Search for the Lepton Flavor Violating B+ → K+ τ (e, μ) decays with the Belle and Belle II detectors

Gaetano de Marino

université
PARIS-SACLAY

Search for the Lepton Flavor Violating $B^+ \to K^+\tau\,(e,\mu)$ decays with the Belle and Belle II detectors

Recherche des désintégrations $B^+ \to K^+\tau\,(e,\mu)$ violant la saveur leptonique auprès des détecteurs Belle et Belle II

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576, Particules, Hadrons, Énergie, 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 l'unité de recherche : **IJCLab (Université Paris-Saclay, CNRS)**, sous la direction de **Karim TRABELSI**, directeur de recherche

Thèse soutenue à Paris-Saclay, le 09 décembre 2022, par

Gaetano DE MARINO

Composition du jury

Membres du jury avec voix délibérative

Sébastien DESCOTES-GENON Directeur de recherche, IJCLab	Président
Tim GERSHON Professeur, University of Warwick	Rapporteur & Examinateur
Stéphane MONTEIL Professeur, Université Clermont Auvergne	Rapporteur & Examinateur
Diego TONELLI Chargé de recherche, INFN Trieste	Examinateur

THESE DE DOCTORAT
NNT : 2022UPASP144

université PARIS-SACLAY

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

Titre : Recherche des désintégrations $B^+ \to K^+\tau\,(e,\mu)$ violant la saveur leptonique auprès des détecteurs Belle et Belle II

Mots clés : Physique du méson B, Violation Saveur Leptonique, Nouvelle physique, Belle, Belle II

Résumé : De nombreuses déviations par rapport au modèle standard ont été observées dans les désintégrations semileptoniques des mésons B, comme par exemple la violation de l'universalité de la saveur leptonique dans les transitions $b \to s\ell\ell$ et $b \to c\tau\nu$. Beaucoup de modèles de Nouvelle Physique (NP) prédisent également une violation de la saveur leptonique (LFV). Dans ce contexte, nous recherchons les quatre modes de désintégration $B^\pm \to K^\pm\tau^\pm(e,\mu)^\mp$ avec l'échantillon de données collectées par l'expérience Belle. Cette analyse n'a jamais été effectuée à Belle ni à Belle II, alors que des limites supérieures sur les taux d'embranchement de ces modes ont déjà été obtenues par les expériences BaBar et LHCb. Pour les modes à énergie manquante – comme c'est le cas dans nos canaux - l'étiquetage du B est couramment exploité dans les usines à B pour réduire le bruit de fond. Nous utilisons la reconstruction des mésons B hadroniques fournie par l'algorithme FEI, le package officiel d'étiquetage de Belle II. La reconstruction d'un méson B nous permet de déduire les propriétés de l'autre B dans l'événement et donc de calculer le M_{recoil}, utilisé pour extraire le signal. En absence de signal, nous obtenons les limites sur les rapports d'embranchement des modes $B \to K\tau\ell$, avec l'ensemble des données de Belle, de l'ordre de quelques 10^{-5}, qui sont les plus contraignantes à ce jour.

Améliorer l'étiquetage hadronique du B est un élément clé pour augmenter la sensibilité expérimentale des futures analyses de ces modes. La correction de la description des désintégrations de B^+ dans la simulation Monte Carlo (MC) pourrait le permettre. En effet, FEI utilise des techniques de *machine learning* pour séparer efficacement le signal du bruit de fond. Toutefois, ces méthodes étant entraînées sur le MC, elles peuvent, en cas d'erreur, conduire à des performances FEI non optimales, engendrant une différence importante d'efficacité entre les données et le MC. Nous révisons la simulation MC en nous concentrant sur les modes les plus pertinents pour l'étiquetage des mésons B^+ et réduisons de manière significative l'écart observé. Les autres directions sont l'ajout de nouveaux modes de désintégration et la récupération des candidats partiellement reconstruits.

Cette thèse explore aussi la possibilité d'utiliser une approche d'étiquetage semileptonique. Dans ce contexte, l'impulsion du méson B ne peut plus être mesurée, augmentant approximativement la résolution sur M_{recoil} d'un facteur 5. Cependant, l'approche d'étiquetage semileptonique offre une grande efficacité de reconstruction en raison des larges rapports d'embranchement des désintégrations $B \to D^{(*)}\ell\nu$, nous étudions comment les différentes conditions en termes de résolution et de composition du fond ont un impact sur la sensibilité finale. Des contraintes supplémentaires peuvent être imposées en se basant, par exemple, sur la connaissance de la cinématique de l'événement. L'utilisation de cette information seule améliore déjà la résolution de M_{recoil}, en particulier pour les désintégrations hadroniques du τ. Nous essayons également d'exploiter l'information du *vertexing* et explorons différents scénarios de résolution avec le détecteur Belle II et l'accélérateur SuperKEKB. Dès lors, l'étiquetage semileptonique devient compétitif par rapport à l'étiquetage hadronique. Ce résultat est encourageant car il permet l'exploitation d'un échantillon de données indépendant et ouvre d'autres perspectives pour les futurs échantillons collectés, sachant que Belle II a deja accumulé 0.6 fois les données de Belle. En supposant que les anomalies ne soient pas bientôt réfutées, un échantillon de données plus important combiné à une stratégie optimisée d'étiquetage du B, pourrait conduire à l'observation de LFV ou à contraindre fortement les modèles de NP qui leur sont associés.

Title : Search for the Lepton Flavor Violating $B^+ \to K^+\tau\,(e,\mu)$ decays with the Belle and Belle II detectors

Keywords : B physics, Lepton Flavor Violation, New Physics, Belle, Belle II

Abstract : For over a decade, many deviations from the Standard Model have been observed in semileptonic B-meson decays, for example, the departure from the lepton flavor universality in $b \to s\ell\ell$ and $b \to c\tau\nu$ transitions. Many New Physics (NP) models trying to explain these results also predict a violation of the leptonic flavor (LFV). In this context, we are searching for the four $B^\pm \to K^\pm\tau^\pm\ell^\mp$ ($\ell = \{e,\mu\}$) decay modes with the data sample collected by the Belle experiment. This analysis has never been performed at Belle nor Belle II, while some upper limits on the rates of such modes were previously set by the BaBar and LHCb experiments. For modes with missing energy – as it is for our channels – B-tagging is commonly exploited at B-factories to reduce the background. We use the hadronic B-meson reconstruction provided by the FEI algorithm, the official exclusive tagging package at Belle II. The reconstruction of one B-meson allows us to infer the properties of the other one and hence compute the M_{recoil}, which is the observable used to extract the signal yield. In the absence of signal, we derive the upper limits on the branching ratios of $B \to K\tau\ell$ modes, obtained with hadronic FEI and the full Belle dataset. The limits are of the order of a few 10^{-5} and are the most stringent to date.

A key point to improve the experimental sensitivity with the available and future Belle II data is to boost the B-tagging performance and obtain higher efficiency. The first step in this direction consists in improving the description of B^+ decays in the Belle Monte Carlo (MC) simulation. In fact, FEI uses machine-learning techniques to efficiently separate the signal from backgrounds. However, such methods are trained on MC, which, if incorrect, can lead to non-optimal FEI performance. One of the consequences is the large discrepancy between data and MC for the FEI tagging efficiency. We revise the MC simulation by focusing on the most relevant modes and significantly reduce the seen discrepancy. Other directions towards a higher efficiency are proposed in this manuscript; they include adding some new decay modes to FEI and recovering only partially reconstructed candidates.

This thesis explores the possibility of using a semileptonic tag approach for the $B^\pm \to K^\pm\tau\ell$ search. As opposed to the hadronic tag, the momentum of the B-meson cannot be measured; hence, the resolution on the M_{recoil} is roughly a factor 5 worse. However, the semileptonic tag approach provides a high reconstruction efficiency because of the large branching ratios of $B \to D^{(*)}\ell\nu$ decays; we study how the different conditions in terms of resolution and background composition impact the final sensitivity. Additional constraints can be imposed based on the knowledge of the event kinematics. The usage of this information-only already improves the resolution of M_{recoil}, especially for hadronic τ decays. We also try to exploit the vertexing information and study different resolution scenarios with the upgraded detector Belle II and accelerator SuperKEKB. Including those improvements, the semileptonic tag approach is competitive with fully reconstructed hadronic tags. This result is encouraging because it allows to exploit an independent data sample and opens other exciting prospects given the larger Belle II final dataset. The Belle II experiment has already accumulated a sample that is more than half that of Belle. Assuming that the anomalies are not disproved within a few years, the larger dataset combined with an optimized B-tagging strategy might lead to observing the LFV decays or strongly constraining the NP models associated with them.

Acknowledgements

This Ph.D. project has been a long and intense journey shared with many people I wish to thank.

Firstly, I would like to express my deepest gratitude to

Prof. Timothy John Gershon and Prof. Stéphane Monteil for accepting the role of *rapporteurs*,

Diego Tonelli and Sébastien Descotes-Genon, who kindly accepted to be part of my jury.

I thank them for the comments on the manuscript and the questions asked during the defense: they were all valuable and greatly tested my knowledge of the subject. It was an honor to interact with such experienced scientists.

I am profoundly grateful to my supervisor Karim Trabelsi for his tireless support during this three-year-long collaboration. His strong motivation has greatly inspired me, and the vast knowledge he shared with me is fundamental for my future work experiences.

I also had the pleasure to work with

Shun Watanuki, who introduced me to the $B \to K\tau\ell$ analysis and with whom I have cooperated very fruitfully over the last years.

Diego Guadagnoli and Chan Beom Park for the numerous exchanges, the hard work and for writing a paper together.

Olcyr Sumensari. *Grazie* for sharing notes and knowledge on the $b \to s\ell\ell'$ transitions.

I want to thank the CNRS for funding my thesis work and the Laboratoire de Physique des 2 Infinis Irène Joliot Curie and Université Paris-Saclay (PHENIICS ED) for the hospitality. I express my gratitude to the Belle II Collaboration for welcoming me into such a large and vibrant research community.

Sincere thanks go to all the Belle II (LAL/)IJCLab colleagues: Philippe Bambade, Valerio Bertacchi, Tak Shun Lau, Francois Le Diberder, Farah Mawas and Aurélien Martens.

Special acknowledgments should go to:

Vidya Sagar Vobbilisetti, I could not have undertaken this journey without his support and friendship. I wish him happiness and success in work and life.

Meihong Liu, my coolest colleague, who added some color and sweetness to my working day.

Pavel Oskin, for the pleasant conversations about physics and life through which I have greatly broadened my views.

Emi Kou, for being a kind colleague and neighbor with whom to share apéros and cooking lessons.

I would like to mention

My intern students, Jordan Gué, Romain Cadas and Slimane Mzerguat, for being an inexhaustible source of questions and pushing me to work harder.

Iolanda Matea, for the friendly chats and support.

Special thanks go to Paula and the Kleij-Cortés family: you have been my 'French' Mexican-Dutch family. Your support in Paris has been integral to my Ph.D. journey: you gave me a roof, fashion tips, fantastic food and invaluable friendship.

Merci P.A., pour m'avoir donné le plus de bonheur durant cette dernière année si difficile. Avec toi, j'ai exploré un peu le monde et moi-même aussi. À de nombreuses nouvelles aventures ensemble !

Infine voglio ringraziare la mia famiglia (compresi voi, nonni dal cielo!), coloro che mi hanno reso la persona che sono oggi, sostenuto e guidato in questo duro percorso. Il vostro amore è la mia linfa vitale.

