained for each of their decay modes after corresponding pre-cuts. The classifiers for $\overline{D}^0$ and other intermediate particles use features in addition to those used for π^0. These include the invariant masses and angles between each combination of daughters, χ^2 based on the vertex fit, the distance to the vertices of the daughters, and the momenta and angles of all daughters in the candidate's rest frame. For the pre-selection of D^* particles, instead of the invariant mass, a cut is applied to the released energy Q, defined as the difference between the mass of the candidate and the masses of all daughters, due to its greater separation power. In order to use the candidates in later stages regardless of their specific decay mode, the candidates from all decay modes are combined into a single list for each particle.

In the final stage of FEI, where the B meson candidates are formed, the beam-constrained mass M_{bc} is often used as a control variable. This makes the use of variables that are strongly correlated with M_{bc}, such as the invariant mass, suboptimal for a pre-selection and classifier feature. To avoid these correlations, this is replaced by very loose pre-selection on ΔE and M_{bc}, and the product of each daughter's classifier outputs is used to select the best candidates before training. Also, all invariant mass combinations are removed from the classifier training variables, although the momenta and angles of the daughters are still used. And since the energy difference ΔE can clearly distinguish correctly reconstructed B meson candidates from those with missing or misidentified particles and is only weakly correlated with M_{bc}, it is added as a classifier input for training and turns out to be the most powerful variable. As a result, the ΔE distribution of the B_{tag} candidates produced by the FEI package is distorted and peaks around zero.

The decays reconstructed with the FEI package for $\overline{D}^0$, $\overline{D}^{*0}$ and B^+ are shown in Tabs. 3.3 and 3.4. When all possible decay modes are considered, through combinations, this corresponds to over $\mathcal{O}(10^3)$ exclusive decay channels. On the one hand, it is important that the list of decay modes for each particle covers a large fraction of the decays. On the other hand, decay modes with low branching fraction can also be useful if the reconstructed candidates are sufficiently pure. Therefore, all aspects such as branching fraction, reconstruction efficiency and associated purity should be considered when adding new decay modes to B-tagging reconstruction. Overall, the number of decay modes for D and B mesons was significantly increased in the FEI package compared to the Full Reconstruction, which affects the effective B-tagging efficiency. In addition, the selection process that candidates must go through has also been significantly improved.

It is important to note that the decays listed in Tab.3.4 are not mutually exclusive. A B decay can be reconstructed in more than one way within the FEI package. For example, the $B^+ \to \overline{D}^{*0}\pi^+$ decay with $\overline{D}^{*0} \to \overline{D}^0\pi^0$ can form two different candidates:

- Correctly reconstructed $B^+ \to \overline{D}^{*0}\pi^+$ with $\overline{D}^{*0} \to \overline{D}^0\pi^0$; or
- $B^+ \to \overline{D}^0\pi^+\pi^0$ with the same final state, but the intermediate $\overline{D}^{*0}$ is not formed.

The FEI package produces both candidates in the event, and it is ideal to choose the first candidate, with the $\overline{D}^{*0}$ constraint whenever possible, with appropriate best candidate selection. This can be done with $\mathcal{P}_{\mathrm{FEI}}$, as will be shown later. Similarly, there is overlap of many modes with the same final state.

The FEI package, as described so far, is centrally trained on a Belle II run-independent MC sample, which corresponds to 200 fb^{-1} and consists of mixed ($B^0\overline{B}^0$) and charged (B^+B^-) events, with generically decaying B mesons. Here, run-independent

ID	Decay Mode	
0.	$D^0 \to K^-\pi^+$	$D^{*0} \to D^0\pi^0$
1.	$D^0 \to K^-\pi^+\pi^0$	$D^{*0} \to D^0\gamma$
2.	$D^0 \to K^-\pi^+\pi^0\pi^0$	
3.	$D^0 \to K^-\pi^+\pi^+\pi^-$	
4.	$D^0 \to K^-\pi^+\pi^+\pi^-\pi^0$	
5.	$D^0 \to \pi^-\pi^+$	
6.	$D^0 \to \pi^-\pi^+\pi^+\pi^-$	
7.	$D^0 \to \pi^-\pi^+\pi^0$	
8.	$D^0 \to \pi^-\pi^+\pi^0\pi^0$	
9.	$D^0 \to K^0_S\pi^0$	
10.	$D^0 \to K^0_S\pi^+\pi^-$	
11.	$D^0 \to K^0_S\pi^+\pi^-\pi^0$	
12.	$D^0 \to K^-K^+$	
13.	$D^0 \to K^-K^+\pi^0$	
14.	$D^0 \to K^-K^+K^0_S$	
0.	$D^+ \to K^-\pi^+\pi^+$	
1.	$D^+ \to K^-\pi^+\pi^+\pi^0$	
2.	$D^+ \to K^-K^+\pi^+$	
3.	$D^+ \to K^-K^+\pi^+\pi^0$	
4.	$D^+ \to \pi^+\pi^0$	
5.	$D^+ \to \pi^+\pi^+\pi^-$	
6.	$D^+ \to \pi^+\pi^+\pi^-\pi^0$	
7.	$D^+ \to K^0_S\pi^+$	
8.	$D^+ \to K^0_S\pi^+\pi^0$	
9.	$D^+ \to K^0_S\pi^+\pi^+\pi^-$	
10.	$D^+ \to K^+K^0_SK^0_S$	

TABLE 3.3: FEI D mesons decay modes.

ID	Decay Mode
0.	$B^+ \to \overline{D}^0\pi^+$
1.	$B^+ \to \overline{D}^0\pi^+\pi^0$
2.	$B^+ \to \overline{D}^0\pi^+\pi^0\pi^0$
3.	$B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$
4.	$B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$
5.	$B^+ \to \overline{D}^0D^+$
6.	$B^+ \to \overline{D}^0D^+K^0_S$
7.	$B^+ \to \overline{D}^{*0}D^+K^0_S$
8.	$B^+ \to \overline{D}^0D^{*+}K^0_S$
9.	$B^+ \to \overline{D}^{*0}D^{*+}K^0_S$
10.	$B^+ \to \overline{D}^0D^0K^+$
11.	$B^+ \to \overline{D}^{*0}D^0K^+$
12.	$B^+ \to \overline{D}^0D^{*0}K^+$
13.	$B^+ \to \overline{D}^{*0}D^{*0}K^+$
14.	$B^+ \to D^+_s\overline{D}^0$
15.	$B^+ \to \overline{D}^{*0}\pi^+$
16.	$B^+ \to \overline{D}^{*0}\pi^+\pi^0$
17.	$B^+ \to \overline{D}^{*0}\pi^+\pi^0\pi^0$
18.	$B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$
19.	$B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$
20.	$B^+ \to D^{*+}_s\overline{D}^0$
21.	$B^+ \to D^+_s\overline{D}^{*0}$
22.	$B^+ \to \overline{D}^0K^+$
23.	$B^+ \to D^-\pi^+\pi^+$
24.	$B^+ \to D^-\pi^+\pi^+\pi^0$
25.	$B^+ \to J/\psi K^+$
26.	$B^+ \to J/\psi K^+\pi^+\pi^-$
27.	$B^+ \to J/\psi K^+\pi^0$
28.	$B^+ \to J/\psi K^0_S\pi^+$
29.	$B^+ \to \bar{\Lambda}_c p\pi^+\pi^0$
30.	$B^+ \to \bar{\Lambda}_c p\pi^+\pi^-\pi^+$
31.	$B^+ \to \overline{D}^0p\bar{p}\pi^+$
32.	$B^+ \to \overline{D}^{*0}p\bar{p}\pi^+$
33.	$B^+ \to D^+p\bar{p}\pi^+\pi^-$
34.	$B^+ \to D^{*+}p\bar{p}\pi^+\pi^-$
35.	$B^+ \to \bar{\Lambda}_c p\pi^+$

TABLE 3.4: FEI B^+ meson decay modes.

MC essentially means that the beam-background in the events is simulated and is known to be overestimated in the calorimeter compared to the data. Further details are given in Sec. 5.1. The continuum events ($e^+e^- \to q\bar{q}$) were not used for the training because it was not clear whether the particle candidates contained in them, such as D, Λ, π^0, should be considered as signal or background. For example, a $\overline{D}^0$ meson in a $B\overline{B}$ event and a continuum event have certain features in common such as invariant mass, but certain features such as momentum would be very different. Therefore, continuum suppression is expected to be applied independently of the FEI package, either through a series of cuts or through other tools specifically designed for this purpose. Once training is complete, the weights of all individual BDTs are made available to the entire collaboration to apply to the reconstruction.

For validation, the hadronic B^+ FEI candidates can be reconstructed using these weights in an independent sample of Belle II run-independent MC of $B^0\overline{B}^0$ and B^+B^- events with a loose cut to the classifier output of B ($\mathcal{P}_{\text{FEI}}$ of $B^+ > 0.001$), discarding some background candidates without significantly affecting the correctly reconstructed particle candidates. As described in Sec. 3.2.1, additional cuts of $-0.15 < \Delta E < 0.1$ GeV and $M_{\text{bc}} > 5.27$ GeV/c^2 can further reduce the trivial combinatorial background, as shown in Fig. 3.6. Since the FEI package can produce many B_{tag} candidates for each collision event, including all candidates would lead to double counting. Therefore, the best candidate selection was applied to the B_{tag} particle lists by selecting only the candidate with the highest classifier output in each event.

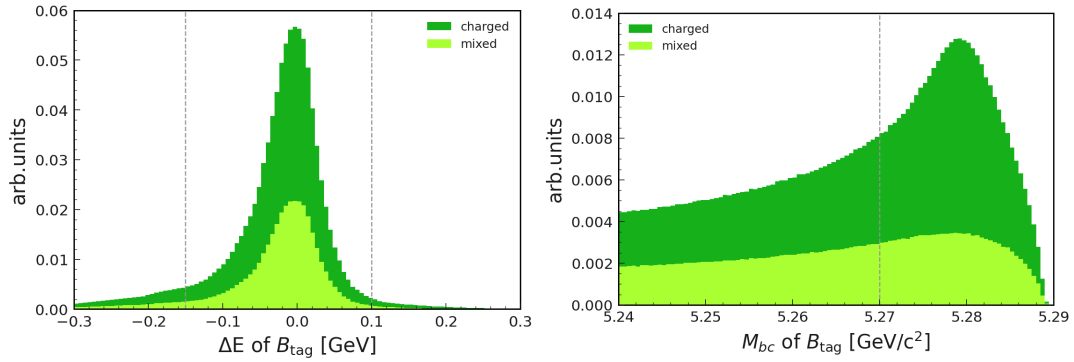


FIGURE 3.6: ΔE and M_{bc} distributions of B_{tag} candidates reconstructed by the hadronic FEI package with $\mathcal{P}_{\text{FEI}} > 0.001$ cut.

For the B_{tag} candidates that pass these cuts, the $\log_{10}$ transformation of the FEI classifier output ($\mathcal{P}_{\text{FEI}}$) is shown in the left plot in Fig. 3.7. Here, the signal is defined as the reconstructed B^+ candidates in the B^+B^- MC that match the generated particles, while the background is the candidates that do not match the generated particles in the B^+B^- MC, along with all candidates in the $B^0\overline{B}^0$ MC. As expected, the signal in this distribution peaks at 0, i.e., $\mathcal{P}_{\text{FEI}}$ equals 1, while the background tends to stay at lower values, corresponding to $\mathcal{P}_{\text{FEI}}$ equals 0. In each of these bins, the purity can be defined as the fraction of events that are signal. This is calculated and shown in the right plot of Fig. 3.7.

This is called the reliability diagram of a classifier, which shows the proportion of items in each class for bands of predicted probability [80]. In this case, it shows that the relationship between purity and $log_{10}(\mathcal{P}_{\text{FEI}})$ is approximately linear. Thus, based on the criteria the:

- The $\mathcal{P}_{\text{FEI}}$ tends to zero for background-like candidates and to one for signal-like candidates;

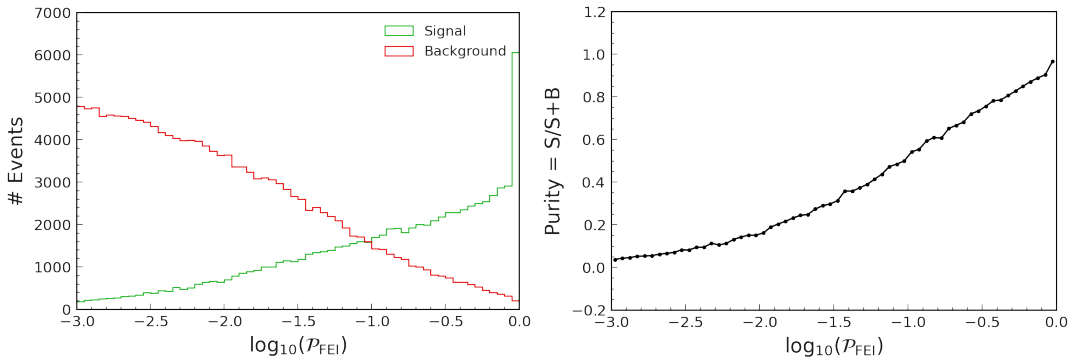


FIGURE 3.7: Left: Log of the classifier output ($\mathcal{P}_{\text{FEI}}$) of B candidates in the hadronic B^+ FEI package after minimal cuts and best candidate selection. Signal represents the truth matched B candidates in the B^+B^- MC, while Background represents the non-truth matched candidates in B^+B^- MC and all the candidates in $B^0\overline{B}^0$ MC.
Right: Purity (the fraction of signal events) calculated in each bin is plotted as a function of the log of classifier output ($\mathcal{P}_{\text{FEI}}$).

- The purity or empirical probability, defined as the fraction of signal events, increases with the predicted probability ($\mathcal{P}_{\text{FEI}}$);

the classifier output ($\mathcal{P}_{\text{FEI}}$) can be interpreted as **signal probability**.

As listed in Tab. 3.4, the FEI package reconstructs B^+ mesons in 36 modes and, as mentioned earlier, can produce multiple candidates of different modes in the same event. Some modes have higher purity than others based on the constraints available from the intermediate and final state particles involved. For example, even with the same final state, the $B^+ \to \overline{D}^{*0}\pi^+$ mode has higher purity than the $B^+ \to \overline{D}^0\pi^+\pi^0$ mode, because of the excellent separating power of the release energy, Q, when $\overline{D}^{*0}$ is reconstructed. This can be observed in Fig. 3.8. The candidates reconstructed as $B^+ \to \overline{D}^{*0}\pi^+$ have significantly higher values for $\mathcal{P}_{\text{FEI}}$, corresponding to higher purity. Thus, when both candidates are provided by the FEI package, choosing the candidate with the highest $\mathcal{P}_{\text{FEI}}$ in the event is equivalent to choosing the purest candidate in the event. As a result, $\mathcal{P}_{\text{FEI}}$ is used from now on to select the best candidate on the tag side.

In fact, all 36 modes have very different $\mathcal{P}_{\text{FEI}}$ distributions, as shown for the modes with the highest efficiency in Fig. 3.9.

Even among candidates of a single FEI B^+ mode, the $\mathcal{P}_{\text{FEI}}$ distribution varies greatly depending on how the event is generated in MC. For illustration, Fig. 3.10 shows the candidates reconstructed as $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$. Here, three of many possible B decays that produce the same final state $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$ are shown. The events generated via the intermediate D^{*-} state have larger $\mathcal{P}_{\text{FEI}}$ compared to events decaying via an ω meson or in a non-resonant way. This is again expected due to the separating power of the release energy Q when reconstructing D^{*-}. This demonstration shows that the FEI package learns the kinematics of the decay and $\mathcal{P}_{\text{FEI}}$ encompasses various selections made on all the classifier input variables. In this way, $\mathcal{P}_{\text{FEI}}$ provides an easy way to clean the tagged B candidates, even if it comes at the cost of efficiency. And since the FEI package is trained on MC, this means that the selection that BDT learns in the form of $\mathcal{P}_{\text{FEI}}$ depends heavily on the decay (kinematic) modelling of the MC. Therefore, correct modelling of B decays in MC is essential for optimal performance of the FEI package.

A realistic analysis should also take into account the continuum background. To demonstrate this, the continuum MC is included with some minimal cuts. Fig. 3.11

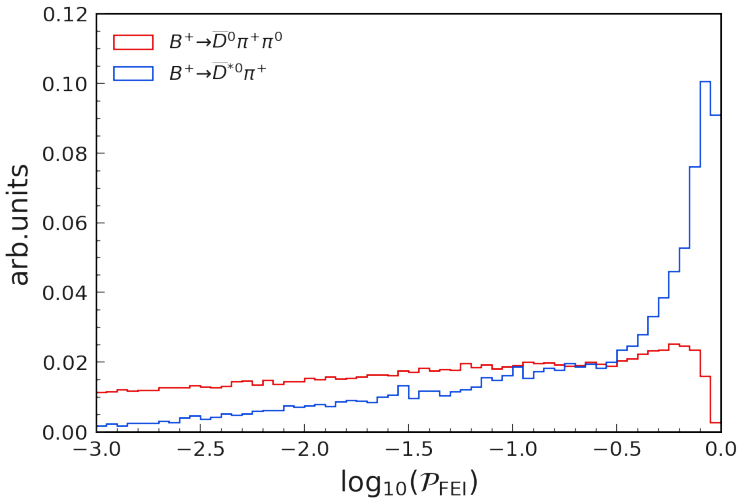


FIGURE 3.8: Log of the classifier output ($\mathcal{P}_{\mathrm{FEI}}$) of B candidates in the hadronic B^+ FEI package, reconstructed in two different modes: $B^+ \to \overline{D}^0\pi^+\pi^0$ and $B^+ \to \overline{D}^{*0}\pi^+$.

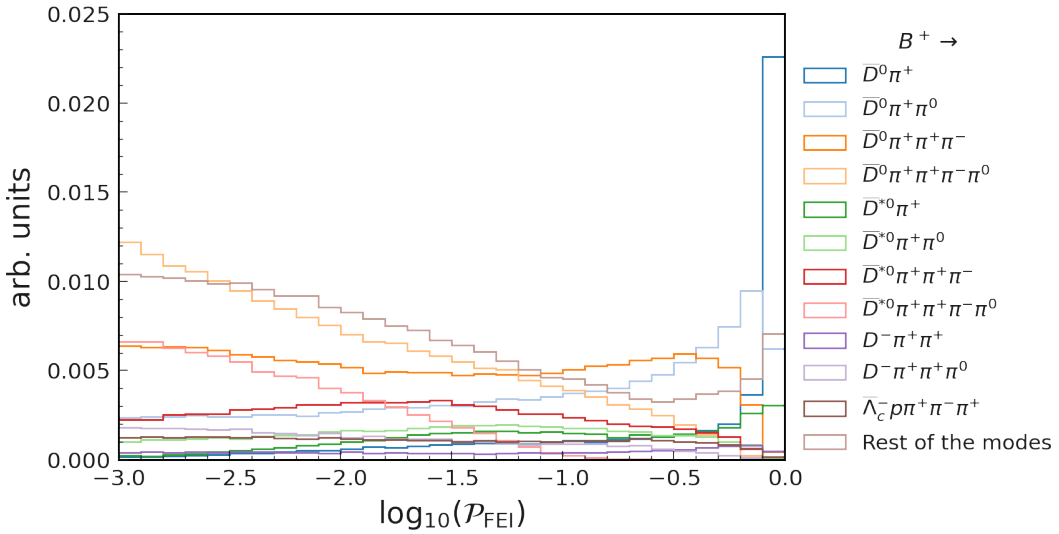


FIGURE 3.9: Log of the classifier output ($\mathcal{P}_{\mathrm{FEI}}$) of B candidates in the hadronic B^+ FEI package, reconstructed in different modes.

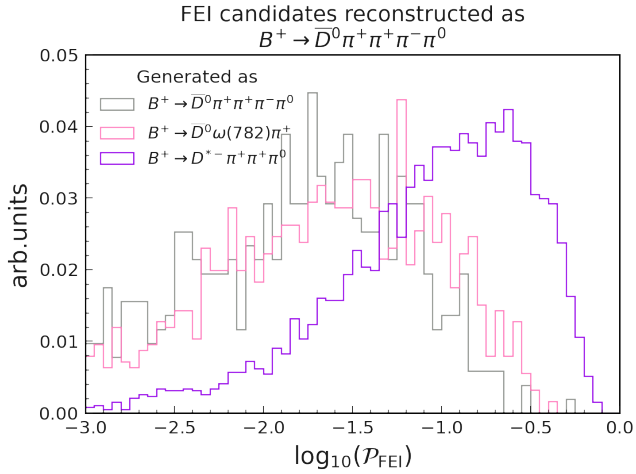


FIGURE 3.10: Log of the classifier output ($\mathcal{P}_{\rm FEI}$) of B candidates in the hadronic B^+ FEI package reconstructed as $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$, split based on how the candidate is generated.

shows the $\mathcal{P}_{\rm FEI}$ distribution with the signal and background candidates (in different types of MC). Although the continuum component dominates the number of candidates, they are shifted to very low values in the $\mathcal{P}_{\rm FEI}$, showing that the FEI package is efficient at selecting $B\overline{B}$ events. Correctly reconstructed candidates, on the other hand, are clustered at high values and can therefore be well separated. The rest of the background consist of candidates from events containing $\Upsilon(4S) \to B\bar{B}$, as shown earlier. When using B^+-tagging, there is a higher chance of producing a bad $B_{\rm tag}$ candidate in a B^+B^- event than in a $B^0\overline{B}^0$ event. Therefore, the charged background is larger than the mixed in Fig. 3.11. This would be reversed if B^0-tagging is used.

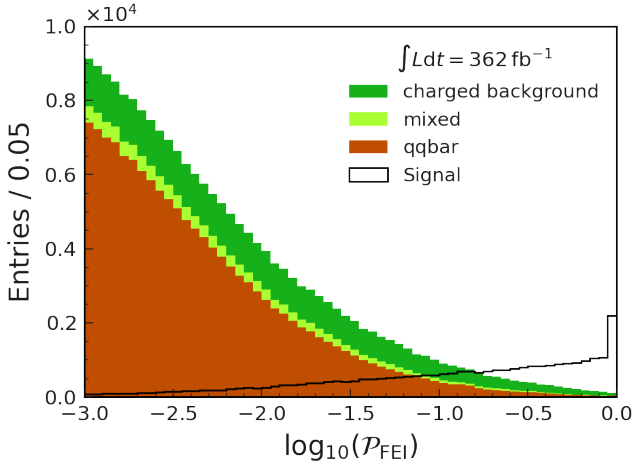


FIGURE 3.11: Log of the classifier output ($\mathcal{P}_{\rm FEI}$) of B candidates in the hadronic B^+ FEI after minimal cuts and best candidate selection. Signal represents the truth matched B candidates in the B^+B^- MC, while Background is the rest of MC components.

As described in Sec. 2.3.1, Belle data and MC can now be converted to Belle II format, which means that all tools (including the FEI package) built in basf2 for Belle II can also be used for Belle data. As with Belle II, the FEI package is centrally trained with $\sim$200 fb^{-1} Belle MC and the weights are made available to the entire

collaboration. However, while the MC used for Belle II contains simulated beam background, Belle MC contains beam background from randomly triggered events from real data. In principle, the performance of FEI is expected to be similar for the Belle and Belle II data sets, except for minor variations due to PID performance and the number of fake tracks.

When the FEI package was introduced [62], a maximum tag-side efficiency of 0.76% for hadronic B^+ was measured in Belle MC by fitting the M_{bc} with a Crystal Ball function for the signal and an ARGUS function for the background. To compare the performance of hadronic FEI with that of FR in a well-defined way that is independent of the signal side, both algorithms are applied to the same set of ten million events randomly sampled from the entire data set of the Belle experiment. By fitting the data, the tag-side efficiency and tag-side purity for different cuts on the $\mathcal{P}_{\mathrm{FEI}}$ can be calculated from the signal and background yields in a window of $5.27 < M_{\mathrm{bc}} < 5.29\ \mathrm{GeV}/c^2$. The resulting ROC curve is shown in Fig. 3.12. The maximum tag-side efficiency in FEI is significantly higher than in FR at 0.28%.

For hadronic B-tagging, the purity can be increased with a cut to $\mathcal{P}_{\mathrm{FEI}}$ to about 95%. However, the purity decreases rapidly with FEI, which may indicate that the FEI package is less stable with respect to background suppression. Note that the definition of efficiency is independent of the presence of continuum background. However, the purity, describing the ability to suppress the background, depends on the continuum background and the corresponding suppression.

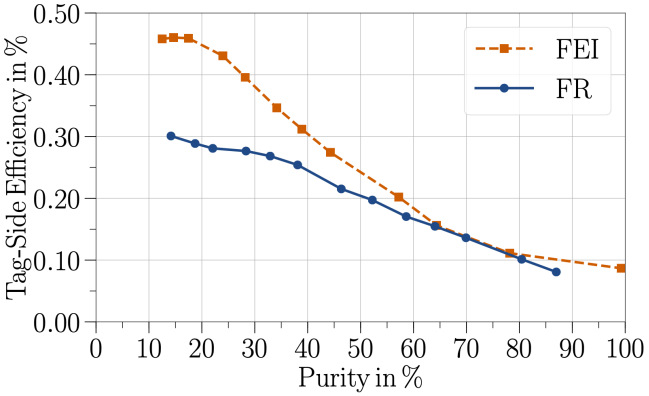


FIGURE 3.12: Tag-side efficiency as a function of the purity of hadronic B^+-tags. The *Full Reconstruction* performance is compared to the *Full Event Interpretation* on the same set of 10M events randomly sampled from the entire data set of the Belle experiment. Taken from Ref. [62].

The fact that the maximum B_{tag} efficiency calculated in the Belle data, shown in Fig. 3.12, is lower than the 0.76% calculated on Belle MC shows that the discrepancy between the data and MC is seen with the FEI package as well as with FR. To estimate the corresponding calibration factor, the hadronic B_{tag} candidates of the FEI package are combined with a set of semileptonic B decays with known branching fractions. Comparing the efficiency of their reconstruction with Belle data and MC, a calibration factor of $0.75 \pm 0.01 \pm 0.05$ was obtained, the first uncertainty being statistical and the second systematic [62].

A later measurement of the calibration factor of the hadronic FEI package with Belle II data and MC of only 35.6 fb^{-1} gave 0.65 ± 0.02 for B^+ [81]. As discussed in Ref. [1] and shown later, the main reason for this discrepancy is the poor MC modelling of the B^+ decays involved. This suggests that the MC simulation of Belle II

has not improved as expected, but has different problems than the MC of Belle. This discrepancy in B_{tag} yield between MC and the data, regardless of signal selection, is considered one of the major systematics in the search for $B^+ \to K^+\tau^+\tau^-$ by the *BABAR* experiment [1]. Therefore, understanding the main sources of this discrepancy and potentially correcting the MC is crucial for the search for $B^+ \to K^+\tau^+\tau^-$ or more generally for the search for decays with missing energy.

3.4 Performance of hadronic FEI at Belle and Belle II

Consider a data set with $N(B\bar{B})$ events of $B\bar{B}$ pairs. If the FEI package produces at least one B_{tag} candidate in T events, it may produce a correctly reconstructed B_{tag} candidate only in S events, called signal events. The rest of the events, $T-S=B$, have only bad B_{tag} candidates and can be called background events. Then the fundamental metrics for evaluating the performance of a B-tagging algorithm can be defined as follows:

$$\text{Tag-side efficiency: } \epsilon_{\text{FEI}} = \frac{S}{N(B\bar{B})}; \tag{3.5}$$

$$\text{Purity: } pur = \frac{S}{T} = \frac{S}{S+B}; \tag{3.6}$$

$$\text{Calibration factor: } CF = \frac{S^{data}}{S^{MC}} = \frac{\epsilon_{\text{FEI}}^{data}}{\epsilon_{\text{FEI}}^{MC}}. \tag{3.7}$$

Here, S^{data} and S^{MC} represent the number of events with correctly reconstructed B_{tag} candidates in the data and MC of the same size (integrated luminosity), respectively.

Individual hadronic B decays have low branching fractions ($\mathcal{B}_i$) and often produce many particles in the final state, leading to a very low reconstruction efficiency (ϵ_i). This means that the overall efficiency of the hadronic FEI package for B^+, given by the sum in Eq. 3.4, should be very low ($\epsilon_{\text{FEI}} \sim 1\%$ or less). However, the purity of this hadronic B-tagging is typically in the $\mathcal{O}(10)\%$ range, much higher than other B-tagging techniques.

Earlier measurements of the calibration factors of the hadronic FEI package already showed large discrepancies between data and MC. Since then, the FEI package has changed considerably: some baryonic B decays are now included, while the MC samples used to train the FEI package were produced with a regularly updated model of B decays. It is therefore imperative that the calibration factors are re-measured along with other metrics – efficiency and purity – for the latest FEI algorithm, the latest MC generation, and the entire data sets available in the Belle and Belle II data sets.

The challenge in evaluating the calibration factors is to determine the number of events with correctly reconstructed B_{tag} candidates, i.e., to separate signal B_{tag} from background candidates. As mentioned in Sec. 3.2.1, the M_{bc} and ΔE variables of B_{tag} are unreliable to obtain this yield due to the presence of peaking background, decay mode dependence, and training bias, respectively. The solution is to reconstruct a known B decay on the signal side together with B_{tag} and use the signal side properties to obtain the yield. The signal sides (samples) commonly used for such calibration measurements are the $B \to D^{(*)}\ell\nu$ decays, where the $D^{(*)}$ decays are exclusively reconstructed (used for Belle [62]) and the $B \to X\ell\nu$ decays with the inclusive charm meson, adopted by Belle II [81]. The large branching fraction of the semileptonic B decays and the inclusive characterization of the charmed hadronic system make this

latter sample very interesting from a statistical point of view. The exclusive reconstruction of $D^{(*)}$ in the $B \to D^{(*)}\ell\nu$ sample, on the other hand, leads to low efficiency but provides very high purity due to the clean recoil mass peak around 0, which corresponds to the mass of the neutrino. While these samples provide calibration factors, they are not sufficient to investigate the sources of discrepancies between the data and the MC. The $B \to D^{(*)}\ell\nu$ sample is statistically limited to study individual B_{tag} modes. And the $B \to X\ell\nu$ sample does not provide a clear separation between signal and background on the signal side, as shown in Fig. 3.14. To overcome these limitations, a novel $B \to D\pi$ sample is introduced in which the signal side is reconstructed using the recoil mass method, as described in Sec. 3.2.2. This sample has a much higher purity than $B \to X\ell\nu$ and a higher efficiency than $B \to D^{(*)}\ell\nu$, making it an optimal middle ground for detailed hadronic B_{tag} studies.

In the following sections, the calibration procedures are described using the $B \to X\ell\nu$ and $B \to D\pi$ samples in Belle II. Combining the calibration factors from these two samples is the current standard in the collaboration and will be used in the future measurements in Belle II. Using the $B \to D\pi$ sample, all performance metrics will be evaluated for comparison in the Belle and Belle II data sets. Further investigation into the causes of the discrepancy between the data and MC is described in the next chapter.

3.4.1 Calibration using the $B^+ \to X\ell^+\nu$ sample

Together with the B_{tag} candidate, the $B^- \to X\ell^-\bar{\nu}_\ell$ on the signal side is said to contain events corresponding to $B^- \to D^0\ell^-\bar{\nu}_\ell$, $B^- \to D^{*0}\ell^-\bar{\nu}_\ell$, $B^- \to D^{**0}\ell^-\bar{\nu}_\ell$, decays in the semileptonic gap, and the $B^- \to X_u\ell^-\bar{\nu}_\ell$, where the ℓ can be either an electron or a muon. Here, the gap is needed to describe the observed difference between the inclusive and the sum of exclusive branching fractions of the semileptonic B decays [82]. The inclusiveness of this sample is achieved by not explicitly reconstructing the X component and attempting to identify the decays from the lepton momentum spectrum alone. Due to the statistical advantage resulting from the high reconstruction efficiency, coupled with the large branching fractions of the semileptonic B decays, this sample became the standard choice in the Belle II collaboration.

First, the B_{tag} candidates are reconstructed by the hadronic B^+ FEI package with the selection $M_{\text{bc}} > 5.27$ GeV/c^2 and $-0.15 < \Delta E < 0.1$ GeV as described in Sec. 3.2.1 together with $\mathcal{P}_{\text{FEI}} > 0.001$. To suppress the continuum background, a cut is applied to the cosTBTO variable. For each event, all tracks and ECL clusters belonging to the B_{tag} candidate can be removed, and the rest of the event should correspond to the activity of the other B, the B_{sig}. For the B_{tag} candidate and the rest of the event, the thrust vectors can be calculated. This is the vector along which the total projection of the collection of momenta is maximized. The cosine of the angle between the two thrust vectors, called *cosTBTO*[3], is a thrust-based discriminator that effectively suppresses the continuum background. In $B\overline{B}$ events, the particles are almost at rest and so the thrust vectors are uniformly distributed. Therefore, cosTBTO is uniformly distributed between 0 and 1. In contrast, in continuum events, the particles are collimated, and the thrust axes point back to back, resulting in a peak at high values of cosTBTO. Such a difference in topology justifies the selection of cosTBTO < 0.9 to reject the continuum events, as shown in Fig. 3.13.

After this selection, an additional lepton (e or μ) is chosen from the rest of the event of the B_{tag}. The additional track must satisfy three conditions: it must originate from the interaction point, be within CDC acceptance, and have $p_t > 0.3$ GeV/c for

[3]Here TB refers to the thrust vector of B_{tag} and TO is the thrust vector of the rest of the event.

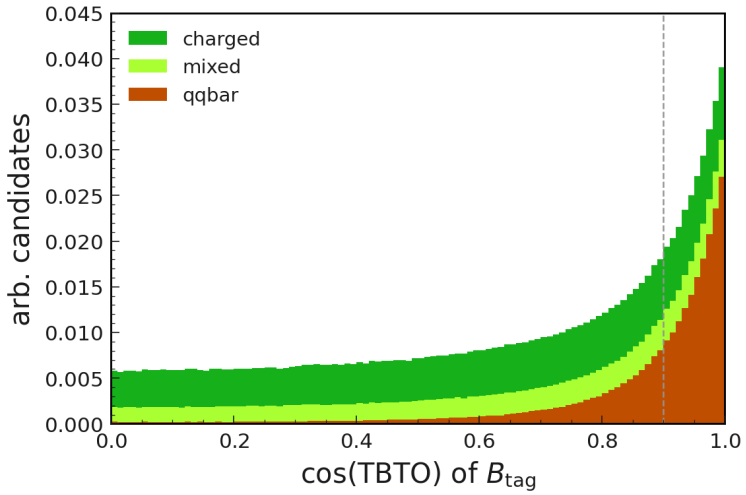


FIGURE 3.13: Distribution of cosTBTO showing the selection, after applying $\mathcal{P}_{\text{FEI}} > 0.001$, $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$ and $-0.15\,\text{GeV} < |\Delta E| < 0.1\,\text{GeV}$ on B_{tag} candidates.

optimal tracking performance. The e and μ candidates are then selected with a cut on the corresponding PID likelihood greater than 0.9. Both samples are treated independently for cross-validation.

Based on the fact that in the CM frame, $\mathbf{p}^*_{B_{\text{sig}}} = -\mathbf{p}^*_{B_{\text{tag}}}$, the momentum of the lepton is evaluated in the rest frame of the B_{sig}, called p^*_ℓ[4]. This serves as a discriminating variable between signal and the backgrounds, which can be divided into three categories: continuum events ($e^+e^- \to q\bar{q}$), fakes (events in which a hadron fakes as a lepton), and secondaries (cases in which the lepton does not arise from the decay of the B meson). Compared to the background, the signal leptons originate directly from the B mesons and therefore tend to have a larger momentum in the B rest frame. This is confirmed by the p^*_ℓ distribution shown in Fig. 3.14. The figure shows the p^*_ℓ in MC, split into signal and background categories. It is important to note that the MC sample used here is independent of the one used for training the FEI package. When superimposed on the p^*_ℓ distribution in the data, the discrepancy is already apparent, with the overall yield in the data being significantly lower than in MC.

Before fitting the distribution to obtain the yields, the selection $p^*_\ell > 1$ GeV/c is applied because the background is dominant, and lepton ID uncertainties are higher at low momentum. Finally, for each event, the best candidate is selected based on the highest $\mathcal{P}_{\text{FEI}}$ of B_{tag} and the highest p^*_ℓ on the signal side. Since the shape of the signal is not clearly defined, a template fitting procedure is used. There are two probability density functions (PDFs): one describing the background ($p^{\text{background}}$) and one describing the signal ($p^{X\ell\nu}$). Accordingly, the fit has two yields, $\nu = (N(\text{background}), N(X\ell\nu))$. Considering each bin of the distribution, a PDF (p^j_i) essentially describes the probability of an event of type j landing in the bin i.

However, the $X\ell\nu$ signal PDF is decomposed into four subcomponents:

$$p^{X\ell\nu} = \left(f^D p^{D\ell\nu} + f^{D^*} p^{D^*\ell\nu} + f^{X_u} p^{X_u\ell\nu} + \left(1 - f^D - f^{D^*} - f^{X_u}\right) p^{gap+D^{**}\ell\nu}\right),$$

corresponding to the contributions of the $\{D, D^*, X_u, D^{**}\}\ell\nu$ decays. Overall, the

[4]Despite the *, note that this variable is defined in the rest frame of B_{sig} and not CM frame, though they do not differ much.

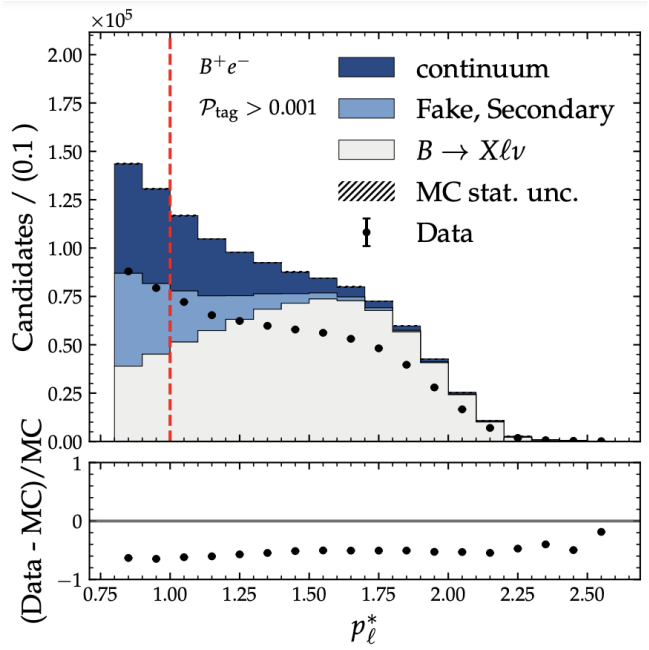


FIGURE 3.14: The distribution of p_ℓ^* (momentum of lepton in B rest frame) in Belle II $B \to X\ell\nu$ sample. Taken from Ref. [83].

$X\ell\nu$ signal PDF has three fractions, which are constrained by the MC expectation based on the latest branching fraction measurements [34].

To account for both MC statistics and additional systematic effects, the sub-PDFs are subject to variations from nuisance parameters, θ_i^j, of the form:

$$p_i^j \to \frac{p_i^j \left(1 + \epsilon_i^j \theta_i^j\right)}{\sum_l p_l^j \left(1 + \epsilon_l^j \theta_l^j\right)}.$$

where ϵ_i^j represents one sigma uncertainty for event type i in bin j. Note that the term in the denominator ensures that the PDFs are always normalized, even though their shape can change due to the nuisance parameters, θ_i^j. The associated bin to bin correlations within and across templates, resulting from systematic uncertainties, are accounted for in the correlation matrix, ρ_θ. This correlation matrix is determined from the total covariance matrix, which is the combination of the covariance matrices for each systematic effect, including MC statistics.

Using all of this, the number of expected events is given by

$$\nu_i^{\text{exp}}\left(\nu^j, \theta_i^j\right) = \sum_j \nu^j p_i^j.$$

Finally, the model is fitted to data by minimizing the expression:

$$\begin{aligned} -2\log\mathcal{L} = -2\log\prod_i \Big[&\text{Poisson}\left(\nu_i^{\text{obs}}, \nu_i^{\text{exp}}\right) \\ &+ \theta^T \rho_\theta^{-1} \theta^T \\ &+ \left(k - k_{\text{constraint}}\right)^T \Sigma_{\text{constraints}}^{-1} \left(k - k_{\text{constraint}}\right) \Big], \end{aligned} \tag{3.8}$$

where ν_i^{obs} is the number of events observed in a given bin i. The last term accounts for the constraints on parameters in the fit, where the $k = (f_D, f_{D^*}, f_{X_u})$ is the vector of fit parameters and the vector $k_{\text{constraints}}$ contains the corresponding nominal values.

Overall, this procedure constraints the fractions of various components in the signal from the simulation while allowing the shapes of the individual PDFs to change. This is illustrated in Fig. 3.15, where on the left, the fit of p_ℓ^* with all tag-side B decay modes in the FEI package is shown. The signal ($X\ell\nu$) component is shown in blue and the background in red. On the right, only the reconstructed $B^+ \to \overline{D}^0\pi^+$ candidates are selected and the fit of p_ℓ^* is performed again. Here, the shape of signal PDF in blue is significantly different.

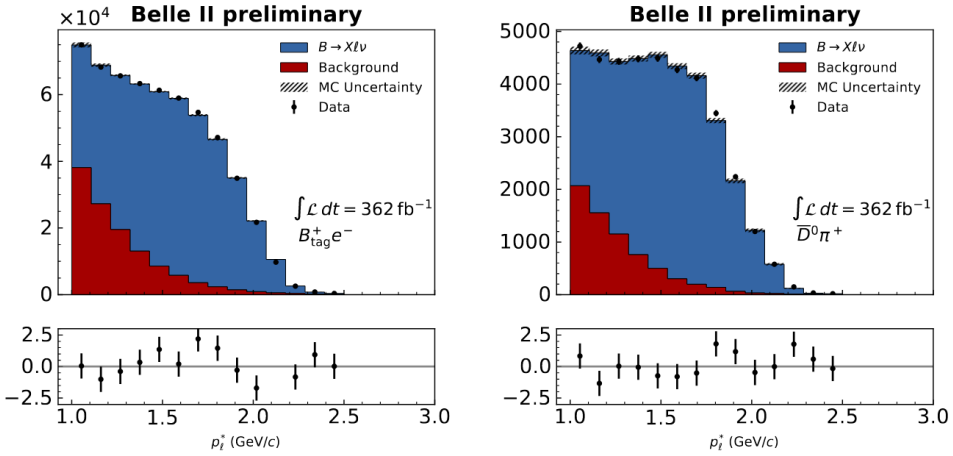


FIGURE 3.15: Template fits of p_ℓ^* for $\ell = e$ with all B decay modes in the FEI package (left) and with $B^+ \to \overline{D}^0\pi^+$ (right). This illustrates that the shapes of PDFs are free to change.