Contents

Synthèse

Introduction

Depuis plus d'une décennie, de nombreuses déviations par rapport au Modèle Standard de la physique des particules (SM) ont été observées dans les désintégrations semileptoniques des mésons B. L'écart par rapport à l'universalité de la saveur des leptons (LFU) dans les transitions $b \to s\ell\ell$ (observables $R_{K^{(*)}}$, voir Eq. 1.6) et $b \to c\tau\nu$ (observables $R_{D^{(*)}}$, voir Eq. 1.7) en est un exemple. La LFU est une symétrie accidentelle du SM et implique que les trois familles de leptons aient les mêmes couplages avec les champs (hormis le champ de Higgs). Une des conséquences de la non-universalité leptonique est la violation de la saveur leptonique (LFV). En effet, même dans un SM étendu incluant des masses non nulles de neutrinos (m_ν) générées par le mécanisme de Higgs, la violation de la saveur leptonique chargée (cLVF) est un phénomène extrêmement rare, supprimé par les puissances de m_ν. La plupart des modèles proposés pour expliquer les anomalies [1–4] prédit des taux de désintégration considérablement accrus dans les processus comme $b \to s\ell\ell^{'}$ ($\ell \neq \ell'$), en particulier les modes $B \to K\tau\mu$ [4].

Parmi les scénarios au-delà du SM (BSM), celui incluant les leptoquarks (LQs) est l'un des plus reconnus. En effet, les LQs peuvent contribuer aux transitions semileptoniques au niveau arbre (Fig. 1.4), montrant ainsi des anomalies importantes, tout en entrant seulement au niveau boucle dans les interactions à quatre quarks ou à quatre leptons, dans lesquelles aucune déviation du SM n'a été trouvée jusqu'à présent [5]. De plus, les masses prédites pour les LQs sont $\sim\mathcal{O}$(TeV), bien au-dessus de l'échelle électrofaible, ceci étant compatible avec le fait qu'aucune nouvelle particule n'a été directement observée durant les expériences LHC depuis la confirmation du boson de Higgs.

Le scénario offert par le LQ vectoriel appelé U_1 a besoin d'un complément ultraviolet, ce qui signifie que le modèle générant les nouvelles particules doit être fourni. Ce problème peut être résolu dans des modèles où le groupe de jauge est brisé à l'échelle du TeV et où U_1 est un boson de jauge [6]. Dans un scénario à deux LQ, les deux scalaires S_1 (ou R_2) et S_3 peuvent fournir une explication aux anomalies B et être également compatibles avec les contraintes de $b \to s\nu\nu$ [7]. Plus de détails sur la façon dont le modèle de nouvelle physique et les contraintes expérimentales se traduisent en couplage réel sont donnés en annexe A. L'observation de la violation de la saveur des leptons dans le secteur des leptons chargés serait un signe univoque de physique au-delà du SM et fournirait des contraintes fortes aux modèles BSM.

Dans ce contexte, nous avons recherché les quatre modes de désintégration $B^\pm \to K^\pm\tau^\pm\ell^\mp$ ($\ell = \{e, \mu\}$), représentant chacun non seulement un ensemble spécifique de couplages pour l'interprétation théorique, mais aussi une nature de fond différente. Nous appelons OS_ℓ les modes où le lepton léger ℓ a la charge opposée par rapport au kaon et SS_ℓ les autres. Nous utilisons l'échantillon de données collecté par l'expérience Belle au collisionneur KEKB entre 1999 et 2010 et correspondant à 772 millions de paires de mésons B de saveur opposée, $B\overline{B}$. Chez les usines à B, les collisions e^+e^- se produisent à l'énergie du centre de masse de 10.58 GeV. À cette énergie, les processus hadroniques les plus abondants sont la production de paires de quarks

légers ($q\bar{q}$, $q = \{u, d, s, c\}$) pour une section efficace totale σ de 3.7 nb. La résonance $\Upsilon(4S)$ est également produite ($\sigma = 1.1$ nb) et se désintègre dans presque 100% des cas en paires de mésons $B\overline{B}$ (chargés ou neutres). La recherche de désintégrations de $B^{\pm} \to K^{\pm}\tau\ell$ n'a jamais été effectuée à Belle ni à Belle II, alors que certaines limites supérieures (UL) sur les rapports d'embranchements (BFs) avaient été fixées précédemment par les expériences *BABAR* et LHC*b* (de l'ordre de quelques 10^{-5}). Pour les désintégrations avec des particules échappant à la detection – les neutrinos provenant du τ dans les modes $K\tau\ell$ – l'étiquetage du B est couramment utilisé dans les usines à B.

Recherche des désintégrations $B^+ \to K^+\tau\ell$ dans l'expérience Belle avec étiquetage hadronique des mésons B

Nous utilisons la reconstruction du méson B hadronique fournie par l'algorithme `FEI` [8], le package d'étiquetage exclusif des mésons B à Belle II. Le *tag* B de l'événement (B_{tag}) est reconstruit dans un ensemble de modes hadroniques connus suivant une approche hiérarchique où les particules intermédiaires sont obtenues en combinant les particules de l'état final et ainsi de suite jusqu'au candidat B_{tag}. L'algorithme reconstruit ~ 35 modes B^+ hadroniques et environ 20 modes D^0, sélectionnant quelques milliers de chaînes de désintégration au total. Chaque étape fait référence à un classificateur BDT entraîné sur des événements simulés (Monte Carlo, MC). $\mathcal{P}_{\text{FEI}}$ est la variable que nous utilisons pour distinguer le signal du bruit de fond. La reconstruction du `FEI` via les modes hadroniques a une efficacité assez faible (environ 1%) mais supérieure à l'algorithme de Belle, appelé FR [9].

Après la reconstruction hadronique exclusive du méson B, nous sélectionnons deux traces comme étant le kaon (K_{sig}) et le lepton (ℓ_{sig}) de signal (ou *prompt*). Comme nous connaissons le quadri-moment initial de l'événement, et donc ceux du B_{tag}, du K_{sig} et du ℓ_{sig}, nous pouvons dériver la cinématique du τ et donc la *masse de recul* du signal B (B_{sig}) M_{recoil}[1] :

$$M_{\text{recoil}} = [m_B^2 + m_{K\ell}^2 - 2(E^*_{\text{beam}}E^*_{K\ell}/c^4 + p^*_{B_{\text{tag}}}p^*_{K\ell}\cos\theta/c^2)]^{\frac{1}{2}} \tag{A}$$

où θ est l'angle entre $\mathbf{p}^*_{B_{\text{tag}}}$ et $\mathbf{p}^*_{K\ell}$. Le remplacement $E^*_{B_{\text{tag}}} \to E^*_{\text{beam}}$[2] conduit à une meilleure résolution sur M_{recoil}, de l'ordre de 30 MeV. La distribution de M_{recoil} est utilisée pour extraire le nombre d'événements de signal (N_{sig}) : alors qu'ils ont un pic à la masse du τ, les événements de fond couvrent une région beaucoup plus large.

L'étape finale consiste à sélectionner une trace supplémentaire (t_τ), avec une charge opposée par rapport à ℓ_{sig}. Cette trace doit être compatible avec un pion, un muon ou un électron. Aucune autre trace provenant du point de collision (IP) n'est autorisée dans le reste de l'événement (ROE). À ce stade, la sélection couvre toutes les désintégrations du τ '1-*prong*' ($\sim 80\%$ de la largeur). La nature du fond obtenu dépend de la configuration de la charge. Dans les deux cas, les désintégrations $B^+ \to \overline{D}^0(\to K^+X^-)X^+$ favorisées par Cabibbo sont les sources dominantes du bruit de fond $B\overline{B}$, qui est réduit à l'aide d'une analyse multivariée sous forme de *boosted decision trees* (BDT). Pour les modes OS, les désintégrations semileptoniques D^0 imitent le côté signal, tandis que pour les modes SS, les désintégrations semileptoniques

[1] Toutes les observations marquées d'un astérisque sont mesurées dans le référentiel du centre de masse.

[2] E^*_{beam} est l'énergie de collision calibrée au centre de masse.

du B fournissent le lepton avec la bonne charge. La distribution $M(K^-X^+)$, définie comme suit :

$$\begin{aligned} OS:\ M(K^-X^+) &\equiv M(K^-_{\rm sig}\ell^+_{\rm sig}) \\ SS:\ M(K^-X^+) &\equiv M(K^-_{\rm sig}t^+_\tau) \end{aligned} \tag{1}$$

est très importante pour l'entraînement du BDT car le système KX provient principalement de désintégrations des mésons D pour le fond et est donc cinématiquement contraint. Nous appliquons également un veto sur une région étroite de $M(K^-X^+)$ afin d'éliminer les événements ayant $D^0 \to K^-\pi^+$.

Nous entraînons d'abord un BDT basé sur certaines caractéristiques utilisant les propriétés de $B^+ \to K^+\tau\ell$ contre les événements $B\overline{B}$ génériques. Celles-ci incluent notamment la variable $M(K^-X^+)$, les propriétés du ROE (par exemple l'énergie déposée dans le calorimètre), les variables liées à la topologie de désintégration de $B_{\rm sig}$ et la qualité du $B_{\rm tag}$. Bien que les variables soient choisies pour supprimer le fond $B\overline{B}$, une bonne séparation est également obtenue pour les événements $q\overline{q}$. Après une coupure sur le score du BDT, qui se traduit par une efficacité du signal à 40-50% et un rejet global du bruit de fond à 90-95%, la composante $q\overline{q}$ devient non négligeable et devons donc procéder à une suppression ultérieure. Pour cela, nous entraînons un second classificateur BDT et exploitons des variables communes pour la suppression du continuum en fonction de la forme de l'événement. En effet, les événements $B\overline{B}$ sont plus sphériques et les particules sont uniformément distribuées alors que les événements $q\overline{q}$ ont deux jets avec des impulsions larges. Les propriétés utilisées pour l'apprentissage ne montrent pas de séparation entre le signal et les événements $B\overline{B}$. En appliquant une coupure assez lâche sur la sortie du BDT (efficacité du signal : 80-90%, rejet du fond $q\overline{q}$: 90%) nous nous débarrassons efficacement du fond $q\overline{q}$ tout en n'améliorant que peu la signification finale, évaluée avec la figure de mérite de Punzi [10].

Afin de mesurer un rapport d'embranchement, nous utilisons l'efficacité du signal final obtenu sur MC, qui doit être corrigé pour les différences possibles par rapport aux données de collision. Nous exploitons différents échantillons de contrôle pour évaluer ces différences et en déduire les erreurs systématiques. Certains termes sont définis comme *multiplicatif* car ils affectent l'estimation du nombre total d'événements reconstruits. Les principales contributions et les procédures permettant de les estimer sont énumérées ci-dessous.

- Efficacité de la suppression $B\overline{B}$. Nous utilisons l'échantillon $B^+ \to D^-\pi^+\pi^+$ en raison de ses propriétés similaires à celles de $B^+ \to K^+\tau\ell$. Le méson D^- est reconstruit dans le canal $K^+\pi^-\pi^-$; un pion est traité comme t_τ tandis que le reste est considéré comme énergie manquante. Étant donné que l'état final $D^-\pi^+\pi^+$ est produit par des désintégrations $\overline{D}^{**0}\pi^+$, les deux pions prompts ont des gammes d'impulsions différentes. Par conséquent, les rôles de $K_{\rm sig}$ et de $\ell_{\rm sig}$ sont randomisés afin d'obtenir un $M(Kt_\tau)$ aussi proche que possible du signal. De bons accords entre le signal et le modèle MC ($\overline{D}^{**0}\pi^+$) et entre les données et le MC sont obtenus.

- Efficacité de la suppression $q\overline{q}$. Comme les caractéristiques utilisées pour entraîner le second classificateur BDT ne permettent pas de séparer le signal des événements $B\overline{B}$, nous utilisons l'échantillon plus pur $B^+ \to J/\psi(\to \ell^+\ell^-)K^+$, qui présente le même nombre de traces de $K\tau\ell$.

- Efficacité de la reconstruction du $B_{\rm tag}$. Un autre terme à prendre en compte est la divergence données/MC liée à la reconstruction du `FEI`, principalement due

à une mauvaise modélisation MC des désintégrations des mésons B^+. Pour y remédier, nous utilisons les facteurs de calibration officiels dépendant du mode `FEI` obtenus avec les événements $B^+ \to \overline{D}^0(\to K^+\pi^-)\ell^+\nu_\ell$ [11]. En repondérant les facteurs en fonction de la composition du signal MC, nous obtenons une correction sur l'efficacité moyenne de 85% avec une erreur de 6%.

Nous considérons ensuite les erreurs *additives*, qui affectent l'estimation de N_{sig} et sont donc exprimées en termes de nombre d'événements. La forme de la fonction de densité de probabilité (PDF) du signal, qui se compose d'une Cristal Ball et d'une gaussienne, doit être calibrée car elle est fixée à partir du MC lors de l'ajustement des données. Nous utilisons un échantillon obtenu en associant un pion au B_{tag}, et le recul du système permet d'extraire les événements $D^{(*)}\pi$. Cette méthode a l'avantage de fournir des statistiques beaucoup plus élevées, car les mésons D et D^* ne sont pas reconstruits. L'écart en N_{sig} dérivé des différences données/MC est utilisé pour estimer l'erreur systématique. Pour la fraction pondérant les deux composantes du signal, nous faisons varier la valeur estimée à partir de MC pour établir l'erreur systématique.

Pour les quatre modes, nous ne trouvons aucune preuve de signal et par conséquence nous dérivons les limites supérieures sur les BFs avec une approche fréquentiste cette phrase est étrange, les je ne vois pas en quoi les deux parties autour du "et" sont liées. Nous générons dix mille pseudo-expériences où le fond a la forme ajustée aux données et le N_{sig} varie dans une gamme pour trouver la valeur correspondant à 90% des pseudo-expériences ayant un N_{sig} obtenu inférieur aux données.

La sélection des événements et la suppression du bruit de fond sont optimisées à l'aide du modèle d'espace de phase à trois corps pour le signal MC. Cependant, les interactions violant la saveur des leptons et entrant dans les transitions de $b \to s\ell\ell$ peuvent produire différentes distributions de $q^2 = M(\ell\tau)$, et ainsi différentes efficacités finales du signal. En suivant la paramétrisation de la Réf. [12], nous considérons le scénario conduisant à la plus basse efficacité et obtenons les limites supérieures suivantes sur les BFs :

$$\begin{aligned}
\mathcal{B}(B^+ \to K^+\tau^+\mu^-) &< 0,65 \times 10^{-5} \\
\mathcal{B}(B^+ \to K^+\tau^+e^-) &< 1.71 \times 10^{-5} \\
\mathcal{B}(B^+ \to K^+\tau^-\mu^+) &< 2.97 \times 10^{-5} \\
\mathcal{B}(B^+ \to K^+\tau^-e^+) &< 2.08 \times 10^{-5}
\end{aligned}$$

qui intègrent les erreurs systématiques.

La limitation actuelle de la mesure provient de l'efficacité, considérablement réduite par l'étiquetage hadronique des mésons B. Par conséquent, nous poursuivons une étude approfondie de la performance du `FEI` afin de mieux exploiter les données disponibles. Ce travail arrive à point nommé puisque le premier long shutdown de Belle II est en cours : nous avons l'opportunité de mieux comprendre notre algorithme d'étiquetage avant l'arrivée de nouvelles données.

Amélioration de l'étiquetage hadronique des mesons B^+

Un élément clé pour améliorer la sensibilité expérimentale avec les données disponibles et futures de Belle II est d'augmenter la performance d'étiquetage et d'obtenir une efficacité plus élevée. La première étape consiste alors à améliorer la description des

désintégrations de B^+ dans la simulation MC de Belle. En effet, `FEI` utilise des techniques multivariées pour séparer efficacement le signal du bruit de fond. Cependant, ces méthodes sont entraînées sur le MC, ce qui, en cas d'erreur, peut conduire à des performances non optimales. L'une des conséquences est l'écart important entre les données et le MC pour l'efficacité du `FEI`.

Si l'on considère les transitions $b \to c$, 1/4 des désintégrations B^+ sont semileptoniques et `FEI` en couvre une grande fraction en raison des larges BFs pour les modes $\overline{D}^{(*)0}\ell^+\nu_\ell$. Les désintégrations hadroniques, qui couvrent les 3/4 restants[3], sont générées avec :

- `EvtGen` En utilisant les mesures collectées par PDG et les principes de symétrie. Les mesures sont souvent très anciennes et mal implémentées dans la simulation.
- `PYTHIA` En effectuant la fragmentation basée sur les quark donnés. Le fait que plus de la moitié des décisions B hadroniques soient décrites avec `PYTHIA` montre à quel point leur connaissance est limitée.

Les contributions dominantes à la largeur hadronique proviennent des transitions $b \to cu\bar{d}$ et $b \to cc\bar{s}$. En particulier, les premières couvrent plus de 60% des désintégrations hadroniques dont une grande fraction est générée par `PYTHIA`. Cette composante est difficile à accorder et peut produire des états finaux erronés. Nous avons étudié ce phénomène de manière plus approfondie à l'aide d'un échantillon pur.

L'échantillon le plus propre dont nous disposons est le $B^+ \to J/\psi K^+$, en raison de la résonance en deux leptons et de la présence d'un kaon ; ce mode permet d'atteindre une pureté de presque 100% et d'isoler l'autres B de l'événement, dont on connaît la saveur. L'inconvénient de cet échantillon est le faible BF, conduisant à seulement $\sim 35k$ événements pour l'ensemble des données Belle.

Nous appliquons l'algorithme `FEI` sur le ROE des B candidats sélectionnés, comptons les $N(B_{\text{tag}})$ pour chaque état final (`FEI` *modes*) et comparons les données avec MC. Les rendements obtenus sont le résultat de deux termes : l'efficacité de reconstruction et le BF effective pour chaque état final. Les modes $\overline{D}^{(*)0}n\pi$ sont responsables de $\sim 90\%$ de l'efficacité finale. Les modes avec deux mésons charmés, malgré des BFs comparables, souffrent des reconstructions exclusives de deux Ds. Nous nous concentrons donc sur le mode 12 `FEI` et sur les principales chaînes de désintégration qui y contribuent selon la simulation. En particulier, les modes à trois et quatre pions présentent les écarts les plus importants et ont un impact fort sur l'étiquetage. Les données expérimentales disponibles sont rares et révèlent le manque de résultats des usines à B avec leurs ensembles de données complets. Nous avons proposé quelques corrections basées sur une lecture attentive des listes PDG [13] et des résultats expérimentaux et théoriques [14–16] tout en essayant de faire un lien avec les désintégrations du τ. Les 6% que nous avons retirés des désintégrations exclusives sont remis dans `PYTHIA`. Ce n'est pas optimal car cela crée de nouveaux modes ou augmente les modes existants, mais nous avons intégré nos connaissances en interdisant à `PYTHIA` de générer certaines désintégrations qui devraient être supprimées.

En plus de réduire la divergence MC/données en termes d'efficacité, ce travail sur le MC devrait être bénéfique pour la pureté de `FEI`, comme le démontrent les distributions cinématiques présentées dans la Fig. 4.20. Ce dernier point est crucial car la cinématique des particules de l'état final composant les B est utilisée pour l'entraînement du `FEI`. Un nouvel entraînement basé sur le MC mis à jour montrera bientôt les améliorations attendues.

[3] Seulement $\sim 30\%$ d'entre elles sont reconstruites par `FEI`.

La principale faiblesse de l'échantillon $J/\psi K$ – la faible puissance statistique pour confirmer nos modifications – peut être surmontée en utilisant l'échantillon inclusif $\overline{D}^{(*)0}\pi^+$. Ce manuscrit propose d'autres directions pour améliorer l'efficacité d'étiquetage, notamment l'ajout de nouveaux modes de désintégration à `FEI` et la récupération de candidats partiellement reconstruits.

Recherche des désintégrations $B^+ \to K^+\tau\ell$ avec étiquetage semileptonique des mésons B

L'étiquetage semileptonique des B n'a jamais été essayé pour la recherche des désintégrations $B \to K\tau\ell$ et présente deux avantages principaux. Le premier est l'efficacité de reconstruction plus élevée due aux larges BFs des désintégrations de $B \to D^{(*)}\ell\nu$. Deuxièmement, seuls quelques modes sont utilisés et sont mieux connus, ce qui rend la description MC plus fiable et la performance `FEI` plus proche des données.

Avec l'étiquetage semileptonique, une plus grande efficacité peut être atteinte mais au prix d'une résolution dégradée sur la variable d'extraction M_{recoil}. En effet, la direction du B_{tag} ne peut pas être récupérée en raison de la présence d'un neutrino. Dans l'expression de la masse de recul (Éq. A), deux termes doivent être remplacés. L'un est la norme de l'impulsion, $p^*_{B_{\text{tag}}}$, qui peut être approximée par $\sqrt{E^{*2}_{\text{beam}}/c^2 - m^2_B c^2}$. L'information de direction, encodée dans le terme $\cos\theta$ est maintenant remplacée par une valeur extraite d'une distribution uniforme entre -1 et 1. Ce changement a pour effet de dégrader la résolution d'un facteur ~ 5. Le but de l'étude est d'avoir une estimation réaliste du fond pour extraire la sensibilité basée sur la nouvelle résolution et l'efficacité pour le signal.

Le processus de sélection commence par l'utilisation de l'étiquetage semileptonique, disponible dans le package `FEI`. Nous utilisons la variable $\cos\theta_{\text{BY}}$ et sélectionnons également les bons candidats B en fonction de $\mathcal{P}_{\text{FEI}}$. La signature des désintégrations semileptoniques correctes de B est la variable $\cos\theta_{\text{BY}}$ car elle définit l'ouverture du cône (autour de la direction du système visible $D\ell$) où le $\mathbf{p}_B$ devrait se trouver. L'ingrédient clé est le fait que la seule particule manquante est un neutrino. Nous prenons également en compte la production de charmonia dans les événements combinatoires où deux leptons de même saveur et de charge opposée sont sélectionnés. Enfin, nous coupons la variable $M(K^-X^+)$ plutôt que de l'utiliser pour l'entraînement BDT, car le résultat de la coupe BDT serait d'écarter la région sélectionnée.

Nous entraînons ensuite un classificateur BDT avec les caractéristiques du B_{tag}, du signal B, de la forme de l'événement et du ROE, dans le but d'éliminer en une seule étape les deux composantes de fond. Le score du BDT sépare très efficacement le fond $q\overline{q}$ grâce aux variables de suppression du continuum, tandis que la composante $B\overline{B}$ avec un B_{tag} semileptonique survit à notre coupe plutôt drastique.

Le résultat au niveau de confiance 90% est prometteur car il se situe à la même valeur que les résultats précédents. En ce sens, la combinaison d'une résolution plus faible et d'une efficacité de reconstruction plus élevée semble plus que réalisable. Nous examinons si l'approche semileptonique peut être encore améliorée.

L'idée est d'appliquer au secteur B une étude déjà publiée concernant les désintégrations du τ à Belle II [17]. Dans cette référence, la masse de recul est vue comme le résultat d'une minimisation contrainte 'M_2' et l'événement est caractérisé par une certaine énergie manquante sur les deux côtés B. La minimisation utilise le fait qu'aux usines à B, le quadri-momentum initial est connu ; par conséquent, la somme des impulsions 'invisibles' est nécessairement égale à l'impulsion manquante de l'événement. Nous appliquons ensuite deux contraintes cinématiques supplémentaires : l'une est

liée à la connaissance de la masse invariante du système, s, et l'autre au fait que de chaque côté nous connaissons la masse des parents, puisque l'événement n'a que deux B (Éq. 5.6). Pour la minimisation, les k_1^2 et k_2^2 sont requis. Le k_1^2 correspond à la masse élevée au carré du neutrino, qui est nulle. Par ailleurs, il faut distinguer les désintégrations hadroniques et leptoniques du τ. Dans le premier cas $k_1^2 = 0$, mais dans le second, une approximation de la masse invariante du système des deux neutrinos est nécessaire. Nous exploitons le fait que les deux B sont presque au repos dans le référentiel du centre de masse, ce qui nous permet de calculer le k_1^2 dans le contexte du laboratoire.