Once templates are established in a MC sample, MC yields are determined from an independent sample. Using the yields in the data and MC, the overall calibration factor (CF) for the B^+ hadronic FEI package in Belle II is calculated to be ($0.721 \pm 0.002 \pm 0.022$) for $\ell = \mu$ with $\mathcal{P}_{\text{FEI}} > 0.001$ cut [83]. The first uncertainty is statistical, the second one is systematic and includes uncertainties due to tracking (0.24%), lepton ID (0.3 - 0.5%), the form factors of $D^{(*)0}\ell\nu$ (0.05 - 0.3%) and, largest of all, the inclusive branching fraction of $B^- \to X\ell^-\nu$ (2.7%). Since the signal side is inclusive, with no constraints on the flavour, a significant cross-feed from B^0 is observed, creating additional systematics. The uncertainty of this measurement is systematically dominated, by the complexity of the fitting procedure and the dependence on the model of semileptonic decays of B mesons. In particular, some of the components, such as $D^{**}\ell\nu$ and the semileptonic gap, are connected to tension observed between the inclusive and exclusive determination of $|V_{cb}|$ [84, 85].

To identify the origin of this discrepancy, it is helpful to consider the calibration factors for each of the B^+ decay modes reconstructed by the hadronic FEI package. Of the 36 B decay modes considered for hadronic FEI B-tagging, only 11 of them, i.e., $B^+ \to \overline{D}^{(*)0}\, n\pi^\pm\, m\pi^0$, $D^-\, n\pi^\pm\, m\pi^0$, $\Lambda_c^- p\pi^+\pi^-\pi^+$, where $1 \leq n \leq 3$ and $0 \leq m \leq 1$, contribute to more than 80% of the efficiency in Belle II. Figure 3.16 shows the calibration factors for these modes. The strong dependence of CF on tag-side modes makes it clear here that some modes are better modelled in MC than others. For some modes, the CF can get as low as 30%. To accommodate for analyses that may choose a tighter $\mathcal{P}_{\text{FEI}}$ cut for higher purity, the calibration factors are also calculated with $\mathcal{P}_{\text{FEI}} > 0.01$ cut.

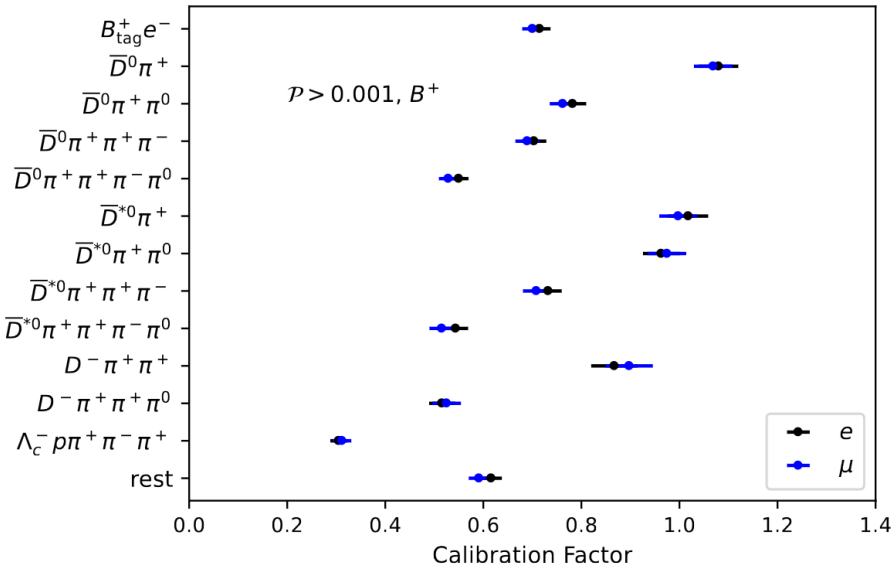


FIGURE 3.16: Calibration factors for B^+ decay modes for $\mathcal{P}_{\text{FEI}} > 0.001$ using the $X\ell\nu$ sample. A slight systematic tension is seen between muon and electron channels, with muon calibration results being below their counterparts. This could be due to larger background for muons due to higher fake rates.

Despite the high statistics, the template fit procedure of the $X\ell\nu$ sample is highly dependent on the semileptonic decay model. Although the uncertainty of the inclusive semileptonic branching ratio is included in the systematics, this does not account for the dependence of the shape and fraction of the PDFs on the individual signal components. This is better illustrated in Fig. 3.17: one of the largest components of the signal PDF corresponds to the D^{**} component and the semileptonic gap decays that have never been measured (shown in grey). These components are either signal or background for some of the semileptonic analyses in Belle II, such as the $R(D^*)$ and $|V_{cb}|$ exclusive versus inclusive measurement, which means that this is not an ideal control sample for them. Therefore, an orthogonal calibration sample is needed to provide calibration factors for all purposes.

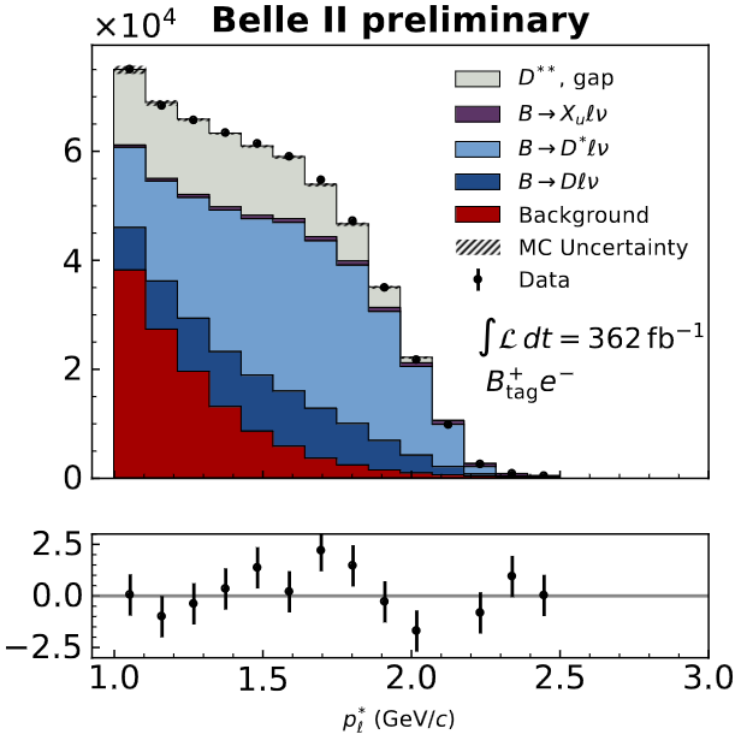


FIGURE 3.17: Template fit of p_ℓ^* for $\ell = e$ with all tag-side B decay modes in the FEI package, showing the different components of the signal PDF.

3.4.2 Performance evaluation using the $B \to D\pi$ sample

The $B \to D\pi$ calibration sample is reconstructed by recoiling B_{tag} candidates with a prompt π track. In this way, the charm mesons are not reconstructed, allowing a higher reconstruction efficiency.

As for the $X\ell\nu$ sample, the B_{tag} candidates provided by the FEI package are selected with the cuts: $M_{\text{bc}} > 5.27$ GeV/c^2; $-0.15 < \Delta E < 0.1$ GeV; and $\mathcal{P}_{\text{FEI}} > 0.001$. The same cut, cosTBTO < 0.9, is also applied. In addition, to further suppress the continuum background, a cut is applied to the variable *sphericity*. First introduced in Ref. [86], the sphericity tensor is constructed from the three-vectors of all particles (both charged and neutral) in the event:

$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i p_i^2} \tag{3.9}$$

where i iterates over all reconstructed particles in the whole event and α and β are the Cartesian components of the momenta. S is not Lorentz invariant, and it is only meaningful when calculated in the CM frame. The eigenvectors of S define the ellipsoid that best fits the particle distribution in the event. The eigenvalues α_1, α_2, α_3, ordered from largest to smallest, are proportional to the lengths of the axes of this ellipsoid. Since S is unitary, only two independent quantities can be constructed from its eigenvalues. One of them, called sphericity, is a linear combination of the second and third eigenvalues, $3(\alpha_2+\alpha_3)/2$. The cut sphericity > 0.2 removes $q\bar{q}$ events with little effect on the $B\bar{B}$ events, as shown in Fig. 3.18.

Among the events passing the above selection, π candidates are selected, requiring impact parameter constraints and vetoing the likelihood of the track being a e, μ or a

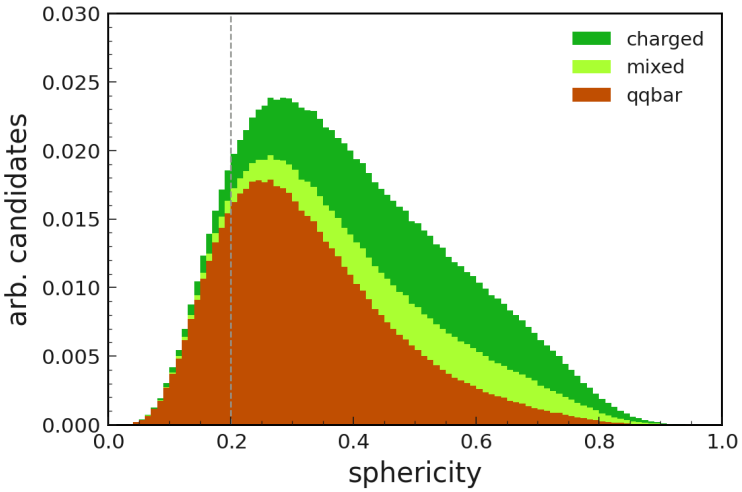


FIGURE 3.18: Distribution of sphericity showing the selection, after applying $\mathcal{P}_{\text{FEI}} > 0.001$, $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$, $-0.15\,\text{GeV} < |\Delta E| < 0.1\,\text{GeV}$ and cosTBTO < 0.9 on B_{tag} candidates.

K. The best candidate in the event is then selected as the one with the highest $\mathcal{P}_{\text{FEI}}$ and the highest p^*_π, the momentum of the pion in the CM frame. Since the pion is a prompt particle coming directly from the B meson, it is expected to have the highest momentum in the CM frame. When $p^{\text{rec}}_{B_{\text{sig}}} = p_\pi$ is used to calculate the M_{recoil} in Eq. 3.3, peaks of D, D^* and D^{**} are expected at their corresponding masses, which can be seen in Fig. 3.19 for the Belle and Belle II data sets. From these plots alone, it is clear that the discrepancy between the data and MC is larger in the Belle II sample than in Belle.

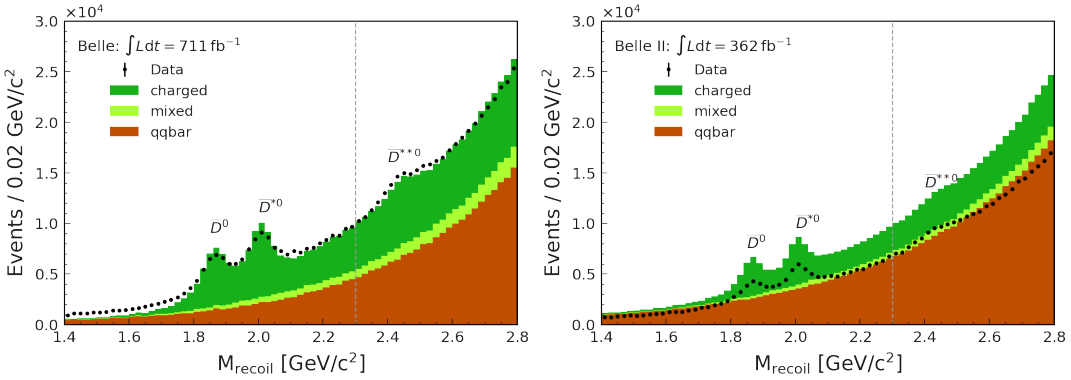


FIGURE 3.19: M_{recoil} of B_{tag} and π with the highest momentum, comparing data with different components of MC in Belle sample (left) and Belle II sample (right).

The peak corresponding to $B \to D^{**}\pi$ decays is barely visible in M_{recoil} beyond 2.4 GeV/c^2, but with a large background. It could be further improved by requiring higher purity with a tighter selection on $\mathcal{P}_{\text{FEI}}$ of the B_{tag} candidates. However, the shape of this would be difficult to model since D^{**} is a collection of four different resonances (as described in Sec. 4.2), two of which have a large mass and are less well understood. To avoid them, an unbinned maximum likelihood fit for the M_{recoil} is performed only in the range of [1.4, 2.3] GeV/c^2.

Most of the background in this fit range is combinatorial and can be modelled by an exponential function, as shown by the qqbar ($q\overline{q}$), mixed ($B^0\overline{B}^0$) and rest of charged (B^+B^-)' components in Fig. 3.20. The only exception is the component of

the $B^+ \to D^{(*)}\rho^+$ decays, where the ρ^+ meson decays to $\pi^+\pi^0$. The π^+ of ρ is used to build up the recoil mass, the missing particles include the charm meson and the additional π^0. This system would have a wider mass distribution peaking at higher mass than just the charmed meson, as shown by the brown components in the left plot of Fig. 3.20, and requires an ad-hoc model.

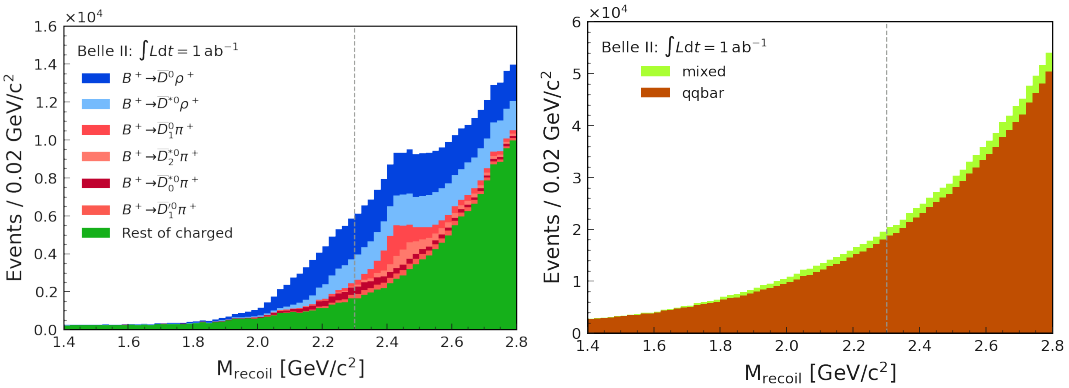


FIGURE 3.20: Background composition in the charged MC sample (left) and the rest of the background (right) in the Belle II sample.

Four components are needed to fit the entire distribution: two signal peaks corresponding to $D\pi$ and $D^*\pi$ events, a combinatorial background, and the contribution of the partially reconstructed $B \to D^{(*)}\rho$ decays. The resolution of the signal peaks depends on the momentum of the signal side π. Nevertheless, MC studies confirm that the two signal shapes are similar and can be defined by a common PDF obtained by combining a Gaussian and a Crystal Ball shape [87], as shown in Fig. 3.21.

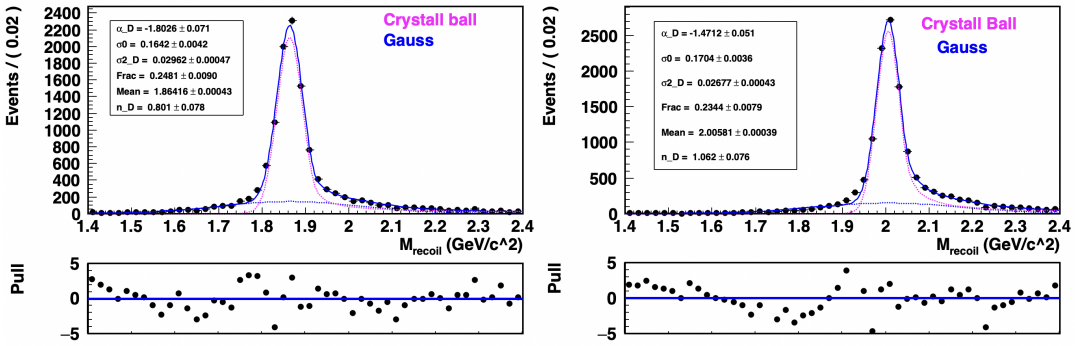


FIGURE 3.21: Signal only fit of $D\pi$ (left) and $D^*\pi$ (right) components.

All parameters in the signal PDF are fixed from MC, except for the mean of the D peak and a fudge factor for the resolution of the Crystall Ball PDF to account for possible differences between the data and MC. The mean of the $D^*\pi$ component is defined as ($D\pi$ mean + ΔM), where ΔM is the mass difference between D^* and D mesons and is set to the known value of 0.142 GeV/c^2 [34]. The combinatorial background is described by an exponential function whose slope is a free parameter in the fit. The partially reconstructed $B^+ \to \bar{D}^{(*)0}\rho^+$ candidates are fitted with a double Gaussian whose parameters are fixed from the MC. The fit of this component is shown in Fig. 3.22. To ensure for stability of the fit, the ratio of the $D^{(*)}\rho$ yield to the $D^*\pi$ yield is fixed from MC.

Figure 3.23 shows the fit to the Belle II MC sample corresponding to 1 ab^{-1} (left) and the Belle II data sample (right). A MC sample larger than the data is used to

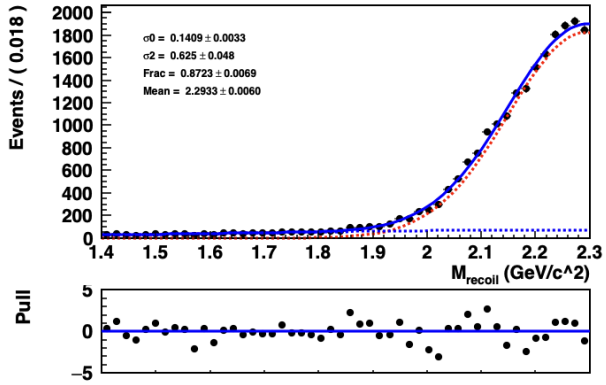


FIGURE 3.22: Fit to the $D^{(*)}\rho$ only component.

reduce statistical uncertainty, but was rescaled to calculate the calibration factors. It can be seen that the $\overline{D}^0$ peak has a smaller background and negligible contribution from the $\overline{D}^{(*)0}\rho^+$ component, unlike the $\overline{D}^{*0}$ peak. Therefore, given the sufficient statistics, only the $\overline{D}^0$ signal yields are used to evaluate the calibration factors.

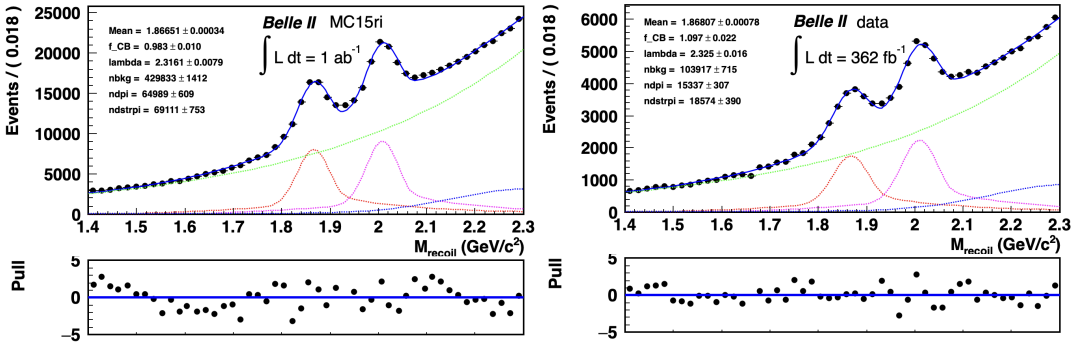


FIGURE 3.23: Fit to M_{recoil} using the latest Belle II MC (left) and Belle II data (right) samples.

The same strategy is used to perform fits for Belle MC and data samples. The sources of the systematics listed in Table. 3.5 are considered. The fit is performed with variations of the fixed parameters within their uncertainty to estimate the systematic uncertainty. The Gaussian tail component of the signal and the $B \to D^{(*)}\rho$ components are completely removed from the fit to evaluate their effect. Only events where the multiplicity is exactly one are used to evaluate the systematics due to the best candidate selection. The difference in PID efficiency between the data and MC is taken into account. Signal peaks are removed during fitting to check for any peaking background from B^0. All these contributions are limited by the available statistics and are reducible, except for the dominant uncertainty due to the branching fraction of the $B^+ \to \overline{D}^0\pi^+$ decay.

The overall calibration factors for B^+ tags are $(0.753 \pm 0.012 \pm 0.025)$ in the Belle sample and $(0.651 \pm 0.015 \pm 0.025)$ in the Belle II sample, the first uncertainty being statistical and the second systematic. As already noted, the overall calibration factor is worse in Belle II. Uncertainties in measuring calibration factors with this sample are statistically limited, so precision will only improve in the future as Belle II takes

Source	Systematic uncertainty (%)	
	Belle	Belle II
Fixed parameters in the fit	0.1	0.1
Gaussian component in signal PDF	0.5	0.6
$B \to D^{(*)}\rho$ component	0.5	0.6
Best candidate selection	0.9	1.3
PID	1	0.4
Cross-feed from B^0	< 0.1	< 0.1
Luminosity (Number of $B\bar{B}$)	0.9	0.6
$\mathcal{B}$ of $B^+ \to \bar{D}^0\pi^+$	1.9	1.9
Total	2.5	2.5

TABLE 3.5: Sources of systematics uncertainty and their contribution to the total calibration factor of B^+-tagging by the FEI package.

more data. The simplicity and reproducibility of the procedure make it an alternative free from the limitations of the systematics in the semileptonic approach.

The $\bar{D}^0\pi^+$ sample yields only about 4000 events per 100 fb^{-1} in data – an order of magnitude less compared to the $X\ell\nu$ sample – but the signal side is much purer, as the signal peaks are clearly visible. In contrast to the $X\ell\nu$ sample, the background level in the signal region is much lower and easy to parametrize without heavy dependence on the simulation. Therefore, the signal extraction procedure should be more reliable compared to the $X\ell\nu$ sample.

Similar to the $X\ell\nu$ sample, the fit can be performed for each tag-side B decay mode of the FEI package, resulting in the calibration factors for the Belle and Belle II samples shown in Fig. 3.24.

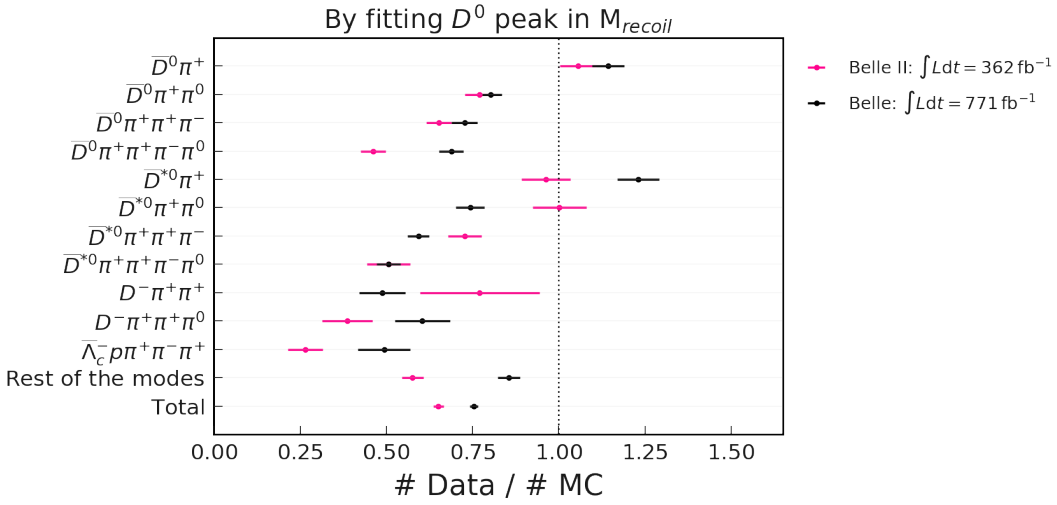


FIGURE 3.24: Calibration factors of different B decay modes produced by the FEI package in Belle and Belle II samples with the selection $\mathcal{P}_{\text{FEI}} > 0.001$ cut. Only statistical uncertainties are shown here.

As observed earlier, there is a strong dependence of the calibration factors on the tag mode. The fact that Belle and Belle II have different CF even for the same mode suggests that there is very poor modelling in MC that is somewhat different in the two samples. This will be described in detail in the next chapter.

In addition to estimating the CF's, this sample can also be used to evaluate other FEI performance metrics. Using these fit yields, the efficiency of the hadronic FEI package can be calculated using Eq. 3.5. Considering the reconstruction efficiency of the signal side in the recoil method, the efficiency definition would be

$$\epsilon_{\text{FEI}} = \frac{\textit{signal yield}}{2\ N(B\bar{B})\ f^{\pm}\mathcal{B}(B^+ \to \bar{D}^0\pi^+)\ \epsilon_\pi}, \tag{3.10}$$

where $N(B\bar{B})$ is the number of $B\overline{B}$ pairs produced, calculated as the product of the integrated luminosity, the cross-section of $e^-e^- \to \Upsilon(4S)$ and $\mathcal{B}(\Upsilon(4S) \to B\overline{B})$; $f^{\pm}$ is the fraction of B^+B^- produced, which is close to 0.5; $\mathcal{B}(B^+ \to \bar{D}^0\pi^+)$ is taken as 0.467% from Ref. [34]; and the efficiency of pion reconstruction ϵ_π is assumed to be 90% here. The FEI efficiency obtained by this method would be independent of the signal side.

The purity of the B-tagging, defined in Eq. 3.6, can be applied here within a signal region of the 3σ window around the $\bar{D}^0$ peak ([1.80, 1.92] GeV/c^2) as:

$$\textit{Purity} = \frac{\textit{signal yield}}{\textit{signal yield}\ +\ \textit{background yield in signal region}}. \tag{3.11}$$

As described in Section 3.3, $\mathcal{P}_{\text{FEI}}$ can be used to increase the purity of B_{tag} candidates at the expense of efficiency. Purity[5] and efficiency for this sample are calculated for different $\mathcal{P}_{\text{FEI}}$ cuts, and the corresponding ROC curves are shown for Belle and Belle II, allowing direct comparison (Fig. 3.25).

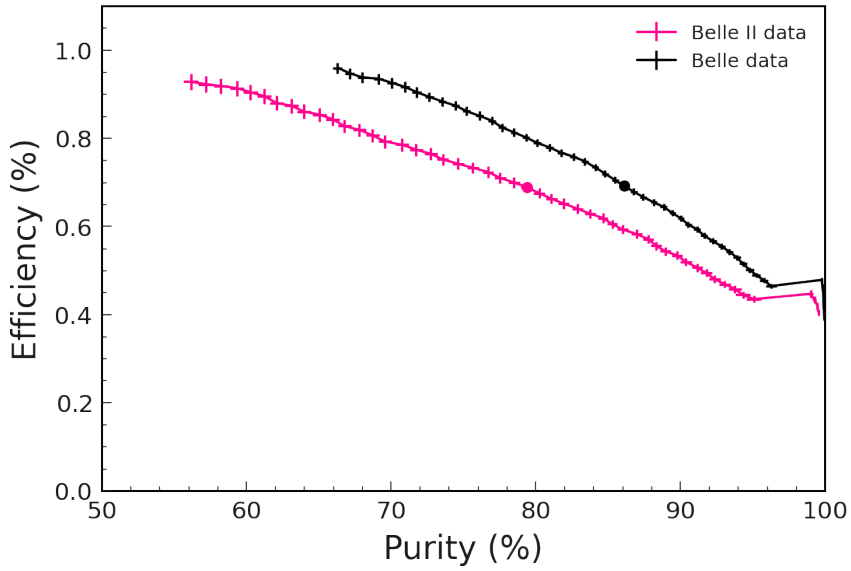


FIGURE 3.25: Efficiency vs purity curve of hadronic FEI computed using $B \to D\pi$ samples in Belle and Belle II. The left most point in the curves represent $\mathcal{P}_{\text{FEI}} > 0.001$ cut, and the shaded circle represents $\mathcal{P}_{\text{FEI}} > 0.01$ cut.

In this way, the $B \to D\pi$ sample allows a direct comparison of FEI performance between the Belle and Belle II data sets. The maximum efficiency obtained is around 1%, which can be reduced to around 0.4% at higher purity. Moreover, the performance in the Belle data set is much better compared to Belle II; this difference is due to the fact that the two experiments use different MC samples for training the FEI package.

[5] Although this definition of purity depends on the signal side, it allows comparison between different cuts in different data sets.

As will be shown in Sec. 4.1, the MC modelling of the B decays plays an important role in the FEI performance. The other reason for the difference is a lower efficiency of the reconstruction of the low-energy photons and the lack of a mass-constraint on π^0 in the FEI package for Belle II, which will be discussed in more detail in Sec.4.4.

Similarly, the overall calibration factors for the Belle and Belle II data sets are calculated at different $\mathcal{P}_{\text{FEI}}$ cuts. This is shown as a function of the purity of the data in Fig. 3.26. The CF is constant (within statistical uncertainty) for most $\mathcal{P}_{\text{FEI}}$ cuts, except when high purity is required (for very tight $\mathcal{P}_{\text{FEI}}$ cuts). This is because different B decay modes have different $\mathcal{P}_{\text{FEI}}$ distributions, as already shown in Fig. 3.9. Thus, when a very tight cut is applied to $\mathcal{P}_{\text{FEI}}$, certain modes are strongly suppressed compared to others, changing the composition of the modes. As shown in Fig. 3.24, different modes (i) have different calibration factors (CF_i), so changing their relative fraction (w_i) changes the overall calibration factor as

$$CF = \sum_{i}^{N} w_i \times CF_i \,.$$

This is verified in Fig. 3.27, which shows the fraction (w_i) of all B_{tag} candidates reconstructed in a particular B mode. With the tighter $\mathcal{P}_{\text{FEI}} > 0.1$ cut, the fractions of $\overline{D}^0\pi^+$ and $\overline{D}^0\pi^+\pi^0$, which have calibration factors close to 1, increase significantly, while the $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$ modes, which have a very low calibration, are strongly suppressed. Therefore, to achieve stable performance of the FEI package, it is recommended not to use a cut tighter than 0.1 on $\mathcal{P}_{\text{FEI}}$.

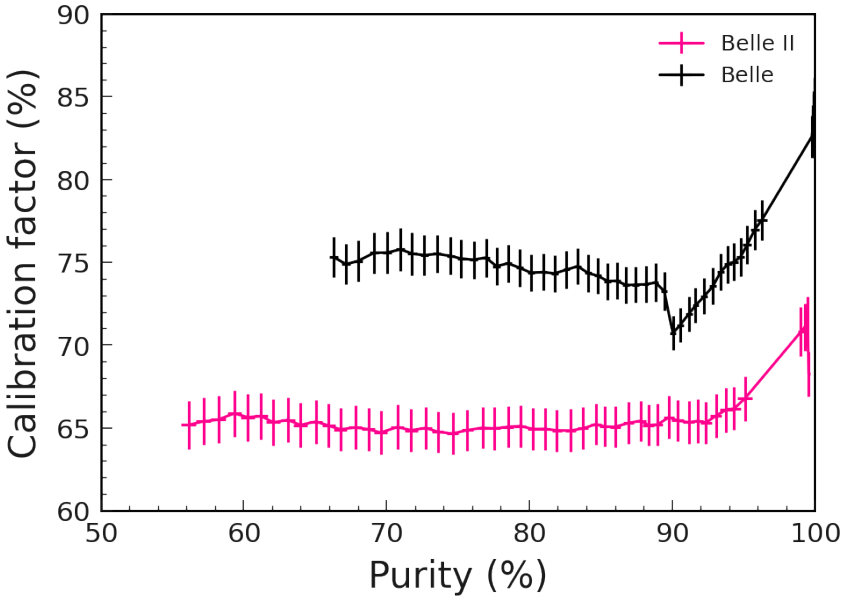


FIGURE 3.26: Calibration factors of hadronic FEI are computed at various $\mathcal{P}_{\text{FEI}}$ cuts and shown as function of purity in data for Belle and Belle II samples.

With sufficient statistics and high purity, the $B \to D\pi$ sample provides the best opportunity to study hadronic B-tagging. The clearly distinguishable signal component allows efficiency and purity to be calculated directly in the data, together with the calibration factors.

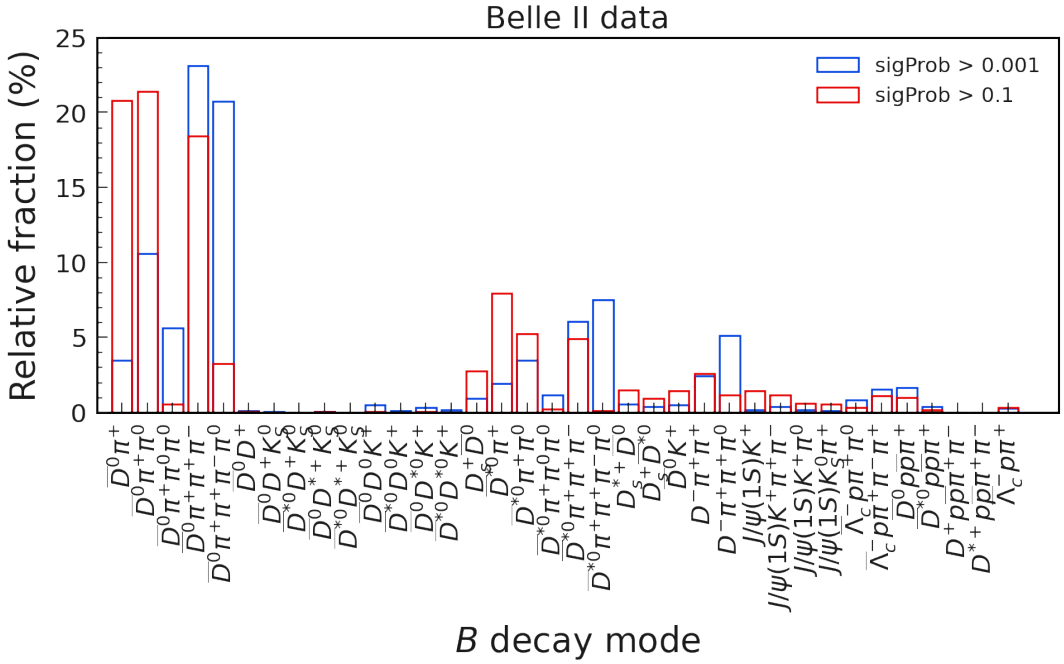


FIGURE 3.27: Relative composition of the different B modes provided by the FEI package at different $\mathcal{P}_{\text{FEI}}$ cuts: 0.001 and 0.1.

3.4.3 Combination of calibration factors

Although the purity of the $B \to D\pi$ sample is higher, it lacks the statistics of $X\ell\nu$ by an order of magnitude. To take advantage of both samples and account for their differences, the calibration factors obtained from the two samples are combined. This results in a single set of tag-side mode-dependent calibration factors with associated uncertainties for measurements with hadronic B-tagging.

For the combination, a fit to $\chi^2 = \chi^2_{e/\mu} + \chi^2_{D\pi} + \chi^2_{NP}$ is used, with nuisance parameters (NP) accounting for the correlated systematic uncertainty. To obtain conservative combined factors, additional uncertainty is added to cover discrepancies between the two samples, with different signal-side purity. The full χ^2 is given as:

$$\begin{aligned}\chi^2 = &\left((\vec{\epsilon},\vec{\epsilon}) + \vec{\nu}_\ell \cdot \vec{\sigma}^{\text{discrep}} + \rho_\ell \sigma_\ell^{\text{sys}} - (\vec{\epsilon}_e,\vec{\epsilon}_\mu)\right)^T \Sigma_\ell^{-1} \left((\vec{\epsilon},\vec{\epsilon}) + \vec{\nu}_\ell \cdot \vec{\sigma}^{\text{discrep}} + \rho_\ell \vec{\sigma}_\ell^{\text{sys}} - (\vec{\epsilon}_e,\vec{\epsilon}_\mu)\right) \\ &+ \left(\vec{\epsilon} + \vec{\nu}_{D\pi} \cdot \vec{\sigma}^{\text{discrep}} + \rho_{D\pi} \sigma_{D\pi}^{\text{sys}} - \vec{\epsilon}_{D\pi}\right)^T \Sigma_{D\pi}^{-1} \left(\vec{\epsilon} + \vec{\nu}_{D\pi} \cdot \vec{\sigma}^{\text{discrep}} + \rho_{D\pi} \vec{\sigma}_{D\pi}^{\text{sys}} - \vec{\epsilon}_{D\pi}\right) \\ &+ \rho_{D\pi}^2 + \rho_\ell^2 + \vec{\nu}_{D\pi} \cdot \vec{\nu}_{D\pi} + \vec{\nu}_\ell \cdot \vec{\nu}_\ell\end{aligned}$$

where,

- $\vec{\epsilon}$ is a vector of tag-side mode-dependent calibration factors to be fitted;
- $\vec{\epsilon}_e, \vec{\epsilon}_\mu$ and $\vec{\epsilon}_{D\pi}$ are the measured calibration factors in the $Xe\nu, X\mu\nu$ and $D\pi$ samples;
- Σ_ℓ and $\Sigma_{D\pi}$ are the corresponding covariance matrices for uncorrelated uncertainties;
- $\rho_\ell \vec{\sigma}_\ell^{\text{sys}}$ and $\rho_{D\pi} \vec{\sigma}_{D\pi}^{\text{sys}}$ describe fully correlated uncertainties such as branching fractions and tracking. The $\vec{\sigma}^{sys}$ are vectors of mode-dependent correlated uncertainties. And the scalar nuisance parameters ρ_ℓ and $\rho_{D\pi}$ are Gaussian constrained with the ρ^2 terms;

- $\vec{\nu}_\ell \cdot \vec{\sigma}^{\text{discrep}}$ and $\vec{\nu}_{D\pi} \cdot \vec{\sigma}^{\text{discrep}}$ add additional mode-dependent uncertainty to account for the absolute differences in calibration factors from the two samples. The $\vec{\sigma}^{\text{discrep}}$ contains these uncertainties, while $\vec{\nu}_\ell$ and $\vec{\nu}_{D\pi}$ are vectors of Gaussian constrained nuisance parameters.

Using the calibration factors evaluated with the Belle II data set, the χ^2 minimisation is performed. A comparison of the individual calibration factors of $Xe\nu, X\mu\nu$, and $D\pi$ samples and the combined values (in yellow) is shown in Fig. 3.28. These combined calibration factors are currently the official values for the collaboration and will be used in upcoming measurements of decays such as $B \to K\nu\overline{\nu}$, $B \to \rho\ell\nu$, $B \to \pi\ell\nu$ and $B \to \tau\nu$ from Belle II. For Belle, the available calibration factors are only from $D\pi$ sample, hence there is no need for a combination.

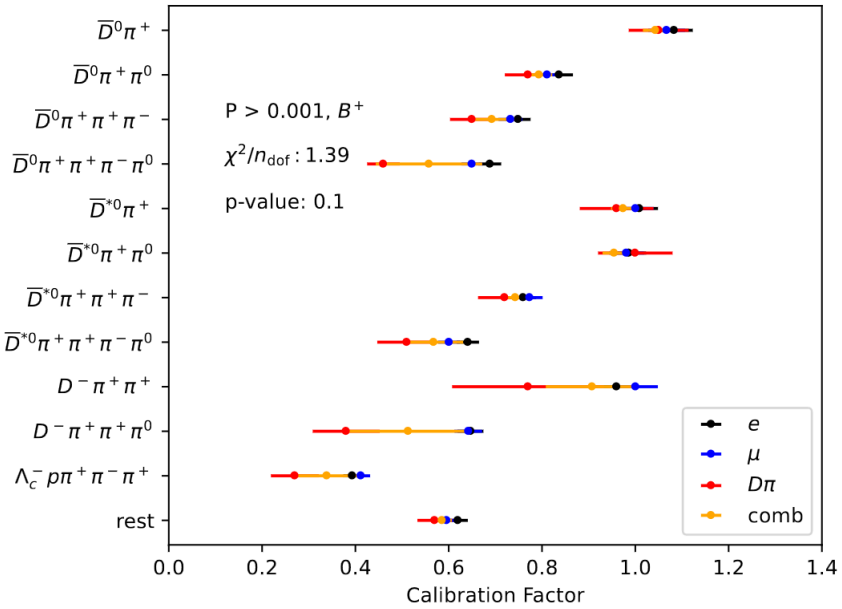


FIGURE 3.28: Calibration factors for the tag-side B^+ decay modes reconstructed by the FEI package with $\mathcal{P}_{\text{FEI}} > 0.001$ in Belle II data set. Combined factors from both samples are shown in yellow.

The calibration factors of B^+-tagging by the FEI package, measured using two completely independent samples, have shown to agree reasonably well with each other. The $X\ell\nu$ results are systematically dominated by the model of semileptonic decays of B mesons. The $D\pi$ sample is statistically limited, and the accuracy can only improve if Belle II collects more data. The CF values of the $D\pi$ sample appear systematically lower than those of the $X\ell\nu$ sample. The largest discrepancies between the two samples are for the B decay modes with the lowest tag-side purity or low statistics. Due to the low signal-side purity of the $X\ell\nu$ sample, its results may be less reliable in such cases. The $D\pi$ sample, on the other hand, has a larger statistical uncertainty. However, the high signal-side purity of the $D\pi$ sample allows the fit to extract the correctly reconstructed events, so that the decay kinematics of the B_{tag} candidates can be studied. Comparing the decay kinematics between data and MC using this sample is key to understanding the impact of MC modelling on the performance of FEI and potentially improving it, as described in the next chapter.

Chapter 4

Understanding and improving the hadronic B tagging

The package FEI, used for hadronic B-tagging in the Belle and Belle II experiments, reconstructs B_{tag} candidates using BDTs. This is done in a hierarchical approach, as shown in Fig. 3.5, by using the decay modes listed in Tabs. 3.3 and 3.4. The BDTs for each of these decay modes are trained on $B\overline{B}$ MC samples with generically decaying B mesons. Thus, the FEI package essentially learns the selection from the decay kinematics of the particles in MC. Therefore, understanding and, if necessary, correcting the MC model of the decays involved is critical for optimal performance of the FEI package. Modelling B mesons is particularly difficult because they are heavier and have a wider range of final states than other intermediate particles, as for example the D mesons. In Section 4.1, we give a general overview of the MC modelling of B decays in the Belle and Belle II experiments. This is followed in Section 4.2 by a detailed description of the B decay modes responsible for most of the efficiency of hadronic B-tagging by the FEI package.

The MC model for these B decays is compared with the latest published measurements to correct them. These changes are made consistently for both Belle and Belle II MC, and new MC samples are produced. In the $B \to D\pi$ control sample presented in Section 3.4.2, after the events with correctly reconstructed signal side are selected by the fit, the kinematics of the final state particles on the tag side can be compared between data and MC. This procedure is used to validate the changes. Overall, the new MC will have a better calibration factor, closer to 1, compared to the current values of about 0.75 (0.65) for B^+-tagging with the Belle (II) data sets. When the FEI package is retrained with the new MC decay kinematics, the learned selections should improve and, consequently, the background suppression capabilities increase, as shown in Section 4.3.

After retraining, the performance of the FEI package is compared between the Belle and Belle II data samples. The observed differences led to the identification of a tight pre-cut that affecting the reconstruction of low-energy photons. Fixing this problem significantly improves the efficiency of FEI in the Belle II data set, as shown in Section 4.4. Finally, Section 4.5 describes some possible future improvements to the B-tagging.