Des contraintes supplémentaires peuvent être imposées sur la base des informations de vertexing et de différentes conditions de résolution liées à Belle II et au collisionneur amélioré SuperKEKB, en exploitant la taille réduite des faisceaux et les excellentes performances de vertexing dues à PXD. Cependant, dans l'implémentation actuelle des contraintes, cet élément d'information n'apporte pas un gain significatif.

Cette étude, documentée dans [18], montre qu'il existe une amélioration significative de la résolution avec les contraintes cinématiques (M_{2sB}) uniquement pour les désintégrations hadroniques du τ en raison de la contrainte précise sur k_1^2. En ce qui concerne les informations de vertexing, pour la configuration testée pour l'article, où seul le vertex B_{sig} est utilisé, il faudrait une résolution d'environ $5\,\mu$m pour atteindre les performances obtenues avec M_{2sB} (en considérant la taille du point d'interaction Belle II de conception). Nous pouvons enfin comparer les ULs en absence de signal obtenues avec un étiquetage hadronique ou semileptonique. Le fait que nous obtenions la même sensibilité est très encourageant car cela permettrait d'exploiter un échantillon indépendant pour élargir les données disponibles ou pour confirmer la présence du signal à tout niveau de signification.

Conclusion et perspectives

Nous avons cherché les désintégrations $B^+ \to K^+\tau\ell$, motivées par les indices de violation du LFU dans les désintégrations B. Nous avons utilisé l'algorithme d'étiquetage B de Belle II sur l'échantillon de données Belle et aucun signal n'a été trouvé pour aucun des quatre modes. Nous avons dérivé les limites supérieures, qui sont les premières établies avec les données de Belle et les plus strictes à ce jour.

En examinant de plus près les performances de l'étiquetage B hadronique, nous nous rendons compte qu'elles sont étroitement liées à la modélisation MC. Nous utilisons l'échantillon $B^+ \to J/\psi(\to \ell^+\ell^-)K^+$ comme le moyen le plus efficace de sonder le B_{tag} et de vérifier les fautes trouvées dans la simulation. Les corrections proposées au MC de Belle apportent une amélioration du facteur de calibration global et des distributions cinématiques des filles de B.

Nous avons exploré la possibilité d'utiliser l'étiquetage semileptoniques des B et il s'est avéré être une alternative viable à Belle II, en particulier si des contraintes cinématiques sont appliquées pour améliorer la résolution de M_{recoil}. Pour l'avenir, les orientations sont claires. Nous pouvons effectuer la mesure avec plus de données et une stratégie améliorée incorporant les connaissances accumulées jusqu'à présent ; l'expérience Belle II ayant déjà accumulé un échantillon de plus de la moitié de celui de Belle et devrait atteindre $5\,\text{ab}^{-1}$ dans approximativement cinq ans. Du côté de la simulation, nous devons clarifier certains modes de désintégration B et obtenir une mesure plus précise du BF. Du côté du `FEI`, il est impératif d'effectuer un nouvel entraînement et d'améliorer l'algorithme, par exemple en révisant la sélection et en ajoutant de nouveaux états finaux. Un travail similaire est en cours sur les mésons

neutres B, ce qui profitera à la recherche de modes comme $B^0 \to K^0_s \tau \ell$. Pour en venir à l'approche semileptonique, nous avons deux approches possibles : la cinématique et le vertexing. Pour en tirer profit au maximum, il est important d'améliorer la sélection du signal et essayer de combiner plusieurs informations faibles pour surcontraindre le système. En conclusion, en supposant que les anomalies ne soient pas réfutées d'ici quelques années, l'ensemble de données plus important combiné à une stratégie d'étiquetage du B optimisée pourrait permettre d'observer les désintégrations LFV ou de contraindre fortement les modèles de nouvelle physique qui leur sont associés.

Introduction

Despite being validated by many experiments, the Standard Model of particle physics (SM) is known to be an incomplete theory. For instance, it provides a very coherent description of the known fundamental particles and allows to make precise predictions about their interactions. On the other hand, experimental observations demand extensions to the SM to obtain a more general theory capable of covering a broader range of energies.

For example, the SM does not describe the gravitational interaction, nor can explain the dynamical origin of the observed matter-antimatter asymmetry. The SM sets to zero the masses of neutrinos, going against the well-established observation of neutrino flavor oscillations, which imply non-zero masses. Furthermore, the SM fails to explain the existence of the 'dark' (neither baryonic nor luminous) matter, which does not interact with ordinary particles with known forces, yet providing the additional gravitational attraction needed to ensure the formation of astrophysical objects.

There are also more conceptual problems that would need extensions of SM to be better understood. One for all, the so-called flavor problem: many SM parameters are poorly constrained by symmetry principles, and measurements have revealed clear hierarchical structures, for example, for the fermion masses or the CKM matrix elements. Additional, unknown symmetries could explain those patterns.

In addition to the *gaps* summarized above, there are also longstanding tensions - significant deviations from SM predictions: the anomalous magnetic moment of the muon $g_\mu - 2$ with $4.2\,\sigma$ discrepancy [19] and an ensemble of observables related to semileptonic B meson decays collectively know as B-anomalies and showing a global discrepancy up to $4.6\,\sigma$ for some specific fits [20].

In $b \to s\ell\ell$ transitions, for example, there is a hint of lepton flavor non-universal interactions that, if confirmed, not only would be a clear sign of physics beyond the SM model (BSM), but also would suggest that lepton flavor violating (LFV) decays could appear in B meson decays with much higher rates than expected. The branching fractions (BFs) of the $B \to K\tau\ell$ ($\ell = e, \mu$) decays, for example, are predicted to be close to the current experimental sensitivities according to some BSM models [4]. Searches for these decays have been performed in the past by *BABAR* [21] and LHC*b* [22] experiments, but no evidence has been found so far.

The work presented here aims at searching for the first time for $B^+ \to K^+\tau\ell$ decays with the data collected by the Belle experiment at the KEKB collider between 1999 and 2010 and corresponding to 772×10^6 $B\overline{B}$ pairs. We have also examined the limitations of the current analysis strategy and proposed paths towards an update of the measurement with the upgraded detector Belle II and collider SuperKEKB.

The manuscript is organized as follows. Chapter 1 briefly introduces the SM basics and some nomenclature, with more emphasis on the aspects related to flavor and the B-anomalies, as well as the theory behind the B meson decays.

Chapter 2 describes the duo machine/experiment (KEKB/Belle) where the used dataset was collected. The key elements about the upgrades are also presented. Special attention has been given to the Belle II High-Level Trigger, as the IJCLab group

has put efforts into its operation since the early stages of data-taking.

In chapter 3 we report about the search of the $B^+ \to K^+\tau\ell$ ($\ell = \{e, \mu\}$) decays using the hadronic B-tagging and the Belle data sample corresponding to $711\,\text{fb}^{-1}$. The analysis procedure on simulated data and the results on collision data are shown. In chapter 4, we describe how the current knowledge on hadronic B-decays has been integrated into the simulation and outline other directions to improve this B-tagging.

In chapter 5 we discuss the feasibility of a search based on a different B-tagging approach, where *semileptonic*, instead of hadronic, B-decays are used. This method has the well-known advantage of being more efficient but the drawback of leading to a degraded resolution. We here explore how the Belle II performance and SuperKEKB parameters could improve the sensitivity.

Lastly, chapter 6 summarizes the achieved results and elaborates on the current limitations of our measurement. We provide possible paths to reach a higher sensitivity for LFV searches in B-decays, not only relying on the larger dataset which will be collected at Belle II but also capitalizing on the experience gained with this piece of research.

Additional material related to the $q^2 = m_{\ell\tau}$ parametrization for the interpretation, in terms of BSM models, of the results obtained in chapter 3 is condensed in Appendix A.

Chapter 1

Flavor in Particle Physics

In particle physics, the word *flavor* is used to denote the species of the SM fermions – 6 quarks and leptons each bearing a specific set of quantum numbers[1] – and *flavor physics* is meant to study the interactions that distinguish between flavors – in the SM, the weak and Yukawa interactions.

In the last century, flavor physics has demonstrated to be a successful path towards scientific advancement and paved the way to many discoveries. For example, the prediction of the existence of a fourth quark (the charm-quark) and the third generation of fermions, but also the prediction of charm and top quark masses. Also, the measurement of the neutrino flavor transitions led to the discovery of neutrino mass [23].

In the context of the present work, the observation of $B \to K\tau\ell$ decays would entail the presence of beyond-SM (BSM) particles with masses that might be beyond the reach of current LHC experiments. In this sense, the 'flavor path', probing and challenging the SM through rare decays, can be seen as complementary to the 'energy path', for which new particles are directly searched among the collision products.

1.1 The Standard Model

At the basis of the Lagrangian field theory there is the concept of symmetry, which denotes an invariance of the Lagrangian $\mathcal{L}$, and thus of the associated equations of motion, under some transformations. Such transformations can be local (coordinate-dependent) or global and they can be continuous or discrete. The invariance may be with respect to coordinate redefinitions, as in the case of Lorentz invariance, or field redefinitions, as in the case of gauge invariance.
Of particular interest are the global discrete invariances known as C or charge conjugation (replacement of a field by its conjugate), P or parity (sign reversal of all spatial coordinates), and T or time reversal (sign reversal of the time coordinate, which reverses the role of *in* and *out* states). When constructing a field theory, locality, Lorentz invariance, and hermiticity of $\mathcal{L}$ are always required. That is sufficient to make any field theory automatically also invariant under the product of operations CPT. In many theories the combination CP, and thus also T, are separately automatic symmetries. This is the reason why the experimental discovery that CP is not an exact symmetry of nature was such a breakthrough event.
The Standard Model (SM) is the unified gauge theory for electroweak and strong interactions, the gauge group being $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$[2]. In this model, a single theory includes the electromagnetic and weak interactions – both neutral current and charged current – and the Quantum Chromodynamics (QCD). In particular,

[1] Electric charge, weak isospin, baryon and lepton number.
[2] Here c denotes the colour, L the chiral component and Y the hyper-charge.

electroweak theory has the symmetry group $SU(2) \otimes U(1)$ and electromagnetic and weak coupling constants are connected by the theory, while QCD and electroweak theories are unified by the theoretical framework keeping their coupling constants independent [24]. The current SM frame describes the interactions between the fermion fields (quarks and leptons in Fig. 1.1) and the bosons as gauge fields:

- the photon γ and the weak bosons ($W^{\pm}$, Z^0) for the electroweak field;
- the gluons, mediators of the strong interactions;
- the Higgs boson, related to the Higgs scalar field, responsible for the mass generation mechanism.

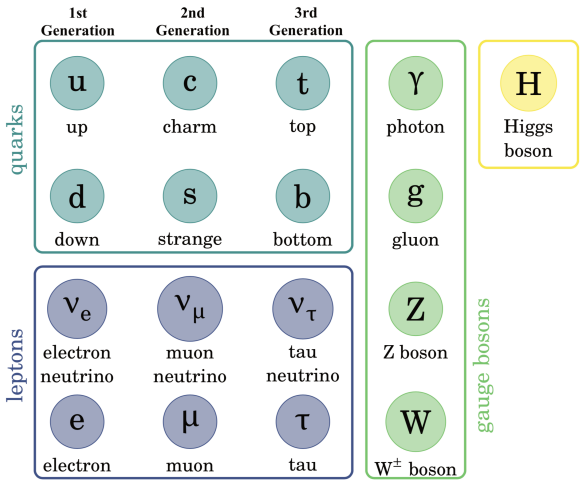


FIGURE 1.1: SM particles [25].