4.1 Modelling of B^+ meson decays: current status at Belle and Belle II

The Belle experiment has stopped recording data over a decade ago and the corresponding MC is frozen since then. Belle II MC, on the other hand, is produced every few years in campaigns using a slightly improved version of the software. This means

that the Belle and Belle II MC samples were produced with different versions of event generators and with different parameters, so the B decay model is different. Many event generators are used by the Belle and Belle II collaborations to generate MC in all the processes listed in Tab. 2.2. More details on these generators can be found in Ref. [88]. The processes of interest here, the decays of $B^{\pm}$ mesons from $\Upsilon(4S) \to B\overline{B}$, are generated by EvtGen[1] [76], complemented by PYTHIA [90, 91].

The known information needed by the EvtGen package to generate MC is compiled into two files called evt.pdl and DECAY.DEC. The evt.pdl file contains the name, identification code, mass, width, charge, spin, and lifetime of all the known particles. The file DECAY.DEC contains for each particle a listing of all decays with three elements: the prompt daughters, the corresponding branching fraction, and the EvtGen model describing the kinematics. For example, a line in DECAY.DEC describing B^+ decays:

```
0.004900    anti-D*0    pi+    SVS
```

would generate $B^+ \to \overline{D}^{*0}\pi^+$ events with 0.49% branching fraction using the EvtGen-model `SVS` for the decays of a pseudoscalar meson into a vector meson (D^*) and a pseudoscalar (π).`SVS: scalar` $\to$ `vector + scalar`.

The decays of B mesons can be roughly divided into: semileptonic and hadronic decays. The semileptonic decays listed in Table 3.1 account for about 25% of the total branching fraction and are explicitly listed in the DECAY.DEC. The remaining 75% are hadronic decays. However, all hadronic $B^{\pm}$ decays measured so far account for less than 30% of the total branching fraction. These have been explicitly listed in the DECAY.DEC. But, for the rest of the decays never measured ($\sim$ 45% of the branching fraction), only the quark-level transition is mentioned and PYTHIA is used for fragmentation. For example, a line in DECAY.DEC of Belle II describing B^+ decays:

```
0.23900997    u    anti-d    anti-c    u    PYTHIA    23
```

would generate decays of $\bar{b} \to \bar{c}$ transition with $W^+ \to u\bar{d}$, and a spectator u quark with a total branching fraction of 23.9%. PYTHIA is then called for quark fragmentation with kinematics described by `modeID`=23.

The Belle MC generation used PYTHIA version 6, while Belle II has been updated to PYTHIA version 8. In both cases, the fragmentation of quarks occurs according to relative rates determined by the parameters of the `StringFlav` class. These factors determine the ratio of vector to pseudoscalar meson production for each flavour, the mixing angles in the vector and pseudoscalar meson sectors, additional suppression of $\eta^{(\prime)}$ production, etc. The Belle (II) DECAY.DEC files use the default values for most of these parameters, with the production of some excited mesons turned off. For example, light mesons with orbital angular momentum (L) = 1, such as $a_1^{\pm}$ and a_1^0, are not produced by our configuration of PYTHIA. Fragmentation compares the final state with the explicitly listed decays, and if the state is listed with a known $\mathcal{B}$, fragmentation is performed again to produce an alternative final state. This check is essential to avoid altering the effective branching fractions of the explicitly listed decays. Therefore, to exclude that a particular decay is generated by PYTHIA, it can be explicitly listed in DECAY.DEC with a branching fraction of 0%. This possibility of excluding decays is not properly exploited in the Belle (II) DECAY.DEC files, which inadvertently inflates the branching fractions of some final states, and it is one of the

[1] The final state radiation is generated by PHOTOS [89]

main causes of the observed difference between the data and MC, as will be shown in Sec. 4.2.

The kinematics of the resulting particles after quark fragmentation is based on the `modeID` set. The relevant `modeID`s in Belle (II) MC, corresponding to the PYTHIA version 6 (8), are:

- **48** (**23**): $n \geq 3$-body weak decays with phase space model; if there is a quark spectator system it collapses to one hadron[2],
- **13** (**43**): $n \geq 2$-body decays with phase space model; turn partons into a random number of hadrons with multiplicity based on a Poisson distribution,
- **23** (**63**), **24** (**64**): colour suppressed baryonic decays of fixed multiplicity with phase space model,
- **32** (**91**): decay to $q\bar{q}$ or $\gamma\gamma$, which should shower and hadronize.

Here again, the Belle and Belle II MC differ in the way the unknown branching fraction is split across different PYTHIA channels, as listed in Tab. 4.1.

Quark transition	`modeID` in PYTHIA v8	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$
u anti-d anti-c u	23	31.23	20.26
u anti-d anti-c u	43	-	3.87
u anti-s anti-c u	43	2.23	2.02
c anti-s anti-c u	43	-	6.66
c anti-d anti-c u	43	-	0.36
u anti-d anti-u u	23	-	0.27
c anti-s anti-u u	23	-	0.36
u anti-u anti-d u	23	-	0.18
d anti-d anti-d u	23	-	$<$0.01
s anti-s anti-d u	23	-	0.01
u anti-u anti-s u	23	-	0.20
d anti-d anti-s u	23	-	0.16
s anti-s anti-s u	23	-	0.13
anti-s u	91	-	0.45
anti-cd_1 uu_1	63	3.40	2.97
anti-cd_1 uu_1	64	1.27	-
anti-cs_0 cu_0	63	0.85	-
anti-cs_1 uu_1	63	0.18	0.81
anti-cs_1 uu_1	64	0.04	-
anti-cd_0 cu_0	63	0.04	-
Total PYTHIA contribution		39.24	38.71

TABLE 4.1: Different PYTHIA channels used to fill the unknown decays of B^+ with corresponding branching fractions in Belle and Belle II.

The components of PYTHIA are added to fill the gap in the total branching fraction (100%) after known decays $\mathcal{B}_{\text{explicit}}$ are explicitly listed. This means that the branching fractions of these components $\mathcal{B}_{\text{PYTHIA}}$ must be adjusted each time such that $\mathcal{B}_{\text{explicit}} + \mathcal{B}_{\text{PYTHIA}} = 100\%$. Since only the quark level transitions are listed in

[2] distributed according to the V−A matrix element

DECAY.DEC for these components, there is no way to find the exact decays that were generated before they were generated. One must actually generate some events using the configuration and examine the generated decay chain, using tools such as TopoAna [92]. TopoAna, a common tool for BESIII, Belle and Belle II experiments, can identify the whole topology, i.e. the decay chain generated at each event. This cycle of generating events and examining the decay chain every time something is changed in the configuration makes debugging tedious.

In the past, there have been efforts to completely remove the PYTHIA contributions in the decays of $D_{(s)}$ mesons [93], in order to make the outcome of the generation predictable and easier to control. However, for B mesons, it is currently impossible to completely remove the PYTHIA contribution because there is a large gap ($\sim 45\%$) of unknown branching fractions that could be filled by numerous possible decays. It is important that the modes that are essential for training the FEI package are correctly modelled and well controlled. To achieve this, the MC model of these B decays should be consistent with the latest measurements; and the contributions from PYTHIA should be checked and controlled.

4.2 Investigating the discrepancies using the $B \to D\pi$ sample

The FEI package uses 36 $B^{\pm}$ decay modes, listed in Tab. 3.4, to reconstruct hadronic B_{tag} candidates. This is more than previous hadronic B-tagging tools because the FEI package makes it easy to add new modes, since all the associated tasks are automated and the selection is learned by training on MC. As mentioned in Sec. 3.4.1, about 90% (80%) of the efficiency comes from a few decays in Belle (II) samples. In particular, the 11 decay modes of the form $B^+ \to \overline{D}^{(*)0}\, n\pi^{\pm}\, m\pi^0$, $D^-\, n\pi^{\pm}\, m\pi^0$, $\Lambda_c^- p\pi^+\pi^-\pi^+$, with $1 \leq n \leq 3$ and $0 \leq m \leq 1$, dominate the efficiency. Therefore, the following validation and improvement of the MC model is limited to these modes. Most of them are multi-body decays, and the full decay kinematics have often been poorly measured or not measured at all. They are thus modelled as a coherent sum of decays through many intermediate particles, with the relative fractions effectively describing the full decay kinematics. Although these intermediate particles are not explicitly required by the FEI package, BDTs could learn these structures during training on MC. Therefore, generating ideal (near-data) decay kinematics of these modes is essential for optimal training, efficiency, and purity of the FEI package, in addition to correct branching fractions affecting calibration factors.

As explained in the previous section, the MC models in Belle and Belle II are very different. To follow the current trend of performing analyses on combined Belle and Belle II data sets [94], it is important to have a compatible B-tagging tool for both experiments. Therefore, the MC model of the 11 decay modes of interest will be made compatible as much as possible. This will facilitate the maintenance, comparison and merging of the samples. To first examine the contents of the DECAY.DEC files, including the PYTHIA components, TopoAna [92] will be applied to the generic MC events and all B^+ decays will be stored along with their corresponding frequencies. Then, a new function based on 'decaylanguage' [95] is developed to find all possible decay chains leading to a given final state. This function uses both DECAY.DEC and the output of TopoAna to distinguish between decays generated by PYTHIA and those explicitly listed. In this way, the current branching fractions of the various contributions to the 11 final states of interest in the Belle and Belle II DECAY.DEC

files are listed in the Tables 4.2, 4.5, 4.7, 4.8, 4.9 and 4.10. In these tables, the decays produced by PYTHIA are shown in blue.

The decay descriptions of these modes are now being extensively studied to validate and update them with correct information to improve the training sample for FEI. Normally, the branching fractions for the explicitly listed decays are regularly updated based on the PDG [34]. However, one of the most common mistakes is to confuse the exclusive and inclusive measurements listed there, as will be shown later. For example, the inclusive (through all possible intermediate states) branching fraction for a mode listed in the PDG is interpreted as the branching fraction of only the non-resonant component of the same mode. Therefore, when necessary, the information in the corresponding publication is used directly instead of the PDG. On this basis, the proposed values for the relevant modes are listed in the tables of this section with further explanation, together with markers indicating the reason for the change according to the legend:

- $\triangle$: Update to the latest PDG or measurement.
- $\bigcirc$: Set to 0 to avoid generating it from PYTHIA.
- $\star$: Based on the update of $\overline{D}^{**}$.
- $\cup$: To fill the inclusive measurement.

Note that both Belle and Belle II MC samples are consistently changed according to these suggested values.

The control sample $B \to D\pi$ used to calculate the calibration factors in Sec. 3.4.2 can be used to validate these changes. The original intent of this sample was to achieve high purity on the signal side in order to study the decay kinematics on the tag side. After fitting the M_{recoil} distribution, the sPlot technique [96] can be used to unfold the signal component in other variables. Using the fit parameters, sPlot weights are calculated for each event using the hepstats Python package [97]. These weights can now be applied to any other variable to unfold different components (signal or background). For example, in Fig. 4.1, the M_{bc} of B_{tag} is shown: clear peak structures can be seen in the $D\pi$ and $D^*\pi$ enhanced components, while the background component has a threshold shape that can be modelled with an ARGUS shape, as expected. This shows that the fit of M_{recoil} is equivalent to the fit of M_{bc} of B_{tag}, i.e., the selected events are correctly reconstructed B_{tag} candidates. With this sPlot procedure, different tag-side kinematics for the signal components (both $B^+ \to \overline{D}^0\pi^+$ and $B^+ \to \overline{D}^{*0}\pi^+$) can be plotted and compared between data and MC. In other words, a tag-and-probe approach is employed: events from the well-measured $B^+ \to \overline{D}^{(*)0}\pi^+$ decays on one side are tagged using the fit, then the properties of B_{tag} (here, the decay kinematics) are probed using the sPlot technique.

To validate the changes, a new MC sample must be generated with the new DECAY.DEC file and processed with the FEI package. Since the FEI package used is still trained on the old MC, it may distort the distributions. If a discrepancy is found despite the bias due to the training, this can be considered as a significant discrepancy. It is important to note that the changes to DECAY.DEC are not tuned to the data, which would not be possible anyway due to the distortion introduced by training with the old MC. All updates are based on previously published measurements or, if necessary, on reasonable choices. The $D\pi$ control sample is used for validation purposes only. In this way, the measurements that are primarily needed to improve the model of hadronic B decays are identified and then reported to the collaboration.

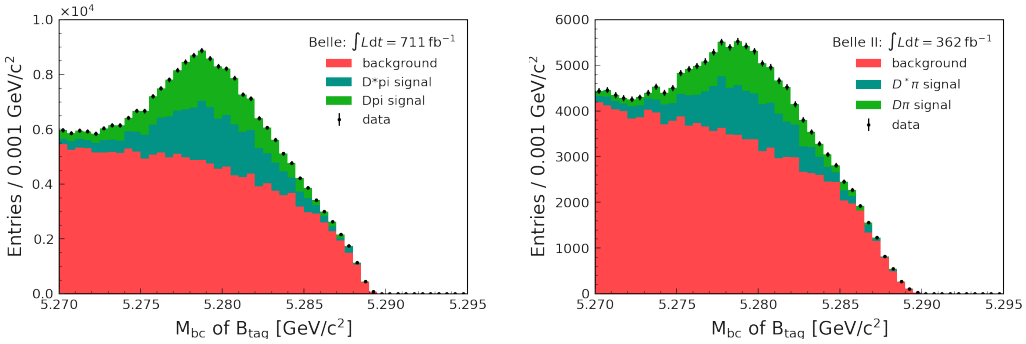


FIGURE 4.1: $M_{\rm bc}$ of $B_{\rm tag}$ distribution with sPlot weights to unfold different categories (signal and background) in Belle (left) and Belle II (right) samples.

Using the new MC, which is closer to the data (though not tuned), would result in calibration factors closer to 1, as shown at the end of this section. However, the FEI package must be trained with the new MC to verify improvements beyond CF, as shown in Sec. 4.3.

FEI decay mode: $B^+ \to \overline{D}^{(*)0}\pi^+$

B^+ FEI mode	Contribution	$\mathcal{B}^{\rm Belle}(\%)$	$\mathcal{B}^{\rm Belle\ II}(\%)$	$\mathcal{B}^{\rm Proposed}(\%)$	Reason
$\overline{D}^0\pi^+$	$\overline{D}^0\pi^+$	0.48	0.47	0.47	△
		0.48	**0.47**	**0.47**	
$\overline{D}^{*0}\pi^+$	$\overline{D}^{*0}\pi^+$	0.52	0.49	0.49	△
		0.52	**0.49**	**0.49**	

TABLE 4.2: Update of branching fractions. The contributions to each of the $B^+ \to \overline{D}^{(*)0}\pi^+$ FEI modes are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column and the proposed common branching fractions in the fifth column. The sum is shown in bold. The marker in the sixth column indicates the reason for the change.

As two-body decays, the $B^+ \to \overline{D}^{(*)0}\pi^+$ modes are the easiest of the 11 decay modes to measure. The branching fractions in Belle II MC are already in agreement with the latest measurements, as shown in Tab. 4.2. Belle MC has been updated accordingly to match.

For the data, the old MC and the new MC of the Belle and the Belle II data sets, the sPlot weights are calculated using the $M_{\rm recoil}$ fit. Using these weights, the signal component is enhanced. In this signal component, the ΔE distribution of the $B_{\rm tag}$ in $B^+ \to \overline{D}^0\pi^+$ mode is shown in Fig. 4.2. As expected, both MC samples in Belle II look the same, and the overall agreement between data and MC is reasonable.

In such cases, where the B modelling is reliable, a second-order effect of D decay modelling can become significant. This is evident from the calibration factors of the $B^+ \to \overline{D}^{(*)0}\pi^+$ modes split based on different $\overline{D}^0$ decay modes, as shown in Fig. 4.3. Again, the modelling of D^0 decay is different in Belle and Belle II MC, and we note that in general Belle II seems to be improved. Further improvements are possible, but are beyond the scope of this thesis.

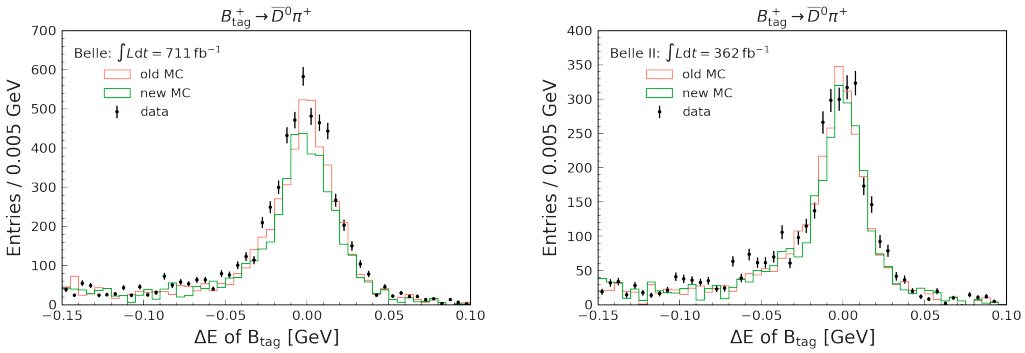


FIGURE 4.2: Comparison of ΔE of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+$ mode of the FEI package in Belle (left) and Belle II (right).

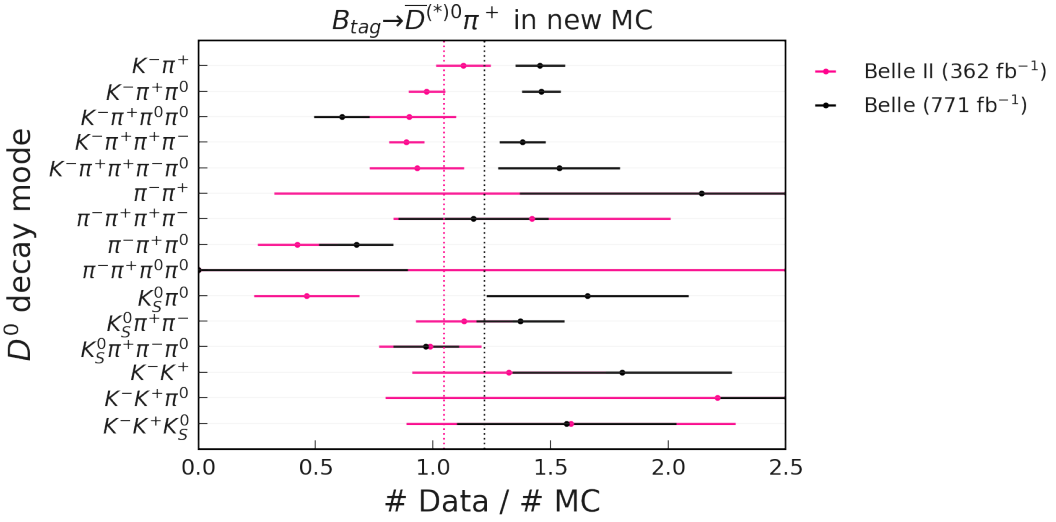


FIGURE 4.3: Calibration factors of $B^+ \to \overline{D}^{(*)0}\pi^+$ decay modes produced by the FEI package split based on different D^0 decay modes, for Belle and Belle II data sets. The dashed lines represent the weighted average.

Interlude: $B^+ \to \overline{D}^{**0}\pi^+$

The decays of the form $B^+ \to \overline{D}^{**0}\pi^+$ are not directly reconstructed by the package FEI. However, they serve as intermediate states for some of the 11 decay modes of interest. Since they were modelled differently in Belle and Belle II MC, a consistent model is created for them.

D^{**} is in fact a collection of four P-wave excitations of D mesons, as shown in Fig. 4.4. Two of them (D_1 and D_2^*) have narrow masses, and the other two (D_0^* and D_1') have broad masses. The decays of the $\overline{D}^{**0}$ particles themselves are first made consistent by propagating the latest Belle II model in Belle MC, as shown in Tab. 4.3.

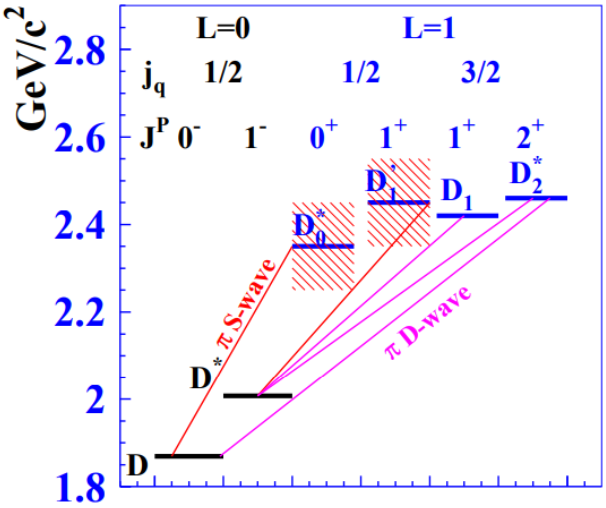


FIGURE 4.4: Spectroscopy of D meson excitations.

The four possible $B^+ \to \overline{D}^{**0}\pi^+$ decays are listed in Table 4.4, together with the proposed branching fractions.

From the measurements to the final state $B^+ \to D^-\pi^+\pi^+$ by the Belle [98], BaBar [99] and LHCb [100] experiments, we obtain a consistent branching fraction for $\mathcal{B}(B^+ \to \overline{D}_2^{*0}[\to D^-\pi^+]\pi^+) = (3.56 \pm 0.24) \times 10^{-4}$. Assuming that $\mathcal{B}(\overline{D}_2^{*0} \to D^-\pi^+) \sim 40\%$, this leads to $\mathcal{B}(B^+ \to \overline{D}_2^{*0}\pi^+) \sim 0.09\%$. The same decay is also measured in the final state $B^+ \to D^{*-}\pi^+\pi^+$ final state by the Belle [98] and LHCb [101] experiments, giving $\mathcal{B}(B^+ \to \overline{D}_2^{*0}[\to D^{*-}\pi^+]\pi^+) = (2.1 \pm 1.0) \times 10^{-4}$. Assuming $\mathcal{B}(\overline{D}_2^{*0} \to D^{*-}\pi^+) \sim 27\%$ leads to $\mathcal{B}(B^+ \to \overline{D}_2^{*0}\pi^+) \sim 0.078\%$. Since the branching fraction set in Belle MC is consistent with the average of these results, it is retained.

In the same measurements, the branching fraction $\mathcal{B}(B^+ \to \overline{D}_1^0[\to D^{*-}\pi^+]\pi^+) = (7.4 \pm 1.0) \times 10^{-4}$ is obtained. Assuming $\mathcal{B}(\overline{D}_1^0 \to D^{*-}\pi^+) \sim 40\%$ leads to $\mathcal{B}(B^+ \to \overline{D}_1^0\pi^+) \sim 0.186\%$. The branching fraction for this mode is lower by a factor of 2 in the Belle (II) MC, and is now corrected.

The decays involving broad states are difficult to measure, but the branching fractions set for $B^+ \to \overline{D}_0^{*0}\pi^+$ and $B^+ \to \overline{D}_1^{\prime 0}\pi^+$ in Belle MC are consistent with the measurements $\mathcal{B}(B^+ \to \overline{D}_0^{*0}[\to D^-\pi^+]\pi^+) = (6.4 \pm 1.4) \times 10^{-4}$ [98, 99] and $\mathcal{B}(B^+ \to \overline{D}_1^{\prime 0}[\to D^{*-}\pi^+]\pi^+) = (3.5 \pm 0.9) \times 10^{-4}$ [98, 101] respectively. Thus, they are simply propagated without any change. This leads to a total branching fraction of 0.42% for the $B^+ \to \overline{D}^{**0}\pi^+$ decays. This is consistent with the inclusive measurement of BaBar [71] $(0.57 \pm 0.12 \pm 0.02)\%$, albeit on the lower side. The effects of all these changes are shown later in the corresponding final states.

$\overline{D}^{**0}$ particle	Decay	$\mathcal{B}^{\text{Belle}}$	$\mathcal{B}^{\text{Belle II}}$	$\mathcal{B}^{\text{Proposed}}$
$\overline{D}_0^{*0}$	$\overline{D}^0\pi^0$	0.33	0.33	0.33
	$D^-\pi^+$	0.67	0.67	0.67
		1	**1**	**1**
$\overline{D}_1^0$	$\overline{D}^{*0}\pi^0$	0.33	0.20	0.20
	$D^{*-}\pi^+$	0.67	0.40	0.40
	$\overline{D}^0\pi^+\pi^-$	-	0.17	0.17
	$\overline{D}^0\pi^0\pi^0$	-	0.11	0.11
	$D^-\pi^+\pi^0$	-	0.11	0.11
		1	**1**	**1**
$\overline{D}_1^{\prime 0}$	$\overline{D}^{*0}\pi^0$	0.33	0.33	0.33
	$D^{*-}\pi^+$	0.67	0.67	0.67
		1	**1**	**1**
$\overline{D}_2^{*0}$	$\overline{D}^{*0}\pi^0$	0.10	0.13	0.13
	$D^{*-}\pi^+$	0.21	0.27	0.27
	$\overline{D}^0\pi^0$	0.23	0.20	0.20
	$D^-\pi^+$	0.46	0.40	0.40
		1	**1**	**1**

TABLE 4.3: Update of branching fractions. The decays of the various $\overline{D}^{**0}$ are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column and the proposed common branching fractions in the fifth column. The sum is shown in bold.

B^+ decay	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$	$\mathcal{B}^{\text{Proposed}}(\%)$
$\overline{D}_0^{*0}\pi^+$	0.083	0.061	0.083
$\overline{D}_1^0\pi^+$	0.093	0.088	0.186
$\overline{D}_1^{\prime 0}\pi^+$	0.062	0.050	0.062
$\overline{D}_2^{*0}\pi^+$	0.085	0.052	0.085
	0.32	**0.25**	**0.42**

TABLE 4.4: Update of branching fractions. The different $B^+ \to \overline{D}^{**0}\pi^+$ decays are shown in the first column, with the corresponding *official* Belle (II) branching fractions in the second (third) column and the proposed common branching fractions in the fourth column. The sum is shown in bold.

FEI decay mode: $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0$

The decay $B^+ \to \overline{D}^{*0}\pi^+$ can also contribute to the FEI final state $B^+ \to \overline{D}^0\pi^+\pi^0$ via $\overline{D}^{*0} \to \overline{D}^0\pi^0$ ($0.49\% \times 0.62 \sim 0.30\%$), and so the update of the latest PDG is also reflected in $\overline{D}^0\pi^+\pi^0$, as shown in Tab. 4.5.

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$	$\mathcal{B}^{\text{Proposed}}(\%)$	Reason
$\overline{D}^0\pi^+\pi^0$	$\overline{D}^0\rho^+$	1.34	1.34	1.34	
	$\overline{D}^{*0}\pi^+$	0.32	0.30	0.30	△
	$\overline{D}^0\pi^+\pi^0$ (NR)	0.05	0.05	0.00	◯
	$\overline{D}_0^{*0}\pi^+$	0.03	0.02	0.03	⋆
	$\overline{D}_2^{*0}\pi^+$	0.02	0.01	0.02	⋆
		1.76	**1.72**	**1.69**	
$\overline{D}^{*0}\pi^+\pi^0$	$\overline{D}^{*0}\rho^+$	1.55	0.98	0.98	△
	$\overline{D}^{*0}\pi^+\pi^0$ (NR)	0.05	0.05	0.00	◯
	$\overline{D}_1^0\pi^+$	0.03	0.02	0.04	⋆
	$\overline{D}_1'^0\pi^+$	0.02	0.02	0.02	⋆
	$\overline{D}_2^{*0}\pi^+$	0.01	0.01	0.01	⋆
		1.66	**1.08**	**1.05**	

TABLE 4.5: Update of branching fractions. The contributions to each of the $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0$ FEI modes are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column and the proposed common branching fractions in the fifth column. The sum is shown in bold. The marker in the sixth column indicates the reason for the change.

The presence of $\overline{D}^{*0}$ is visualized in $\Delta M = M(\overline{D}^{*0}) - M(\overline{D}^0)$ (corresponding to the released energy, Q) due to its better resolution. In Fig. 4.5, the contribution of $\overline{D}^{*0}$ in the $B^+ \to \overline{D}^0\pi^+\pi^0$ is shown using the sPlot technique. Ideally, all $B^+ \to \overline{D}^{*0}\pi^+$ decays should be reconstructed as $\overline{D}^{*0}$, but some events leak into the $B^+ \to \overline{D}^0\pi^+\pi^0$ final state, seen as the small peak structure in the distribution of the Belle sample.

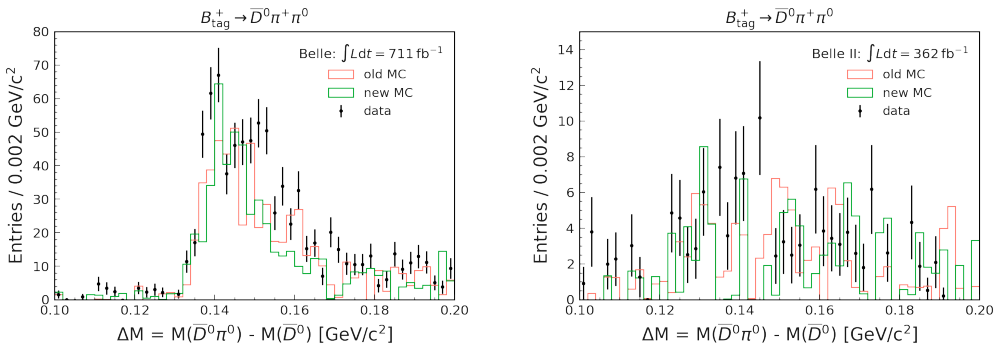


FIGURE 4.5: Comparison of ΔM of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

Apart from this, the main contribution to this final state comes from $B^+ \to \overline{D}^0\rho^+$, which was measured to be $(1.34 \pm 0.18)\%$ by CLEO in 1994 [102]. This can be illustrated in the $\text{M}(\pi^+\pi^0)$ distribution shown in Fig. 4.6. Here, the contribution from $\overline{D}^{*0}$ is vetoed with a cut on the ΔM. The rho component still looks overestimated in MC, which could be explained by the large uncertainty of the measurement due to the

small statistics (0.9 fb^{-1}) used. Therefore, a new measurement with larger statistics is needed for this mode.

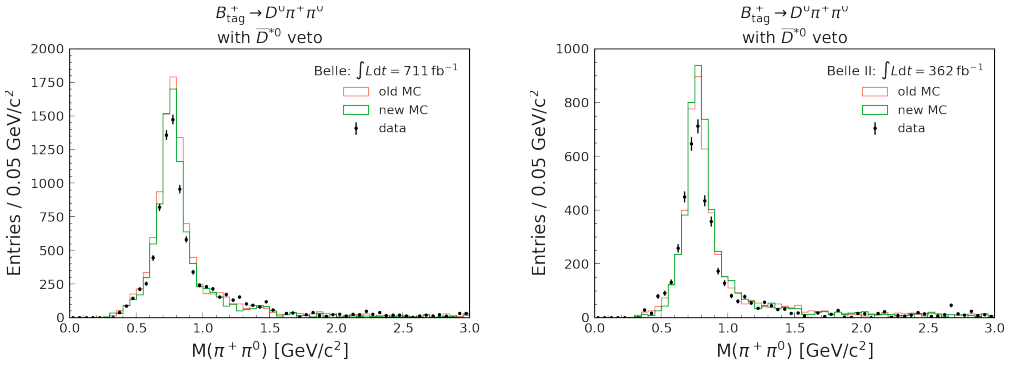


FIGURE 4.6: Comparison of M($\pi^+\pi^0$) of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

Using a larger data set of 9 fb^{-1}, CLEO [103] updated the measurement of $\mathcal{B}(B^+ \to \overline{D}^{*0}\rho^+)$ in 2003 and obtained $(0.98 \pm 0.17)\%$. This last result was not updated in Belle MC and is therefore now corrected. The effect of this update can be seen in the distribution of M($\pi^+\pi^0$) in the Belle samples (left) in Fig. 4.7.

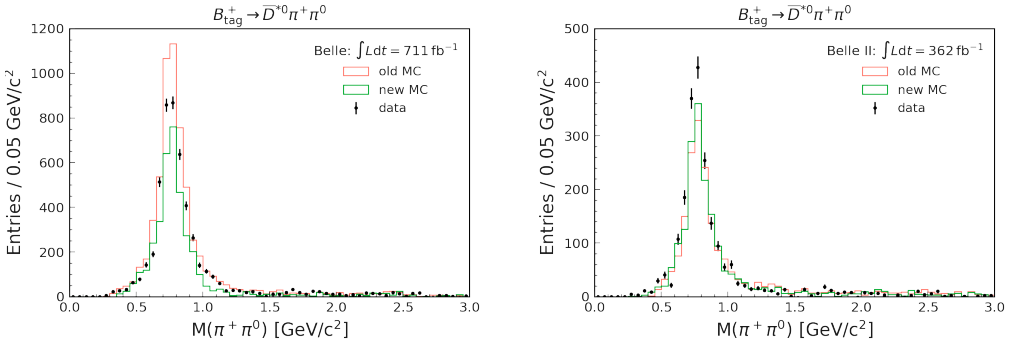


FIGURE 4.7: Comparison of M($\pi^+\pi^0$) of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

The contributions to these final states by $\overline{D}^{**}\pi^+$ are statistically too small to be seen. The arbitrarily set non-resonant component in both final states is now explicitly set to 0 to avoid being generated by PYTHIA.

Interlude: $B^+ \to \overline{D}^{**0}\pi^+\pi^0$

Belle MC assumed that the branching ratios of $B^+ \to \overline{D}^{**0}\rho^+$ are the same as those of $B^+ \to \overline{D}^{**0}\pi^+$. On the other hand, Belle II MC applied some predictions based on some old recommendations of BaBar. Following the prediction [104]

$$\frac{\mathcal{B}(B \to D\rho)}{\mathcal{B}(B \to D\pi)} \sim 2$$

for both D and D^* in both B^+ and B^0, we assume that it also holds for D^{**}. Therefore, all $B^+ \to \overline{D}^{**0}\rho^+$ branching fractions are assumed to be twice the $B^+ \to \overline{D}^{**0}\pi^+$ branching fractions, see the fourth column in Tab. 4.6. And just as in the $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0$ modes, all the contribution is assumed to come from ρ, and the non-resonant components are set to 0 to avoid being generated by PYTHIA.

B^+ decay	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$	$\mathcal{B}^{\text{Proposed}}(\%)$
$\overline{D}_0^{*0}\rho^+$	0.083	-	0.166
$\overline{D}_1^{0}\rho^+$	0.093	0.080	0.372
$\overline{D}_1^{\prime 0}\rho^+$	0.062	0.220	0.124
$\overline{D}_2^{*0}\rho^+$	0.085	0.380	0.170
$\overline{D}_0^{*0}\pi^+\pi^0$	0.106	-	0
$\overline{D}_1^{0}\pi^+\pi^0$	0.040	-	0
$\overline{D}_1^{\prime 0}\pi^+\pi^0$	0.115	-	0
$\overline{D}_2^{*0}\pi^+\pi^0$	0.117	-	0
	0.70	**0.68**	**0.84**

TABLE 4.6: Update of branching fractions. The different $B^+ \to \overline{D}^{**0}\pi^+\pi^0$ decays are shown in the first column, with the corresponding *official* Belle (II) branching fractions in the second (third) column and the proposed common branching fractions in the fourth column. Contributions from PYTHIA are shown in blue, while the sum is shown in bold.

The effect of these changes will be shown later in the corresponding final states.

FEI decay mode: $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-$

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$	$\mathcal{B}^{\text{Proposed}}(\%)$	Reason
$\overline{D}^0\pi^+\pi^-\pi^+$	$\overline{D}^0\pi^+\pi^-\pi^+$ (NR)	0.46	0.51	0	◯
	$\overline{D}^0\rho^0\pi^+$	0.39	0.42	0	◯
	$\overline{D}^0 a_1^+$	0.18	0.26	0.58	△
	$\overline{D}_1^0\pi^+$	0.04	0.04	0.08	⋆
	$\overline{D}_1'^0\pi^+$	0.03	0.02	0.03	⋆
	$\overline{D}_2^{*0}\pi^+$	0.01	0.01	0.01	⋆
	$D^{*-}\pi^+\pi^+$	-	0.09	0	◯
		1.11	**1.36**	**0.70**	
$\overline{D}^{*0}\pi^+\pi^-\pi^+$	$\overline{D}^{*0}\pi^+\pi^-\pi^+$ (NR)	1.03	0.20	0	◯
	$\overline{D}^{*0}\rho^0\pi^+$	0	0.04	0	◯
	$\overline{D}^{*0} a_1^+$	0.91	1.06	1.06	△
	$\overline{D}^{*0} f_0\pi^+$	0.07	-	0	◯
		2.01	**1.25**	**1.01**	

TABLE 4.7: Update of branching fractions. The contributions to each of the $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-$ FEI modes are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column and the proposed common branching fractions in the fifth column. Contributions from PYTHIA are shown in blue, while the sum is shown in bold. The marker in the sixth column indicates the reason for the change. Contributions smaller than 10^{-4} are omitted.

The branching fractions of the contributions to the $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-$ final state are given in Tab. 4.7. Due to the absence of π^0 in the $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ final state, this is an ideal mode to study the contributions of $\overline{D}^{**0}$ decays. In Belle II MC, an additional mode $D^{*-}\pi^+\pi^+$ is added. This is a classic error in interpreting the PDG listing of the inclusive measurement of $\overline{D}^{**0}$ [98, 101] as an exclusive non-resonant contribution, leading to double counting. Therefore, this contribution is set to 0 to avoid being generated by PYTHIA. Such errors are more common in Belle II because in the past an automatic PDG parser was used to update the DECAY.DEC file. Fig. 4.8 shows the effects of removing this non-resonant contribution in Belle II (right).

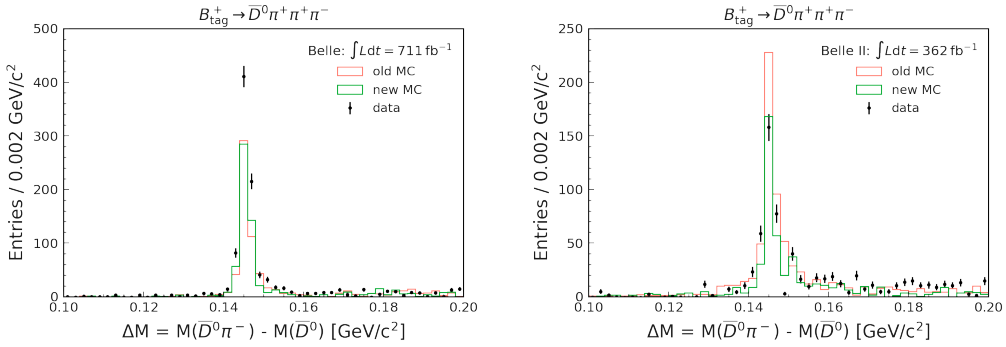


FIGURE 4.8: Comparison of ΔM of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

This D^{*-} contribution is in turn expected to originate from $\overline{D}^{**0}$ decays, as shown by the $\overline{D}_1^0$ peak in Fig. 4.9. Since there are two π^+'s in the final state, the combination that gives the minimum $M(\overline{D}^0\pi^+\pi^-)$ is used, as we expect the pion from the $D^{**} \to D^*\pi$ transition to have a low momentum.

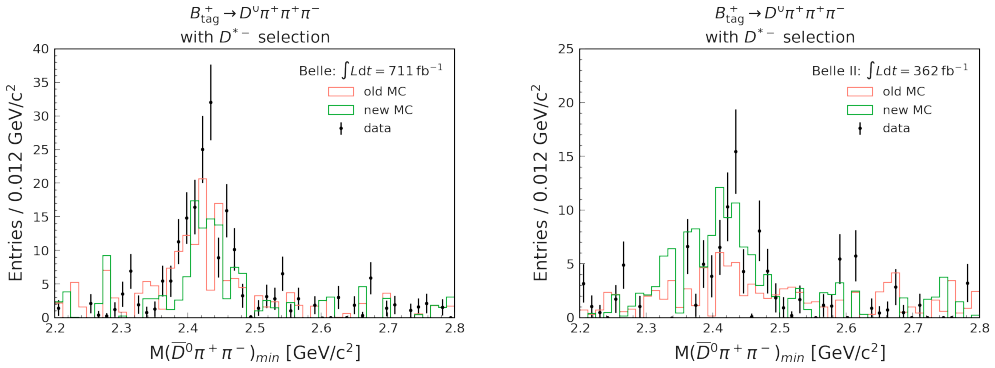


FIGURE 4.9: Comparison of $M(\overline{D}^0\pi^+\pi^-)_{\text{min}}$ with D^{*-} selection of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

Regarding the predominant contributions, an early paper from CLEO [105] in 1992, using only 212 pb^{-1} data, measured the four-body $\overline{D}^0\pi^+\pi^+\pi^-$ rate as $(1.15 \pm 0.29 \pm 0.21)\%$ and attempted to disentangle the three subcomponents $\overline{D}^0\pi^+\pi^+\pi^-$, $\overline{D}^0\rho^0\pi^+$, and the resonant $\overline{D}^0a_1^+$, assigning them the fractions 48%, 34%, and 18%, respectively. The only other measurement comes from an early data set of LHCb [106], corresponding to 35 pb^{-1}, which gave the ratio of the branching fractions $\mathcal{B}(B^+ \to \overline{D}^0\pi^-\pi^+\pi^-)/\mathcal{B}(B^+ \to \overline{D}^0\pi^+) = 1.27 \pm 0.06 \pm 0.11$. It contributes to the $\mathcal{B}(B^+ \to \overline{D}^0\pi^+\pi^+\pi^-) = (0.56 \pm 0.21)\%$ quoted in PDG [34], where the uncertainty is scaled considerably (by a factor of 3.7 to account for the large difference between CLEO and LHCb measurements). The LHCb study shows a prominent structure at low mass – consistent with the $a_1(1260)^+$ resonance – and a long tail extending up to 3 GeV/c^2. In the higher $M(\pi^+\pi^+\pi^-)$ regions, the main contributions are expected from $\overline{D}^0_{1,2}\pi^+$ decays as described in Fig. 4.10 from LHCb data.

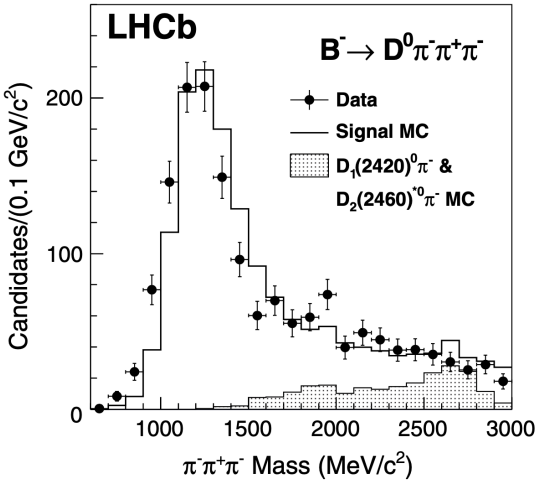


FIGURE 4.10: 3π-system invariant mass for $B^- \to D^0\pi^-\pi^-\pi^+$ candidates. Data points are overlaid with MC simulation (solid and shaded lines). Taken from Ref. [106].

Unfortunately, no quantitative information is given, except for the total $\mathcal{B}$ of

$\overline{D}^0\pi^+\pi^+\pi^-$ (or, more precisely, its ratio to $\bar{D}^0\pi^+$). But the ideal way to model this mode with the known information would be through a_1, setting the $\rho\pi$ and 3π non-resonant (NR) components to 0, as shown in Fig. 4.11. To focus on these components, the contribution from the D^{*-} shown in Fig. 4.8 is vetoed with a selection on ΔM.

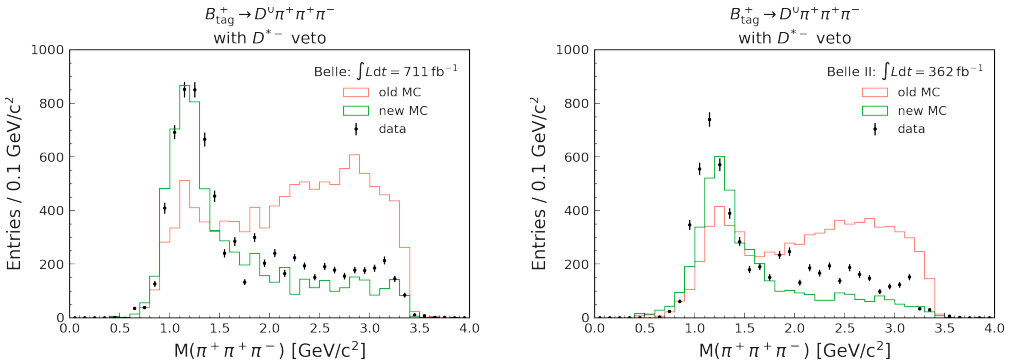


FIGURE 4.11: Comparison of $M(\pi^+\pi^+\pi^-)$ of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

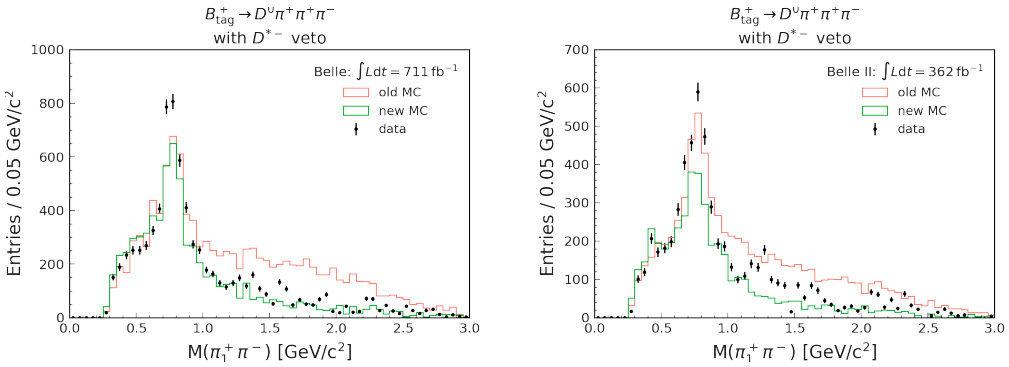


FIGURE 4.12: Comparison of $M(\pi_1^+\pi^-)$ of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

These plots confirm the dominance of Da_1 in the data, while an exaggerated contribution of the NR $D3\pi$ and $D\rho\pi$ components is present in MC. The effects of these changes on the invariant masses of 2π systems are shown in Figs. 4.12 and 4.13, which are clearly dominated by ρ^0 peaks from a_1^+ decays.

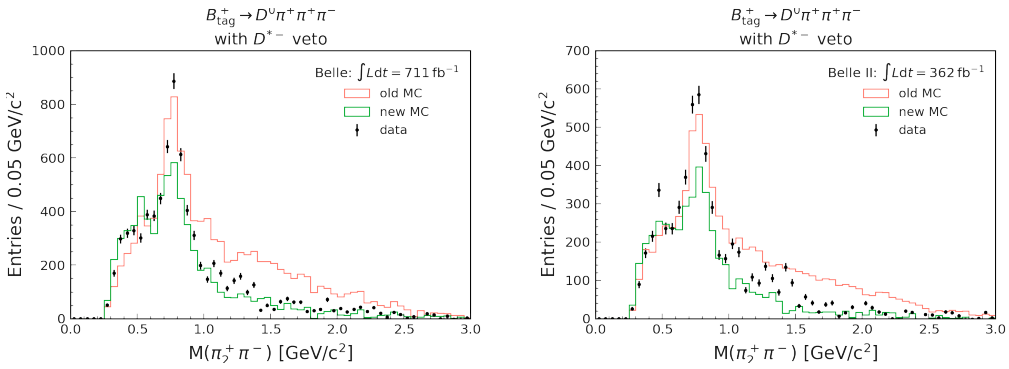


FIGURE 4.13: Comparison of $M(\pi_2^+\pi^-)$ of $B_{\rm tag}$ between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

In 1994, CLEO [102] measured $\mathcal{B}(B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-) = (0.94 \pm 0.20 \pm 0.17)\%$ with a sample of $0.9\,\text{fb}^{-1}$ and found that the mass of the three pions is predominantly in the a_1^+ mass range $(1.0, 1.6)$ GeV/c^2. This result was misinterpreted in the Belle decay file and attributed to a non-resonant 4-body contribution. In 2004, Belle [107] measured $\mathcal{B}(B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-) = (1.06 \pm 0.05 \pm 0.13)\%$ and confirmed the dominance of a_1^+, but unfortunately without measuring anything other than the total branching fraction. It is decided to set the non-resonant $\overline{D}^{*0}\pi^+\pi^+\pi^-$ contribution to zero and also to remove the PYTHIA contribution (see Table 4.7). It is worth noting that in the decay table of Belle II the non-resonant $\overline{D}^{*0}\pi^+\pi^+\pi^-$ contribution was commented out with the correct intention of removing it; however, since its branching fraction was not explicitly set to zero, PYTHIA generates it with a $\mathcal{B}$ of 0.2%, resulting in double counting. Fig. 4.14 shows the dominance of the a_1^+ contribution in the data, consistently with the new MC.

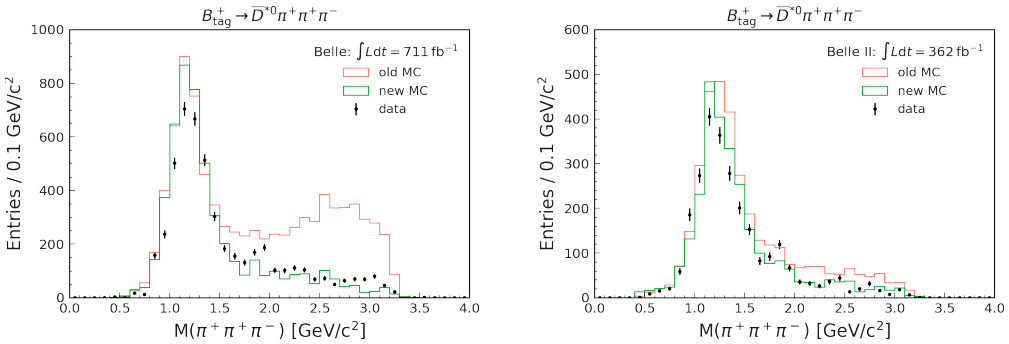


FIGURE 4.14: Comparison of $M(\pi^+\pi^+\pi^-)$ of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

The 2π intermediate states are shown in Figs. 4.15 and 4.16. The agreement between the data and MC in these plots is clear but still not prefect. In the data, most of the events seem to come from $a_1^+ \to \rho_0\pi^+$, but in MC there is a significant contribution from $a_1^+ \to \sigma_0\pi^+$ in addition to this. However, changing the a_1^+ decay model is beyond the scope of this project, as it has implications for various particle decays beyond the hadronic B decays.

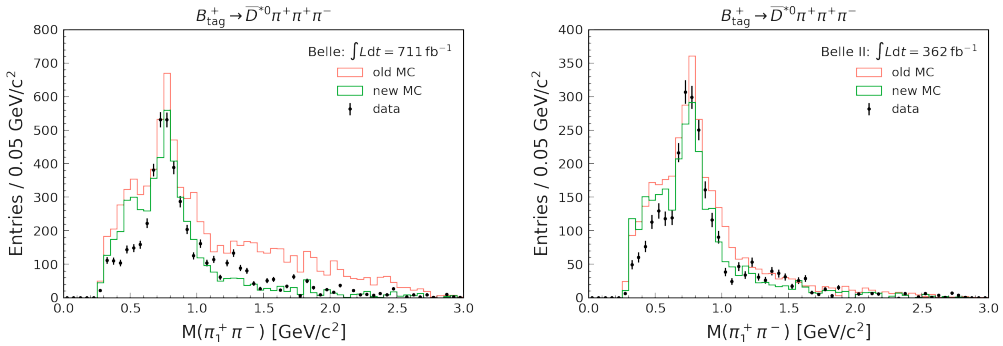


FIGURE 4.15: Comparison of $M(\pi_1^+\pi^-)$ of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

As a side effect of these changes, the symmetric modes $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0\pi^0$, which are also reconstructed by the FEI package though with a lower efficiency due to π^0s, are also modified. Following the same pattern, the non-resonant $\pi^+\pi^0\pi^0$ and $\rho^+\pi^0$ components are set to 0, and the contribution of $B^+ \to \overline{D}^{(*)0}a_1^+$ is expected to

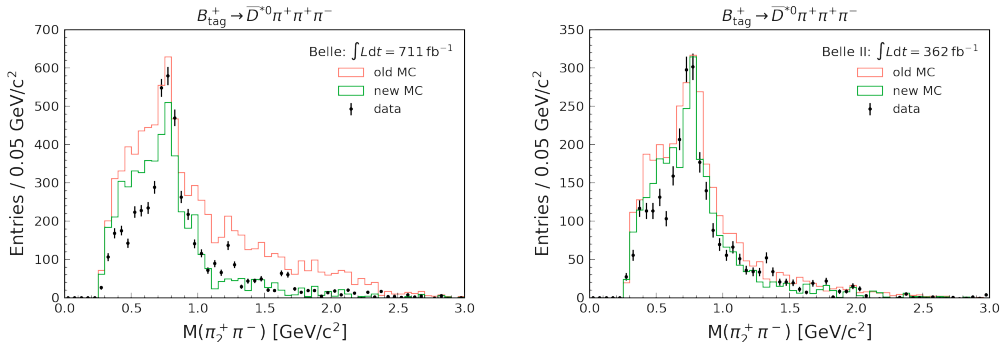


FIGURE 4.16: Comparison of $M(\pi_2^+\pi^-)$ of B_{tag} between the data, the old MC and the new MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$ mode of the FEI package in Belle (left) and Belle II (right).

dominate. There will also be some inconsistencies in these modes as well because the decay model of a_1^+ is not perfect and differs between Belle and Belle II MC.

FEI decay mode: $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$

As multiplicity increases, there are more possible intermediate states, and the decay model becomes more complicated, as shown in Tab. 4.8. The main contribution for $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ comes from a 1990 measurement of $D^{*-}\pi^+\pi^+\pi^0$ from ARGUS [108], which gives $\mathcal{B}(B^+ \to D^{*-}\pi^0\pi^+\pi^+) = (1.5 \pm 0.7)\%$, which is set as non-resonant contribution in MC. Unfortunately, there is a large uncertainty due to the small statistic used (229 pb^{-1}), and there are no other measurement on this subject. This is clearly overestimated, and it should happen through $\bar{D}^{**0}\rho^+$ with $\bar{D}^{**0} \to D^*(2010)^-\pi^+$. After setting it to zero along with $D^{*-}\rho^+\pi^+$, the comparison of ΔM in the data can be seen in Fig. 4.17. As expected, the $\bar{D}^{**0}$ contributions are sufficient to agree with the data.

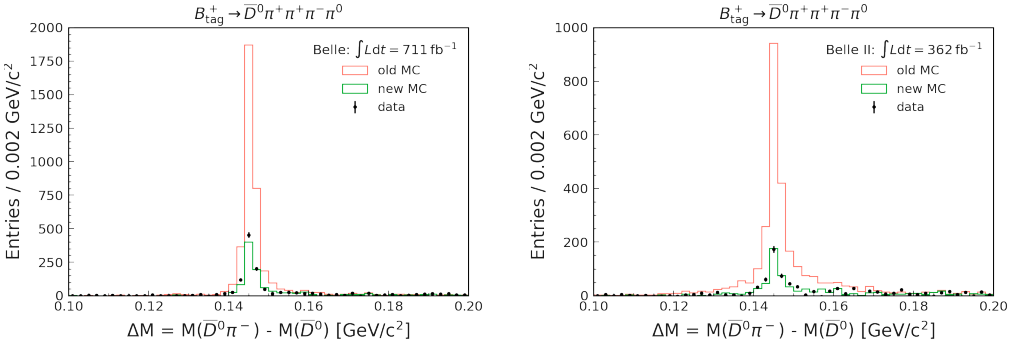


FIGURE 4.17: Comparison of ΔM between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

The decay mode $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$, also produces this final state through $\overline{D}^{*0} \to \overline{D}^0\pi^0$. Hence, the changes implemented for that final state are also reflected here. The only other known contribution is $\mathcal{B}(B \to D^0\omega(782)\pi^-) = (0.41 \pm 0.07 \pm 0.06)\%$ measured by CLEO [109], which also claims a first observation of $B \to D^{(*)}\rho'^-, \rho'^- \to \omega\pi^-$. However, the ρ'^- meson is currently not well studied, and its decay model cannot yet be implemented. The ω peak can be seen in Fig. 4.18 at 0.782 GeV/c^2, where there is good agreement with the data.

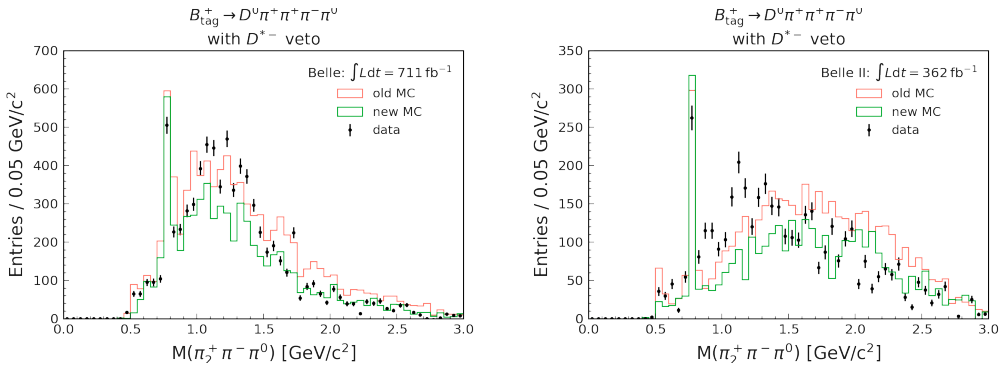


FIGURE 4.18: Comparison of $M(\pi_2^+\pi^-\pi^0)$ between data, old MC and new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

This is only one combination since there are two π^+'s in the final state, but the other combination looks similar. In the same plot, another peak corresponding to η mass can be seen at 0.547 GeV/c^2 in MC, produced entirely by PYTHIA. However, it is not clear whether the corresponding peak in the data is biassed by the FEI training. There is an ongoing measurement in Belle II to find the branching fraction of this contribution, but in the meantime, a branching fraction of 0.04% is assigned to it to control PYTHIA.

In the absence of better measurements for this final state, the remaining contributions are generated by PYTHIA. One of them, namely $\rho^+\rho^0$, is constrained (set to 0) because there is no significant signal in the data, while it is overestimated in the old MC(see Fig. 4.19). As mentioned in Section 4.1, the a_1^+ and a_1^0 particles are not

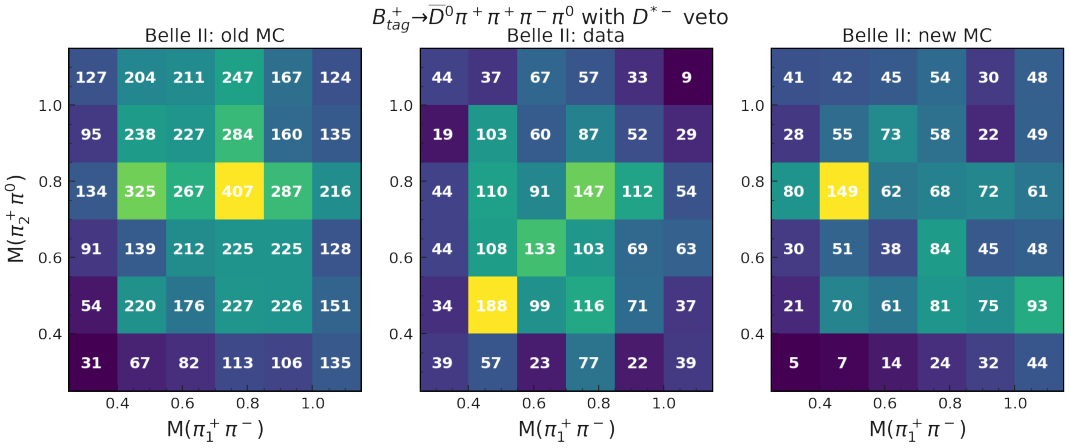


FIGURE 4.19: Comparison of a 2D histogram between $M(\pi_1^+\pi^-)$ vs $M(\pi_2^+\pi^0)$ between the old MC (left), data (middle) and the new MC (right) in the $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in the Belle II data set. The yield in each bin is shown for reference. The $\rho^+\rho^0$ component is overestimated in the old MC, as can be seen at the ρ mass (0.77 GeV/c^2). A similar effect is seen in the Belle data set, as well as in the $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$ mode.

produced by PYTHIA in Belle (II) MC, which might better represent the peak visible in the data in Fig. 4.20.

The contributions of PYTHIA to $\rho\pi\pi$ modes appear to be consistent with Belle's data, but not with those of Belle II, underlining the differences caused by different versions and configurations of PYTHIA. A similar effect is seen for the $M(\pi^+\pi^+\pi^-\pi^0)$ in Fig. 4.21. Unlike all previous figures in this section, this $M(\pi^+\pi^+\pi^-\pi^0)$ distribution is different in the data itself between the Belle and Belle II data sets. This is clear

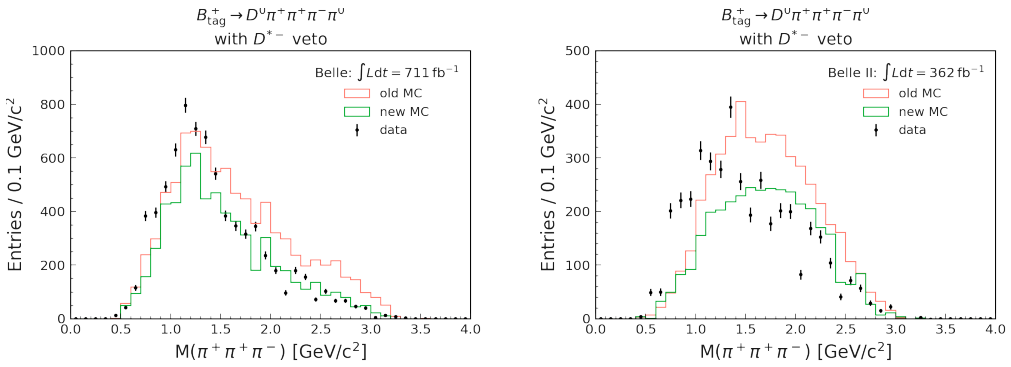


FIGURE 4.20: Comparison of $M(\pi^+\pi^+\pi^-)$ between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

evidence of the bias introduced by training on MC. The events at low mass in the Belle II data are striking, despite the training on MC, which is dominated by high mass. The phase-space model of EvtGen would distribute all particles uniformly, but the kinematic model generated by PYTHIA distributes the particles from the W^+ chain uniformly. Therefore, the contributions of PYTHIA in Belle, which seem to describe the data better, cannot be replicated simply by mentioning them explicitly. The lack of control over PYTHIA contributions leads to different sum of contributions

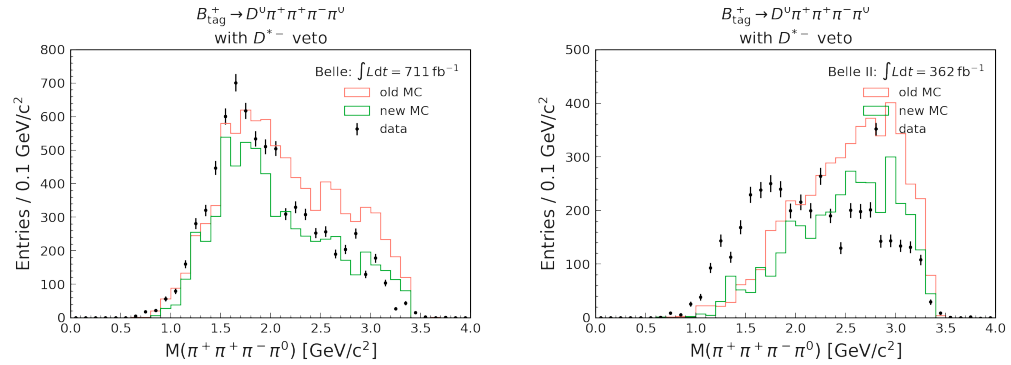


FIGURE 4.21: Comparison of $M(\pi^+\pi^+\pi^-\pi^0)$ between the data, the old MC and the new MC in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

in Belle and Belle II MC. And these sums change over time as the total quark-level PYTHIA contribution changes in the DECAY.DEC file.

These two modes, $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$, were not part of earlier hadronic B-tagging packages. In the FEI package, they account for 20% (16%) of the total efficiency in Belle (II) samples. The lack of control over the model is not ideal for use in FEI reconstruction, but generation with PYTHIA is the only option, given the poor information available. There is a great need for better measurements of these final states, especially to describe the intermediate resonance structures such as the ρ' meson.

The CLEO analysis [109] also gives results for the final state $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$, with $\mathcal{B}(B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0) = (1.80 \pm 0.24 \pm 0.27)\%$ and $\mathcal{B}(B^+ \to \overline{D}^{*0}\omega\pi^+) = (0.45 \pm 0.10 \pm 0.07)\%$. The first measurement here is the inclusive branching fraction, and again from PDG, it is misinterpreted as the branching fraction of the non-resonant component in Belle (II) MC. This is now corrected, by setting it to zero. The ω contribution is clearly present in one of the combinations of 3π invariant mass, as shown in Fig. 4.22. The other combination looks similar.

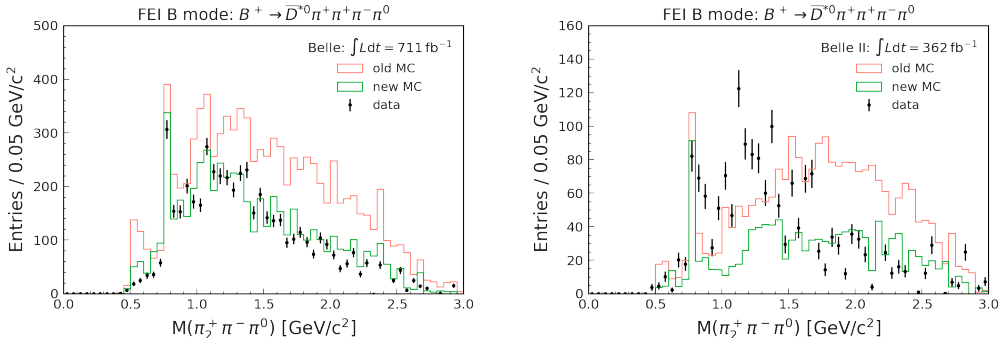


FIGURE 4.22: Comparison of $M(\pi_2^+\pi^-\pi^0)$ between the data, the old MC and the new MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

Following the same strategy as before, the $\overline{D}^{*0}\eta\pi^+$ and $\overline{D}^{*0}\rho^+\rho^0$ contributions are constrained, and the remainder is generated by PYTHIA in the form of $\rho\pi\pi$ intermediate states. Although the inclusive branching fraction is measured here, it cannot be constrained due to the lack of control over the contributions from PYTHIA, but the obtained sums are within the uncertainty of the measurement. Figures 4.23 and 4.24 show that the PYTHIA configuration of Belle is in better overall agreement with the data than Belle II.

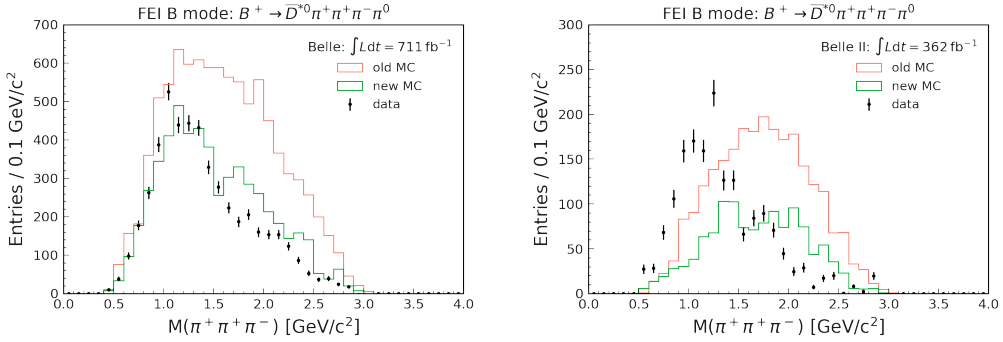


FIGURE 4.23: Comparison of M($\pi^+\pi^+\pi^-$) between data, old MC and New MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

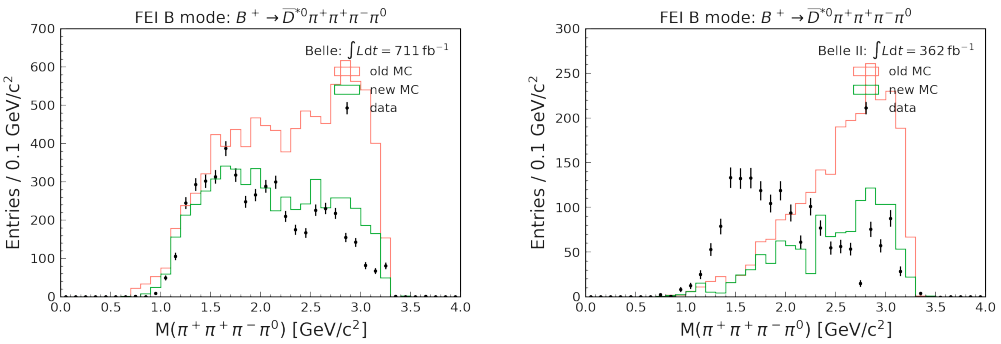


FIGURE 4.24: Comparison of M($\pi^+\pi^+\pi^-\pi^0$) between data, old MC and New MC in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$	$\mathcal{B}^{\text{Proposed}}(\%)$	Reason
$\overline{D}^0\pi^+\pi^-\pi^+\pi^0$	$D^{*-}\pi^+\pi^+\pi^0$	1.02	1.02	0	◯
	$D^{*-}\rho^+\pi^+$	0.14	0.14	0	◯
	$\overline{D}^{*0}\pi^+\pi^+\pi^-$	0.64	0.13	0.00	◯
	$\overline{D}^{*0}a_1^+$	0.56	0.69	0.69	△
	$\overline{D}^{*0}f_0\pi^+$	0.03	-	0	◯
	$\overline{D}^0\omega\pi$	0.37	0.37	0.37	
	$\overline{D}^0\eta\pi^+$	0.05	0.07	0.04	-
	$\overline{D}^0\rho^0\rho^+$	0.20	0.20	0	-
	$\overline{D}^0\rho^+\pi^+\ \pi^-$	0.20	0.30	-	
	$\overline{D}^0\rho^-\pi^+\pi^+$	0.10	0.10	-	
	$\overline{D}^0\rho^0\pi^+\pi^0$	0.10	0.20	-	
	$\overline{D}_1^0\pi^+\pi^0$	0.02	-	0	⋆
	$\overline{D}_1'^0\pi^+\pi^0$	0.05	-	0	⋆
	$\overline{D}_2^{*0}\pi^+\pi^0$	0.02	-	0	⋆
	$\overline{D}_1^0\rho^+$	0.04	0.03	0.16	⋆
	$\overline{D}_1'^0\rho^+$	0.03	0.10	0.06	⋆
	$\overline{D}_2^{*0}\rho^+$	0.01	0.07	0.03	⋆
	$\overline{D}_2^{*0}\rho^0\pi^+$	0.02	-	-	
	$\overline{D}_0^{*0}\rho^0\pi^+$	0.03	-	-	
		3.63	**3.42**	**$\sim 1.7/\sim 1.9$**	
$\overline{D}^{*0}\pi^+\pi^-\pi^+\pi^0$	$\overline{D}^{*0}\pi^+\pi^-\pi^+\pi^0$	1.80	1.80	0	◯
	$\overline{D}^{*0}\omega\pi$	0.41	0.41	0.41	
	$\overline{D}^{*0}\eta\pi^+$	0.14	0.07	0.04	-
	$\overline{D}^{*0}\rho^0\rho^+$	0.49	0.20	0	-
	$\overline{D}^{*0}\rho^0\pi^+\pi^0$	0.40	0.20	-	
	$\overline{D}^{*0}\rho^+\pi^-\pi^-$	0.40	0.20	-	
	$\overline{D}^{*0}\rho^-\pi^+\pi^+$	0.20	0.10	-	
	$\overline{D}_2^{*0}\rho^0\pi^+$	0.01	-	-	
	$\overline{D}_1'^0\rho^0\pi^+$	0.03	-	-	
		3.89	**2.99**	**$\sim 1.9/\sim 1.2$**	

TABLE 4.8: Update of branching fractions. The contributions to each of the $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$ FEI modes are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column and the proposed common branching fractions in the fifth column. Contributions from PYTHIA are shown in blue. The sum is shown in bold. The marker in the sixth column indicates the reason for the change. Contributions smaller than 10^{-4} are omitted.

FEI decay mode: $B^+ \to D^-\pi^+\pi^+(\pi^0)$

The measurements leading to the final state $B^+ \to D^-\pi^+\pi^+$ have already been described in the introduction of the model for the $B^+ \to \overline{D}^{**0}\pi^+$ modes. In the Belle (II) simulation, the entire inclusive measurement is incorrectly interpreted as a non-resonant component, which is now corrected. However, a non-resonant component is still included to fill the gap in the inclusive measurement until a clearer picture of the higher resonances of D mesons is obtained.

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$	$\mathcal{B}^{\text{Proposed}}(\%)$	Reason
$D^-\pi^+\pi^+$	$D^-\pi^+\pi^+$	0.17	0.11	0.03	$\cup$
	$\overline{D}_0^{*0}\pi^+$	0.06	0.04	0.06	$\star$
	$\overline{D}_2^{*0}\pi^+$	0.03	0.02	0.03	$\star$
		0.26	**0.17**	**0.12**	
$D^-\pi^+\pi^+\pi^0$	$D^-\pi^+\pi^+\pi^0$	0.20	0.20	0	◯
	$D^-\rho^+\pi^+$	0.20	0.20	0	◯
	$\overline{D}_1^0\pi^+$	0.02	0.02	0.04	$\star$
	$\overline{D}_1'^0\pi^+$	0.01	0.01	0.01	$\star$
	$\overline{D}_2^{*0}\pi^+$	0.01	0.00	0.01	$\star$
	$\overline{D}_0^{*0}\rho^+$	0.06	-	0.11	$\star$
	$\overline{D}_2^{*0}\rho^+$	0.04	0.15	0.07	$\star$
	$\overline{D}_0^{*0}\pi^+\pi^0$	0.07	-	0	$\star$
	$\overline{D}_2^{*0}\pi^+\pi^0$	0.05	-	0	$\star$
		0.66	**0.58**	**0.24**	

TABLE 4.9: Update of branching fractions. The contributions to each of the $B^+ \to D^-\pi^+\pi^+(\pi^0)$ FEI modes are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column and the proposed common branching fractions in the fifth column. Contributions from PYTHIA are shown in blue. The sum is shown in bold. The marker in the sixth column indicates the reason for the change.

As for the final state $D^-\pi^+\pi^+\pi^0$, the only known input is the $\overline{D}^{**}\pi^+$ mentioned above. The D^{*-} equivalents of $D^-\pi^+\pi^+\pi^0$ and $D^-\rho^+\pi^+$ are already set to 0, as mentioned in the description of $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$, and it is assumed that the entire contribution comes from the $\overline{D}^{**}$. So far, there are only upper limits on the $\overline{D}^{**}\rho$ which makes it difficult to validate. However, if we consider the invariant masses of the $D^-\pi^+$ and $D^-\pi^+\pi^0$ combinations by selecting the combination with the lowest mass for each event (we expect the pion from the $D^{**} \to D^*\pi$ transition to have low momentum) (Fig. 4.25), the contribution from $\overline{D}_2^*$ at the corresponding mass of 2.46 GeV/c^2 becomes visible, albeit statistically limited.

Here, selecting the events from $\overline{D}_2^*$ with the selection $2.38 < M(D^-\pi^+)_{\text{min}} < 2.52$ GeV/c^2 and considering the corresponding $M(\pi^+\pi^0)$ in Fig. 4.26, it is obvious that the contribution is dominated by the ρ intermediate resonance, which justifies setting all non-resonant $\overline{D}^{**}\pi^+\pi^0$ contributions to zero. The higher statistics of the Belle data set helps validate such modes, which would not be possible with the Belle II data set alone.

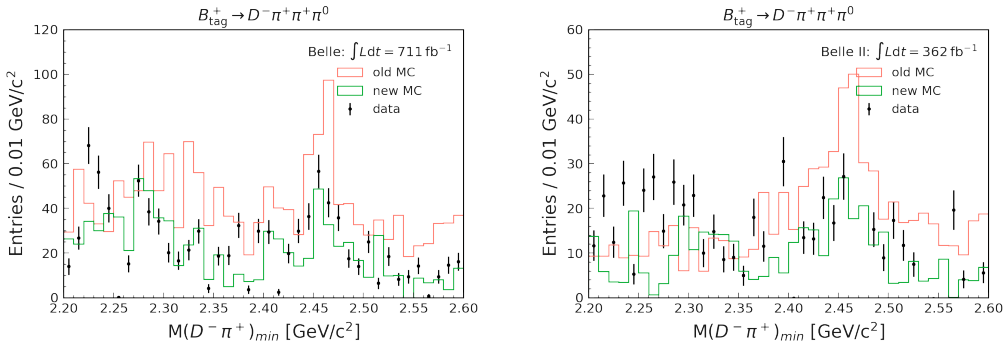


FIGURE 4.25: Comparison of $M(D^-\pi^+)_{\text{min}}$ between the data, the old MC and the new MC in $B^+ \to D^-\pi^+\pi^+\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

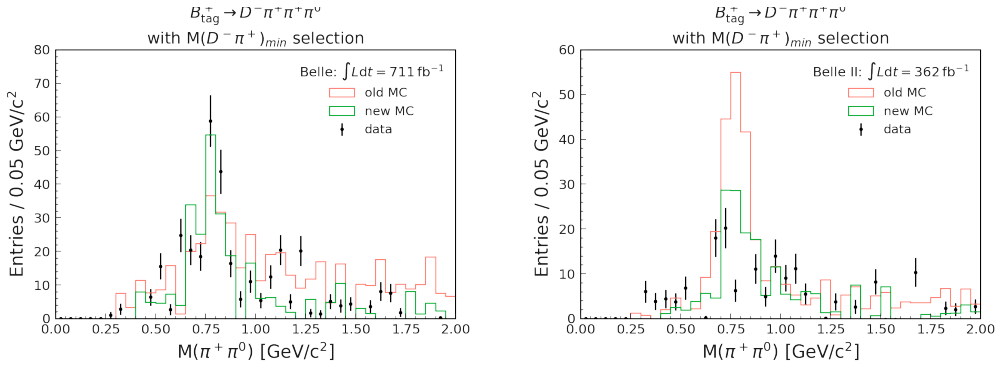


FIGURE 4.26: Comparison of $M(D^-\pi^+)_{\text{min}}$ between the data, the old MC and the new MC in $B^+ \to D^-\pi^+\pi^+\pi^0$ mode of the FEI package in Belle (left) and Belle II (right).

FEI decay mode: $B^+ \to \overline{\Lambda}_c^- p\pi^+\pi^+\pi^-$

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{Belle}}(\%)$	$\mathcal{B}^{\text{Belle II}}(\%)$
$B^+ \to \overline{\Lambda}_c^- p\pi^+\pi^+\pi^-$	$B^+ \to \overline{\Lambda}_c^- p\pi^+\pi^+\pi^-$	-	0.22
	$\overline{\Sigma}_c(2520)^0 p\rho(770)^0$	0.05	0.04
	$\overline{\Sigma}_c(2455)^{--} p\pi^+\pi^+$	0.01	-
		0.06	**0.26**

TABLE 4.10: Investigation of branching fractions. The contributions to the $B^+ \to \overline{\Lambda}_c^- p\pi^+\pi^+\pi^-$ FEI modes are shown in the second column, with the corresponding *official* Belle (II) branching fractions in the third (fourth) column. Contributions from PYTHIA are shown in blue. The sum is shown in bold. Contributions smaller than 10^{-4} are omitted.

The most recent measurement for this final state is from CLEO using a 9 fb^{-1} data set and yields an inclusive branching fraction of $(0.22 \pm 0.07)\%$ [110]. Due to the automatized update in Belle II by parsing PDG, this inclusive measurement is set as a non-resonant component in Belle II decay file, which is incorrect. This CLEO measurement is a factor of two larger than the previous measurement [111], which calls into question its reliability. The measurement also showed that this decay occurs over several intermediate states, such as $\overline{\Sigma}_c^0$, $\overline{\Sigma}_c^{--}$ and $\overline{\Lambda}_{c1}^-$. Even in the non-resonant component, the 3π system could originate from a_1^+, as in the case of $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-$ final states. All these possibilities are difficult to model without a better measurement with larger statistics. At the moment, PYTHIA generates some

contributions to this final state, as shown in Tab. 4.10, but this is unpredictable and uncontrollable.

Belle's [112] and BaBar's [113] studies of $B^+ \to \overline{\Lambda}_c^- p\pi^+$ revealed a "resonance"-like (threshold) structure in the lower mass region of $M(\overline{\Lambda}_c^- p)$, as shown in Fig. 4.27. This is not modelled in MC and is strongly suppressed in data, as shown by the absence of a similar peak in $M(\overline{\Lambda}_c^- p)$ in Fig. 4.28. It is not clear whether this threshold effect is also present in the $B^+ \to \overline{\Lambda}_c^- p\pi^+\pi^+\pi^-$ final state, highlighting the need for better measurement and better modelling of baryonic B decays in MC. Currently, no changes can be proposed to MC, which leaves a large discrepancy between the data and MC for this mode, especially in the Belle II data set.

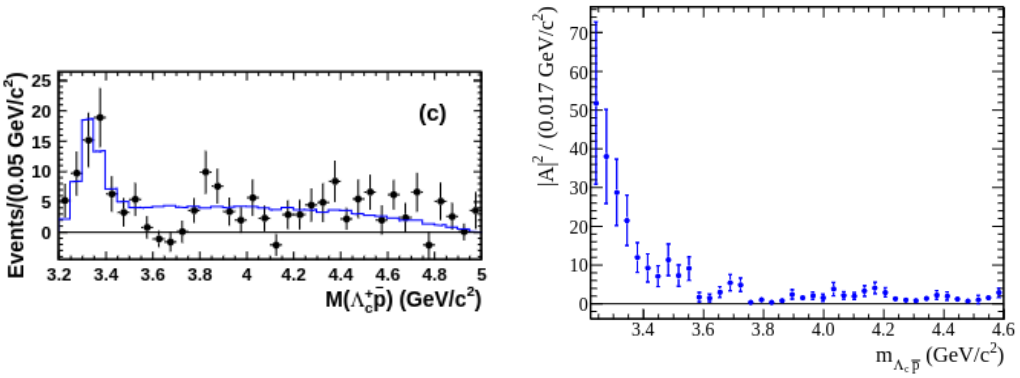


FIGURE 4.27: The distribution of $M(\overline{\Lambda}_c^- p)$ in the measurements of $B^+ \to \overline{\Lambda}_c^- p\pi^+$ in Belle (left) and BaBar (right). Taken from [112] and [113] respectively.

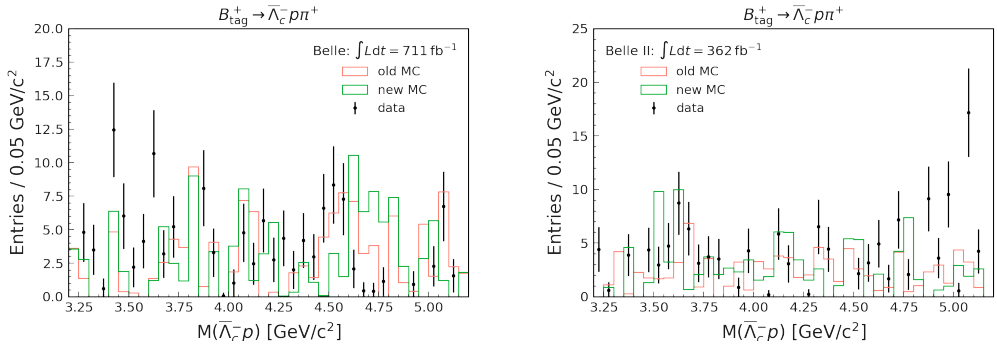


FIGURE 4.28: Comparison of $M(\overline{\Lambda}_c^- p)$ between the data, the old MC and the new MC in $B^+ \to \overline{\Lambda}_c^- p\pi^+$ mode of the FEI package in Belle (left) and Belle II (right).

Summary of the MC modifications

Overall, an attempt is made to compare the eleven B decay modes that dominate the efficiency of the FEI package with the most recent measurements. The corresponding branching fractions in MC are corrected, without fine-tuning for ten of these modes. The baryonic decay $B^+ \to \overline{\Lambda}_c^- p\pi^+\pi^+\pi^-$ could not be modified with current knowledge. In most cases, the branching fractions covered by the individual B modes are reduced by eliminating double counts and constraining the overestimated PYTHIA components, as shown in Fig. 4.29. This has the effect of reducing the sum of the

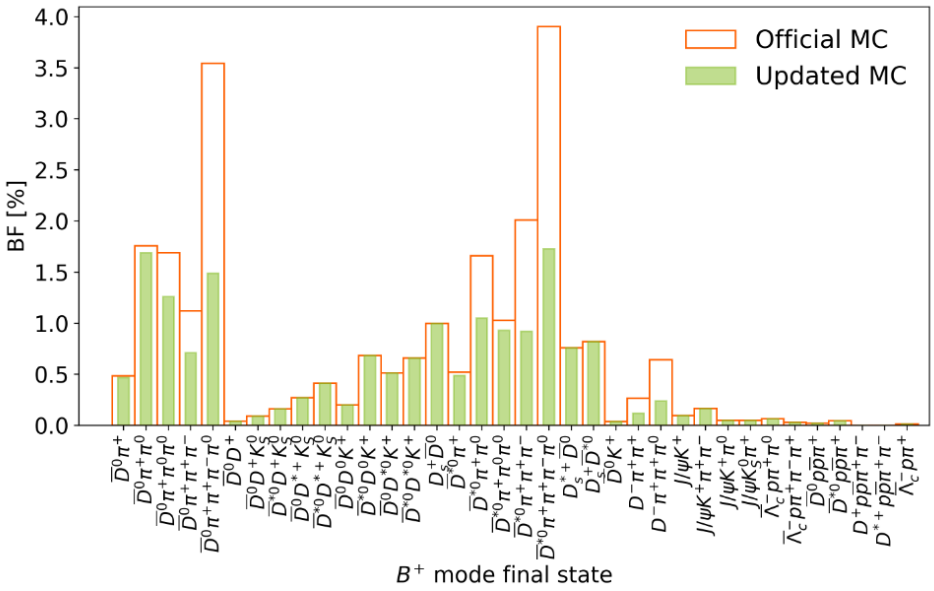


FIGURE 4.29: The branching fraction coverage per B mode in the FEI package, compared between the official and the new (updated) MC of Belle II.

branching fractions of the decays explicitly listed in the DECAY.DEC file. To account for this, an additional 4% of the branching fraction is given back to PYTHIA. This shows that almost half of the branching fraction of the B decays is completely unknown, and we rely on PYTHIA to overcome this issue. It is important that we close this gap by using the large data sets of modern B-factories, rather than relying on measurements from CLEO and ARGUS experiments.