The fermion fields of the SM, written in terms of Weyl spinors in a $(c, L)_Y$ notation, are:

Q_i	$(3,2)_{1/6}$	left-handed quark doublet
U_i^c	$(3,1)_{2/3}$	right-handed quark-up
D_i^c	$(3,1)_{-1/3}$	right-handed quark-down
L_i	$(1,2)_{-1/2}$	left-handed lepton doublet
E_i	$(1,1)_{-1}$	right-handed charged lepton ,

where $i = 1,2,3$ denotes the generation index. For example, $Q_i(3,2)_{1/6}$ denotes the left-handed (chiral) component of a field which is a triplet under $SU(3)$ colour group and weak isospin doublet, whose hyper-charge is $1/3$. The weak hyper-charge is defined as $Y_W = (Q - T_3)$, being Q the electric charge and T_3 the third component of weak isospin ($SU(2)$ component).

The renormalizable Lagrangian can be written in the form:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yuk}} \, .$$

Here $\mathcal{L}_{\text{kin}}$ describes the free propagation in the spacetime and contains the covariant derivatives and the fields' strengths; $\mathcal{L}_{\text{Higgs}}$ gives the Higgs scalar potential:

$$\mathcal{L}_{\text{Higgs}} = \mu^2 |\Phi|^2 - \lambda |\Phi|^4 \tag{1.1}$$

where $\Phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}$.

Finally, the interaction of the quarks and leptons with the Higgs field, introduced to give masses to them and to the gauge bosons, is described by

$$\mathcal{L}_{\text{Yuk}} = y^d_{ij}\bar{Q}^i\Phi D^j + y^u_{ij}\bar{Q}^i\Phi^\dagger U^j + \text{lepton term} + \text{h.c.}\,.$$

Choosing $\mu^2 < 0$ and $\lambda > 0$ and defining $v^2 = -\frac{\mu^2}{\lambda}$, the scalar potential of Eq. 1.1 is (up to a constant term)

$$\mathcal{L}_{\text{Higgs}} = -\lambda\left(\Phi^\dagger\Phi - \frac{v}{2}\right)^2,$$

which implies that the field acquires a *vacuum expectation value*, or *vev*, $\langle\phi\rangle = v/\sqrt{2}$ for at least one direction of Φ, for example in the real direction of the down component:

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}.$$

The Higgs *vev* breaks the $SU(2)\otimes U(1)$ symmetry down to a $U(1)$ subgroup and, writing the doublet fields out explicitly

$$Q^i_L = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix},$$

gives mass terms of the form

$$\mathcal{L}_{\text{mass,q}} = m^d_{ij}\bar{d}^i_L d^j_R + m^u_{ij}\bar{u}^i_L u^j_R. \tag{1.2}$$

The mass matrices are related to the Yukawa couplings and the Higgs *vev* $\langle\phi\rangle = v/\sqrt{2}$ by

$$m^q_{ij} = \frac{v}{\sqrt{2}}y^q_{ij}\,.$$

In general the m^q matrices are not diagonal; one can move to the mass basis and have the m^q's diagonal through a unitary transformation

$$\hat{m}^q_{ij} = (V^q_L)_{ik}(m^q)_{kl}(V^{q\dagger}_R)_{lj}\,,$$

with the notation $\hat{m}^q_{ij}$ indicating the diagonalized matrix. In principle, the above L, R indices are distinct from the L, R of the previous notation (Eq. 1.2), but the i index is associated with the left-chiral quark while the j index with the right-chiral quark (see Eq. 1.2). One can thus rotate the left-chiral and right-chiral fields by V_L and V_R accordingly:

$$\begin{aligned} q^i_L &= (V^q_L)^i_j q^{'j}_L \\ q^i_R &= (V^q_R)^i_j q^{'j}_R\,, \end{aligned}$$

having on the left the interaction-basis field and the linear combination of mass-basis fields on the right. From here on, the mass basis is used and the prime is dropped. In the mass basis the Yukawa interactions are diagonal since mass matrices are proportional to the Yukawa matrices but now the coupling of the W bosons contains off-diagonal terms. Indeed this change of the basis leads to:

$$\mathcal{L}_{Wqq} \propto \frac{g}{\sqrt{2}}\bar{u}_L i\gamma_\mu d_L W^\mu \to \frac{g}{\sqrt{2}}\bar{u}_L i\gamma_\mu (V_{uL}V^\dagger_{dL})d_L W^\mu\,. \tag{1.3}$$

The element $(V_{uL}V_{dL}^{\dagger})$ indicates the Cabibbo-Kobayashi-Maskawa (CKM) unitary matrix:

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} ,$$

where the quarks are ordered by increasing mass and the mass eigenstates are named after their dominant flavor. As an $N \times N$ unitary matrix, $V_{\rm CKM}$ has N^2 real parameters, $2N-1$ of which can be reabsorbed by the quark fields on the right and left sides of the matrix in Eq. 1.3 by globally redefining the arbitrary phases. Of the remaining $(N-1)^2$ free parameters, $N(N-1)/2$ are the Euler mixing angles while the others $\frac{N}{2}(N-3)+1$ are unremovable complex phases – just one, for $N=3$. The standard way to parametrize the matrix in terms of mixing angles and the irreducible phase is the following

$$V_{\rm CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} , \quad (1.4)$$

given that $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$.

The measured values of the modules of the elements are reported below [26]:

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

At the price of loosing unitarity, Wolfenstein introduced a different parametrization guided by the fact that the matrix elements can be expressed in terms of the four parameters λ, A, ρ, η, defined as:

$$\lambda \equiv s_{12} = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}$$
$$A\lambda^2 \equiv s_{23} = \lambda \left| \frac{V_{cb}}{V_{us}} \right|$$
$$A\lambda^3(\rho + i\eta) \equiv s_{13}e^{i\delta} = V_{ub}^* = \frac{A\lambda^3(\bar{\rho} + i\bar{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1-\lambda^2}[1 - A^2\lambda^4(\bar{\rho} + i\bar{\eta})]}$$

which leads to:

$$V_{\rm CKM} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) . \quad (1.5)$$

Experimental values of the four parameters are [26]:

$$\lambda = 0.22453 \pm 0.00044, \qquad A = 0.836 \pm 0.015, \qquad \bar{\rho} = 0.122^{+0.018}_{-0.017}, \qquad \bar{\eta} = 0.355^{+0.012}_{-0.011}.$$

The power of the Wolfenstein representation is that it captures the relevant physics of CKM matrix: the upper left 2×2 matrix is the Taylor expansion in θ_{12} of the

Cabibbo matrix[3]. Furthermore, at the first order in λ (which is < 1) the first two generations are decoupled from the third one and only at third order in λ the complex phase shows up. The relations of Eq. 1.5 ensure that $\bar{\rho} + i\bar{\eta} = -(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)$ is phase convention independent and the CKM matrix written in terms of λ, A, $\bar{\rho}$ and $\bar{\eta}$ is unitary to all orders in λ.

A theory can be *CP* violating if and only if it has some set of couplings for which rephasing of all fields cannot remove all the phases, while *CP* conservation is automatic for any theory for which the most general form of the Lagrangian allows all complex phases to be removed. If the Lagrangian involves a sufficient number of fields, either fermions or scalars, so that there are more couplings than the possible phase redefinitions, not all couplings can be made real by rephasing the fields. The three generation SM with a single Higgs doublet has only one *CP*-violating (CPV) parameter.

1.2 $b \to s\ell\ell'$: a window onto the New Physics

1.2.1 Lepton flavor and B-anomalies

The expression 'B-anomalies' is commonly referred to the tensions with the SM observed in B meson semileptonic decays. They started to appear almost ten years ago and since then the number of observables showing discrepancies has grown, especially in the $b \to s\ell\ell$ sector, increasing the overall significance.

Recent measurements performed at B-factories but mainly at LHCb [29–32] have provided experimental indications of lepton flavor universality (LFU) violation - deviations from μ/e universality in loop-induced, neutral-current (NC) transitions $b \to s\ell\ell$ and τ/ℓ universality in tree-level, charged-current (CC) transitions $b \to c\ell\nu$. The LFU is an accidental symmetry of the SM, implying that the three lepton families interact with the same strengths with the fields (except for the Higgs, where the different couplings are responsible for the mass hierarchy $m_\tau > m_\mu > m_e$).

The $b \to s\ell\ell$ transitions are studied via the decays $B \to H\ell\ell$, where H is a hadron with a strange quark. The BFs of those decays are difficult to compute because of the strong forces among the quarks composing the hadrons. These local contributions related to the $B \to H$ transitions are encoded in the form factors. However, the QCD effects do not involve the two leptons ℓ's and are identical between different ℓ species. This makes the prediction on the ratio

$$R_H^{e/\mu} = \frac{\Gamma\left(B \to He^+e^-\right)}{\Gamma\left(B \to H\mu^+\mu^-\right)}\Big|_{q^2 \in (q^2_{\min}, q^2_{\max})} \tag{1.6}$$

very precise and allows to directly test the LFU. In fact the ratio is predicted to be $1 \pm \mathcal{O}(1\%)$ [33], the uncertainty being due to electromagnetic corrections and effects related to the mass of the e, μ (small compared to the hadrons). The $q^2 = m^2_{\ell\ell}$ of Eq. 1.6 is chosen in a range $(q^2_{\min}, q^2_{\max})$ to exclude the $B \to X_{q\bar{q}}K$ resonant states. For example, when $X_{c\bar{c}} \equiv J/\psi$ and $q^2 = 9.59\,\mathrm{GeV}^2/c^4$ the $B \to KJ/\psi(\to \ell^+\ell^-)$ cross section is much larger because of the favored $b \to c\bar{c}s$ transition.

Turning to the CC anomalies, the discrepancies are found in the ratios

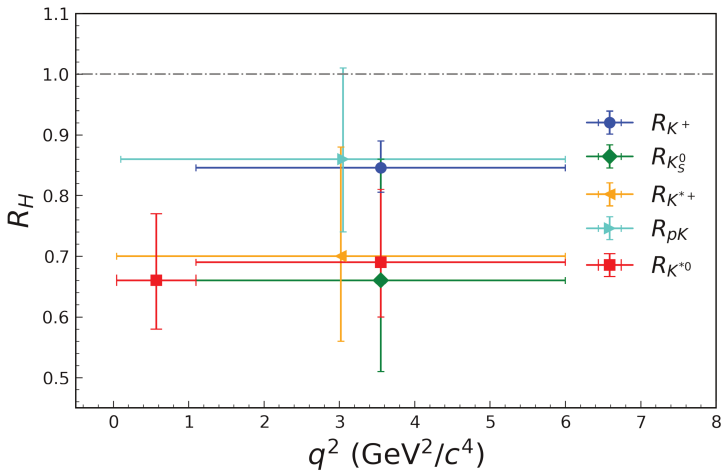
$$R_{D^{(*)}}^{\tau/\ell} = \frac{\Gamma\left(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau\right)}{\Gamma\left(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell\right)} \qquad (\ell = e, \mu)\,, \tag{1.7}$$

[3]The first form in which the matrix was postulated took into account only the first two generations. This proposal, made by Gell-Mann and Levy and later Cabibbo [27, 28], was the basis of the anticipation of the existence of the charm quark.