Correcting MC to the best interpretations of the latest measurements is expected to improve the agreement between the data and MC. This is reflected in the improvement in the calibration factors of the hadronic B-tagging by the FEI package, as shown in Fig. 4.30 for both Belle and Belle II data sets. The calibration factors for most modes approach 1, although not exactly, since no fine-tuning was done. Overall, the calibration factor for Belle improved from 0.75 to 1.04. For Belle II, the calibration factor improves from 0.65 to 0.83, but most of the remaining discrepancy comes from the "rest of the modes" category, excluded from our investigation. If one considers only the top 10 modes where MC is significantly improved, the calibration factor increases from 0.68 to 0.92. At this stage, the calibration factors are limited by the uncertainties in the corresponding measurements, as described in Sec. 4.5.1, and can only be further improved by better measurements in the future. The best example is the $B^+ \to \overline{D}^0\pi^+\pi^0$ mode, where the measurement has an uncertainty of 13% due to the low statistics in the CLEO sample. The two modes $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$, which account for 20% (16%) of the FEI efficiency in the Belle (II) data, are still mostly generated by PYTHIA, which needs to be better controlled for reliable performance in the future.

Rather than just scaling the MC branching fractions to match the data, the other advantage of improving the MC model is the better agreement of the $\mathcal{P}_{\text{FEI}}$ distribution between the data and MC, as shown for the mode $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ in Fig 4.31. Here, the signal components of the $D\pi$ control sample are selected using the sPlot technique and the $\mathcal{P}_{\text{FEI}}$ is compared between the old MC and the data in the left plots, while the new MC and the data are compared in the right plots. Such an improvement has great importance because $\mathcal{P}_{\text{FEI}}$ is used to select B_{tag} candidates in the analysis.

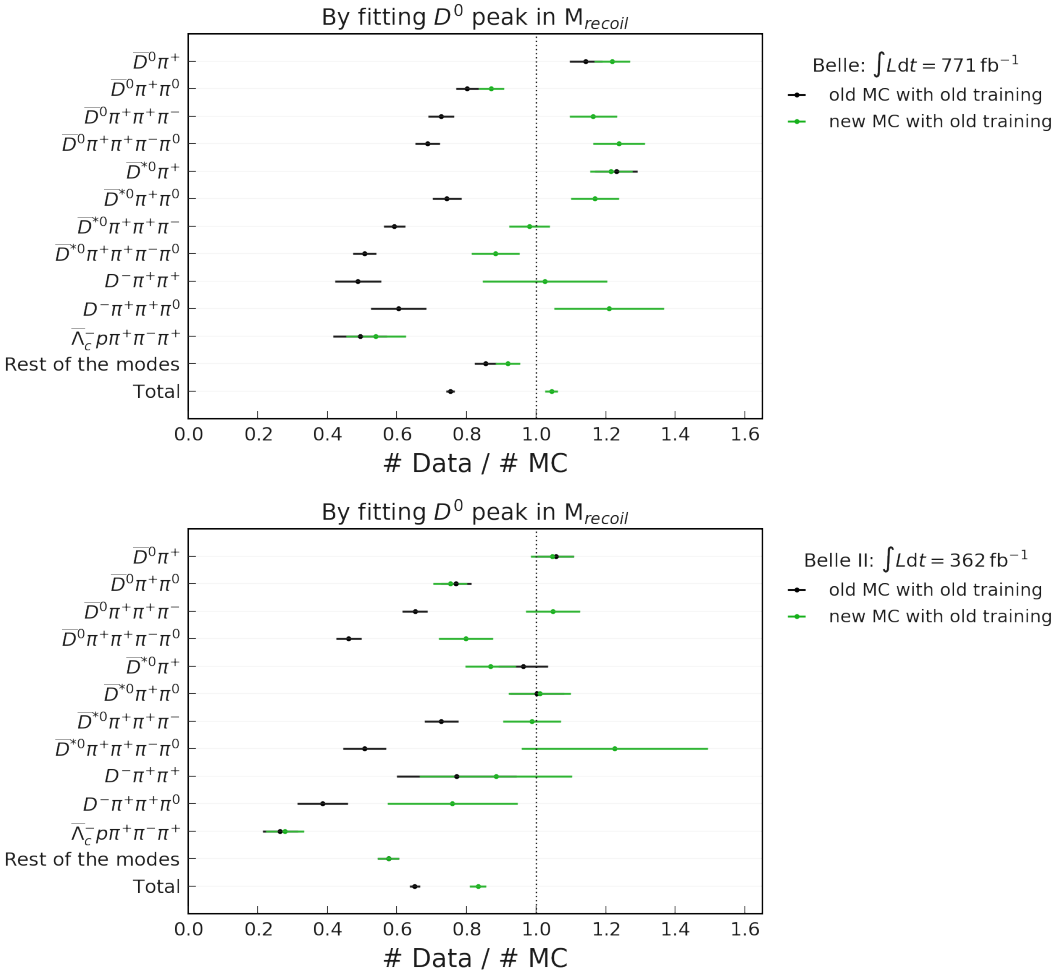


FIGURE 4.30: Calibration factors by fitting the D^0 peak in the M_{recoil} in Belle sample (top) and Belle II sample (bottom) with the new MC produced using $\mathcal{B}^{\mathrm{Proposed}}(\%)$. Only statistical uncertainties are shown here.

Therefore, a good agreement will ensure that the optimizations performed on MC are also effective on the data. This also means that when the FEI package is retrained with the new MC, the learned selections improve and consequently the background suppression capabilities increase, as described in Sec. 4.3

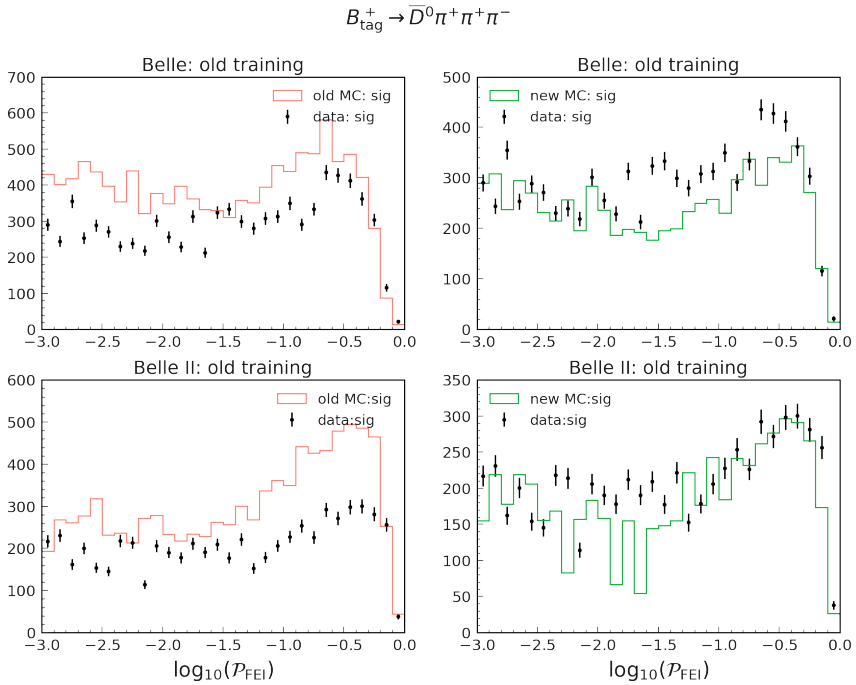


FIGURE 4.31: Comparison of $\mathcal{P}_{\text{FEI}}$ distribution between the old (new) MC and the data on left (right). The plots for Belle (II) data set are on top (bottom).

4.3 Training the FEI package with new MC for Belle II

The new MC model, obtained through many iterations, can be further improved in the future with more and better measurements. However, it is the best possible model in the short time span. It has now been approved by the Belle II collaboration and will be used for the next MC production and consequent FEI training. Using this new MC, an unofficial FEI training with a 200 fb^{-1} sample was performed for our study, called "v1 training". Using the newly trained FEI package, the M_{recoil} of the $D\pi$ sample is reconstructed in the Belle II data, and a comparison with the current standard training is shown in Fig. 4.32. The signal efficiency appears to be almost unchanged, while the background level in the whole M_{recoil} region has decreased significantly.

Using the $D\pi$ sample, the efficiency and purity in the signal region for the new training can be calculated using Eq. 3.11. The comparison of the ROC curves is shown in Fig. 4.33. Here, training with a MC sample closer to the data shows an increase in purity with about the same efficiency in the data.

The increase in purity can be compared separately for each of the B decay modes, as shown in Fig. 4.34. As expected, the largest increase in purity is seen for the modes where the MC model is changed, like $\overline{D}^0\pi^+\pi^+\pi^-$, $\overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$ and $D^-\pi^+\pi^+\pi^0$. In contrast, for the modes where the MC model is identical (such as $B^+ \to \overline{D}^{(*)0}\pi^+(\pi^0)$), there is no change in purity, which can be seen as a validation to our procedure. The change in the "Rest of the modes" component is due to the changed MC model of

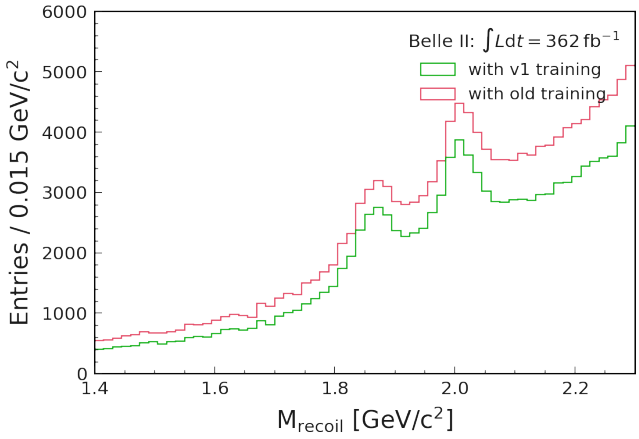


FIGURE 4.32: The M_{recoil} distribution of the $D\pi$ control sample with the v1 training in comparison with the current standard training in the Belle II data.

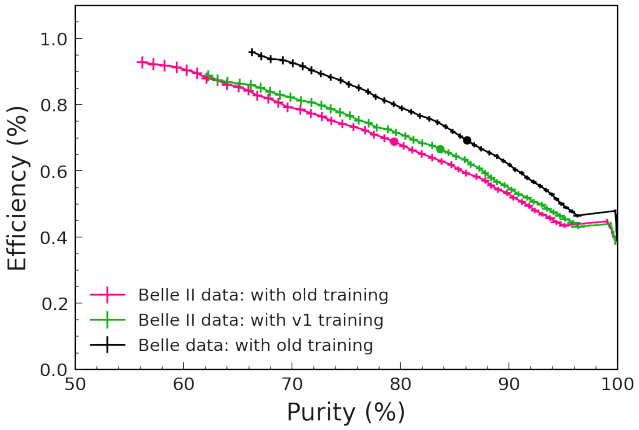


FIGURE 4.33: Efficiency as a function of purity in data, computed at various $\mathcal{P}_{\mathrm{FEI}}$ cuts. A comparison is shown between using standard FEI training and v1 FEI training with new MC in Belle II samples. The standard FEI training of Belle sample is also shown as reference.

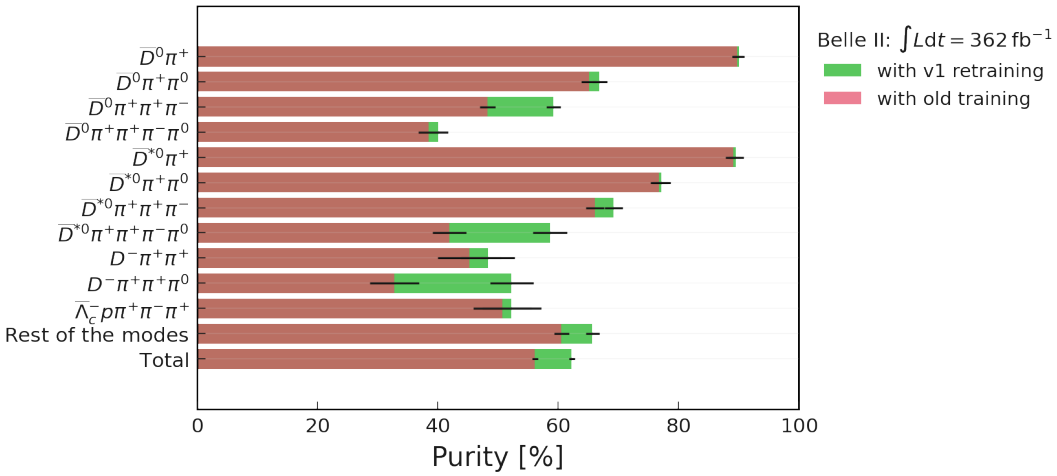


FIGURE 4.34: Purity in the $D\pi$ sample of the Belle II data with the v1 FEI training performed on the new MC.

modes $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0\pi^0$, as a side effect of the change in $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-$ explained earlier.

For example, considering the mode $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$, the sPlot technique can unfold the signal and background components in the $D\pi$ sample. Comparing the signal and background components with the new training in Fig. 4.35, it is clear that not only the background level has decreased significantly, but also the agreement of background between MC and the data in the v1 training is improved compared to the standard training. This is due to the fact that the FEI package now learns

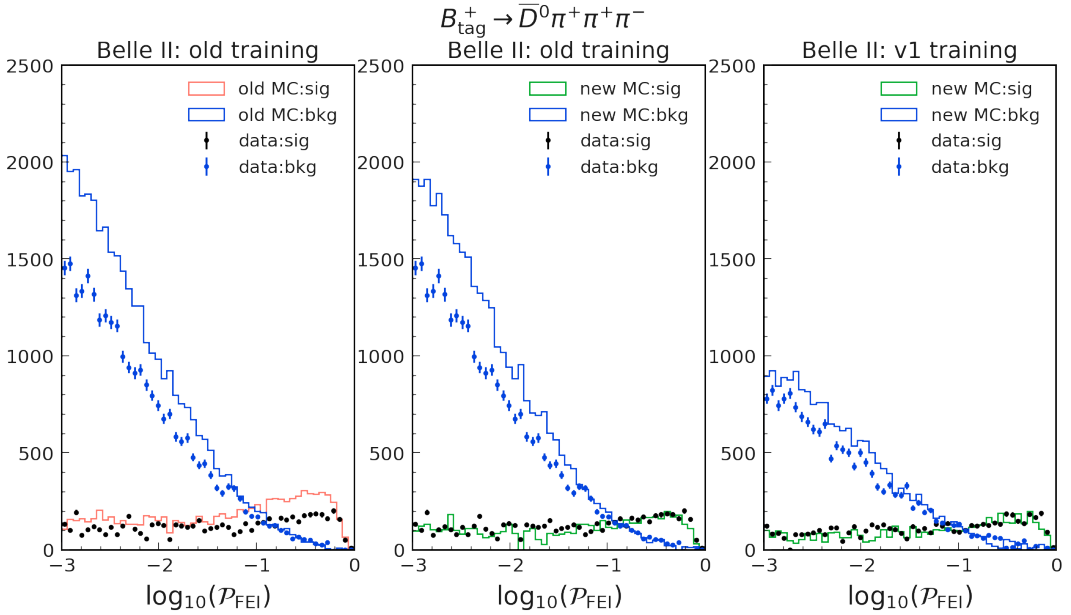


FIGURE 4.35: Comparison of the $\log_{10}\mathcal{P}_{\text{FEI}}$ distribution between the old MC and the data (left); the new MC and the data (middle); the new MC and the data, now using v1 training (right).

the kinematics from the new MC, such as the a_1^+ shape, as seen in the background enhanced sPlot of the $M(\pi^+\pi^+\pi^-)$ in Fig. 4.36. In summary, simply changing the MC

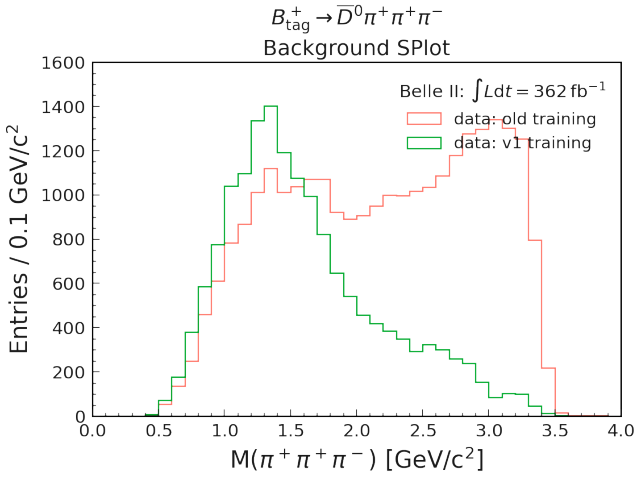


FIGURE 4.36: Comparison of $M(\pi^+\pi^+\pi^-)$ of the background component in the $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ mode between the standard and v1 trainings in the Belle II data.

model improves the calibration factors and leads to better agreement of $\mathcal{P}_{\text{FEI}}$ for the

signal component. In addition, training the FEI package with the new MC provides better background suppression while improving the agreement between the data and MC for $\mathcal{P}_{\text{FEI}}$ of the background component.

4.4 D^{*0} FEI reconstruction in Belle II

Despite the improvement in purity by training the FEI package for Belle II using the new MC, the performance in the Belle II data set is lower than that of Belle, as already represented in Fig. 4.33. At the same purity, the efficiency of the FEI package seems to be lower in Belle II, which is surprising. To understand the reason for this difference, the yield in the Belle and Belle II data sets can be normalized to 100 fb^{-1} and compared for each of the B decay modes. This is shown in Fig. 4.37, which

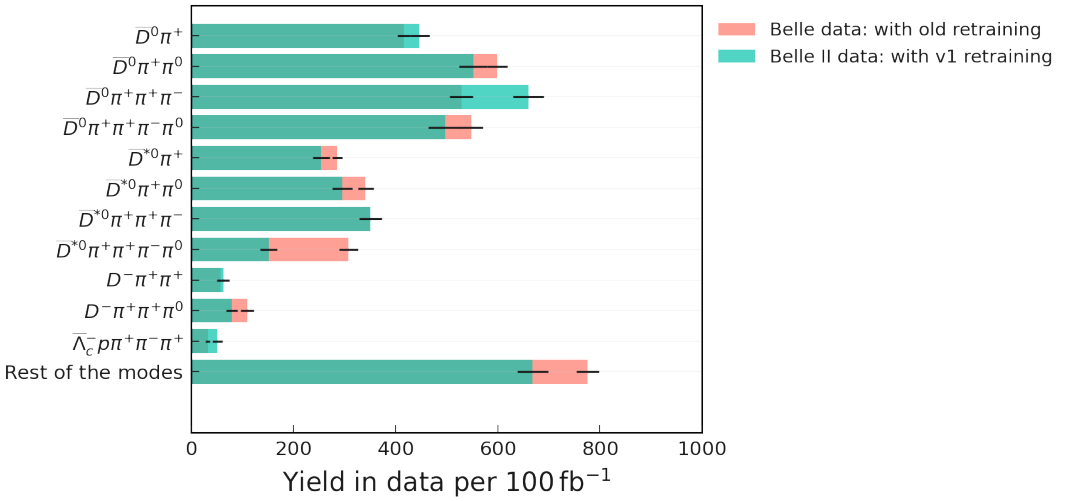


FIGURE 4.37: Comparison of yield in Belle and Belle II data sets for the top B tag-side decays modes, normalized to 100 fb^{-1}.

suggests that the performance of the π^0 reconstruction (in particular, as part of the $\overline{D}^{*0}$) is worse in Belle II than in Belle.

The $\overline{D}^{*0}$ mesons can decay to either $\overline{D}^0\pi^0$ or $\overline{D}^0\gamma$, and FEI reconstructs both decays. Looking at the $\Delta M = M(\overline{D}^{*0}) - M(\overline{D}^0)$ comparison of these two modes in Fig. 4.38, it is clear that the problem in the Belle II reconstruction stems primarily from the $\overline{D}^0\pi^0$ mode. This deficit is largely due to the more stringent cuts applied in the γ preselection in FEI in Belle II, as listed in Tab. 3.2. The requirement of higher energy for γ means that the reconstruction of slow π^0, which decay into 2 γ of low energy could be affected. This can be seen in Fig. 4.39: the π^0 momentum in Belle II clearly has a much higher threshold than in Belle. Part of these lost events where $\overline{D}^{*0} \to \overline{D}^0\pi^0$ can be partly recovered by reconstructing only of the γ with sufficiently high energy from the π^0, and FEI reconstructs them as $\overline{D}^{*0} \to \overline{D}^0\gamma$, which can be seen as an excess in the tail in Belle II in the right plot of Fig. 4.38.

To fix this problem, and increase the efficiency of the FEI package in Belle II samples, the pre-selection cuts for γ are relaxed to match those of Belle. Furthermore, the FEI package is trained with this new selection for Belle II, called "v2 training". Although the changes lead to increased efficiency, it should be noted that the ΔM distribution has not improved. The reason for this is the lack of a mass constraint on π^0 particles, which makes the ΔM variable (or Q variable, which is used as a FEI training variable for $\overline{D}^{*0}$) less powerful. This highlights the fact that including

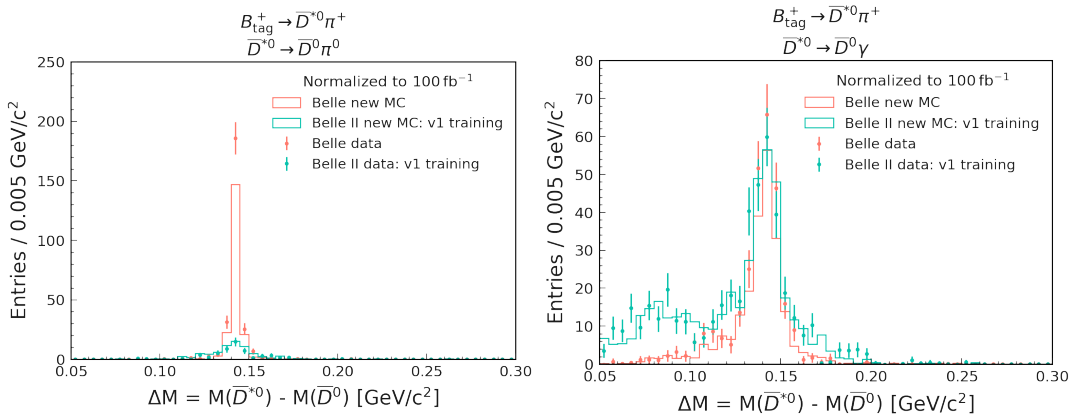


FIGURE 4.38: Comparison of $\Delta M = M(\overline{D}^{*0}) - M(\overline{D}^0)$ between the Belle and Belle II samples, normalized to 100 fb^{-1}, reconstructed by FEI mode $B^+ \to \overline{D}^{*0}\pi^+$ where $\overline{D}^{*0} \to \overline{D}^0\pi^0$ on left and $\overline{D}^{*0} \to \overline{D}^0\gamma$ on right.

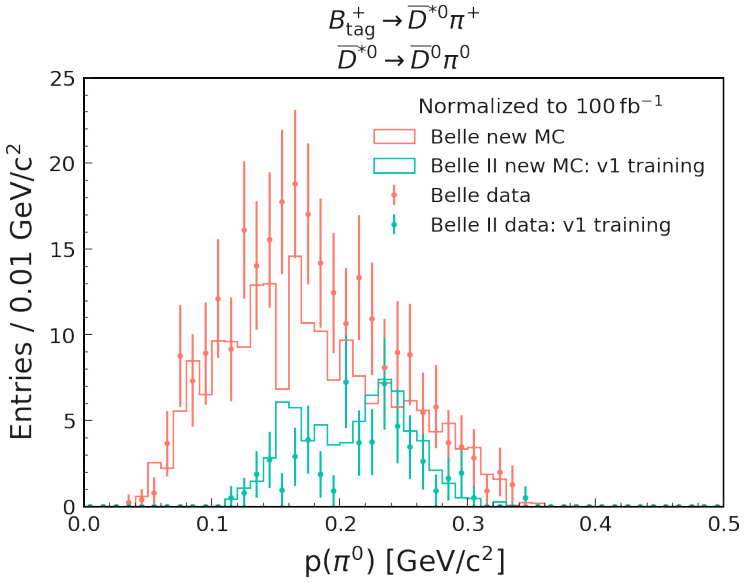


FIGURE 4.39: Comparison of the momentum of the π^0 coming from the $B^+ \to \overline{D}^{*0}[\to \overline{D}^0\pi^0]\pi^+$ mode between the Belle and Belle II samples, normalized to 100 fb^{-1}, showing the inefficiency to reconstruct slow π^0 in Belle II.

ΔM as a training variable in FEI is an inefficient way to use such a powerful variable, rather instead of applying a cut on ΔM. In Belle mdst, the mass constraint is already applied to the π^0 particles, and is therefore read directly into basf2 via the b2bii package. On the Belle II side, only a vertex fit is performed in the FEI package for π^0's. To fix this, the mass constraint is implemented for the π^0 particles in the FEI package and another training is performed, called "v3 training". Now, the ΔM distribution shown in Fig. 4.40 looks as expected. The efficiency and purity, calculated

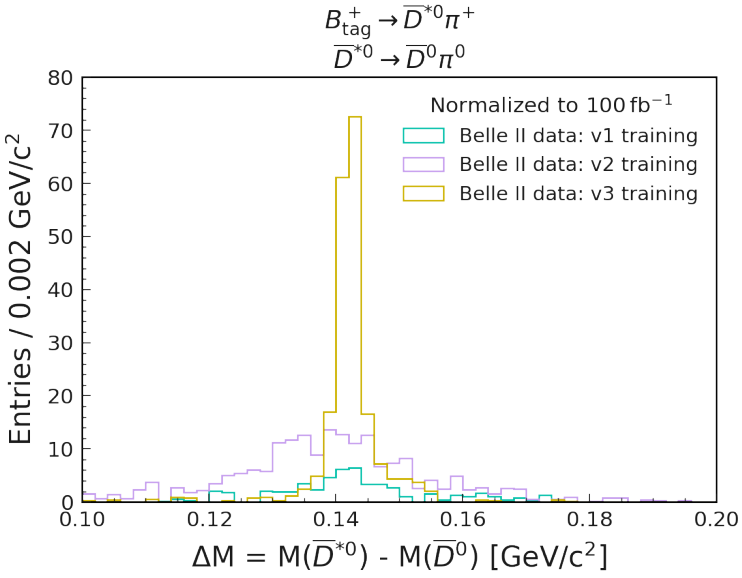


FIGURE 4.40: Comparison of $\Delta M = M(\overline{D}^{*0}) - M(\overline{D}^0)$ between normalized Belle II samples reconstructed by different training of the FEI package in the mode $B^+ \to \overline{D}^{*0}\pi^+$ where $\overline{D}^{*0} \to \overline{D}^0\pi^0$.

again with the v3 training, are compared with the previous trainings as well as with Belle in Fig. 4.41. The performance in Belle II is better than in the Belle sample. The

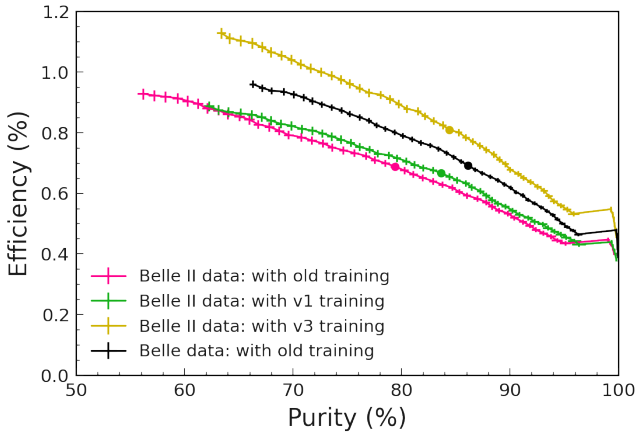


FIGURE 4.41: Efficiency as a function of purity in data, calculated with different $\mathcal{P}_{\text{FEI}}$ cuts. A comparison is shown between the use of the standard FEI training, the v1 FEI training and the v3 FEI training in Belle II samples. The standard FEI training of Belle sample is also shown for reference.

calibration factors with the v3 training agree with those of the v1 training, within the statistical uncertainty, as shown in Fig. 4.42.

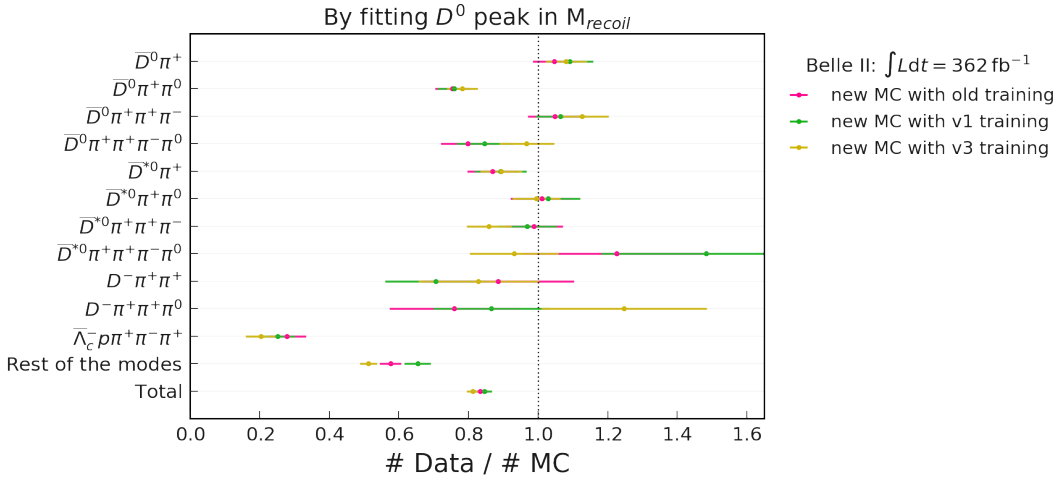


FIGURE 4.42: Calibration factors of different B decay modes reconstructed by the FEI package in new MC with different trainings in the Belle II data set.

Although hadronic B-tagging can be further improved in several ways, as it will be described in Sec. 4.5, this is the best that could be done within the short time frame. Therefore, for the study of $B^+ \to K^+\tau^+\tau^-$, the new MC will be used. This leads to an improvement of the calibration factors for Belle and Belle II data sets from 0.75 to 1.04 and 0.65 to 0.81, respectively. And for Belle II, the v3 training is used. Using the new MC for training improves purity (56.2% to 63.4%) and looser γ preselection improves efficiency (0.93% to 1.13%). Thus, the expected performance of the FEI package would be as shown in Tab. 4.11.

	$\mathcal{P}_{\mathrm{FEI}} > 10^{-3}$		$\mathcal{P}_{\mathrm{FEI}} > 10^{-2}$	
	Belle	Belle II	Belle	Belle II
Efficiency in data	0.96%	1.13%	0.69%	0.81%
Purity in data[3]	66.3%	63.4%	86.1%	84.4%
Calibration Factor	1.04	0.81	1.00	0.75

TABLE 4.11: Performance metrics of hadronic B-tagging by the FEI package, with different $\mathcal{P}_{\mathrm{FEI}}$ selection, used to study $B^+ \to K^+\tau^+\tau^-$. The standard training is used for the Belle data set, while the v3 training is used for the Belle II data set. The new MC is used in both cases.

4.5 Possible future improvements

The efforts described in this chapter have improved all the performance metrics for the hadronic B-tagging in Belle (II) experiments, but further improvements are possible. The B decay model can be improved in the future with better measurements, which could also lead to the inclusion of new modes in the FEI package. For some of the modes involving D^* mesons, including double charm B decays, a more efficient reconstruction technique described in Sec. 4.5.2 can be used.

4.5.1 PDG uncertainties and future measurements

As described in Section. 4.2, most measurements for the decays required for the hadronic B-tagging were obtained by ARGUS and CLEO experiments with M_{bc} fit on small statistics, and are therefore likely to be over-estimated. Consequently, after updating the B decay model, the deviations of the calibration factors from one are dominated by the uncertainties in the measurements (PDG), as shown in Fig. 4.43. These uncertainties (in red) consider only the cases where a measurement is available,

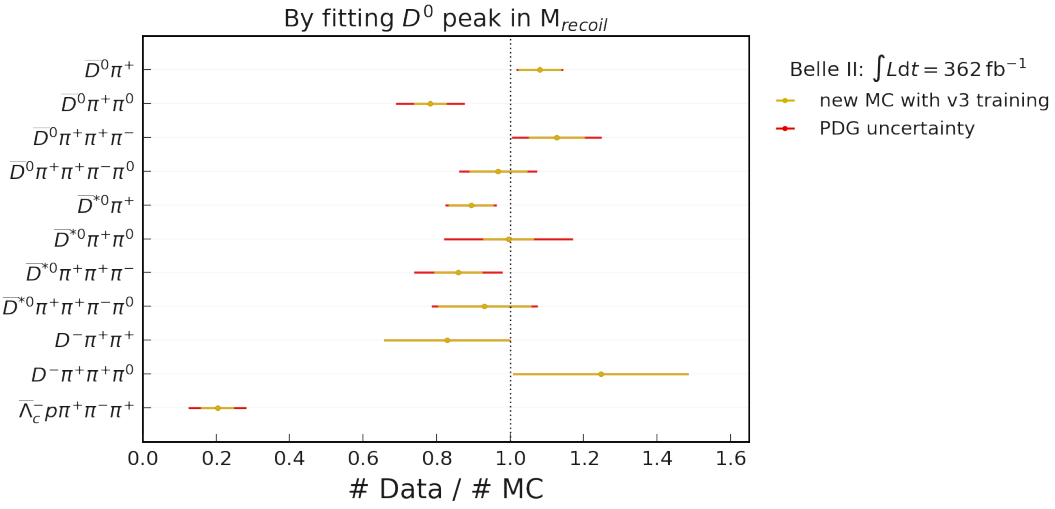


FIGURE 4.43: Calibration factors of different B decay modes reconstructed by the FEI package with the latest MC and training in the Belle II data set, showing additional 1σ uncertainty from measurements (PDG) in red.

which is not always the case. As shown earlier, the decay model becomes more complex with increasing multiplicity, so that the contribution of PYTHIA increases, as for $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$. Resorting PYTHIA biases the BDT training, leading to poor performance. Therefore, one could either adopt a cut-based approach here and choose, for example, $B^+ \to \overline{D}^{(*)0}\omega\pi^+$ (which covers about 25% of the final state). Or the model in MC needs to be improved with future measurements: identify intermediate states and provide a simple model that can be implemented in EvtGen instead of complex interferences. A similar model is needed for baryonic decays and even for modes such as $B^+ \to D^{*-}(4\pi)^{++}$ and, $B^+ \to \overline{D}^{*0}(5\pi)^+$, for which large inclusive branching fractions have been measured [107] but whose decay kinematics are poorly known to be included in FEI. This must be investigated independently of the FEI package and by exclusive reconstruction, since the FEI selection is already biased by the MC model. Once the B modelling issues are improved, the effect of D decay modelling becomes significant, as shown in Fig. 4.3, which should also be

updated. Since the inclusive B-tagging is also trained on the same MC, improvements to MC modelling also affect its performance.

Apart from the 11 modes considered in this chapter, the remaining of the B decay modes used in the FEI package do not contribute much to efficiency, but could increase the background on the tag side (with the exception of pure modes such as $B^+ \to J/\psi K^+$). These modes should either be removed completely, or reconstructed using a different approach, as described in the next section.

4.5.2 Partial reconstruction from shifted ΔE

Most of the hadronic B-tagging efficiency comes from the $B \to D\ n\pi\ m\pi^0$ decays (i.e., the $W \to ud$ transitions). The $W \to cs$ contribution to the inclusive branching fraction of the B decays is about half that of the $W \to ud$ transitions [114], but currently does not contribute much to the B-tagging. The FEI package contains modes such as $B \to \overline{D}^{(*)0} D^{(*)} K$ and three of the $B \to \overline{D}^{(*)0} D_s^{(*)+}$ combinations. However, in the case of exclusive reconstruction, the efficiency is very low: because of the low coverage of the D branching fraction and the low reconstruction efficiency at higher multiplicity. For example, in the FEI package, D^0 (D_s^+) has a branching fraction coverage of about 40% (20%), but the final efficiency is only half as high. It is even worse for the D^* mesons, which have a coverage of 100%, but the effective efficiency is less than $\sim 10\%$ due to the low momentum pion or photon emitted. The low reconstruction efficiency of D^* mesons is the reason why the $B^+ \to \overline{D}^{*0} D_s^{*+}$ decay in the FEI package is not used, although it has the largest branching fraction among the four combinations.

One way to recover these events is to partially reconstruct the B, without the low-momentum pion or photon, and select these events in the shifted region of the ΔE (around -200 MeV). As shown in Fig. 3.2, the ΔE of a partially reconstructed B candidate would be shifted to the left. However, this is currently not possible because ΔE is used as a training feature, which skews the distribution so that the peak is zero. Once ΔE is removed as a training feature, only the $B \to \overline{D}^0 D_s^+$ can be reconstructed, and the other three combinations can be selected from the corresponding regions in ΔE. Similarly, ΔM (or Q) should also be removed as a training feature and replaced by a tighter cut in the pre-selection to avoid the behaviour observed in Sec. 4.4. The decays $B^+ \to \overline{D}^{(*)0} K K^{(*)}$, which also have large branching fractions [115], can be included in the partial reconstruction to cover more $W \to cs$ transitions.

Chapter 5

$B^+ \to K^+\tau\tau$ search with Belle and Belle II data sets

In this chapter, we describe the strategy for performing a sensitivity study of the $B^+ \to K^+\tau^+\tau^-$ search using hadronic B-tagging with the data set of the Belle and Belle II experiments. The final states $\tau \to \ell\nu\overline{\nu}$ are chosen similarly to the earlier search for this mode in the *BABAR* experiment described in Section 1.3.2.

To take advantage of all the improvements described in Chapter 4, the latest FEI training is used to reconstruct the B_{tag} candidates with the new MC. This is followed by the selection of signal events and the best candidate. The primary sources of background are characterized and the BDT-based background suppression procedure is described in Section 5.3.

After removing all tracks and clusters from the signal and tag-side B candidates of the event, the remaining energy deposited in the calorimeter, E_{ECL}, is used as an observable for signal extraction. For signal events, no additional energy is expected to be deposited in the electromagnetic calorimeter, so the E_{ECL} distribution of the signal should peak at zero. However, the resolution of E_{ECL} is broadened by contamination of the beam background photons and neutral clusters from hadronic split-off processes. To minimize this, certain requirements, called a mask, can be applied to the clusters used to calculate E_{ECL}. The optimisation performed to choose the mask that reduces contamination while preserving separation between signal and background is described in Section 5.4. As described in Section 5.5.2, control samples are used to examine the data-MC agreement of the E_{ECL} distribution with different masks.

For the selected mask, the signal efficiency and the expected background yield are estimated in the defined signal window. Considering for the data-MC discrepancies in FEI efficiency, an upper limit is derived from MC in case of no observed signal. We estimate that the current upper limit on $B(B^+ \to K^+\tau^+\tau^-)$ can be improved by an order of magnitude.

5.1 Simulated samples

Signal MC

To study selection and signal efficiency, samples of the signal MC are generated using the generator EvtGen [76]. Each event in this sample contains a signal candidate ($B^+ \to K^+\tau^+\tau^-$) and a generically decaying B meson. For the generically decaying B, the new decay model described in Sec. 4.1 is used to have results closer to data (overall calibration factor closer to 1) and to take advantage of the latest and more optimal FEI training in the case of Belle II. The BTOSLLBALL generator model is used to generate the kinematics of the signal candidate based on the form-factor

calculations according to the Standard Model [116]. The τ's on the signal side are generated as decaying generically.

For Belle signal MC, the mcproduzh package is used. It performs generation (via EvtGen), the simulation of the detector response via Geant3 [117], and the reconstruction of events, and provides output in Belle mdst format. The beam energy, IP profile, and detector configurations here are experiment-dependent, but not run-dependent[1]. The beam background obtained from the randomly triggered data is overlaid on the simulated MC events. Here, the experiment-dependent background files are created from the average of all the runs of the experiment. With this configuration, 10 million signal MC events are generated for Belle.

For Belle II, the signal MC is generated in a run-independent (RI) way. For RI MC, simulated beam backgrounds and static detector conditions are used. The beam backgrounds are simulated based on the nominal expectations in the early phase of the experiment, when most of the data from Belle II are collected. However, it is found that this simulation is overestimated in the calorimeter compared to the beam background in the data. The simulation of the detector response for Belle II MC is performed by Geant4 [118]. In this configuration, 10 millions signal MC events are generated for Belle II. Again, with the new model for the generic B^+ on one side and $B^+ \to K^+\tau^+\tau^-$ with BTOSLLBALL model on the other side, with generically decaying τ's.

The quark-level diagram with the penguin $\bar{b} \to \bar{s}\tau^+\tau^-$ transition in the and the potential NP contributions are shown in Fig. 5.1. The kaon always has the same charge as the parent B and the τ decays to all the possible final states, either *leptonic* ($f \equiv \ell\nu$) or *hadronic* ($f \equiv h$).

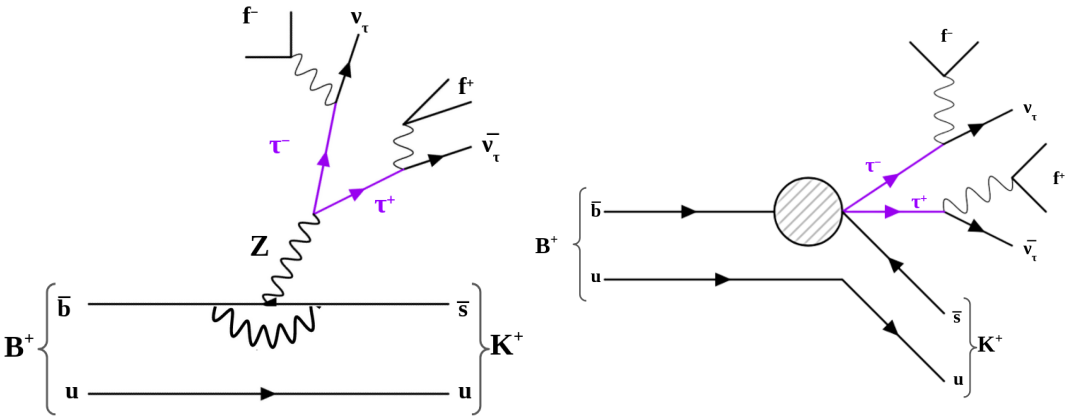


FIGURE 5.1: Schematic diagram for $K\tau^+\tau^-$ decays showing an electroweak penguin transition in standard Model (left) and a potential new physics transition (right). The New Physics (NP) operator coupling to the four fermions is represented as a blob.

As seen in Fig. 1.7, an important kinematic variable is q^2, which corresponds to the invariant mass of the di-tau system:

$$q^2 = (p_B - p_K)^2 = (p_{\tau^+} + p_{\tau^-})^2 = m^2_{\tau^+\tau^-} \tag{5.1}$$

where p is the four-momentum of the particle. The range of q^2 is given by the 4−momentum conservation $(m_\tau^+ + m_\tau^-)^2 \leq q^2 \leq (m_B - m_K)^2$ and the shape obtained with the BTOSLLBALL model of EvtGen generator is shown in Fig. 5.2.

[1]The same holds for the boost vector components. The definitions of 'experiment' and 'run' are given in Sec. 2.2.8.

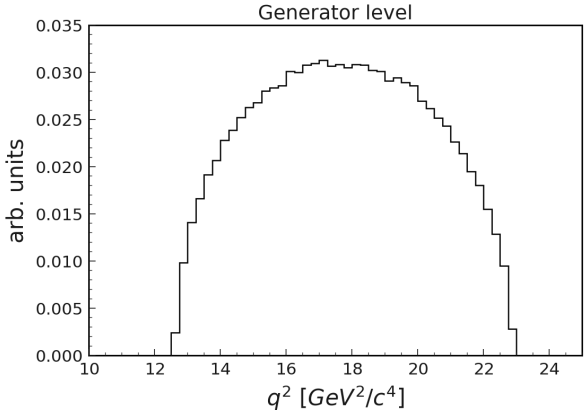


FIGURE 5.2: $q^2 = m^2_{\tau\tau}$ distribution for $B^+ \to K^+\tau^+\tau^-$ events generated with the BTOSLLBALL model.

Generic MC

For background studies, the officially produced *generic* samples with run-dependent beam background were used in Belle. They are generated through different processes, which are listed in the Table. 5.1, aiming to resemble the recorded data set in both size and composition. Like the signal MC, the decays of all particles in charged and mixed samples are simulated with EvtGen[2], and the continuum samples are generated with KKMC interfaced to PYTHIA 6. The detector response is simulated with Geant3.

Name	Process
mixed	$\Upsilon(4S) \to B^0\bar{B}^0$, with generic B^0 decays
charged	$\Upsilon(4S) \to B^+B^-$, with generic B^+ decays
charm	continuum $e^+e^- \to c\bar{c}$
uds	continuum $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}$

TABLE 5.1: Composition of the generic MC samples.

For charged MC, the official MC is replaced by the new MC produced based on the corrections described in Sec. 4.1 in order to use the latest calibration factors. It is worth noting that the generic B samples in Belle contain only decay modes with $b \to c$ quark transitions. The suppressed *charmless* B decays are simulated in specific MC samples. These include the $B \to X_u\ell\nu_\ell$ decays where X_u originates from a $b \to u$ transition, e.g., a π or ρ meson. And another such sample is called *rare* and contains other known and predicted rare decays, such as the $b \to s/d$ penguin transitions. The $B \to K^{(*)}\ell^+\ell^-$ modes are included in this data set with the branching fractions based on SM predictions, e.g. the branching fraction used for $B^+ \to K^+\tau^+\tau^-$ is 1.3×10^{-7} based on the prediction from Ref. [116]. Because of their low rates, these samples are produced with a much larger luminosity: $20\times$ for $b \to u\ell\nu$ and $50\times$ for other rare decays. Nevertheless, the total amount of signal candidates in this sample would not be sufficient for our goal of studying the sensitivity, so the dedicated signal MC samples described above are essential.

All of these official Belle MC are stored in Belle `mdst` format. As mentioned in Sec. 2.3.1, the b2bii framework is used to convert the PANTHER [60] format (used in BASF) into ROOT-objects [61]. After such a conversion, all Belle II analysis tools can

[2]The final state radiation (FSR) is simulated with the PHOTOS package [89].

be used in basf2 and the same scripts can be used for studies on Belle II samples. This is essential to perform any Belle+Belle II measurement, such as the present one. The b2bii conversion is performed by three main modules: `B2BIIMdstInput`, which opens and reads Belle `mdst` files, `B2BIIFixMdst`, which corrects the read-in mDST files, and the `B2BIIConverMdst` module, which is responsible for the actual conversion. After that, Belle II-type objects (Tracks, ECLClusters, PIDLikelihoods...) are available in the framework.

The `B2BIIFixMdst` also applies the selection of *HadronBJ* skim. Events are discarded based on track multiplicity and visible energy: the event must have at least three charged tracks with $p_t > 0.1$ GeV/c and the visible energy (sum of the energy of the charged tracks and the reconstructed photons) must be greater than 20% of total energy, $\sqrt{s}$. Additional selection criteria remove most beam background and two-photon events, with the former reduced by requiring that the position of the primary vertex of the event be near the IP. Background events from QED and *tau*-pairs are suppressed by cutting on the total energy measured in the ECL and the HJM variable[3]. Such combined cuts do not remove light quark pair production events ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$), but are not efficient for inclusive ψ events. Therefore, the events with J/ψ and $\psi(2S)$ candidates are explicitly added to *HadronBJ*. The *HadronBJ* selection makes the data set almost free of low-multiplicity processes (QED and $\tau^+\tau^-$). Although reconstructing only the exclusive reconstruction of hadronic B-decays would naturally discard those events, the skim prevents us from running on them and thus saves computational time.

In Belle II, we use the generic MC from the 15th official MC production campaign, with the same composition as in Table. 5.1. Just as in Belle, the official charged MC is replaced by the new charged MC generated based on the corrections described in Sec. 4.1. Here, the continuum generator is updated to PYTHIA 8 and the detector simulation is performed using Geant4. In each campaign, two variants of MC are produced: run-independent (RI) and run-dependent (RD). RD MC uses beam background overlay from data taken with random triggers and evolving detector conditions. In our case, an RI generic MC equivalent to 400 fb^{-1} is used for background estimation to match the RI Signal MC used for Belle II. However, the relevant differences between the RI and RD MC samples are further investigated using control samples.

Control sample

A centrally produced $B^+ \to J/\psi K^+$ MC corresponding to 100 times the total integrated luminosity of Belle is used as a control sample to investigate the agreement of the data-MC of the $E_{\rm ECL}$ observable. There is no equivalent production in Belle II, so the events with $B^+ \to J/\psi K^+$ are skimmed from the charged MC for similar validation studies. Here, both RI and RD charged MC are used for comparison.

5.2 Selection

This section describes the selection of the $B_{\rm tag}$ candidates and the tracks that form the $B_{\rm sig}$ candidates for the Belle and Belle II data sets. The $B_{\rm tag}$ selection described in Sec. 5.2.1 is the same as in Ch. 3 and 4, so that the performance measured there can be used directly. Section 5.2.3 describes the additional selection used to remove certain trivial background contributions.

[3]The *Heavy Jet Mass* is the invariant mass of particles found in hemispheres perpendicular to the event thrust axis.

5.2.1 B_{tag} reconstruction

Event selection begins with the reconstruction of hadronic B_{tag} candidates using the FEI package, as described in Sec. 3.3. By default, the FEI package can produce up to 20 B_{tag} candidates per event, requiring some selection of the best candidate. To take full advantage of all the improvements implemented by our B-tagging studies described in Ch. 4, we use the new charged MC for Belle and Belle II. For Belle II we use the v3 training, trained on the new MC with a looser gamma selection and including the π^0 mass constraint as described in Sec. 4.4. Thus, with the same selection, we can expect the performance described in Tab. 4.11.

The final classifier output of the FEI package, $\mathcal{P}_{\text{FEI}}$, ranges from zero to one, representing candidates identified as background and signal, respectively. It is shown that a selection on $\mathcal{P}_{\text{FEI}}$ can be used to increase the purity of the tag side at the expense of efficiency. In the $D\pi$ sample used in Chapter 4, the signal and background components can be unfolded using the sPlot technique. A comparison of these components in Belle and Belle II MC in Fig. 5.3 shows that the background, especially at low $\mathcal{P}_{\text{FEI}}$, is much higher in Belle II. While $\log_{10}(\mathcal{P}_{\text{FEI}}) > -3$ is used for Belle, a

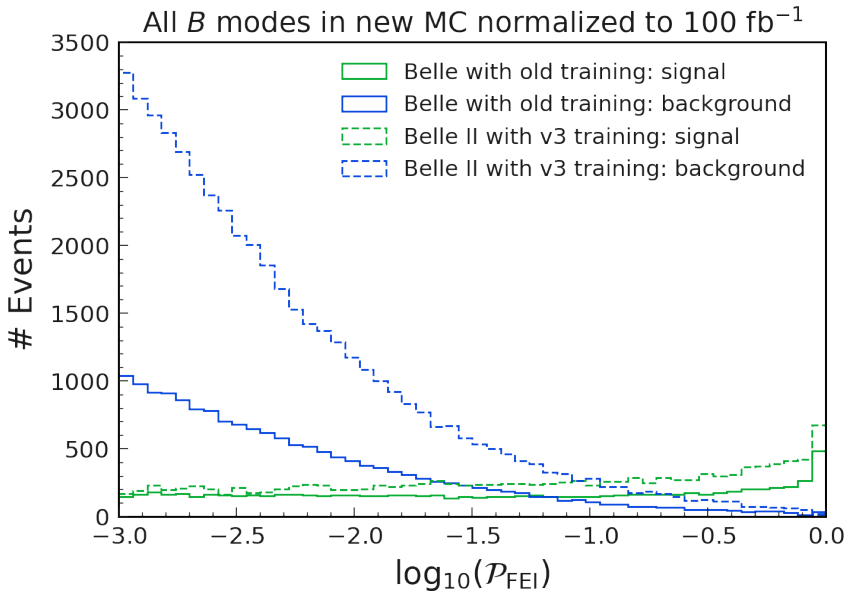


FIGURE 5.3: Comparison of $\log_{10}(\mathcal{P}_{\text{FEI}})$ between Belle and Belle II MC, normalized to 100 fb^{-1}, for both signal and background components unfolded using the sPlot technique on the $D\pi$ sample.

more stringent selection of $\log_{10}(\mathcal{P}_{\text{FEI}}) > -2$ is used for Belle II to achieve comparable performance in both samples. The rest of the selection is kept consistent for both data sets, corresponding to:

- $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$,
- $-0.15\,\text{GeV} < \Delta E < 0.1\,\text{GeV}$,
- cosTBTO < 0.9,
- sphericity > 0.2.

Hadronic B-tagging provides powerful constraints on two variables described in Sec. 3.2.1. M_{bc} represents the mass of the B meson, independent of the final state

particle mass hypotheses; and ΔE represents the deviation of the reconstructed energy from the beam energy, which is instead sensitive to misidentification. The spread of the $M_{\rm bc}$ distribution is dominated by the spread in the beam energy, while the spread in ΔE is dominated by the resolution of the reconstructed B meson energy. However, in the case of the $B_{\rm tag}$ candidates produced by the FEI package, the ΔE distribution is skewed because it is used as a training variable. The selection applied to $M_{\rm bc}$ and ΔE of the $B_{\rm tag}$ candidates is shown in Fig. 5.4.

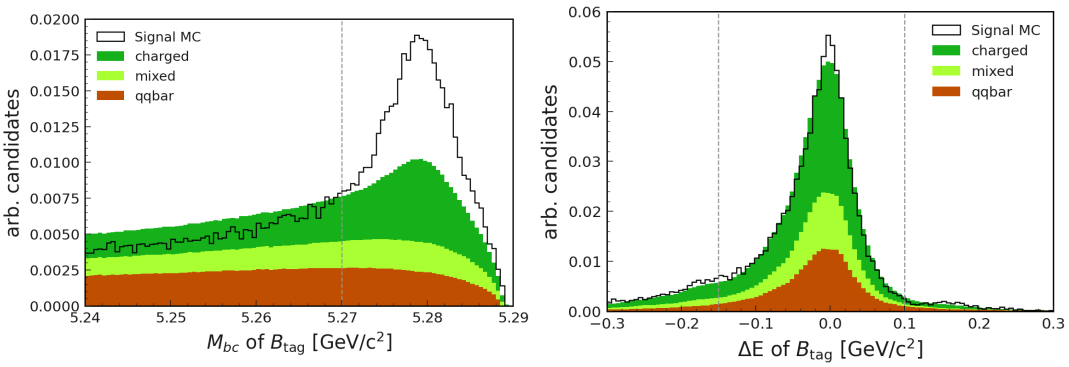


FIGURE 5.4: $M_{\rm bc}$ and ΔE distributions of the $B_{\rm tag}$ candidates reconstructed by the FEI package. The cuts $\mathcal{P}_{\rm FEI} > 0.001$ and $-0.15\,\text{GeV} < \Delta E < 0.1\,\text{GeV}$ ($M_{\rm bc} > 5.27\,\text{GeV}/c^2$) are applied for $M_{\rm bc}$ (ΔE) distribution. Only Belle MC is shown here, the distributions for Belle II MC look similar.

To further suppress the continuum background, a selection is applied to cosTBTO and sphericity variables. The cosTBTO is the cosine of the angle between the thrust vectors of the $B_{\rm tag}$ candidate and the corresponding rest of the event. As shown in Fig. 3.13, a cut at high values is applied to this thrust-based discriminant. The sphericity is calculated from the second and third eigenvalues of the sphericity tensor described in Eq. 3.9. The $B\overline{B}$ events are at rest in the CM frame, and their resulting momentum is uniformly distributed, resulting in higher values of sphericity. The momentum of the continuum events, on the other hand, are generally collimated back-to-back, and have low values of sphericity, which can be removed by a cut, as shown in Fig. 3.18.

At this stage, about 1.7% and 1.2% of the events are retained in the Belle and Belle II signal MC, respectively, which is mostly the FEI efficiency. The corresponding average multiplicities of $B_{\rm tag}$ candidates are 2.12 ± 0.01 and 1.68 ± 0.01. This can be reduced by selecting only the candidate with the highest $\mathcal{P}_{\rm FEI}$. This ranking is performed after the signal side has been selected to achieve optimum efficiency.

5.2.2 $B_{\rm sig}$ reconstruction

After the reconstruction of the $B_{\rm tag}$ candidate, the charged tracks on the signal side are combined to form the $B_{\rm sig}$ candidate. The charge of the kaon track is uniquely determined from the charge of the $B_{\rm tag}$ candidate. As shown in Fig. 5.5, the leptonic τ decays account for only ∼35% of the total branching ratio, but provide the highest purity.

The inclusion of $\tau \to \pi$ decays will lead to large background. Therefore, only leptonic τ decays are chosen for this study, in three different combinations: e^+e^-, $\mu^+\mu^-$ and $e^\pm\mu^\pm$.

FIGURE 5.5: Visualization of the τ decay branching fractions (%). The leptonic decays account for ∼35% of the total, the rest coming from hadronic decays.

To reject secondary tracks created by the interaction with the detector, the tracks are required to originate from the interaction point through cuts on the two impact parameter variables defined as:

i d_r (transverse impact parameter): the absolute distance in the transverse plane between the IP and the point of closest approach (POCA) of the helix describing the track,

ii d_z (longitudinal impact parameter): the signed distance along the z-axis between the POCA and the IP.

With the improved nano-beam scheme used in SuperKEKB, the resolution of the IP in the z-direction is greatly reduced in Belle II. This allows a tighter constraint on d_z for Belle II, as shown in Fig. 5.6. A cut on the variable `thetaInCDCAcceptance` ensures that the theta of the track is in the detector acceptance. And requiring $p_t > 0.2$ GeV/c ensures that the track is well reconstructed in the CDC.

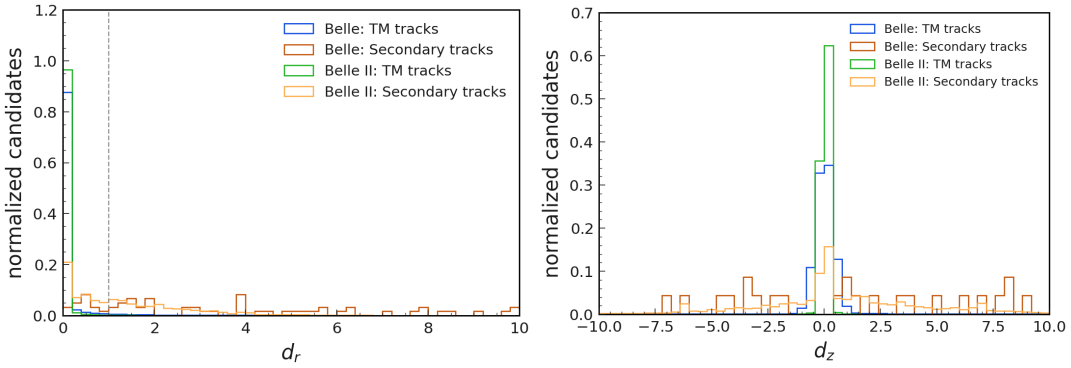


FIGURE 5.6: Distributions of d_r and d_z showing the selection. TM stands for truth-matched tracks, shown in contrast to secondary tracks in the signal MC.

As described in Sec. 2.2, particle-ID variables are computed for each charged track and are applied here accordingly. The kaon track is selected using the $K\text{ID} = L(K : \pi) > 0.6$ criterion (see Eq. 2.7, while for the two leptons, the corresponding lepton-ID should be above 0.9. The typical performance of these particle-ID cuts is summarized in Tab. 5.2, which is evaluated over the entire momentum spectrum. Since the leptons (from τ decays) in this analysis are kinematically confined, they tend to have lower momentum, leading to lower particle-ID performance, as shown in Figs. 2.8 and 2.10.

In Belle, the μID is calculated using only the KLM information. A track needs a momentum of at least 0.6 GeV/c to reach the KLM detector, so using $\mu\text{ID} > 0.9$ only retains tracks with $|\vec{p}| > 0.6$ GeV/c. However, in Belle II, likelihoods calculated individually from the inner detectors' information: CDC, TOP, ARICH, ECL, is included on top of the KLM to obtain μID to improve the muon identification efficiency at low momentum. However, in the low momentum region, the fake rate is much higher, as shown in Fig. 5.7. In Belle II, the tracks that match the generated muons,

Selection	Efficiency		Fake rate from π	
	Belle	Belle II	Belle	Belle II
kID>0.6	92%	89%	5%	7%
eID>0.9	87%	86%	0.2%	0.4%
μID>0.9	89%	88%	1.5%	7%

TABLE 5.2: Average performance of particle-ID selection in Belle and Belle II data sets.

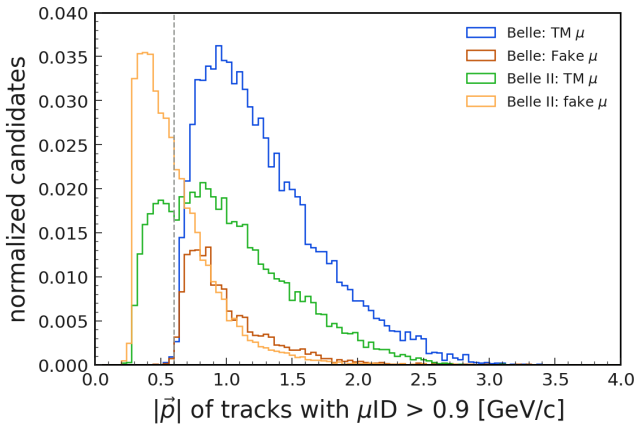


FIGURE 5.7: Momentum distribution of tracks with the condition μID>0.9 showing the truth matched (TM) muons and fake muons in Belle and Belle II signal MCs. A cut at $|\vec{p}| > 0.6$ is applied for Belle II in order to contamination from pions.

labelled as "truth matched" (TM) muons, are shown in green. And the tracks that meet the selection criteria but are not generated as muons are shown in orange. In the low momentum region, it is clear that the fake tracks dominate the muon tracks and provide more background than signal. Therefore, an additional constraint of $|\vec{p}| > 0.6$ is applied to muon candidates in Belle II.

The entire selection criteria for the charged tracks from signal side for Belle and Belle II are summarized in Tab. 5.3.

Particle	Selection Cuts in Belle (II)
All tracks	$\lvert d_r\rvert < 1$cm, $\lvert d_z\rvert < 2$ (1)cm
	thetaInCDCAcceptance
	$p_t > 0.2$ GeV/c
$K^\pm$	kID > 0.6
$e^\pm$	eID > 0.9
$\mu^\pm$	μID > 0.9 and $\lvert\vec{p}\rvert > 0.6$ GeV/c

TABLE 5.3: Signal side FSP selection criteria. Except for the impact parameter cuts, the rest are the same for Belle and Belle II.

The three tracks can be combined into a B_{sig} candidate, even if not fully reconstructed. Before combining with the B_{tag} candidate, some of the trivial background on the signal side can be removed by simple cuts. For example, if both leptons are of the same family, there may be some background contribution from $B^+ \to K^+ J/\psi$, where the J/ψ decays to two leptons. These events can be easily discarded by vetoing the $3.00 < M(\ell^+\ell^-) < 3.19$ GeV/c^2 region, as shown in Fig. 5.8. These discarded events are later used as control sample for this analysis, as described in Sec. 5.5.2.

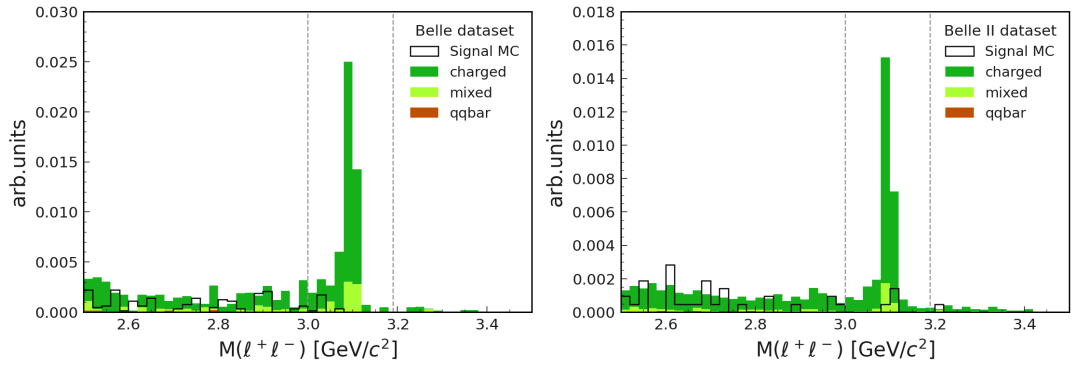


FIGURE 5.8: Invariant mass of the leptons, showing the selection to discard background from $B^+ \to K^+ J/\psi$ in Belle (II) MC on left (right).

Some contributions may arise from photon conversion. Most of this background is already suppressed by the fact that the electrons must have a transverse momentum greater than 0.2 GeV/c. It can be further suppressed by requiring the invariant mass of the electron with any other oppositely charged track in the event to be greater than 100 MeV/c^2, as shown in Fig. 5.9.

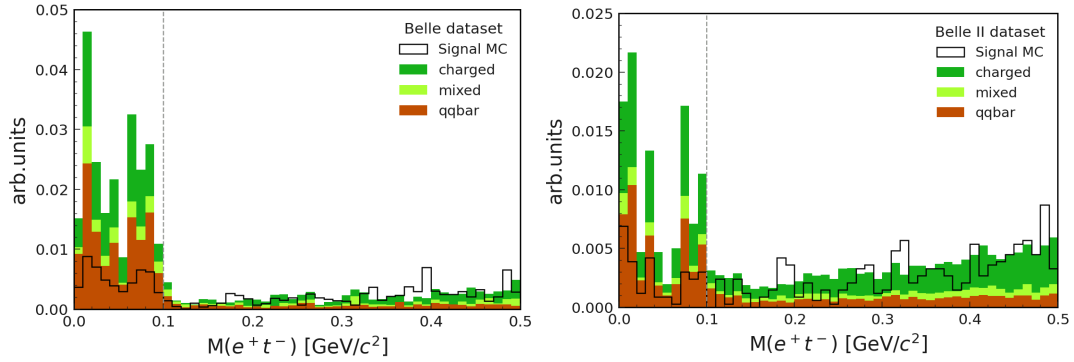


FIGURE 5.9: Invariant mass of the electron with an oppositely charged track, illustrating the selection applied to discard the background contribution from photon conversion for Belle (II) on left (right).

Another source of background is $D^0 \to K^-\pi^+$ decays in which the π has been misidentified as a lepton, usually a muon. To reject this contribution, the invariant mass of the $K^\pm$ with the oppositely charged lepton must be outside the D^0 mass region, i.e. $M(K^+\ell^-) < 1.80$ GeV/c^2 or $M(K^+\ell^-) > 1.90$ GeV/c^2, as illustrated in Fig. 5.10.

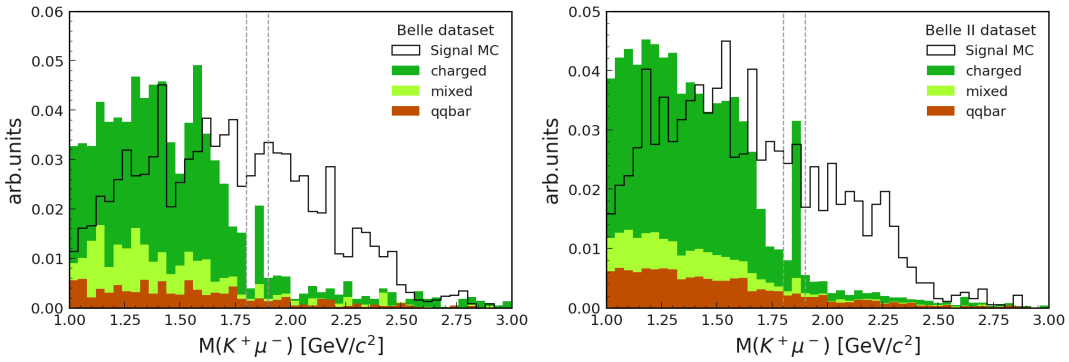


FIGURE 5.10: Invariant mass of the $K^{\pm}$ with the oppositely charged lepton, illustrating the selection applied to discard background from $D^0 \to K^-\pi^+$ decays for Belle (II) MC on left (right).

5.2.3 Event and best candidate selection

After selecting the signal candidates, all compatible B_{tag} and B_{sig} candidates can be combined to form $\Upsilon(4S)$ candidates. For each candidate, q^2 is calculated according to Eq. 5.1. The value of q^2 must be at least $(m_\tau + m_\tau)^2$, so a cut of $q^2 > 12$ can remove some trivial background, as shown in Fig. 5.11.

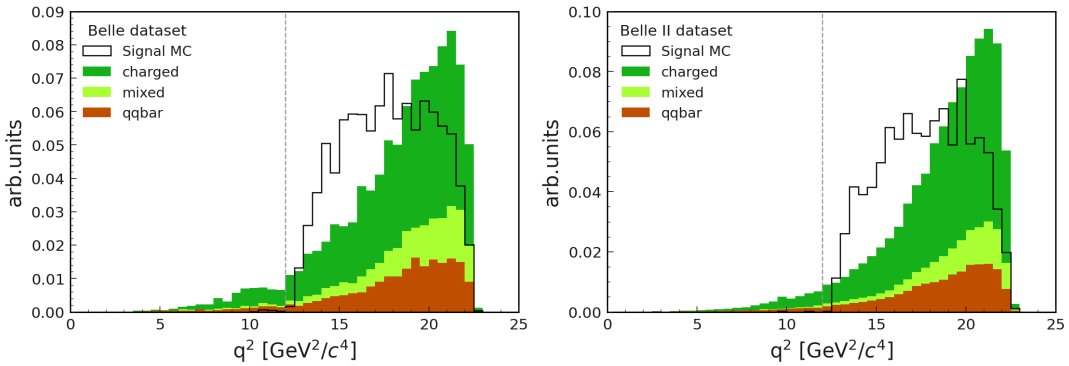


FIGURE 5.11: q^2 distribution, showing the $q^2 > 12$ cut, in Belle (II) MC on left (right).

For all $\Upsilon(4S)$ candidates that satisfy the selection criteria described so far, the Rest of Event (ROE) is constructed with all tracks and clusters of the event that were not associated to B_{tag} and B_{sig} candidates. For correctly reconstructed signal events, no additional tracks are expected in the event. Additional tracks in the ROE can occur in signal MC, especially if a pion originating from a τ decay with higher multiplicity is misidentified as a lepton track. Since this is undesirable, the number of additional tracks satisfying the requirements given in Tab. 5.3 is explicitly required to be 0. This significantly reduces the background, as shown in Fig. 5.12.

The entire selection requirements are summarized in Table. 5.4. At this stage, there may still be multiple candidates per event, essentially from the B_{tag} candidates. Thus, keeping only the candidate with the highest $\mathcal{P}_{\text{FEI}}$ ensures that there is only one $\Upsilon(4S)$ candidate per event.

After selecting the best candidate, the extra energy deposited in the calorimeter can be calculated as the sum of the energies of all neutral clusters in the ROE. The distribution of E_{ECL} is shown in Fig. 5.13 for the Belle and Belle II data sets. Here the signal MC corresponds to the generated 10^7 events and the generic MC is normalized to 100 fb^{-1}, allowing a comparison between the two experiments. The mean of the E_{ECL}

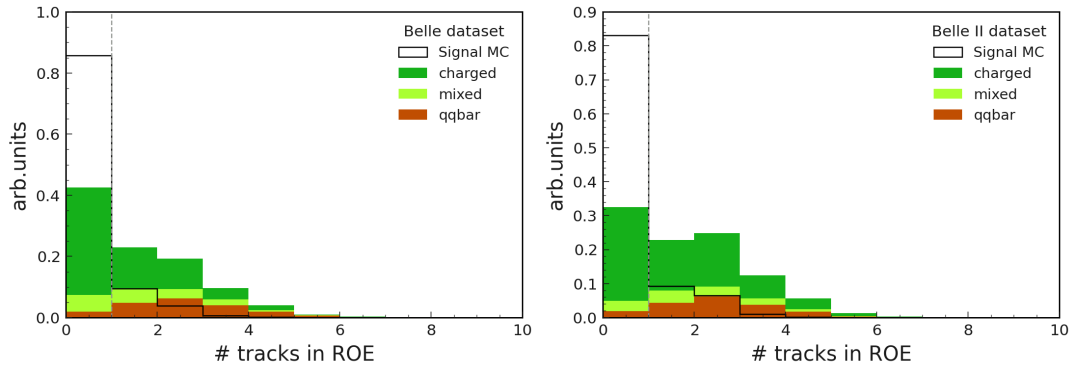


FIGURE 5.12: Distribution of number of extra tracks in the Rest of Event (ROE) of the $\Upsilon(4S)$ candidates, showing the cut requiring exactly 0 track, in Belle MC (left) and Belle II MC (right).

Selection	Cut
J/ψ veto	not$(3.00 < M(\ell^+\ell^-) < 3.19$ GeV$/c^2)$
Photon conversion veto	not$(M(e^+t^-) < 0.1$ GeV$/c^2)$
D^0 veto	not$(1.80 < M(K^+\ell^-) < 1.90$ GeV$/c^2)$
Kinematic constraint on $m_{\tau^+\tau^-}$	$q^2 > 12$
No extra tracks	# tracks in ROE of $\Upsilon(4S) = 0$

TABLE 5.4: Selection applied to remove trivial background.

distribution for the signal is lower than the background. However, the distribution does not peak at the expected value of 0, due to the contributions from the beam background and hadronic split-off clusters in the E_{ECL} calculation of signal events. The mean is higher in the case of Belle II MC due to the larger beam background, which is simulated in RI MC as described in Sec. 5.1. This observable can be improved by applying certain requirements (called *mask*) to the clusters that are summed to obtain E_{ECL}, as described in detail in Sec. 5.4.

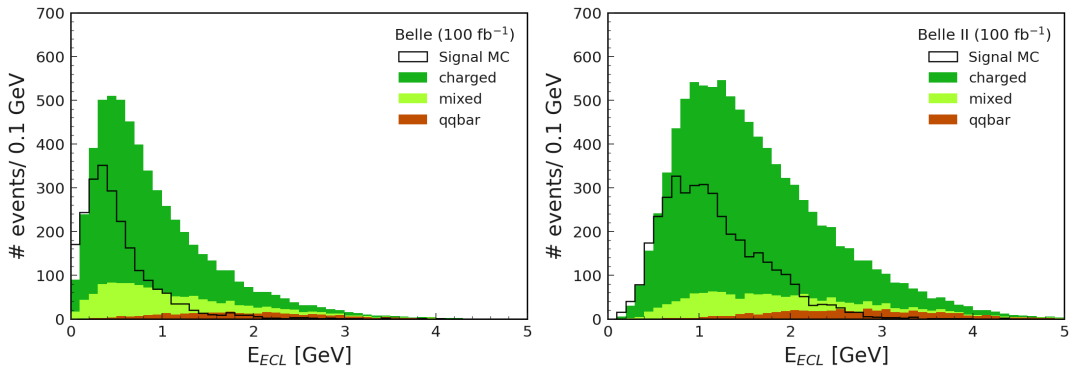


FIGURE 5.13: Distribution of the extra energy in calorimeter, with no constraints on the clusters, in Belle (II) MC on left (right).

The signal efficiency at this stage is 2.3×10^{-4} and 3.9×10^{-4} for Belle and Belle II, respectively. The difference between the two arises from the differences in the FEI and in the PID performance of the signal side tracks. At this stage, the background yields are still very high. The composition and the strategy to suppress the background are described in the next section. Note that these are calculated in the whole region of

E_{ECL} and will definitely improve when measured in the appropriate signal window of E_{ECL} with an optimal mask.

5.3 Background studies

After removal of the trivial background by cuts, the background yields are 5.7×10^3 and 9.5×10^3 per 100 fb^{-1} in Belle and Belle II, respectively. For each of the final states that depend on the τ decays, the background composition can be analysed as described in Sec. 5.3.1. Once the primary background contributions are known, BDTs are trained with appropriate variables to suppress them while maintaining the signal efficiency.

5.3.1 Sources of background

The signal efficiencies and background yields are given in Tab. 5.5 for the four possible final states. The $e\mu$ state is divided into two categories, depending on whether the charge of μ is the same as or opposite to K. The final states with μ have lower signal efficiency due to the poor performance of μ-ID at low momentum. The differences in performance between the two experiments lead to the signal efficiency and background yield in Belle II being about a factor two higher, especially for the modes with a μ in the final state.

	$K^+e^+e^-$		$K^+e^+\mu^-$		$K^+\mu^+e^-$		$K^+\mu^+\mu^-$	
	B	BII	B	BII	B	BII	B	BII
Signal efficiency (10^{-5})	8.7	12	5.5	10	5.2	9.3	3.4	7.6
Background yields	2010	2884	1059	2082	1689	2661	954	1897

TABLE 5.5: Signal efficiency and background yields in Belle (represented by B) and Belle II (represented by BII) MC samples. Signal efficiency is calculated as the number of events retained after all the cuts in Signal MC divided by 10^7 events generated. The background yields are calculated per 100 fb^{-1} for direct comparison between experiments.

For all the final states, the fractions of the different MC types in the background yield are given in Tab. 5.6. The small remaining $q\bar{q}$ background can be suppressed using the normalized Fox-Wolfram moment $R_2 = H_2/H_0$ [119], which has values close to one for events with strongly collimated jets. Efforts are devoted to suppress the background in charged MC, which is dominant in both data sets, as can also be seen in Fig. 5.13.

Background fraction (%)	B	BII
charged	77	84
mixed	18	11
$q\bar{q}$	6	5

TABLE 5.6: Fraction (in %) of different MC types in the background yields in Belle (represented by B) and Belle II (represented by BII) samples.

For the charged (B^+B^-) events, the primary background is expected to have the same final state. This means that the B_{tag} is correctly reconstructed, accompanied by $B^+_{\text{sig}} \to \overline{D}^{(*)0}\ell^+\overline{\nu}_\ell$, where $\overline{D}^0$ can either decay semileptonically ($\overline{D}^0 \to K\ell'\overline{\nu}_{\ell'}$) or sometimes decay hadronically ($\overline{D}^0 \to K^+\pi^-X$, where X is missing). To analyse this, the charged events are categorized by the nature of decays of B and D, as in Tab. 5.7.

Fraction of charged bkg (%)	$K^+e^+e^-$		$K^+e^+\mu^-$		$K^+\mu^+e^-$		$K^+\mu^+\mu^-$	
	B	BII	B	BII	B	BII	B	BII
SL B x SL D	82	71	69	56	84	74	69	56
SL B x Had D	9	15	21	33	8	14	23	34
Had B x SL D	6	9	5	6	6	8	5	6
Had B x Had D	3	5	5	5	2	4	3	4

TABLE 5.7: Fraction (in %) of charged background in different configurations of B and D decays in Belle (represented by B) and Belle II (represented by BII) samples. SL stands for semileptonic decays and Had stands for hadronic decays

As expected, the semileptonic B decays make up about 90% of the background in charged MC. In such a background, the lepton with the same charge as B_{sig} would be the prompt lepton originating from B, and would therefore have a higher momentum than the lepton originating from the τ decay in signal events. This variable, the momentum of the lepton with the same sign in the CM frame ($p^*_{\ell^+}$), as shown in Fig. 5.14, can be used to suppress the dominant background. The effect of $p_\mu > 0.6$ GeV/c can be seen in the figure as a jump in the distribution at the corresponding momentum.

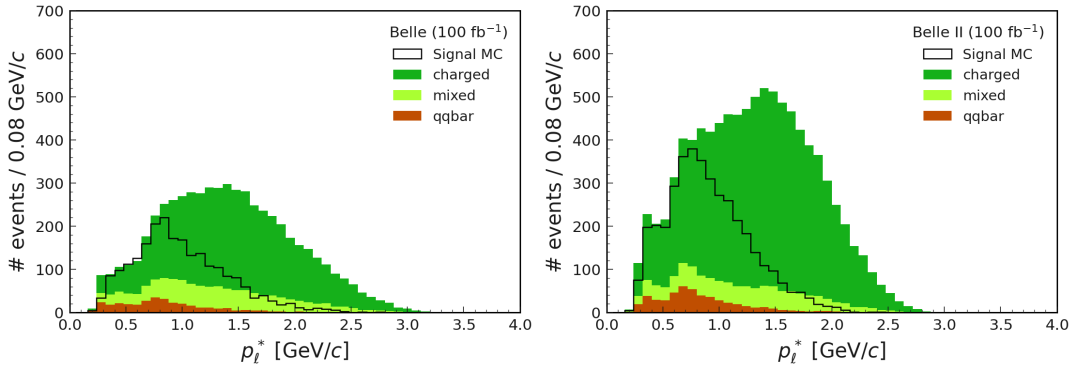


FIGURE 5.14: Distribution of the momentum of the lepton with the same charge as B_{sig} in the CM frame in Belle (II) MC on left (right).

Since μ are more often misidentified from π, the contribution of hadronic D decays to the background is larger when a μ with opposite charge than K is present in the final state. And systematically, the contribution of hadronic D decay is larger in Belle II due to the higher fake rate. Regardless of the D decays, the invariant mass of $K^+\ell^-$ for these backgrounds would be less than the D mass because of the missing energy, which can be seen as a large background in the range $M(K^+\ell^-) < 1.8$ GeV/c^2 in Fig. 5.10.

The kinematic variable q^2 shown in Fig. 5.11, as well as $missingMass^2$, may also contribute to background suppression. The $missingMass^2$ is an event kinematic variable that is defined independently of the reconstruction of the B candidates. The visible four-momentum is first reconstructed from all the clusters and tracks (with π mass hypothesis) and subtracted from the total four-momentum of $\Upsilon(4S)$ to obtain the missing four-momentum. The $missingMass^2$ shown in Fig. 5.15 generally has higher values for signal events than for background events due to higher neutrino multiplicity.

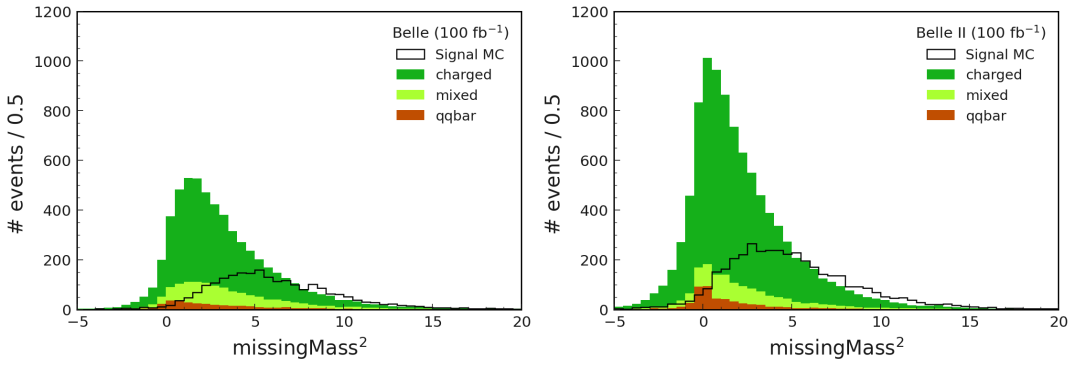


FIGURE 5.15: Distribution of the $missingMass^2$ in Belle (II) MC on left (right).

5.3.2 Background suppression

Using the five features described so far, individual BDTs are trained for Belle and Belle II MC samples for background suppression. Due to the low statistics of the signal MC at this stage, a common BDT is trained with all final states. In the future, with a larger signal MC, an individual BDT for each final state can be trained and optimized, which could improve the background suppression capabilities. It is important that there is minimal correlation between the input features and the observable E_{ECL} so that the BDTs don't introduce any bias. And the mask for E_{ECL} can be chosen independently of the BDT-based background suppression. Fig. 5.16 shows the linear Pearson correlation constants between the training variables, in addition to E_{ECL}: on the upper (lower) side of the diagonal the values refer to the background (signal) events. For the background events, the largest correlation with the E_{ECL} is about -40% for p^*_{l+}, which is expected because for the $B^+_{\text{sig}} \to X\ell^+\overline{\nu}_\ell$ events a lower p^*_{l+} indicates that X could be heavier than $\overline{D}^{(*)0}$ and contain photons or π^0 entering the ROE and increasing the E_{ECL}.

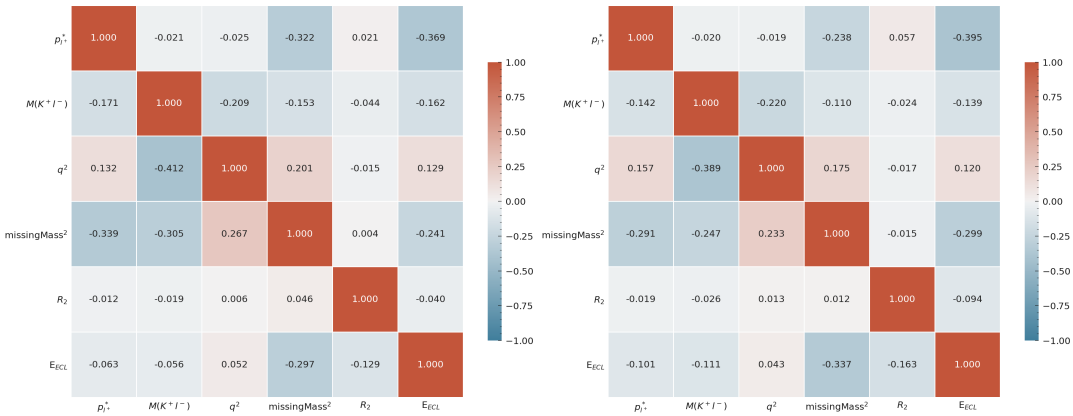


FIGURE 5.16: Correlation matrix of the variables used plus the E_{ECL} for Belle (II) MC samples on left (right). On the upper side of the diagonal, the values refer to background events, while on the lower side, signal events are shown.