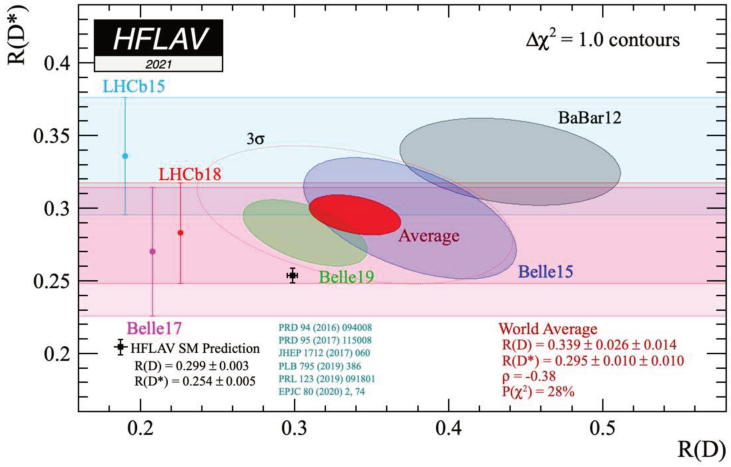
where a deviation of 3.4σ from the most recent calculations is observed [34], and also in [35]

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

It is assumed that the BSM mediator mainly alters the $b \to c\tau^-\bar{\nu}_\tau$ rate and could be due to the tree-level exchange of a W' boson (the partner of Z' in the vector-boson model), a leptoquark or a charged Higgs. In any case, a large tree-level contribution is needed - of the order of $\mathcal{O}(10\%)$ at amplitude level. When combining the NC and CC anomalies, decay rates like $b \to s\tau^+\tau^-$ are expected to be much higher than the SM prediction [36]. Fig. 1.2 summarizes the experimental results for the LFU tests, comparing them with the theoretical expectations.



(A)



(B)

FIGURE 1.2: LFU tests. A: R_H ($H = \{K^{(*)+}, K^{*0}, K^0_S, pK\}$) measurements from LHC$b$. B: $R(D) - R(D^*)$ [34] with the recent measurements from *BABAR*, Belle and LHCb.

1.2.2 Theoretical interpretation

The low-energy effective Hamiltonian for the semileptonic $b \to s$ transitions (allowing for LFV) can be written in the form[4]:

$$\mathcal{H}_{\text{eff}}(b \to s\ell_1^-\ell_2^+) = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_{i=7,9,10,S,P} \Big(C_i^{12}(\mu)\mathcal{O}_i^{12}(\mu) + C_i'^{12}(\mu)\mathcal{O}_i'^{12}(\mu) \Big) + \text{h.c.} \tag{1.8}$$

where the operator product expansion allows to separate the short-distance and long-distance contributions [12]. The $C_i(\mu)$ functions are called Wilson coefficients (WCs) and are scale-dependent couplings (μ denotes the renormalization scale) to the vertices described by the $\mathcal{O}_i$'s, local operators governing the interaction. In particular, those relevant for the $b \to s\ell\ell$ decays are[5]:

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

$$\mathcal{O}_9^{(\prime)12} = \frac{\alpha}{4\pi}(\bar{s}\gamma_\mu P_{L/R}b)(\bar{\ell}_1\gamma^\mu\ell_2), \qquad \mathcal{O}_{10}^{(\prime)12} = \frac{\alpha}{4\pi}(\bar{s}\gamma_\mu P_{L/R}b)(\bar{\ell}_1\gamma^\mu\gamma^5\ell_2),$$
$$\mathcal{O}_S^{(\prime)12} = \frac{\alpha}{4\pi}(\bar{s}P_{R/L}b)(\bar{\ell}_1\ell_2), \qquad \mathcal{O}_P^{(\prime)12} = \frac{\alpha}{4\pi}(\bar{s}P_{R/L}b)(\bar{\ell}_1\gamma^5\ell_2),$$

respectively describing the electromagnetic $\mathcal{O}_7^{(\prime)}$, the (axial)vector $\mathcal{O}_{(10)9}^{(\prime)}$ and the (pseudo)scalar $\mathcal{O}_{(P)S}^{(\prime)}$ interactions. In particular, for

- $\ell_1 = \ell_2 = \ell$ ($C_i^{12} \equiv C_i^\ell$), all the C' coefficients are zero in the SM – except C_7' which has a small SM contribution of $\mathcal{O}(m_s/m_b)$ [37]. Furthermore, the non-vanishing WCs are lepton-flavor universal but BSM mediators can contribute differently to each flavor: $C_i^\ell = C_{i\text{SM}}^\ell + C_{i\text{BSM}}^\ell$, where $C_{i\text{BSM}}^\ell$ is responsible for the LFU-breaking.
- $\ell_1 \neq \ell_2$ all the WCs are zero in the SM. A specific model needs to be set in order to produce non-zero values and allow for LFV. In the case where only scalar operators generate LFV, $C_{S,P}^{(\prime)12} \neq 0, C_{9,10}^{(\prime)12} = 0$ and the opposite situation holds with only vector operators.

There are two ways of inferring the properties of the possible BSM model from the measurements collectively forming the B-anomalies. In the *model-independent* approach, a fit to the experimental data is performed to find the best values of the WCs related to the effective Hamiltonian with all the operators describing the process (either $b \to s\ell\ell$ or $b \to c\tau\nu$). The fit is 1D, 2D or nD if one, two or all the parameters – the Wilson Coefficients – are allowed to vary according to a given BSM hypothesis. For example, in the 1D fit a single WC (or a constrained combination of WCs) is tested to explain the anomalies. In the other approach called *model-dependent*, the theorist builds models that are compatible with the constraints from the measurements.

Usually these two approaches can provide complementary information that can be combined to get the best interpretation given the current experimental status.

[4] $G_F = 1.166\,378\,7 \times 10^{-5}\,\text{GeV}^{-2}$.

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

Model-independent approach

In addition to the LFU tests (R_K, R_{K^*}), which are the most valuable because of the small theoretical uncertainties and high sensitivity to short-distance contributions, the $b \to s\ell\ell$ transitions provide many other observables. To this category belong $\mathcal{B}(B_s \to \mu^+\mu^-)$, which is particularly interesting due to its reduced hadronic sensitivity and its dependence on a reduced subset of WCs. Other less reliably predicted (clean) quantities are the differential BFs of $B \to K^{(*)}\mu^+\mu^-$, $\Lambda_b \to \Lambda\mu^+\mu^-$, $B_s \to \phi\mu^+\mu^-$ and the observables parametrizing their four-body angular distributions (e.g. P_5' [38] or Q_1 [39]). In fact, they are affected by larger theoretical uncertainties due to charm-quark loop contributions that are difficult to disentangle from possible BSM effects [40].

A few observables commonly used as input for the model-independent procedure are listed in Tab. 1.1, where the experimental values and the theoretical expectations are compared.

Observable	Measurement	SM exp.	References
$R_K^{[1.1,6.0]}$	$0.846^{+0.042+0.013}_{-0.039-0.012}$	1.00 ± 0.01	[29], [33]
$R_{K^*}^{[1.1,6.0]}$	$0.69^{+0.11}_{-0.07} \pm 0.05$	1.00 ± 0.01	[30], [33]
$\mathcal{B}(B_s \to \mu^+\mu^-)$	$(3.01 \pm 0.35) \times 10^{-9}$	$(3.66 \pm 0.14) \times 10^{-9}$	[13], [41]
(R_D, R_{D^*})	(0.337(30),0.298(14))	(0.299(3),0.258(5))	[42], [43]
$\mathcal{B}(B \to \tau\nu_\tau)$	$1.09(24) \times 10^{-4}$	$0.812(54) \times 10^{-4}$	[13], [44]

TABLE 1.1: Low-energy observables: measurements and SM predictions. The upper part collects the most clean observables. Table adapted from Ref. [20].

Several groups provide regular updates – the so-called *global fits*, each time incorporating to their calculations the newly available experimental results. Recent global fits of the anomalies have been performed after the 2021 LHC*b* results [37, 45, 46] and show the best-fit values in 1D, 2D or 6D scenarios. For example, Figure 1.3 shows one of the possible outputs of the fit performed for Ref. [37]: 2D contours for pairs of WCs, given a set of input data.

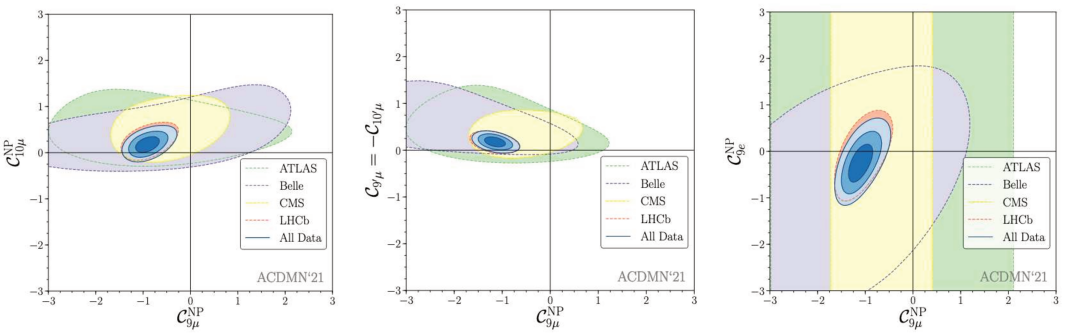


FIGURE 1.3: 2D-Fits to 2021 $b \to s\ell\ell$ data. Each plot corresponds to a pair of WCs that is allowed to float in the process of fitting 254 $b \to s\ell\ell$ observables. Picture from Ref. [37].

Model-dependent approach

Numerous extensions to the SM have been proposed to explain the discrepancies observed in B-decays [1–4] and leptoquarks (LQs) represent one of the most popular options nowadays. One of the main reasons is that they can contribute to the

semileptonic transitions at tree-level (Fig. 1.4), thus showing sizeable anomalies, while entering only at the loop level into four-quark or four-lepton interactions, in which no deviation from the SM has been found so far [5]. Furthermore, the predicted masses for the LQs are $\sim\mathcal{O}$(TeV), well above the electroweak scale, and this is compatible with the fact that no new particles have been directly observed at the LHC experiments since the confirmation of the Higgs boson.

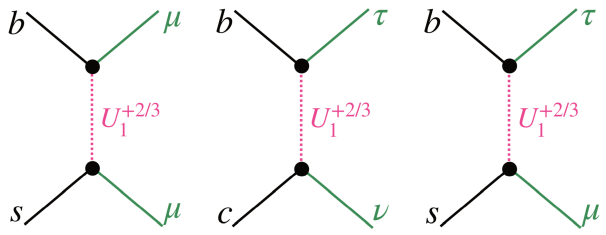


FIGURE 1.4: $b \to s\mu\mu$, $b \to c\tau\nu$, $b \to s\tau\mu$ transitions with a LQ mediator.

LQs are colored states [47] that can mediate interactions between quarks and leptons. In general, a LQ can be a scalar or a vector field[6] and can behave as singlet, doublet or triplet under $SU(2)_L$. LQs are designated by their quantum numbers with respect to the SM gauge group: $(SU(3)_c, SU(2)_L)_Y$, where the electric charge, $Q = Y + T_3$ is the sum of the hyper charge Y and the third weak isospin component T_3 and can assume the values Q $= \pm 5/3, \pm 4/3, \pm 2/3, \pm 1/3$. They also have a well-defined fermion number $F = 3B + L \in \{-2, 0\}$ depending if the coupling is quark-lepton or antiquark-lepton while the single lepton L and baryon B numbers are defined only for states without di-quark couplings (i.e. with F=0)[7].

When considering single LQ solutions, the $U_1 \sim (\mathbf{3}, \mathbf{1}, 2/3)$ is currently the preferred one [20], as it can explain both sets of anomalies[8], surviving the low-energy constraints and the direct searches at LHC.

In fact the pp collisions can provide complementary information via the searches of:

i $q_i\bar{q}_j \to \ell_\alpha\bar{\ell}'_\beta$ processes. The right-edge of the $m_{\ell\ell}$ spectrum ('high p_T'-tails) can show an excess due to the exchange of BSM mediators in t-channel in $b\bar{b}/\bar{s}b/\bar{s}s$-initiated processes [48].

ii $pp \to LQLQ^* \to \ell\ell qq$ (LQ pair production) or $pp \to LQ\ell \to \ell\ell q$ (single-LQ production), like the $b\tau t\nu$ search at CMS [49].