FastBDT [120], the default MVA algorithm in the Belle II software, implements a uniform stochastic gradient boosting algorithm. The algorithm uses an equal-frequency binning of the input data, which allows to map the continuous input features to integers, allowing to fit faster than other algorithms. The individual MC components are scaled down proportionally to match the size of the signal being trained. To

avoid over-training, the following hyper-parameters are carefully chosen: the number of decision trees, the maximum depth for each tree, the number of points within the range of each variable used to determine the optimal cut (2^N bins are used for each feature), the fraction of events randomly selected to train each tree, and the level of learning rate for the gradient boosting method. 70% of the sample is used for training and the remaining 30% for testing. The BDT output of the training and testing samples along with the Receiver Operating Characteristic (ROC) curve is shown in Fig. 5.17. Within the statistical uncertainty, no overtraining is observed.

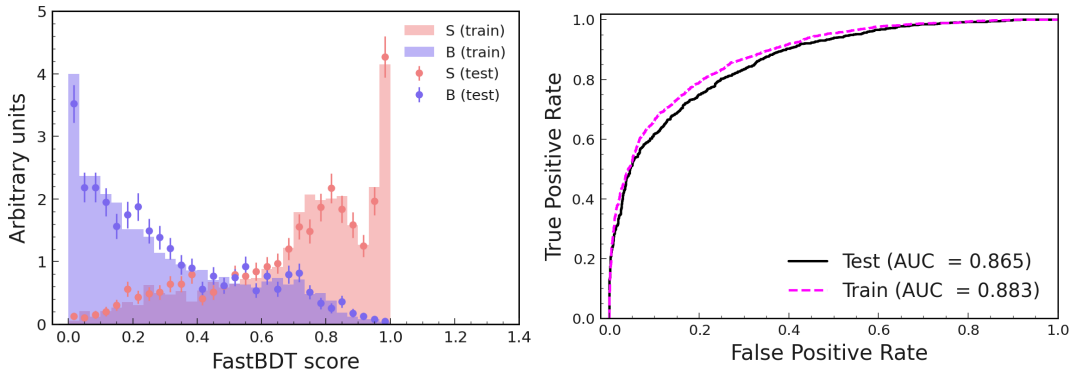


FIGURE 5.17: BDT output of test and training samples for the supervised BDT training and performance in terms of ROC in Belle II MC. The plots for Belle MC are similar.

FastBDT calculates the importance of features based on the information gain of each applied cut by summing the separation gain of each feature in a loop over all trees and nodes. The ranking of the variables used in terms of importance for the training is shown in Fig. 5.18 for both Belle and Belle II MC samples.

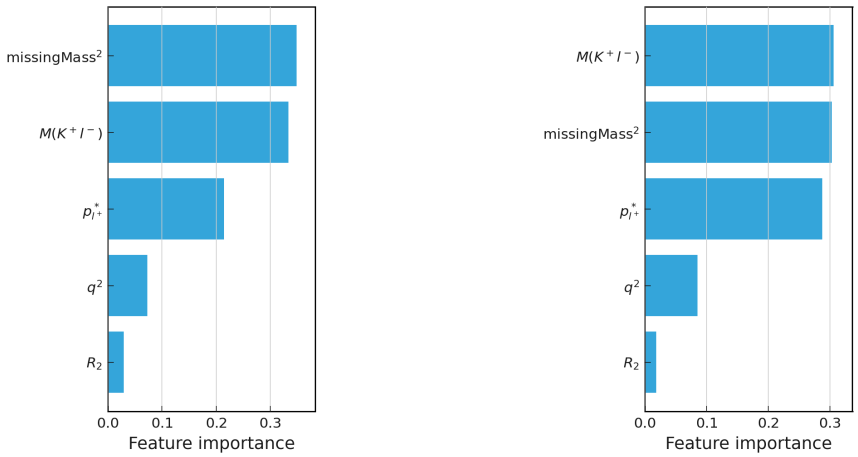


FIGURE 5.18: The ranking of the used variables in terms of their importance for the training in Belle (left) and Belle II (right) MC samples.

The cuts on the BDT output are optimized using the Punzi figure of merit (FOM), which is defined in Ref. [121] as:

$$\mathcal{F}(a,t) = \frac{\varepsilon_{\text{sig}}(t)}{\frac{a}{2} + \sqrt{N_{\text{bg}}(t)}} \tag{5.2}$$

where t is the chosen threshold and $\varepsilon_{\text{sig}}(t)$ and $N_{\text{bg}}(t)$ are the corresponding signal efficiency and number of background events. a is the number of sigmas corresponding

to a one-sided Gaussian test at significance α. For our search, the value $a = 3$ is chosen, although the optimisation point does not strongly depend on the value for a. This FOM is more appropriate than others because the branching fraction of the searched process is unknown.

The distributions of the BDT output along with the chosen cut are shown for Belle and Belle II MC in Fig. 5.19. Overall, the BDTs retain about 26% of the signal

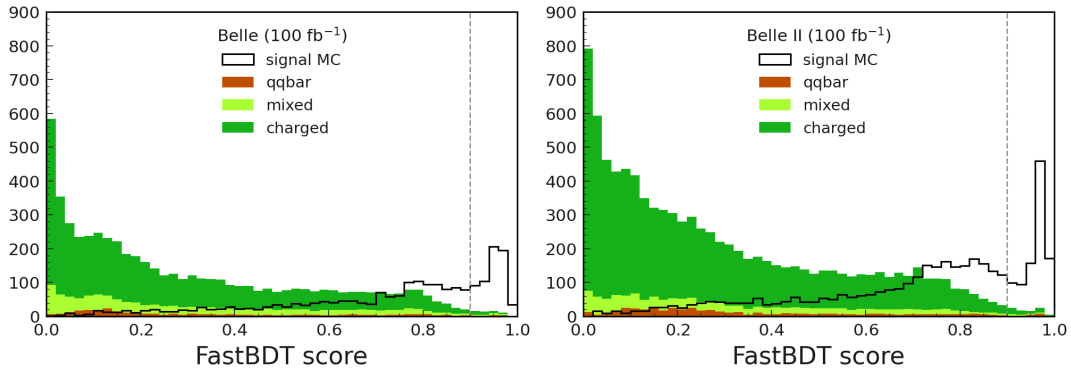


FIGURE 5.19: The BDT outputs of the signal and generic MC, showing the chosen cut based on Punzi FOM in Belle (left) and Belle II (right) MC samples.

events while rejecting about 99% of the background events in both data sets. The E_{ECL} distributions computed with all clusters in the ROE after this BDT cut are shown in Fig. 5.20. For easy comparison with the background shape, the signal MC is

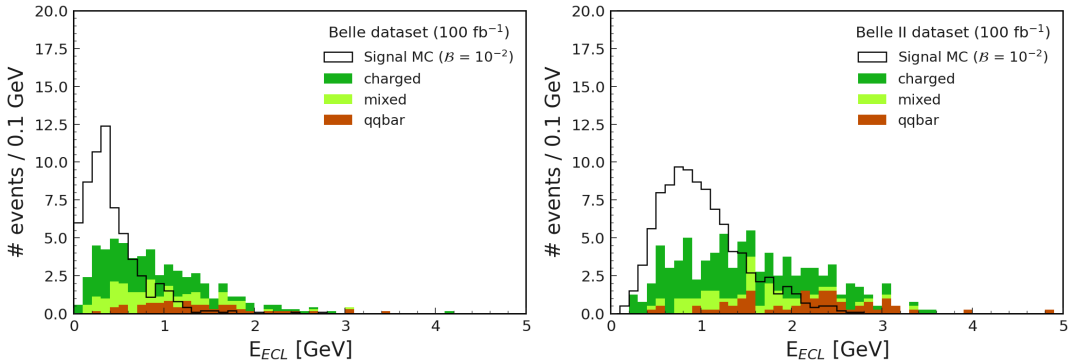


FIGURE 5.20: Distribution of the extra energy in calorimeter, with no constraints on the clusters, in Belle (II) MC on left (right). The generic MC is normalized to 100 fb^{-1} while the signal MC is scaled to the assumption of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) = 10^{-2}$ for visual purposes.

scaled down to the assumption $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) = 10^{-2}$. In order to define a signal window for estimating the upper limit, the resolution of the E_{ECL} must be improved by choosing an appropriate selection on the clusters, as described in the next section.

5.4 Signal extraction: E_{ECL}

After background suppression, the extra energy deposited in the calorimeter, E_{ECL}, is the observable used to identify the signal of $B^+ \to K^+\tau^+\tau^-$. For an $\Upsilon(4S)$ candidate, it is essentially the sum of the energies of all clusters in the corresponding ROE:

$$E_{\text{ECL}} = \sum_{i=1}^{N} E_i \tag{5.3}$$

where N denotes the number of clusters in the ROE. By default, this includes the electronic noise of the calorimeter, the energy deposited by beam background photons and neutral clusters from hadronic split-off processes, etc. Beam background photons are photons produced by various processes involving the particles of the beams in the collider, e.g., synchrotron radiation, pair production, radiative Bhabha scattering, and beam gas Bremsstrahlung with residual gas molecules in the beam pipe. As the beam currents at SuperKEKB will increase in the future, the beam background contribution is expected to increase. When a hadron interacts in the ECL and produces a shower of secondary particles that deposit their energy in another region of the ECL; or in general when energy deposits that are split into multiple clusters, they are called fake photons. To minimize contamination from such sources, certain minimum requirements, called mask, can be applied to the clusters used to calculate the E_{ECL}. Therefore, the effects of different masks on resolution, separation between signal and background as well as the data-MC agreement of the E_{ECL} must be carefully studied.

5.4.1 Photon discriminating variables

The primary variables commonly used to separate background photons from the "good" photons originating from B-decays are:

- **clusterE**: the energy deposited in the calorimeter. The beam background photons are usually of very low energy and are more abundant in forward and backward endcap regions because they are the result of showers from particles from the beam, starting often far from the IP. Most of the beam background can be removed by simple cuts on the clusterE in different detector regions. Figures 5.21 and 5.22 show the clusterE distribution in different detector regions in Belle and Belle II Signal MC. The specific configuration of the cuts, called goodGamma are also shown, which is also the same used for gamma preselection in the FEI package, see Sec. 3.3.

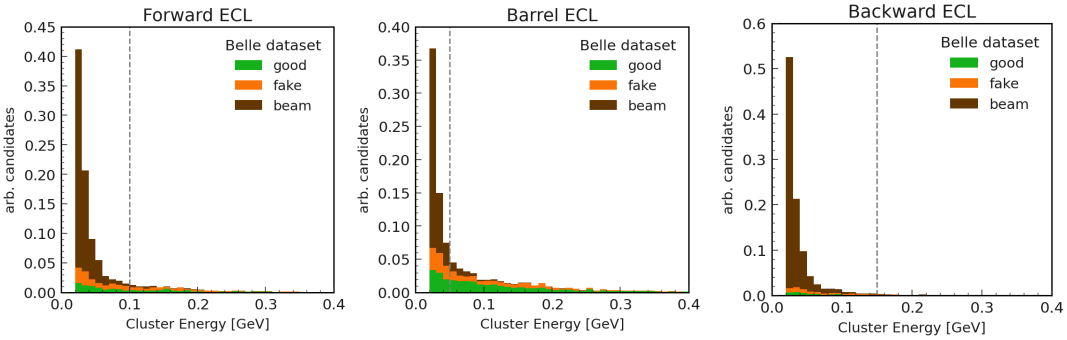


FIGURE 5.21: Cluster energy distributions of the clusters in the ROE in the Signal MC of Belle. The cuts applied to remove the low energy beam background called as goodGamma is shown as dashed line. The plot is normalized to show the relative composition of different types of photons at this level.

- **clusterTiming**: In Belle II, clusterTiming is defined and recorded as the time difference, $t-t_0$, between the time of collision (t_0) and the cluster time (t). Beam background photons can be generated whenever beams travel through the beam pipe, so their time distribution is broad and independent of the collision time t_0. However, for signal photons and fake photons generated shortly after collision, the clusterTiming distribution is expected to peak at zero, as shown in Fig. 5.23. Since most of the beam background can already be removed by the goodGamma

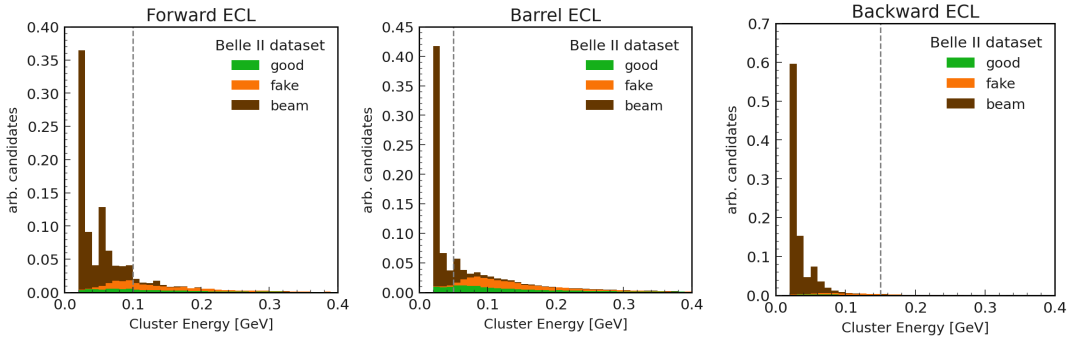


FIGURE 5.22: Cluster energy distributions of the clusters in the ROE in the Signal MC of Belle II. The cuts applied to remove the low energy beam background called as goodGamma is shown as dashed line. The plot is normalized to show the relative composition of different types of photons at this level.

cut described above, this does not do much. But as the beam currents in Belle II are increasing, and consequently the beam background contributions increase, the clusterTiming information will become soon crucial. Therefore, it is used as a training feature in a BDT specifically designed to suppress beam background, as will be explained later.

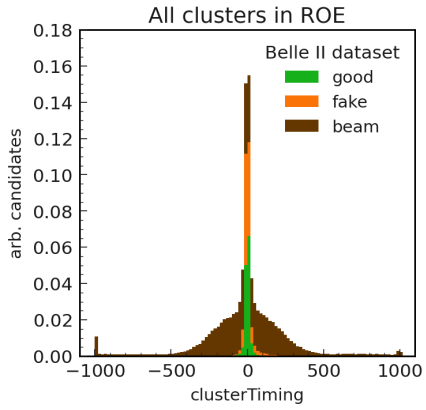


FIGURE 5.23: Cluster Timing distribution of the clusters left in the ROE in the Signal MC of Belle II. The cuts applied to remove the beam background is shown as dashed lines. The plot is normalized to show the relative composition of different types of photons at this level.

However, this information is not available in the Belle data set. As an alternative, the trigger cell (TC) time of the ECL cluster is explored. A TC is a 4×4 crystal matrix in the calorimeter, and a threshold of 0.1 GeV is applied to record the time of each TC. A cluster is built using one or more TCs and when a low energy photon hits the calorimeter, the cluster might be reconstructed, but none of the TCs might cross the threshold to record the time. Therefore, this information is only useful for clusters with energy higher than 0.2 GeV, which is checked using the $J/\psi K^{*+}$ control sample described in Sec. 5.5.1. Fig. 5.21 shows that the beam background photons usually have an energy lower than this, so that the TC time is not useful.

- **minC2TDist**: The distance (in cm) between the cluster and its nearest track. Since the fake photons correspond to clusters that are close to the original tracks, minC2TDist is expected to peak at low values for them. In Belle II, this is measured by the point of closest approach for the particle track extrapolated

into the calorimeter during the reconstruction, taking into account the material effects. In Belle, however, such extrapolated hits of the tracks are not available, so the center of the cluster and the nearest tracks are extrapolated to the inner surface of the calorimeter to calculate this distance. This results in poorer resolution in Belle, but still provides good separation from the fake photons, either by a simple cut, as demonstrated in Fig. 5.24, or as a training feature in a BDT dedicated for fake photon suppression.

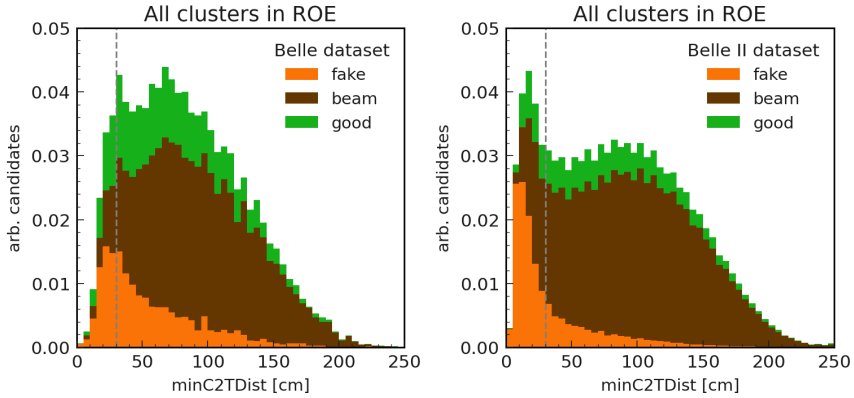


FIGURE 5.24: The distribution of minC2TDist of the clusters in the ROE in the Signal MC of Belle (II) on left (right). The cuts applied to remove the fake photons is shown as dashed line. The plot is normalized to show the relative composition of different types of photons at this level.

In addition to these three variables, there are other variables that provide a non-linear separation. Therefore, simple cuts cannot be applied to them, and instead dedicated BDTs are trained individually for the separation of beam background and fake photon. The full list of features used for training can be found in Tab. 5.8. The **clusterTheta**, the polar angle of the cluster describing the region of the cluster (barrel or endcaps), as shown in Fig. 5.25, is very powerful in distinguishing the beam background and is used in BDTs for both Belle and Belle II.

Feature	Beam background		Fake photon	
	Belle	Belle II	Belle	Belle II
clusterE	✓	✓	✓	✓
clusterTiming		✓		✓
minC2TDist	✓		✓	✓
clusterTheta	✓	✓	✓	✓
clusterPulseShapeDiscriminationMVA		✓		✓
clusterZernikeMVA		✓		✓
clusterNHits	✓		✓	
clusterPhi	✓		✓	
clusterE9E25	✓		✓	
clusterHighestE	✓		✓	
clusterLAT	✓		✓	

TABLE 5.8: Input features for BDTs trained to suppress beam background photons and fake photons.

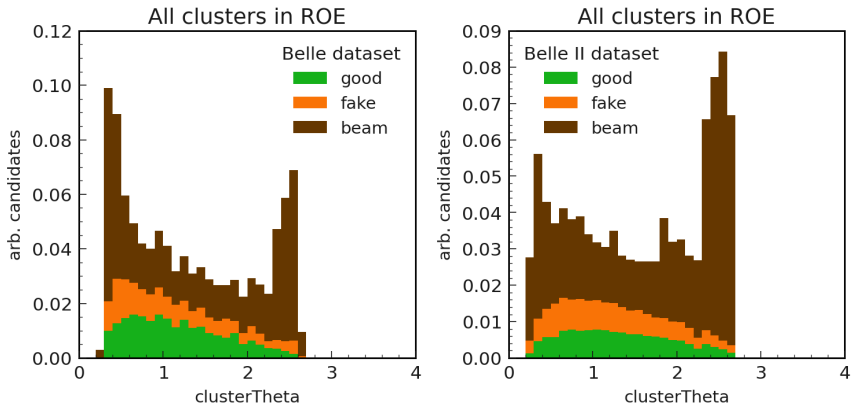


FIGURE 5.25: The distribution of clusterTheta of the clusters in the ROE in the Signal MC of Belle (II) on left (right). The plot is normalized to show the relative composition of different types of photons at this level.

In addition, for the BDTs trained on the Belle II data sets, two variables are used: the output of the MVA trained on the pulse-shape information of the cluster [122], **clusterPulseShapeDiscriminationMVA**; and the output of the MVA trained on the shower-shape information (eleven Zernike moments) of the cluster [123], **clusterZernikeMVA**. They are both distributed between 0 for hadronic showers and 1 for electromagnetic showers, as shown in Fig. 5.26. For Belle II data sets, the BDTs are trained centrally, and the weights are available to the collaboration for use in individual analyses [124]. The clusterPulseShapeDiscriminationMVA has the highest separation power, followed by clusterTiming for separating beam background photons and minC2TDist for separating fake photons. The results of these BDTs are shown in Fig. 5.27 along with the recommended cuts.

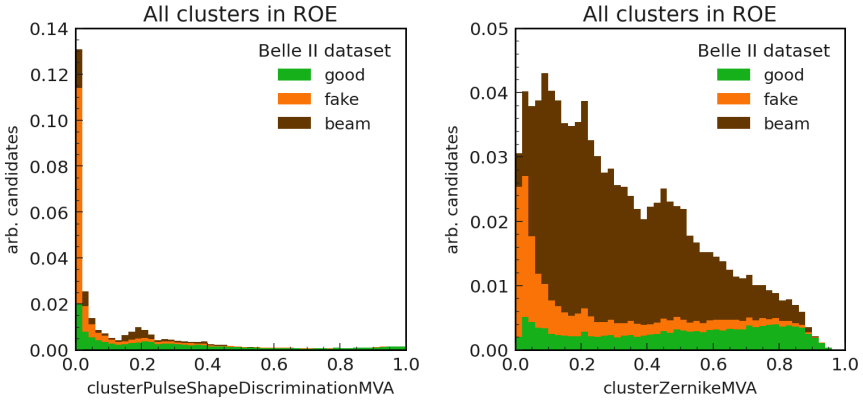


FIGURE 5.26: The distributions of clusterPulseShapeDiscriminationMVA and clusterZernikeMVA of the photons in the ROE in the Signal MC of Belle II. The plot is normalized, to show the relative composition of different types of photons at this level.

These MVA-based variables are not available for Belle and cannot be easily reproduced. Therefore, a set of less powerful features are used as input features for BDTs, shown in Fig. 5.28:

- **clusterNHits**: the sum of the weights w_i ($w_i \leq 1$) of all crystals in an ECL cluster. For non-overlapping clusters, this is equal to the number of crystals in the cluster. In the case of energy splitting between adjacent clusters, this can

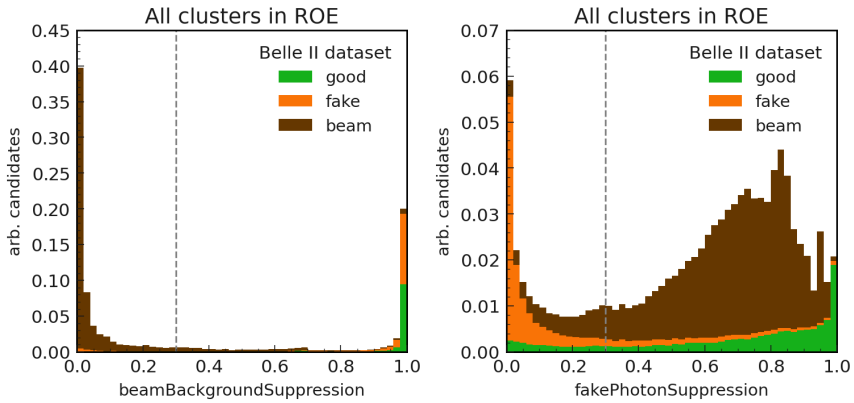


FIGURE 5.27: Outputs of the BDTs trained to classify beam background and fake photons clusters in Belle II of the photons in the ROE in the Signal MC of Belle II. The cuts applied are shown as dashed line. The plot is normalized, to show the relative composition of different types of photons at this level.

be a non-integer value. As can be seen from Fig. 5.28, the background photons tend to have a slightly lower number of hits than the good photons.

- **clusterPhi**: the azimuthal angle ϕ of the ECL cluster, calculated with respect to the centroid position of the cluster. This value does not necessarily coincide with the photon azimuthal angle, since the centroids are generally biased towards the center of the highest-energy crystal. This effect is larger for low energy photons, so the position resolution for beam background photons is poor.

- **clusterE9E25**: the ratio of energies in the inner 3×3 crystals, $E9$, and 5×5 crystals around the central crystal, $E25$. Since $E9 \leq E25$, this ratio is less than 1 and tends to larger values for photons and smaller values for hadrons, describing the shower shape. As shown in Fig. 5.28, the fake photons tend to have lower values than the signal and the beam background.

- **clusterHighestE**: the energy of the highest energy crystal in the ECL cluster. Similar to clusterE, the beam background and the fake photons tend to have lower clusterHighestE values. Usually, the crystal with the highest energy deposit is the center of the cluster, i.e., E1, so clusterHighestE/clusterE can also serve as an alternative to clusterE1E9, which is not directly available to Belle.

- **clusterLAT**: the lateral energy distribution of the cluster defined as:

$$S = \frac{\sum_{i=2}^{n} w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1)\, r_0^2 + \sum_{i=2}^{n} w_i E_i r_i^2}$$

where $E_i = (E_0, E_1, \ldots)$ are the sorted single-crystal energies in descending order; w_i is the crystal weight; r_i is the distance of the i-th digit to the shower centre, projected on a plane perpendicular to the shower axis; $r_0 \approx 5$ cm is the distance between two crystals. clusterLAT has a range of 0 - 6 and peaks at about 4 for radially symmetric electromagnetic showers and is larger for hadronic events and electrons with a close-by radiative or Bremsstrahlung photon.

For our analysis, the BDT trainings for Belle MC are performed locally with FastBDT, and the corresponding outputs with the recommended cuts are shown in Fig. 5.29. In both BDTs, clusterE and clusterHighestE have the greatest feature

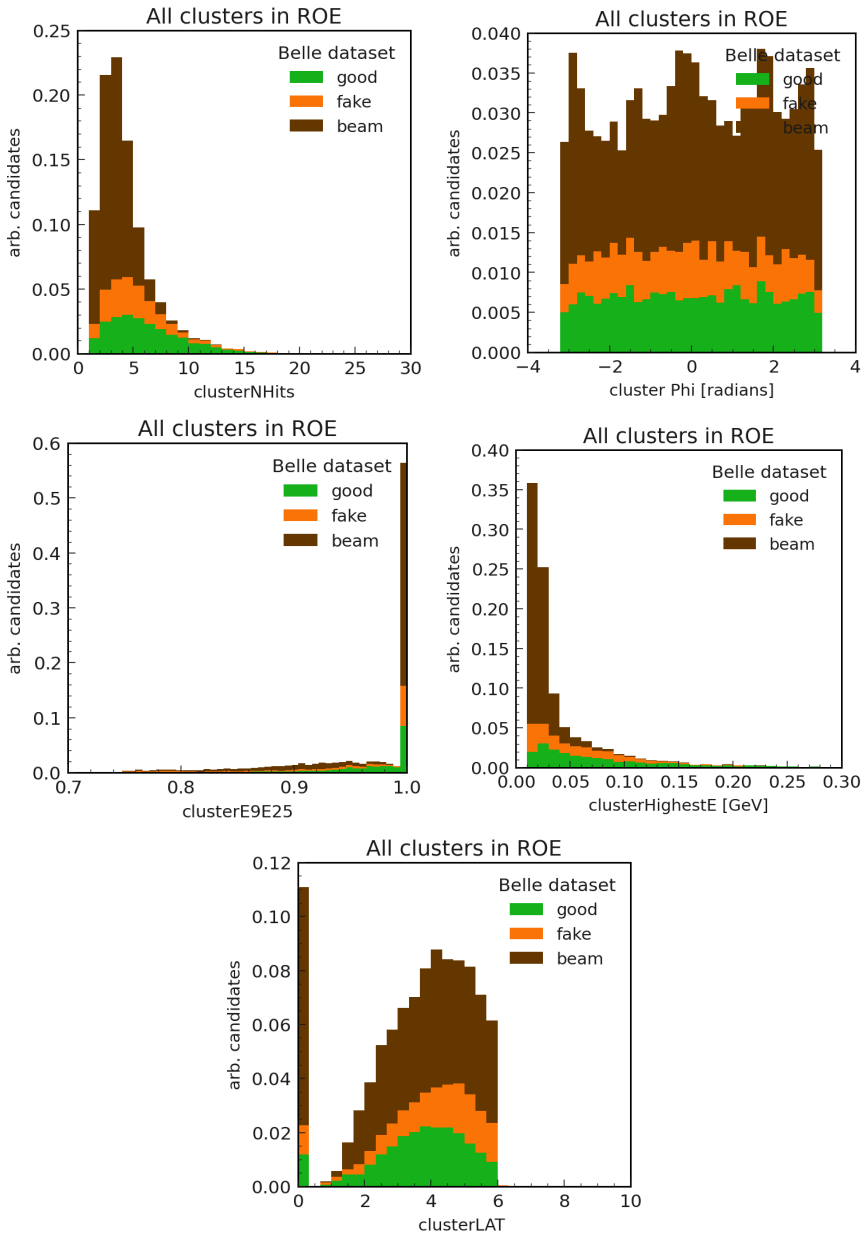


FIGURE 5.28: The distributions of the cluster shower-shape based features of photons in the ROE in the Signal MC of Belle, which are used for BDT training. The plot is normalized, to show the relative composition of different types of photons at this level.

importance, followed by minC2TDist for suppressing fake photons and clusterNHits for suppressing the beam background.

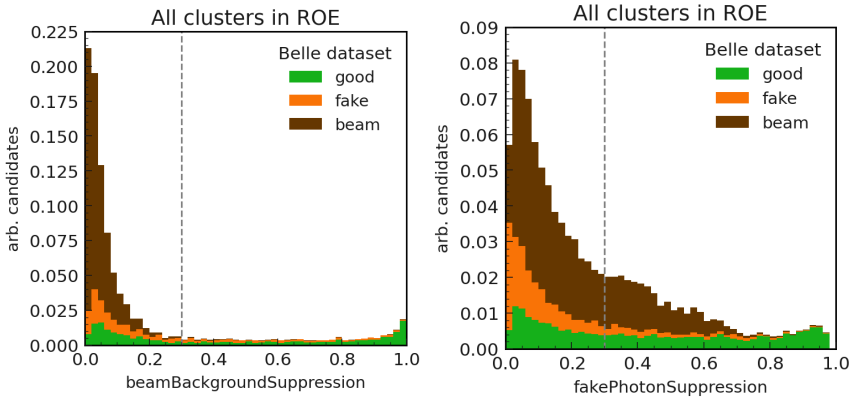


FIGURE 5.29: Outputs of the BDTs trained to classify beam background and fake photons clusters in Belle of the photons in the ROE in the Signal MC of Belle. The cuts applied are shown as dashed line. The plot is normalized, to show the relative composition of different types of photons at this level.

5.4.2 Choosing the ROE mask

Using all the variables described in the previous section, various combinations of cuts can be applied to the clusters to clean up the E_{ECL} distribution. The masks examined here, in order of complexity, are as follows:

- Mask 0: no mask
- Mask 1: goodGamma
- Mask 2: goodGamma + minC2TDist
- Mask 3: goodGamma + beamBackgroundSuppressionBDT + minC2TDist
- Mask 4: goodGamma + beamBackgroundSuppressionBDT + fakePhotonSuppressionBDT

Here, goodGamma refers to the fact that the cluster energy must be greater than 50 MeV in the barrel and 100, 150 MeV in the forward and backward endcaps respectively; minC2TDist cut refers to $>$ 30 cm; and beamBackgroundSuppressionBDT and fakePhotonSuppressionBDT must be greater than 0.3.

The E_{ECL} distribution after background suppression shown in Fig. 5.20 has no requirement on the clusters and therefore corresponds to mask 0. Fig. 5.30 shows the evolution of the E_{ECL} distribution for the same events with progressively tighter masks. Applying the goodGamma requirement (mask 1) would remove most of the beam background photons. In addition, fake photons can be removed by cutting to minC2TDist or the specific BDT. As expected, the resolution improves with a tighter mask, resulting in a signal MC more peaking at zero. When the E_{ECL} is dominated by "good" photons that are independent of the experimental conditions, the distributions of Belle and Belle II MC get closer, as with mask 4. However, this cleaning also shifts the background distribution toward lower E_{ECL} values, which can affect the separation of signal and background.

To choose the mask that provides the best possible resolution and at the same time the best possible signal-background separation, the expected upper limit (UL) is used as the figure-of-merit. For each mask, several signal windows are considered in E_{ECL} in the range $(0 - X)$ GeV, where X ranges from 0.05 GeV to 2 GeV in steps of 0.05 GeV. The signal efficiency (ϵ_s) and expected background are measured in each

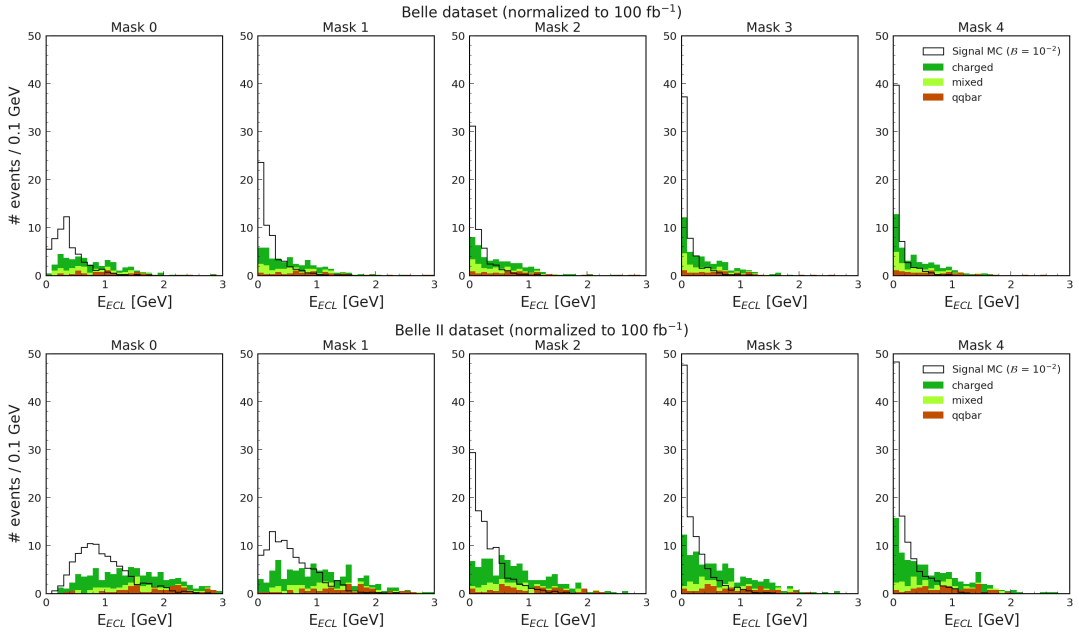


FIGURE 5.30: Distribution of the extra energy in calorimeter, with different masks in Belle (II) MC on top (bottom) after background suppression. The generic MC is normalized to 100 fb^{-1} while the signal MC is scaled to the assumption of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) = 10^{-2}$ for visual purposes.

of these signal windows, and the expected upper limits are calculated assuming that no signal is observed. The upper limit on branching fraction would be given by:

$$\mathcal{B}\left(B^+ \to K^+\tau^+\tau^-\right) < \frac{N^{UL}}{\epsilon_s \cdot 2 \cdot f^{\pm} \cdot N(B\overline{B})}$$

where $N(B\overline{B})$ is the number of recorded B meson pairs in the data set and $f^{\pm} = (51.4 \pm 0.6)\%$ [34] is the fraction of B^+B^-pairs. An upper limit on the observed candidates, N^{UL}, is derived with a certain confidence, assuming the absence of a signal by:

$$1 - \alpha = \sum_{k=0}^{n} P(k;\mu),$$

where $100(1-\alpha)\%$ is the chosen confidence level, μ is the unknown true expectation value, and P is the underlying probability density function. The implementation of the frequentist approach proposed by W. Rolke [125] in the ROOT [61] is used for this derivation. It uses a profile likelihood method with a Gaussian background model and a Poisson signal distribution:

$$L(\mu, b, \sigma \mid x, y) = \frac{(\mu+b)^x}{x!}\exp(-(\mu+b)) \times \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-b)^2}{2\sigma^2}\right),$$

where b is the expected background with uncertainty σ; x and y are the measured values for signal and background, respectively. The uncertainties on x and y are treated as nuisance parameters and modelled with a Gaussian probability density function.

The resulting UL estimates at 90% confidence level (CL) for the branching fraction as a function of the defined signal window are shown for all different mask choices in

Fig. 5.31 for Belle and Belle II data sets. It is obvious that the optimal signal window is different for each mask. In both data sets independently, the best upper limit is around $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 8 \times 10^{-4}$ within a (0 - 0.25) GeV signal window for tighter masks. Note that this is the estimate in MC, and the calibration factor of FEI efficiency must be applied to obtain the expected sensitivity in the data. Mask 4 appears to provide a good upper limit, while having a similar distribution in both data sets, making it easier to merge. Before doing so, however, it is essential to verify that the E_{ECL} distribution observed in MC with mask 4 is consistent with the data. To check this, the $B^+ \to J/\psi K^+$ control sample is used, as described in Sec. 5.5.2. Then the two data sets can be merged with appropriate calibration factors and the best possible upper limit in the data can be estimated.

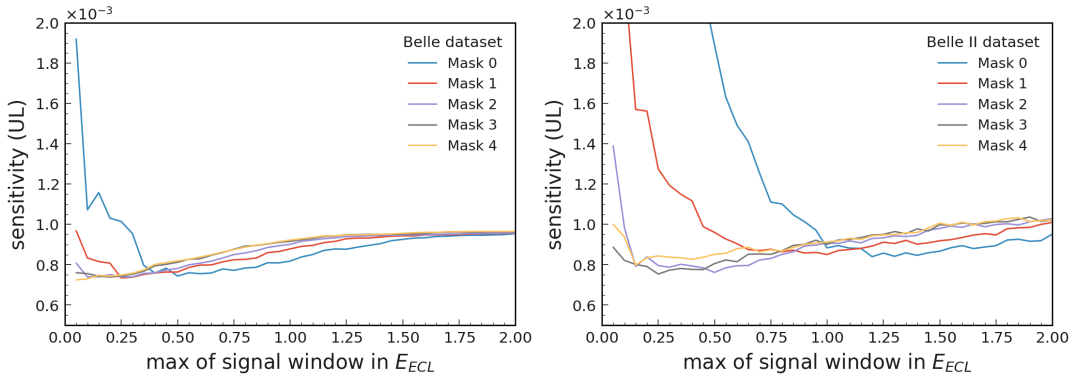


FIGURE 5.31: UL estimation at 90% CL as a function of the defined signal window for different mask choices in Belle (left) and Belle II (right) MC samples.

5.5 Control samples

In the previous section, it was shown that the distribution of signal and background components in E_{ECL} strongly depends on the mask applied to the clusters. Therefore, in order to estimate the UL sensitivity from the data, it is essential to confirm a good agreement exists between the data and MC for the E_{ECL} distribution with an appropriate mask. To check this, a pure sample without background $B^+ \to J/\psi K^+$ is used. As with the $B^+ \to K^+\tau^+\tau^-$, once the entire signal side and tag side are reconstructed, the E_{ECL} is expected to peak at zero, which is the region we are interested in.

The use of TC time in Belle is also validated. Ideally, the beam background photons must be rejected while the "good" photons are preserved. To verify that the selection does not reject "good" photons, a source of "good" photons is required. For this purpose, a tag-and-probe strategy with $B^+ \to J/\psi K^{*+}[\to K^+\pi^0]$ decays is used, where one of the photons from the π^0 decay can be constrained, and the effects of different cuts on the other photon can be studied.

5.5.1 $B^+ \to J/\psi K^{*+}$

To reconstruct the $B^+ \to J/\psi K^{*+}$ decays in the Belle data set, charged tracks in the final state are selected with the same requirements as those in Tab. 5.3. The J/ψ candidates are then reconstructed from lepton pairs within the $2.947(2.847) < M_{\mu^+\mu^-(e^+e^-)} < 3.177$ GeV/c^2 range. For the ee mode, a looser cut is applied on the lower side to account for possible energy losses due to Bremsstrahlung. The

constraints on π^0 and the corresponding K^{*+} are kept loose to allow for a larger photon background in the form of $0.08 < M_{\gamma\gamma} < 0.18$ GeV/c^2 and $0.8 < M_{K^+\pi^0} < 1.0$ GeV/c^2. However, to restrict the background from other B decay modes, the selection $M_{\rm bc} > 5.27$ GeV/c^2 and $|\Delta E| < 0.05$ GeV is applied.

For the events that pass through this selection, one of the photons from the π^0 is selected with the goodGamma requirement, and the effects of different cuts on the second photon are examined. If the second photon is a "good" photon, this would lead to a peak in the $M(\gamma\gamma)$ distribution at the π^0 mass. The cuts investigated for the second photon are: goodGamma; TC time cut and beamBackgroundPhotonSuppression BDT cut. Clusters produced at the collision time have TC time in the arbitrary range of 9000-11000, while the clusters from the beam background have TC time around 0, so TC cut > 0 is applied to reject beam background photons. A caveat is that the TC time information is only available in Belle data since experiment 31 and is not simulated in the Belle MC. To mimic the effect of the TC time cut in MC, all photons that match with the beam background are removed. And the corresponding TC cuts are applied only to events from experiments larger than 30, while goodGamma and beamBackgroundPhotonSuppression BDT cuts can be applied to all events. Fig. 5.32 shows the distributions of $M(\gamma\gamma)$ in the Belle data with different selection of the second photon. The reduction of the background in the sidebands of the π^0 peak is obvious with tighter cuts.

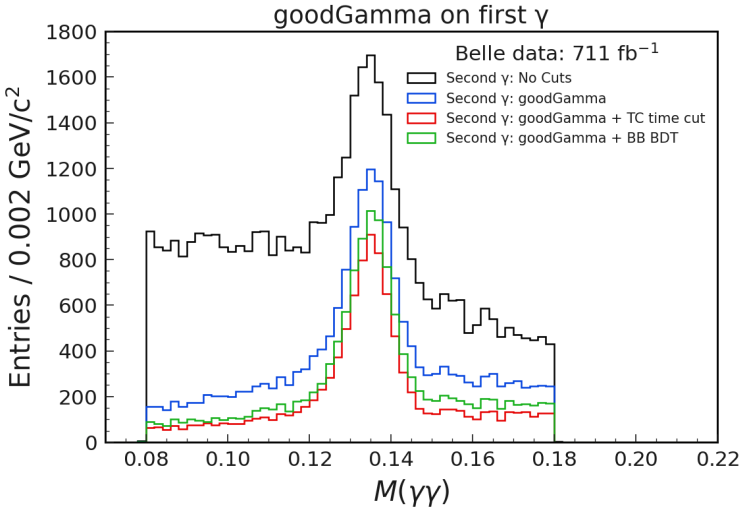


FIGURE 5.32: $M(\gamma\gamma)$ of the π^0 candidate in $B^+ \to J/\psi K^{*+}$ in Belle data where the first photon is constrained with goodGamma and different selections are applied on the second photon.

To quantify the effects on the "good" photons, the yield of the π^0 peak is estimated. The background yield is first estimated from the sideband, which is subtracted in the signal region ($0.12-0.15$ GeV/c^2) to obtain the signal yield. The ratio of the obtained signal yield in the data and MC is shown in Fig. 5.33 for different selections of the second photon in different bins of the cluster energy. The figure shows that the yield of "good" photons at low cluster energy decreases rapidly with the TC cut. This is due to the threshold of 0.1 GeV applied to TC to record the time, as described in Sec. 5.4.1. If no time is recorded, the TC time is zero by default and a "good" photon mimics a beam background photon and therefore removed. If a low energy photon hits more than one TC, the cluster may be formed, but it is possible that none of the TCs have crossed the threshold to store time. Therefore, the TC time information in Belle

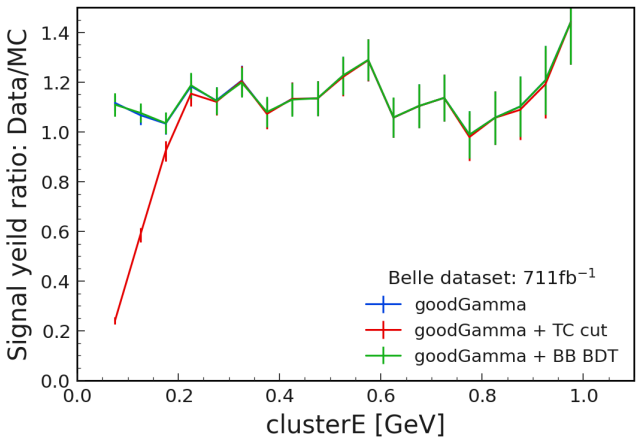


FIGURE 5.33: The ratio of signal yield of π^0 in data and MC in the $B^+ \to J/\psi K^{*+}$ control sample, with different selection on the photon, in bins of cluster energy of the photons.

cannot be used to distinguish beam background photons from "good" photons at low energy, which is where beam background photons are populated. It is also confirmed that the beamBackgroundPhotonSuppression BDT cut does not suffer from such a problem, the data/MC ratio is constant over the cluster energy and therefore used in the mask.

As SuperKEKB increases the beam currents to increase the instantaneous luminosity, the beam background of the Belle II experiment is expected to increase. Therefore, clusterTiming in Belle II is expected to be critical to control the E_{ECL} in the future, and thus is used as a feature in the beamBackgroundPhotonSuppression BDT of Belle II.

5.5.2 $B^+ \to J/\psi K^+$

To check the agreement of the E_{ECL} distribution in the signal window between data and MC, the control sample $B^+ \to J/\psi K^+$ is used. The charged track selection and the J/ψ reconstruction are the same as described in the previous section. The B_{sig} candidates are selected with $M_{\text{bc}} > 5.27$ GeV/c^2 and $|\Delta E| < 0.1$ GeV. On the other hand, the B_{tag} candidates are selected using the same selection as described in Sec. 5.2.1. After this selection, about 50 events with $B^+ \to J/\psi K^+$ candidates without background are retained, as verified in MC. For this study, a centrally generated $B^+ \to J/\psi K^+$ MC corresponding to 100 times the total integrated luminosity is used in Belle. In Belle II, $B^+ \to J/\psi K^+$ are skimmed from the charged (B^+B^-) MC: both run-dependent and run-independent are used for comparison. In all figures in this section, the FEI calibration factors from Tab. 4.11 are applied to the MC for comparison with the data.

The E_{ECL} distribution with mask 0 is shown in Fig. 5.34. In the case of Belle II, the agreement between the data and the RD MC is better than the RI MC. The difference between the two is that the beam background in the RD MC is retrieved from data and overlaid, while in the RI MC, it is simulated. This shows that the beam background in the calorimeter is overestimated in the RI MC. With a goodGamma selection on photons in mask 1, most of the beam background photons are removed, and the discrepancy between both MC would decrease, as shown in Fig. 5.35. Therefore, the RI MC can be used for the estimation, with mask 1 or tighter.

The E_{ECL} distribution of the selected mask 4 is shown in Fig. 5.36. Within the statistical uncertainty of this control sample, the agreement between data and MC is

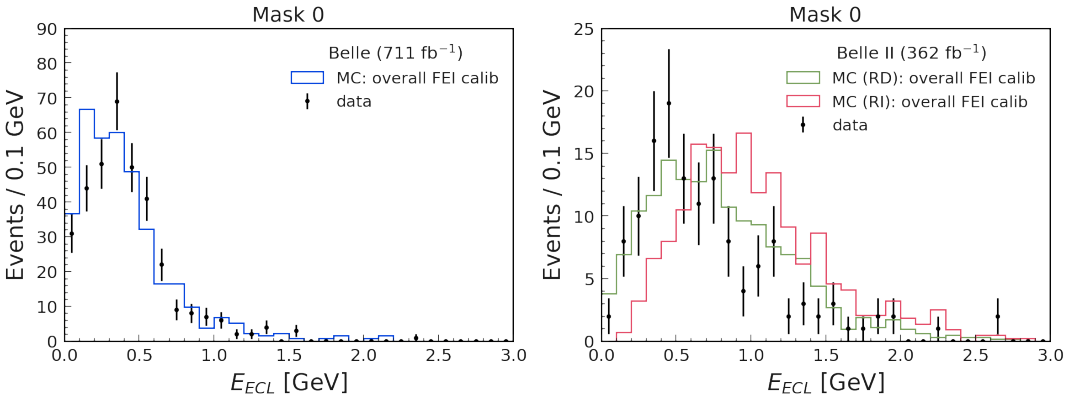


FIGURE 5.34: The distribution of E_{ECL} with mask 0 comparing data and MC in Belle (left) and Belle II (right) data sets. In Belle II, both RI and RD MC are used for comparison.

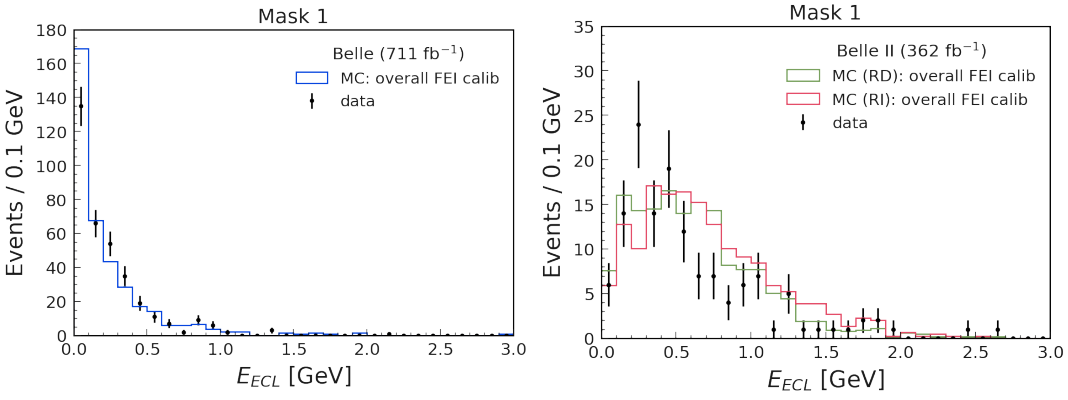


FIGURE 5.35: The distribution of E_{ECL} with mask 1 comparing data and MC in Belle (left) and Belle II (right) data sets. In Belle II, both RI and RD MC are used for comparison.

reasonable. Therefore, both data sets can be merged with appropriate FEI calibration factors, and the UL of $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ can be calculated as in data.

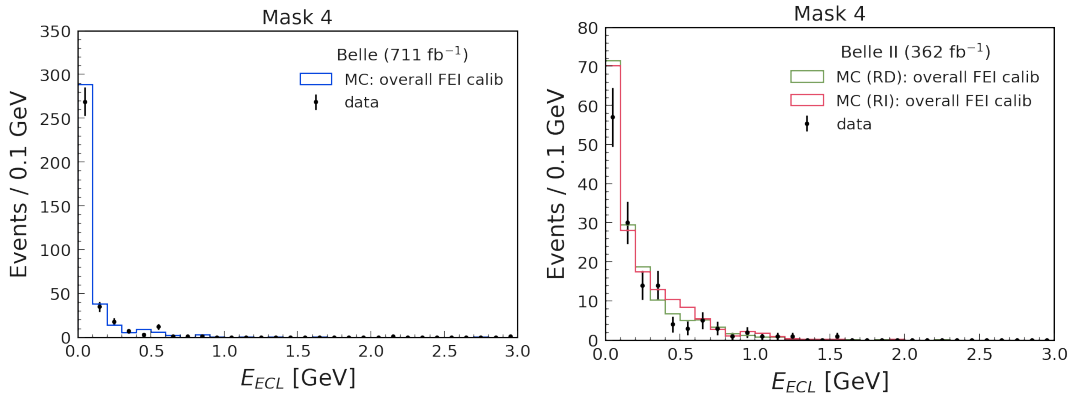


FIGURE 5.36: The distribution of E_{ECL} with mask 1 comparing data and MC in Belle (left) and Belle II (right) data sets. In Belle II, both RI and RD MC are used for comparison.

5.6 UL estimation on $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$

With mask 4 in the signal window of (0-0.25) GeV in the E_{ECL} distribution, the signal efficiencies in the Belle and Belle II data sets are 5.4×10^{-5} and 7.0×10^{-5}, respectively. While the Belle efficiency in MC is slightly underestimated and requires a FEI calibration factor of 1.04, the Belle II efficiency in MC is overestimated and requires a FEI calibration factor of 0.75, as listed in Tab. 4.11. Taking into account for this difference and the individual integrated luminosities, the expected signal efficiency in the merged data sets of both experiments is 5.5×10^{-5}, split into different final states, as shown in Tab. 5.9. The expected background yield in the signal window

	$K^+e^+e^-$	$K^+\mu^+\mu^-$	$K^+e^\pm\mu^\mp$
Signal efficiency (10^{-5})	1.6 ± 0.3	1.1 ± 0.2	2.8 ± 0.3
Background yields	82.2 ± 9.1	64.7 ± 8.0	153.2 ± 12.4

TABLE 5.9: Signal efficiency and expected background yields, in the merged Belle and Belle II data sets, in different final states. To estimate signal efficiency in data, FEI calibration factors are applied on MC. The background yields correspond to the total integrated luminosity of 1073 fb^{-1}.

in the individual data sets of Belle and Belle II is 27.7 and 28.5 events per 100 fb^{-1}, respectively. In the merged sample, the total expected background yield would be 300 events for the entire data set of 1073 fb^{-1}. This would result in an estimated upper limit of 6×10^{-4} at 90% confidence level for the $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ in the data.

The previous, and only existing, measurement by the BaBar experiment gave the upper limit of 2.25×10^{-3} at 90% confidence level with 424 fb^{-1} [1]. In contrast, the signal efficiency improved from 4.5×10^{-5} to 5.5×10^{-5}, and the expected background yield is significantly lower, from 36 events to 28 events per 100 fb^{-1}. This allows for an overall better upper limit, due in part to the larger data sample size and in part to the improved purity of the B_{tag} and signal side selection.

5.7 Prospective improvements

Although the upper limit is significantly improved compared to the earlier measurement, there is still room for improvement. Purity, especially from the tag side, can be improved by optimising the selection depending on the B_{tag} mode. Individual background suppression BDTs can be trained for different signal side final states. And the overall efficiency can be increased by including $\tau \to \pi$ decays.

5.7.1 B_{tag} dependence

When estimating the upper limit in data, the overall FEI calibration factors calculated for well-reconstructed B_{tag} candidates are applied to the signal efficiency. But a tag-side B decay mode dependent calibration factors, that are available and provided in Chapters 3 and 4, can be used for a more accurate estimate. The optimisation of the analysis can also be performed in a B_{tag} mode dependent way, for example, a mode dependent ΔE selection can be used, along with all the other potential improvements to B_{tag} described in chapters 3 and 4. In addition, only certain B decay modes that give the best signal-to-background ratio could be used in the reconstruction, rather than using all modes provided by the FEI package.

5.7.2 Including $\tau^+ \to \pi^+$ decays

Including the $\tau^+ \to \pi^+$ decays in the reconstruction of the signal side in the form of the $K^+\pi^\pm e^\mp$ and $K^+\pi^\pm\mu^\mp$ final states would increase the signal efficiency, but this would introduce contributions from $\tau^+ \to \rho^+$. The additional π^0 of these ρ^+ particles would contaminate the E_{ECL} distribution, increasing the tail and degrading the resolution, as shown in Fig. 5.37. Compared to purely leptonic final states (shown in yellow), the

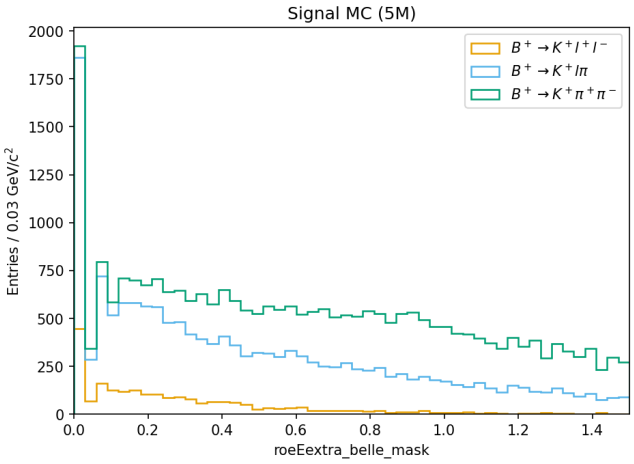


FIGURE 5.37: E_{ECL} distribution with mask 2 in Belle control sample split into purely leptonic final states (yellow), one pion (blue) and two pions (green) in the final state. The use of pion in the final state increases the efficiency, but degrades the resolution of E_{ECL}.

inclusion of a pion in the final state (shown in blue) significantly increases the E_{ECL} in the higher energy region of the signal MC itself. As shown using the control sample $B^+ \to J/\psi K^{*+}$ in Sec. 5.5.1, the γ cleaning variables summarized in Sec. 5.4.1 can optimize the π^0 reconstruction. This will help to separate $\tau \to \pi$ events from $\tau \to \rho$ events. Assuming the same efficiency and background yield as leptonic τ decays, the inclusion of $K^+\pi^\pm e^\mp$ and $K^+\pi^\pm\mu^\mp$ final states can extend the upper limit to 4×10^{-4} at 90% confidence level with a 1073 fb^{-1} data set.

Chapter 6

Conclusion and outlook

This thesis addressed the characterization of the performance of the hadronic B-tagging by FEI package, significant improvements for the same, and used these improvements to search for $B^+ \to K^+\tau^+\tau^-$ decays in Belle and Belle II.

Characterizing the performance of hadronic B-tagging

The main outcome of Chapter 3 is the measurement of the three metrics that describe the performance of hadronic B-tagging: efficiency, purity and calibration factor. Reconstruction of B_{tag} candidates by the FEI package is described, followed by the calibration procedure with the standard sample for Belle II, $B \to X\ell\nu$. The corresponding results are systematically dominated by the model of semileptonic decays of B mesons. As an alternative, a new $B \to D\pi$ sample is using the recoil-mass procedure, where the D mesons are not reconstructed. This sample allows for a simpler fit but is statistically limited, so the accuracy will improve as Belle II collects more data in the future. Calibration factors measured for individual B_{tag} decay modes using these two completely independent samples show reasonably good agreement. To account for the differences, they are combined, which is the current official method used in the latest Belle II measurements.

The overall calibration factors for B^+ tags are significantly different in the Belle sample ($0.753 \pm 0.012 \pm 0.025$) and in the Belle II sample ($0.651 \pm 0.015 \pm 0.025$), suggesting that the problems in MC modelling of the B decays are different in the two experiments. The high signal purity of the $D\pi$ sample allows us to investigate the sources of these discrepancies between the data and MC, which was the original intent of this sample. To do this, we first select the correctly reconstructed events from the signal side fit and then compare the decay kinematics of the B_{tag} candidates between data and MC. It is also shown that the efficiency and purity of hadronic B-tagging is different in the two data sets, indicating possible improvements.

Understanding and improving hadronic B-tagging

The hadronic B-tagging algorithm learns the selection from the decay kinematics of the particles in MC via BDTs. A reliable MC model is therefore essential, not only for a better calibration factor, but also for optimal performance of the FEI package. A general overview of the MC modelling of B decays in the Belle and Belle II experiments is given in Sec. 4.1, highlighting the differences, particularly in the way PYTHIA is configured and used to fill the gap of unknown hadronic decays. The poorly measured multi-body decays are modelled as a coherent sum of decays through many intermediate particles, with the relative fractions effectively describing the full decay kinematics.

Section 4.2 examines the MC model for the eleven decay modes that contribute most to the efficiency of the hadronic B-tagging, in both Belle and Belle II. The information is then compared to the most recent published measurements to uniformly correct and update it in MC for both experiments. This includes removing double counting caused by misinterpreting inclusive measurements as non-resonant contributions, and unintended PYTHIA contributions by explicitly setting them to zero. With these updates, a new MC is generated and validated using the sPlot technique in the $B \to D\pi$ sample. Overall, the new MC has a better calibration factor, improved from 0.75 to 1.04 for the Belle data set and from 0.65 to 0.83 for the Belle II data set, with most of the remaining discrepancy coming from the unmodified modes. Similar work is underway for the B^0-tagging.

When the FEI package is retrained with these new MC decay kinematics, the learned selections and consequently the background suppression capabilities improve, as shown in Section 4.3. This increases the purity of tag side in Belle II from 56.2% to 63.4%. By comparing the performance of the Belle and Belle II data sets, a tight preselection is identified that affects the reconstruction of low-energy photons in Belle II. By correcting this problem and introducing a mass constraint on π^0 in Belle II, the reconstruction efficiency of B^+ tags is increased from 0.93% to 1.13%, as described in Section 4.4. The improvements obtained will be useful for many searches for missing energy modes in Belle (II). In particular, the improved MC model will also serve as a better training sample for the inclusive B-tagging.

Better measurements are needed to further improve the hadronic B-tagging. The two modes $B^+ \to \overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$, which account for one-fifth of the total efficiency, are still dominated by PYTHIA, which is unpredictable and uncontrollable. Better understanding through future measurements: Identifying intermediate states and providing a simple model that can be implemented in EvtGen is essential for baryonic B decays used in B tagging, and also for adding additional modes such as $B^+ \to D^{*-}(4\pi)^{++}$ and $B^+ \to \overline{D}^{*0}(5\pi)^+$. For modes containing D^* mesons, an alternative and more efficient approach of partial reconstruction in the shifted ΔE region is proposed in Section 4.5.

$B^+ \to K^+\tau^+\tau^-$ search with Belle and Belle II data sets

In Chapter 5, a sensitivity study for the $B^+ \to K^+\tau^+\tau^-$ search using hadronic B-tagging is performed with the data sets from the Belle and Belle II experiments. All the improvements described in the previous chapters: new MC and latest FEI training, are used to obtain the best possible upper limit. On the signal side, the final states $\tau \to \ell\nu\overline{\nu}$ are chosen similarly to the previous search in the *BABAR* experiment. To suppress the background, which comes mainly from semileptonic B decays, a BDT with five variables is used as described in Section 5.3.

The extra energy in the calorimeter, E_{ECL}, calculated as the sum of the energy of the clusters not associated with any of the B candidates, is used to identify the signal. Any information that can be used to reduce contamination of the E_{ECL} by beam background photons and neutral clusters from hadronic split-off processes is described in Section 5.4. A common photon selection (mask 4) is chosen for both data sets, which mainly retains "good" photons for the E_{ECL} calculation. The data-MC agreement of the E_{ECL} distribution with this mask is validated using the $B^+ \to J/\psi K^+$ sample (see Section 5.5.2). In the combined data set of 1073 fb^{-1}, the signal efficiency after correction for the FEI calibration factors is 5.5×10^{-5}, with an expected background yield of 300 events. This results in an upper limit for $\mathcal{B}(B^+ \to K^+\tau^+\tau^-)$ of 6×10^{-4} at 90% confidence level.

Although the upper limit is significantly improved compared to the previous measurement, there is still room for improvement. Background suppression can be improved by optimising the selection depending on the tag-side decay mode as well as the signal side final state. Inclusion of the $K^+\pi^\pm e^\mp$ and $K^+\pi^\pm\mu^\mp$ final states should improve the upper limit to 4×10^{-4} at 90% confidence level.

Bibliography

[1] J. P. Lees et al., BaBar, *Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment*, Phys. Rev. Lett. **118** (2017) no. 3, 031802, arXiv:1605.09637 [hep-ex].

[2] Y. Nir, *Flavour physics and CP violation*, 2015. https://cds.cern.ch/record/2019737.

[3] Y. Grossman and P. Tanedo, *Just a Taste: Lectures on Flavor Physics*, in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics.*, pp. 109–295. 2018. arXiv:1711.03624 [hep-ph]. (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016.

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[5] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[6] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.

[7] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533.

[8] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[9] M. Tanabashi et al., Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (08, 2018) 030001. http://pdg.lbl.gov/2019/reviews/rpp2018-rev-ckm-matrix.pdf.

[10] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, Prog. Theor. Phys. **28** (1962) 870–880.

[11] Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, Phys. Rev. D **107** (2023) 052008, arXiv:2206.07501 [hep-ex].

[12] R. Aaij et al., LHCb, *Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$*, Phys. Rev. Lett. **120** (2018) no. 12, 121801, arXiv:1711.05623 [hep-ex].

[13] M. Bordone, G. Isidori, and A. Pattori, *On the standard model predictions for R_K and R_**, aug, 2016The European Physical Journal C **76** (aug, 2016) , https://doi.org/10.1140/epjc/s10052-016-4274-7.

[14] LHCb Collaboration, R. Aaij, et al., *Test of lepton universality in beauty-quark decays*, 2021. https://arxiv.org/abs/2103.11769.

[15] LHCb Collaboration, R. Aaij, et al., *Test of lepton universality with $B^0 \to K^{*0}\ell\ell$ decays*, Journal of High Energy Physics **2017** (aug, 2017) . https://doi.org/10.1007/JHEP08(2017)055.

[16] LHCb Collaboration, R. Aaij, et al., *Tests of lepton universality using $B^0 \to K^0_S\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays*, 2021. https://arxiv.org/abs/2110.09501.

[17] R. Aaij et al., LHCb, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, Phys. Rev. Lett. **131** (2023) no. 5, 051803, arXiv:2212.09152 [hep-ex].

[18] R. Aaij et al., LHCb, *Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays*, Phys. Rev. D **108** (2023) no. 3, 032002, arXiv:2212.09153 [hep-ex].

[19] A. Abdesselam et al., Belle, *Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle*, Phys. Rev. Lett. **126** (2021) no. 16, 161801, arXiv:1904.02440 [hep-ex].

[20] S. Choudhury et al., BELLE, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, JHEP **03** (2021) 105, arXiv:1908.01848 [hep-ex].

[21] T. Aaltonen et al., CDF, *Measurements of the Angular Distributions in the Decays $B \to K^{(*)}\mu^+\mu^-$ at CDF*, Phys. Rev. Lett. **108** (2012) 081807, arXiv:1108.0695 [hep-ex].

[22] B. Aubert et al., BaBar, *Angular Distributions in the Decays B —> K* l+ l-*, Phys. Rev. D **79** (2009) 031102, arXiv:0804.4412 [hep-ex].

[23] J. T. Wei et al., Belle, *Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. Lett. **103** (2009) 171801, arXiv:0904.0770 [hep-ex].

[24] R. Aaij et al., LHCb, *Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$*, JHEP **08** (2013) 131, arXiv:1304.6325 [hep-ex].

[25] S. Watanuki et al., Belle, *Search for the Lepton Flavor Violating Decays B+→K+τ±ℓ∓ (ℓ=e, μ) at Belle*, Phys. Rev. Lett. **130** (2023) no. 26, 261802, arXiv:2212.04128 [hep-ex].

[26] BaBar Collaboration, J. P. Lees, et al., *Search for the decay modes $B^\pm \to h^\pm\tau\ell$*, Phys. Rev. D **86** (Jul, 2012) 012004. https://link.aps.org/doi/10.1103/PhysRevD.86.012004.

[27] LHCb Collaboration, R. Aaij, et al., *Search for the lepton flavour violating decay $B^+ \to K^+\mu^-\tau^+$ using B^{*0}_{s2} decays*, Journal of High Energy Physics **2020** (2020) no. 6, 129. https://doi.org/10.1007/JHEP06(2020)129.

[28] R. Aaij et al., LHCb, *Search for the lepton-flavour violating decays $B^0 \to K^{*0}\tau^\pm\mu^\mp$*, JHEP **06** (2023) 143, arXiv:2209.09846 [hep-ex].

[29] C. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, HPQCD, *Standard Model Predictions for $B \to K\ell^+\ell^-$ with Form Factors from Lattice QCD*, Phys. Rev. Lett. **111** (2013) no. 16, 162002, arXiv:1306.0434 [hep-ph]. [Erratum: Phys.Rev.Lett. 112, 149902 (2014)].

[30] C. Cornella, D. A. Faroughy, J. Fuentes-Martín, G. Isidori, and M. Neubert, *Reading the footprints of the B-meson flavor anomalies*, Journal of High Energy Physics **2021** (2021) no. 8, 50. https://doi.org/10.1007/JHEP08(2021)050.

[31] A. Angelescu, D. Becirevic, D. A. Faroughy, F. Jaffredo, and O. Sumensari, *Single leptoquark solutions to the B-physics anomalies*, Phys. Rev. D **104** (2021) no. 5, 055017, arXiv:2103.12504 [hep-ph].

[32] I. Dorsner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Kosnik, *Physics of leptoquarks in precision experiments and at particle colliders*, *Phys. Rept.* **641** (2016) 1–68, arXiv : 1603.04993[hep − ph].

[33] R. Aaij et al., LHCb, *Observation of the decay* $\Lambda_b^0 \to \Lambda_c^+ \tau^- \overline{\nu}_\tau$, Phys. Rev. Lett. **128** (2022) no. 19, 191803, arXiv:2201.03497 [hep-ex].

[34] P. D. Group, P. A. Zyla, et al., *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (08, 2020) , https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf. https://doi.org/10.1093/ptep/ptaa104. 083C01.

[35] C. Alpigiani, A. Bevan, M. Bona, M. Ciuchini, D. Derkach, E. Franco, V. Lubicz, G. Martinelli, F. Parodi, M. Pierini, L. Silvestrini, V. Sordini, A. Stocchi, C. Tarantino, and V. Vagnoni, *Unitarity Triangle Analysis in the Standard Model and Beyond*, 2017. https://arxiv.org/abs/1710.09644.

[36] G. Aad et al., ATLAS, *Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Rev. Lett. **125** (2020) no. 5, 051801, arXiv:2002.12223 [hep-ex].

[37] A. Tumasyan et al., CMS, *Searches for additional Higgs bosons and for vector leptoquarks in* $\tau\tau$ *final states in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **07** (2023) 073, arXiv:2208.02717 [hep-ex].

[38] J. Aebischer, G. Isidori, M. Pesut, B. A. Stefanek, and F. Wilsch, *Confronting the vector leptoquark hypothesis with new low- and high-energy data*, Eur. Phys. J. C **83** (2023) no. 2, 153, arXiv:2210.13422 [hep-ph].

[39] W. Altmannshofer et al., Belle-II, *The Belle II Physics Book*, PTEP **2019** (2019) no. 12, 123C01, arXiv:1808.10567 [hep-ex]. [Erratum: PTEP 2020, 029201 (2020)].

[40] T. V. Dong et al., Belle, *Search for the decay B0→K*0τ+τ- at the Belle experiment*, Phys. Rev. D **108** (2023) no. 1, L011102, arXiv:2110.03871 [hep-ex].

[41] S. Kurokawa and E. Kikutani, *Overview of the KEKB accelerators*, Nucl. Instrum. Meth. A **499** (2003) 1–7.

[42] K. Akai, K. Furukawa, and H. Koiso, SuperKEKB, *SuperKEKB Collider*, Nucl. Instrum. Meth. A **907** (2018) 188–199, arXiv:1809.01958 [physics.acc-ph].

[43] Y. Ohnishi et al., *Accelerator design at SuperKEKB*, PTEP **2013** (2013) 03A011.

[44] K. Wille, *The physics of particle accelerators: An introduction*. 2000.

[45] KEKB Commissioning Group, Y. Funakoshi, *Operational experience with crab cavities at KEKB*, in *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, pp. 27–36. 2014. arXiv:1410.4036 [physics.acc-ph].

[46] J. Brodzicka et al., Belle, *Physics Achievements from the Belle Experiment*, PTEP **2012** (2012) 04D001, arXiv:1212.5342 [hep-ex].

[47] V. Aulchenko et al., *Electromagnetic calorimeter for Belle II*, J. Phys. Conf. Ser. **587** (2015) no. 1, 012045.

[48] Z. Natkaniec et al., *Status of the Belle silicon vertex detector*, Nucl. Instrum. Meth. A **560** (2006) 1–4.

[49] B. Wang et al., Belle-II DEPFET, PXD, *Operational experience of the Belle II pixel detector*, Nucl. Instrum. Meth. A **1032** (2022) 166631.

[50] A. Bevan et al., *The Physics of the B Factories*, The European Physical Journal C **74** (2014) no. 11, 3026.

[51] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, *Electron identification in Belle*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **485** (jun, 2002) 490–503. https://doi.org/10.1016/S0168-9002(01)02113-1.

[52] C. P. Group and T. B. I. Collaboration, *Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2022*, .

[53] A. Abashian et al., *Muon identification in the Belle experiment at KEKB*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **491** (2002) no. 1, 69–82. https://www.sciencedirect.com/science/article/pii/S0168900202011646.

[54] T. Koga and H. Nakazawa, *Trigger Summary in experiment 24,25,26*, .

[55] J. Rauch and T. Schlüter, *GENFIT — a Generic Track-Fitting Toolkit*, J. Phys. Conf. Ser. **608** (2015) no. 1, 012042, arXiv:1410.3698 [physics.ins-det].

[56] R. Itoh, *BASF - BELLE AnalysiS Framework*, in *9th International Conference on Computing in High-Energy and Nuclear Physics*. 4, 1997.

[57] A. Moll, *The Software Framework of the Belle II Experiment*, Journal of Physics: Conference Series **331** (dec, 2011) 032024. https://doi.org/10.1088/1742-6596/331/3/032024.

[58] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, *The Belle II Core Software*, Computing and Software for Big Science **3** (2018) no. 1, 1. https://doi.org/10.1007/s41781-018-0017-9.

[59] M. Gelb, T. Keck, M. Prim, H. Atmacan, J. Gemmler, R. Itoh, B. Kronenbitter, T. Kuhr, M. Lubej, F. Metzner, C. Park, S. Park, C. Pulvermacher, M. Ritter, and A. Zupanc, *B2BII: Data Conversion from Belle to Belle II*, Computing and Software for Big Science **2** (Nov, 2018) . http://dx.doi.org/10.1007/s41781-018-0016-x.

[60] N. Katayama, R. Itoh, A. Manabe, and T. Sasaki, *Belle computing model*, Computer Physics Communications **110** (1998) no. 1, 22–25. https://www.sciencedirect.com/science/article/pii/S0010465597001483.

[61] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86.

[62] T. Keck et al., *The Full Event Interpretation*, Computing and Software for Big Science **3** (2019) no. 1, 6. https://doi.org/10.1007/s41781-019-0021-8.

[63] A. Matyja et al., Belle, *Observation of B0 —> D*- tau+ nu(tau) decay at Belle*, Phys. Rev. Lett. **99** (2007) 191807, arXiv:0706.4429 [hep-ex].

[64] A. Bozek et al., Belle, *Observation of $B^+ \to \bar{D}^*0\tau^+\nu_\tau$ and Evidence for $B^+ \to \bar{D}^0\tau^+\nu_\tau$ at Belle*, Phys. Rev. D **82** (2010) 072005, arXiv:1005.2302 [hep-ex].

[65] F. Abudinén et al., Belle-II, *Search for B+→K+νν ¯ Decays Using an Inclusive Tagging Method at Belle II*, Phys. Rev. Lett. **127** (2021) no. 18, 181802, arXiv:2104.12624 [hep-ex].

[66] I. Adachi et al., Belle, *Evidence for $B^- \to \tau^-\bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle*, Phys. Rev. Lett. **110** (2013) no. 13, 131801, arXiv:1208.4678 [hep-ex].

[67] B. Kronenbitter et al., Belle, *Measurement of the branching fraction of $B^+ \to \tau^+\nu_\tau$ decays with the semileptonic tagging method*, Phys. Rev. D **92** (2015) no. 5, 051102, arXiv:1503.05613 [hep-ex].

[68] O. Lutz et al., Belle, *Search for $B \to h^{(*)}\nu\bar{\nu}$ with the full Belle $\Upsilon(4S)$ data sample*, Phys. Rev. D **87** (2013) no. 11, 111103, arXiv:1303.3719 [hep-ex].

[69] A. Heller et al., Belle, *Search for $B^+ß\ell^+\nu_\ell\gamma$ decays with hadronic tagging using the full Belle data sample*, Phys. Rev. D **91** (2015) no. 11, 112009, arXiv:1504.05831 [hep-ex].

[70] H. Atmacan et al., Belle, *Search for $B^0 \to \tau^\pm\ell^\mp$ ($\ell = e, \mu$) with a hadronic tagging method at Belle*, Phys. Rev. D **104** (2021) no. 9, L091105, arXiv:2108.11649 [hep-ex].

[71] B. Aubert et al., BaBar, *Measurement of the absolute branching fractions B to Dpi, D*pi,D**pi with a missing mass method*, Phys. Rev. D **74** (2006) 111102, arXiv:hep-ex/0609033.

[72] A. J. Bevan et al., BaBar, Belle, *The Physics of the B Factories*, Eur. Phys. J. C **74** (2014) 3026, arXiv:1406.6311 [hep-ex].

[73] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander, and A. Zupanc, *A Hierarchical NeuroBayes-based Algorithm for Full Reconstruction of B Mesons at B Factories*, Nucl. Instrum. Meth. A **654** (2011) 432–440, arXiv:1102.3876 [hep-ex].

[74] C. Pulvermacher, T. Keck, M. Feindt, M. Heck, and T. Kuhr, *An automated framework for hierarchical reconstruction of B mesons at the Belle II experiment*, Journal of Physics: Conference Series **608** (apr, 2015) 012048. https://dx.doi.org/10.1088/1742-6596/608/1/012048.

[75] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **559** (2006) no. 1, 190–194.

https://www.sciencedirect.com/science/article/pii/S0168900205022679. Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research.

[76] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) no. 1, 152–155. https://www.sciencedirect.com/science/article/pii/S0168900201000894. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[77] A. Sibidanov et al., Belle, *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) no. 3, 032005, arXiv:1306.2781 [hep-ex].

[78] T. Keck, *FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, CoRR **abs/1609.06119** (2016) , 1609.06119. http://arxiv.org/abs/1609.06119.

[79] C. Pulvermacher, *Analysis Software and Full Event Interpretation for the Belle II Experiment*. PhD thesis, 2015.

[80] A. Niculescu-Mizil and R. Caruana, *Predicting Good Probabilities with Supervised Learning*, in *Proceedings of the 22nd International Conference on Machine Learning*, ICML '05, p. 625–632. Association for Computing Machinery, New York, NY, USA, 2005. https://doi.org/10.1145/1102351.1102430.

[81] F. Abudinén et al., Belle II, *A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \to X \ell \nu$ decays*, arXiv:2008.06096 [hep-ex].

[82] F. U. Bernlochner, Z. Ligeti, and S. Turczyk, *A Proposal to solve some puzzles in semileptonic B decays*, Phys. Rev. D **85** (2012) 094033, arXiv:1202.1834 [hep-ph].

[83] W. Sutcliffe and F. Bernlochner, *Calibration of hadronic tagging with $B \to X \ell \nu$ decays with the LS1 Dataset*, .

[84] P. Colangelo and F. De Fazio, *Tension in the inclusive versus exclusive determinations of $|V_{cb}|$: A possible role of new physics*, Phys. Rev. D **95** (Jan, 2017) 011701. https://link.aps.org/doi/10.1103/PhysRevD.95.011701.

[85] F. Abudinén et al., Belle-II, *Measurement of lepton mass squared moments in $B \to X c \ell \nu_e \ell \ell$ decays with the Belle II experiment*, Phys. Rev. D **107** (2023) no. 7, 072002, arXiv:2205.06372 [hep-ex].

[86] J. D. Bjorken and S. J. Brodsky, *Statistical Model for electron-Positron Annihilation Into Hadrons*, Phys. Rev. D **1** (1970) 1416–1420.

[87] J. E. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'*, other thesis, 8, 1982.

[88] T. Ferber and P. Urquijo, *Overview of the Belle II Physics Generators*, .

[89] E. Barberio and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections. Version 2.0*, Comput. Phys. Commun. **79** (1994) 291–308.

[90] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[91] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, arXiv:2203.11601 [hep-ph].

[92] X. Zhou, S. Du, G. Li, and C. Shen, *TopoAna: A generic tool for the event type analysis of inclusive Monte-Carlo samples in high energy physics experiments*, Comput. Phys. Commun. **258** (2021) 107540, arXiv:2001.04016 [hep-ex].

[93] R. A. Briere and E. R. Oxford, *Update of Charm Decay Tables*, .

[94] F. Abudinén et al., Belle, Belle-II, *Combined analysis of Belle and Belle II data to determine the CKM angle ϕ_3 using $B^+ \to D(K^0_S h^- h^+)h^+$ decays*, JHEP **02** (2022) 063, arXiv:2110.12125 [hep-ex]. [Erratum: JHEP 12, 034 (2022)].

[95] E. Rodrigues and H. Schreiner, *DecayLanguage*, https://doi.org/10.5281/zenodo.8102036, June, 2023.

[96] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A **555** (2005) 356–369, arXiv:physics/0402083.

[97] M. Marinangeli, J. Eschle, E. Rodrigues, A. Desai, H. Schreiner, and B. Pollack, *scikit-hep/hepstats: Python 3.11 support*, https://doi.org/10.5281/zenodo.8070070, June, 2023.

[98] Belle Collaboration, K. Abe, et al., *Study of $B^- \to D^{**0}\pi^-$ ($D^{**0} \to D^{(*)+}\pi^-$) decays*, Phys. Rev. D **69** (2004) 112002, arXiv:hep-ex/0307021.

[99] BaBar Collaboration, B. Aubert, et al., *Dalitz Plot Analysis of $B^- \to D^+\pi^-\pi^-$*, Phys. Rev. D **79** (2009) 112004, arXiv:0901.1291 [hep-ex].

[100] R. Aaij et al., LHCb, *Amplitude analysis of $B^- \to D^+\pi^-\pi^-$ decays*, Phys. Rev. D **94** (2016) no. 7, 072001, arXiv:1608.01289 [hep-ex].

[101] R. Aaij et al., LHCb, *Determination of quantum numbers for several excited charmed mesons observed in $B^- \to D^{*+}\pi^-\pi^-$ decays*, Phys. Rev. D **101** (2020) no. 3, 032005, arXiv:1911.05957 [hep-ex].

[102] M. S. Alam et al., CLEO, *Exclusive hadronic B decays to charm and charmonium final states*, Phys. Rev. D **50** (1994) 43–68, arXiv:hep-ph/9403295.

[103] S. E. Csorna et al., CLEO, *Measurements of the branching fractions and helicity amplitudes in $B \to D^*\rho$ decays*, Phys. Rev. D **67** (2003) 112002, arXiv:hep-ex/0301028.

[104] T. Huber, S. Kränkl, and X.-Q. Li, *Two-body non-leptonic heavy-to-heavy decays at NNLO in QCD factorization*, JHEP **09** (2016) 112, arXiv:1606.02888 [hep-ph].

[105] D. Bortoletto et al., CLEO, *Inclusive and exclusive decays of B mesons to final states including charm and charmonium mesons*, Phys. Rev. D **45** (1992) 21–35.

[106] R. Aaij et al., LHCb, *Measurements of the Branching fractions for $B_{(s)} \to D_{(s)}\pi\pi\pi$ and $\Lambda_b^0 \to \Lambda_c^+\pi\pi\pi$*, Phys. Rev. D **84** (2011) 092001, arXiv:1109.6831 [hep-ex]. [Erratum: Phys.Rev.D 85, 039904 (2012)].

[107] G. Majumder et al., Belle, *Observation of $B^0 \to D^{*-}(5\pi)^+$, $B^+ \to D^{*-}(4\pi)^{++}$ and $B^+ \to \bar{D}^{*0}(5\pi)^+$*, Phys. Rev. D **70** (2004) 111103, arXiv:hep-ex/0409008.

[108] H. Albrecht et al., ARGUS, *Exclusive Hadronic Decays of B Mesons*, Z. Phys. C **48** (1990) 543–552.

[109] J. P. Alexander et al., CLEO, *First observation of* $\overline{B} \to D^{(*)}\rho'^{-}, \rho'^{-} \to \omega\pi^{-}$, Phys. Rev. D **64** (2001) 092001, arXiv:hep-ex/0103021.

[110] S. A. Dytman et al., CLEO, *Measurement of exclusive B decays to final states containing a charmed baryon*, Phys. Rev. D **66** (2002) 091101, arXiv:hep-ex/0208006.

[111] X. Fu et al., CLEO, *Observation of exclusive B decays to final states containing a charmed baryon*, Phys. Rev. Lett. **79** (1997) 3125–3129.

[112] N. Gabyshev et al., Belle, *Study of decay mechanisms in B- —> Lambda/c+ anti-p pi- decays and observation of low-mass structure in the Lambda/c+ anti-p system*, Phys. Rev. Lett. **97** (2006) 242001, arXiv:hep-ex/0409005.

[113] B. Aubert et al., BaBar, *Measurements of B(anti-B0 —> Lambda(c)+ anti-p) and B(B- —> Lambda(c)+ anti-p pi-) and Studies of Lambda(c)+ pi- Resonances*, Phys. Rev. D **78** (2008) 112003, arXiv:0807.4974 [hep-ex].

[114] F. Krinner, A. Lenz, and T. Rauh, *The inclusive decay* $b \to c\bar{c}s$ *revisited*, Nucl. Phys. B **876** (2013) 31–54, arXiv:1305.5390 [hep-ph].

[115] V. Bertacchi and K. Trabelsi, *Measurement of the* $B \to D^{(*)}K^{-}K^{(*)0}$ *branching fractions using the 2019–2022 Belle II data sample*, .

[116] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays* $B \to (K, K^{*})\ell^{+}\ell^{-}$ *in standard model and supersymmetric theories*, Phys. Rev. D **61** (2000) 074024, arXiv:hep-ph/9910221.

[117] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, *GEANT3*, .

[118] S. Agostinelli et al., GEANT4, *GEANT4–a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250–303.

[119] G. C. Fox and S. Wolfram, *Event Shapes in e+ e- Annihilation*, Nucl. Phys. B **149** (1979) 413. [Erratum: Nucl.Phys.B 157, 543 (1979)].

[120] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, Computing and Software for Big Science **1** (2017) no. 1, 2. https://doi.org/10.1007/s41781-017-0002-8.

[121] G. Punzi, *Sensitivity of searches for new signals and its optimization*, 2003. https://arxiv.org/abs/physics/0308063.

[122] S. Longo et al., *CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders*, Nucl. Instrum. Meth. A **982** (2020) 164562, arXiv:2007.09642 [physics.ins-det].

[123] A. Hershenhorn, T. Ferber, and C. Hearty, *ECL shower shape variables based on Zernike moments*, .

[124] P. Cheema, R. Cheaib, B. Yabsley, C.-L. Hsu, and S. Longo, *Development of FastBDT Classifiers to Suppress Beam Background Clusters and Fake Photons*, .

[125] W. A. Rolke, A. M. Lopez, and J. Conrad, *Limits and confidence intervals in the presence of nuisance parameters*, Nucl. Instrum. Meth. A **551** (2005) 493–503, arXiv:physics/0403059.