The appealing scenario offered by the U_1 needs an ultraviolet completion, i.e. the model generating the new particles must be provided. In fact the U_1 is a vector LQ and it is not renormalizable without a Higgs sector, making impossible to compute the loop-level contributions - which might not be negligible. This problem can be fixed within models where the gauge group is broken at the TeV scale and U_1 is a gauge boson [6]. In a two-LQ scenario, the two scalars S_1 (or R_2) and S_3 can provide explanation of the B-anomalies and also be compatible with the constraints from $b \to s\nu\nu$ [7]. More details on how the model and the experimental constraints translate into actual couplings are given in Appendix A.

It is important to stress that in most of the models proposed to explain the anomalies, the lepton flavor conservation is no longer guaranteed, predicting significantly

[6] There are six spin-0 and as many spin-1 states.
[7] Such coupling should be strongly suppressed as they can mediate the proton decay.
[8] Owing to the absence of tree-level constraint from $\mathcal{B}(B \to K\nu\bar{\nu})$ [4].

enhanced decay rates in processes like $b \to s\ell\ell'$, especially the $B \to K\tau\mu$ modes [4], which we aim to search for. The observation of lepton flavor violation in the charged lepton sector would be a univocal sign of physics beyond the SM and would provide strong constraints to those models. In fact, even in an extended SM including non-zero neutrino masses generated by the Higgs mechanism, the charged lepton flavor violation (cLVF) is an extremely rare phenomenon – suppressed by m_ν^4[9].

1.2.3 Experimental status on LFV $b \to s\ell\ell'$

Many searches of LFV decays involving $b \to s\ell\ell'$ transitions have been performed over the last ten years, providing stringent results for the $b \to se\mu$ modes. The obtained sensitivities at LHCb for the $B^+ \to K^+\mu e$, $B^0 \to K^{*0}\mu e$ and $B_s^0 \to \phi\mu e$ [50,51] decays are in the range $(6 \div 16) \times 10^{-9}$ and they rely on the possibility of fully reconstructing the B meson. Concerning the modes with τ's, the limits are not as constraining due to presence of missing energy: among the τ decay products there is at least one neutrino, which escapes the detection as it does not interact with the detector materials. In addition, τ's lack distinctive signatures, making the measurement of their decays experimentally more challenging. This can be easily explained from the terms used when deriving an upper limit (UL) on the BF:

$$\mathcal{B}^{(\mathrm{UL})} = \frac{N_{\mathrm{sig}}^{(\mathrm{UL})}}{\varepsilon \times N_{\mathrm{B}}} \,. \tag{1.9}$$

The reconstruction (and selection) efficiency for signal events at the denominator (ε) determines the attainable sensitivity for a given measured N_{sig} and available data (N_B). In fact, the higher the efficiency, the more stringent the limit is. At B-factories, the exclusive reconstruction of $B \to K\ell^+\ell^-$ has an efficiency of the order of 30%. In contrast, it goes down to $\mathcal{O}(0.1\%)$ for tagged analyses, i.e., when the full reconstruction of the companion B is required. However, unlike other SM processes involving τ's (e.g. $B \to \tau\nu$, $B \to D^*\tau\nu$), the $K\tau\ell$ channels have the unique property of having the (one or two) neutrino(s) coming from the τ itself, thus allowing to extract the signal yield using the recoil mass[10], which should peak at the mass of the τ lepton. Such variable is easily obtained at B-factories, because of the well-known initial kinematics and the full reconstruction of the other B in the event, while at LHCb [22] a different strategy was adopted. The analysis consists in tagging the B^+ candidates by looking for $B_{s2}^{*0} \to B^+K^-$ decays (see Fig. 1.5); such mode is responsible for only $\sim 1\%$ of B^+ production but provides useful constraints. In fact the prompt kaon allows to determine the flavor of the companion B (using the K charge sign) and to compute the direction of the B^+ and its energy up to a quadratic ambiguity. For that the momentum of the K^- is needed, as well as the decay vertex of the B^+ (from the $K^+\mu^-$ pair). Such method does not only provide the missing-mass discriminating variable[11], but also a control sample (obtained with the wrong combinations B^+K^+) which the selection can be optimised on.

BABAR provided the first result on eight $B^+ \to h^+\tau^\pm\ell^\mp, (h = \pi, K)$ modes with the usage of hadronic B-tagging [21], while LHCb obtained an independent UL for the $B^+ \to K^+\tau^+\mu^-$ mode only using the B_{s2}^*-tagging described above. The current experimental status is summarized in Tab. 1.2.

[9] For example, BFs for $\mu \to e\gamma$ and $\tau \to \mu\gamma$ are expected to be lower than 10^{-54} and 10^{-53}, respectively.

[10] Invariant mass of the system recoiling against the $B_{\mathrm{tag}}K\ell$ system.

[11] $m^2_{\mathrm{miss}} = (p_B - p_{K\mu})^2$.

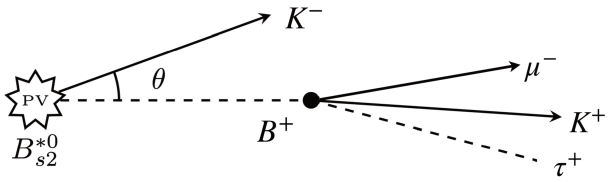


FIGURE 1.5: Decay chain reconstructed (LHCb [22]). B^+ candidates are tagged by the presence of a K^-, which is compatible with a B^{*0}_{s2} decay. The $K^+\mu^-$ vertex and the B^{*0}_{s2} vertex provide the direction of the B meson. Picture adapted from Ref. [52].

Mode	*BABAR* (433 fb^{-1}) Had. B-tag	LHCb (9 fb^{-1}) B^*_{s2}-tag
$K^+\tau^+\mu^-$	2.8×10^{-5}	3.9×10^{-5}
$K^+\tau^-\mu^+$	4.5×10^{-5}	
$K^+\tau^+e^-$	1.5×10^{-5}	
$K^+\tau^-e^+$	4.3×10^{-5}	

TABLE 1.2: Summary of the experimental limits for the $B^+ \to K^+\tau\ell$ modes at 90% confidence level [21, 22].

It must be stressed that, even once the flavor of the light lepton is fixed, there is still a charge ambiguity, depending if the charge of the heavier lepton (the τ) is the same as the B (or the kaon) - in this case we refer to an $\mathbf{OS}_\ell$ mode - or opposite, leading to an $\mathbf{SS}_\ell$ mode. In fact the differential decay rate expression (Eq. 9 in Ref. [12]) is not symmetric between the processes $B \to K\ell_1^-\ell_2^+$ and $B \to K\ell_1^+\ell_2^-$, thus making of more difficult interpretation the average of two measurements related to the (lepton) charge-conjugated modes.

Recently the LHCb collaboration has shown preliminary results on the two $B^0 \to K^{*0}\tau^\pm\mu^\mp$ modes, providing the first UL measurement. The reconstruction exploits the kinematic constraints given by the $K^{*0} \to K^+\pi^-$ resonance and the usage of the vertexing information from the $\tau^- \to \pi^+\pi^-\pi^-(\pi^0)\nu_\tau$ decay mode. The observed ULs at 90% confidence level (C.L.) are 1.0×10^{-5} for the $\mathbf{OS}_\mu$ mode and 0.8×10^{-5} for the $\mathbf{SS}_\mu$.

1.3 Theory on $b \to c$ transitions: charged B meson decays

1.3.1 B-tagging at B-factories

A significant part of the B-physics program at Belle and Belle II targets modes involving neutrinos which are produced either 'directly' ($D^{(*)}\ell\nu$, $K^{(*)}\nu\bar{\nu}$) or indirectly through τ's ($K\tau\ell$, $K\tau\tau$) . These decays are powerful probes of either crucial SM model parameters (like $|V_{cb}|$ and $|V_{ub}|$) or BSM couplings as they are related to current flavor anomalies.

At B-factories many different strategies are used and explored to study the modes with missing energy with the common goal of identifying the B-decays with a desired purity. This is combined with the exploitation of the specific properties of the channels under study in order to have a handle to control the background level. The identification of B-decays, or *B-tagging*, consists of various degrees of reconstruction of one of the two B's in order to infer the properties of the second one. On one side (see Fig. 1.6) of the spectrum there is the full reconstruction of B's decaying

hadronically (i.e. without neutrinos) and on the opposite side, the so-called 'inclusive' approach and they correspond respectively to high purity and low efficiency and vice versa. The *semileptonic tag*, where semileptonic B-decays are reconstructed, can be seen as an intermediate approach as it allows to get a higher reconstruction efficiency (compared to hadronic B-tagging) while loosing some information on the B meson's kinematic properties due to the presence of neutrinos. The hadronic and semileptonic tag applied to the search of $B \to K\tau\ell$ decays are described in chapter 3 and chapter 5 respectively.

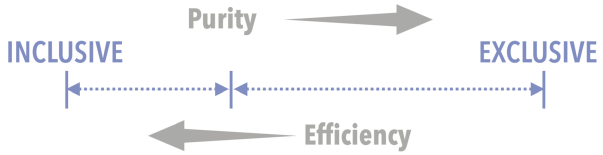


FIGURE 1.6: B-tagging strategies. The adopted method can be seen almost as a continuous cursor between inclusive tag and full hadronic reconstruction, depending on the desired level of purity and efficiency.

1.3.2 Knowledge of the charged B hadronic decays

Regardless of the chosen B-tagging approach, machine-learning techniques are commonly used to boost the performance of the selections. Such methods require the training of classifiers and simulated Monte Carlo data are used as inputs for the learning process. Unfortunately our current modeling of B-decays is non-optimal, due to the lack of experimental inputs or clear prescriptions from theory. The use of a wrong MC leads to non-optimal B-tagging performance and thus lower experimental sensitivities.

The status on the knowledge on hadronic B-decays is presented in this section while an in-depth study of the interplay between MC simulation and B-tagging performance are presented in chapter 4.

Approximately 1/4 of the B^+ meson width is coming from $b \to c\ell\nu$ transitions [13]; consequently, the remaining 75% (Γ_{had}) must be related to hadronic decays (if one neglects the small contribution coming from leptonic decays $b \to \ell\nu$, which are $\mathcal{O}(10^{-4})$). However, most of this large Γ_{had} has unknown source: only a small fraction of it has been measured via exclusive decays, mainly comprising $\bar{D}^{(*)0}(n\pi)\,(1 \leq n \leq 3)$, $\bar{D}^{(*)0}D_s^{(*)+}$ and J/ψ modes and covering only $\sim$20%.

The b-quark predominantly decays via the $b \to cW^{*-}$ transition, resulting in a c-quark and a virtual W^-. The weak boson can in turn decay into $\bar{u}d$ or a $\bar{c}s$ pair and both decays are Cabibbo-favored (CF) as $|V_{ud}| \sim |V_{cs}| \sim 1$ but the second is suppressed by a phase-space factor. The first process, $b \to c\bar{u}d$, mainly produces $\overline{D}^0 n\pi$ decays while the $b \to c\bar{c}s$ transitions produces a flavor-anticorrelated c quark, which can either lead to a second 'open-charm' meson, as is it the case for $B \to D^{(*)}\bar{D}^{(*)}K^{(*)}(n\pi)$ decays, or to $(c\bar{c})X$ states (which have lower BFs due to color-suppression). As a result, the inclusive rate $\mathcal{B}(B^+ \to \overline{D}^0 X)$ is roughly 80% [13, 53]; the situation is less unbalanced for the B^0 meson, where the inclusive $\overline{D}^0 X$ and $D^- X$ rates are respectively 48% and 37%. This is due to the fact that $D^{*0} \to D^-\pi^+$ channel is kinematically forbidden, while the correlated D^{*-} can decay in both $\overline{D}^0$ and D^- mesons.

There has been quite a large effort in improving the theoretical description of B-decays; for example, Ref. [15] uses the heavy quark expansion (HQE) to make

predictions on the inclusive production. The decays rates into 0, 1 and 2 charm quarks are provided as a function of the most-reliable semileptonic decay rate Γ_{sl}:

$$\mathcal{B}_{b\to X} = \frac{\Gamma_X}{\Gamma_{sl}} \times \frac{\Gamma_{sl}}{\Gamma_{tot}} = r_X \times \mathcal{B}_{sl}^{exp} \tag{1.10}$$

which eliminates Γ_{tot} in favor of Γ_{sl} but needs as input the experimental semileptonic BR[12].

One has $r_{c\ell\nu} = 2.22 \pm 0.04$ for all the possible semileptonic decays and $r_{c\bar{u}d} = 4.1 \pm 0.4$ [54,55] and $r_{c\bar{c}s} = 1.95 \pm 0.5$ [55–57], which translate into rates of $(44.3 \pm 4)\%$ and $(21.1 \pm 2)\%$, respectively[13].

It is worth noting that the average multiplicity of charged final-state particles for hadronic B^+ decays is 5.8 ± 0.1 [58], which contains the contribution from the decay of the $\bar{D}^0$ meson. Therefore we expect a sizeable decay rate into final states with multiple pions $\bar{D}^{(*)}(n\pi)^+$. For $n \leq 3$, the decays are predominantly two-body ones, where $B^+ \to \rho^+(a_1^+)$ dominates the width for $n = 2$ ($n = 3$). They result in a *spectator* mechanism where the virtual W^+ decays into a single hadron: π^+, ρ^+ or a_1^+. In this context, the corresponding rates can be explained within a *factorization* model: they are obtained by the product of the currents between the B and the D mesons, and the other associated to the W^{*+} decaying into the hadronic system X. The corrections to this simple model, due to perturbative QCD effects, are expected to grow with m_X/E_X[14] since this ratio characterizes the deviation of the X system from the light cone (low $q^2 = m_X^2$ regime) [59]. In the limit where the pions are not emitted in the $c\bar{b}$ current (the charmed state is a $D^{(*)}$):

$$\frac{m_X c^2}{E_X} = \frac{2 m_X m_B}{m_B^2 + m_X^2 - m_{D^{(*)}}^2}.$$

Given these points, it is clear that an understanding of the $B \to Dn\pi$ production beyond $n = 3$ is fundamental to fill the gap between what has been measured so far and the expected 44% of the Γ_{tot} related to $b \to c\bar{u}d$. The task is hard both theoretically, because of the reasons explained above, and experimentally, as the size of combinatorial background grows very rapidly with n, especially when the pion system does not have any kinematic constraint. There are few exceptions though, for example the searches of $B \to D^*\pi^+\pi^-\pi^-\pi^0$ performed at CLEO [60] or the measurement of $B \to D^* n\pi$ ($n = 4, 5$) at Belle [61]. A more systematic effort would be needed from current experiments to shed light on these modes and provide theorists with valuable inputs.

In order to better understand the composition of the hadronic system coming from the $W^{*+} \to u\bar{d}$, a parallel can be made with the τ lepton phenomenology. The τ is the heaviest lepton and the second heaviest fundamental fermion and provides the perfect laboratory to study the structure of the weak charged current and the hadrons produced from the QCD vacuum. The probability of creating hadrons out of the QCD vacuum is encoded in spectral functions that depend on the q^2 (i.e. on the squared mass of the hadronic system), on the spin J of the hadronic system X_{had} and on its vector (V) or axial-vector (A) nature. Due to the $V - A$ nature of the charged weak interaction, the allowed states have $J^P = 0^{\pm}, 1^{\pm}$; furthermore the weak vector and axial-vector currents produce an even or odd number of pseudoscalars, corresponding

[12] We will use $\mathcal{B}(B^+ \to X_c e^+ \nu_e) = 10.8\%$ [13].
[13] When more than one r_X prediction is available, we take average of them.
[14] E_X denotes the energy of the system X in the B-rest frame.

to a G-parity[15] $G = 1$ and $G = -1$, respectively. Weak hadronic currents with spin-parity J^P are defined as of *first* or *second* class depending on their properties under G parity. The first-class decays, corresponding to $J^{PG} = 0^{++}, 0^{--}, 1^{+-}, 1^{-+}$ dominate hadronic τ decays whereas the combinations $J^{PG} = 0^{+-}, 0^{-+}, 1^{++}, 1^{--}$ refer to the second class and are suppressed (BFs of $\mathcal{O}(10^{-5})$) as their matrix element is proportional to $m_u - m_d$, which vanishes in the limit where the up- and the down-quark have the same mass (perfect isospin symmetry). The decay $\tau^- \to \eta\pi^-\nu_\tau$ belongs to the second-class family, the possible J^{PG} values for the $\eta\pi$ system being $0^{+-}, 1^{--}$. This mode has never been observed and *BABAR* set an UL on its BR: $\mathcal{B}(\tau^- \to \eta\pi^-\nu_\tau) < 9.9 \times 10^{-5}$ [62].

Depending on the number of pions n_π in the X_{had} we have:

- $n_\pi = 2$. The $\pi\pi^0$ spectral function is dominated by the broad ρ resonance.
- $n_\pi = 3$. The decay $\tau \to 3\pi\nu_\tau$ is the cleanest mode to study axial-vector resonance structure. The spectrum is dominated by the $J^P = 1^+$ a_1 state, known to decay essentially through $\rho\pi$.

Concerning $n_\pi = 4$, there has been an attempt [59] to compare the $4\pi^-$ invariant mass spectrum of $B \to D(4\pi)$ decays with $\tau \to \pi^+\pi^-\pi^-\pi^0\nu$ data (see Fig. 1.7). In fact, a large BF of $(1.72 \pm 0.14 \pm 0.24)\%$ was measured by CLEO [60] for the decay $B^0 \to D^{*+}\pi^+\pi^-\pi^-\pi^0$. Using a model based on factorization, Ref. [59] shows that the two 4π spectra data agree up to a mass-squared of $2.9\,\text{GeV}/c^2$, within a precision of about 15%. This can be explained by the fact that in both cases the simplest diagram has the four-pions emitted from the virtual W. Despite the good agreement, the authors are cautious about describing it as a success success of factorization.

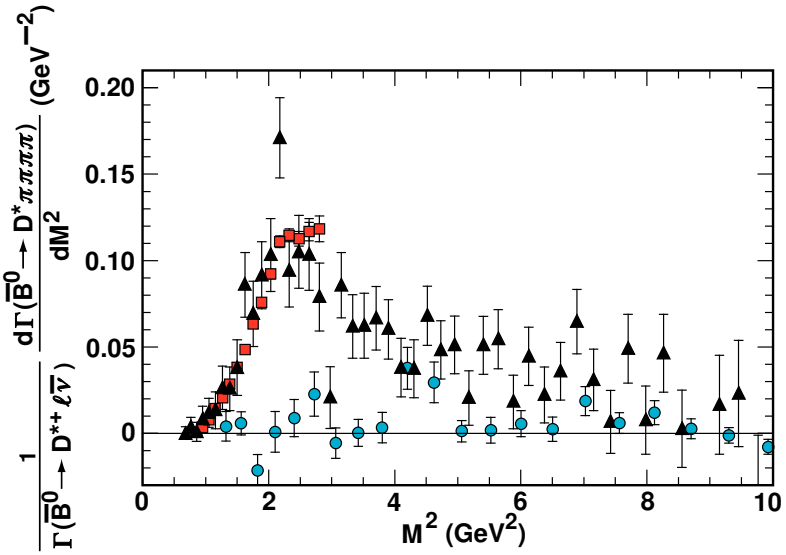


FIGURE 1.7: $\frac{d\Gamma}{dM^2}(\bar{B}^0 \to D^*\pi\pi\pi\pi)$, where M^2 is the squared invariant mass of the 4π-system, normalized to the semileptonic width $\Gamma(\bar{B}^0 \to D^*\ell^-\bar{\nu})$. The triangles are the CLEO data for $\bar{B}^0 \to D^{*+}\pi^+\pi^+\pi^-\pi^0$, the circles for $\bar{B}^0 \to D^{*0}\pi^+\pi^+\pi^-\pi^-$ and the squares the model prediction of Ref. [59].

The τ sector is not the only laboratory where spectral functions can be studied in detail: the e^+e^- data provide a complementary source owing to the fact that in the limit of isospin invariance the vector current is conserved (CVC). This implies that

[15] A combination of charge conjugation and isospin rotation.

the spectral function of a vector τ decay mode in a given isospin state for the $X_{\rm had}$ is related to the e^+e^- annihilation cross section of the corresponding isovector final state X_0 [63]. The advantage of e^+e^- data compared to τ's is that in the first case larger invariant masses for the $X_{\rm had}$ can be explored,

The problem of finding the contributions to $B \to DX$ is not only related to the hadronic system coming from the W^*: there are D^{**} states that contribute to the $Dn\pi$ system and are not always well modeled. With D^{**} we denote the ensemble of $c\bar{u}$ states, or D_J mesons, where the charm quark and the light quark are in a P-wave state ($L = 1$). They are called D_0^*, D_1', D_1 and D_2^*; the first two correspond to a $j = 1/2$[16] and, because of parity and angular momentum conservation in strong interactions is conserved, they can decay to S-wave $D\pi$ and $D^*\pi$ states. D_1 and D_2^*, instead, have $j = 3/2$ and decay through a D-wave ($D_1 \to D\pi$ and $D_2^* \to D^{(*)}\pi$) and have much narrower widths ($\mathcal{O}(10\,\text{MeV}/c^2)$) with respect to the other two.

The properties of these states, as well as the ansatz on the possible decay modes and their BFs are listed in Tab. 1.3. The BFs, based on the conservation of quantum numbers, phase-space constraints and isospin considerations do not include possible states involving ρ's or η's; such possibility is instead explored in Ref. [64] and represented in Fig. 1.8.

D^{**}	Mass (MeV/c^2)	J^P	Γ (MeV/c^2)	Dec. channel	BF
$D_0(2300)^0$	2343 ± 10	0^+	222 ± 16	$D^0\pi^0$	0.3333
				$D^+\pi^-$	0.6667
$D_1(2430)^0$	2412 ± 9	1^+	314 ± 29	$D^{*+}\pi^-$	0.6667
				$D^{*0}\pi^0$	0.3333
$D_1(2420)^0$	2422.1 ± 0.6	1^+	31.3 ± 1.9	$D^{*0}\pi^0$	0.1997
				$D^{*+}\pi^-$	0.3994
				$D^0\pi^+\pi^-$	0.1719
				$D^0\pi^0\pi^0$	0.1145
				$D^+\pi^-\pi^0$	0.1145
$D_2^*(2460)^0$	2461.1 ± 0.8	2^+	47.3 ± 0.8	$D^{*0}\pi^0$	0.1334
				$D^{*+}\pi^-$	0.2669
				$D^0\pi^0$	0.1999
				$D^+\pi^-$	0.3998

TABLE 1.3: Decay channels of D^{**}. Masses and widths are taken from Ref. [13].

The problem of determining the precise BFs for these resonances and their decay modes has also been studied with the aim to find a the source of the *gap* between the measured inclusive $B \to X_c\ell\nu$ rate and the sum of exclusive $B \to D^{((*)*)}\ell\nu$ modes [64].

The elements presented in this chapter will be invoked several times throughout this manuscript. In particular, the theory on $b \to s\ell\ell'$ transitions will be further explored in Appendix A, in conclusion to the results obtained in chapter 3, while the theory on B-decays will be extensively exploited in chapter 4.

[16] j is the sum of the light quark spin and the orbital angular momentum $\vec{L}$.

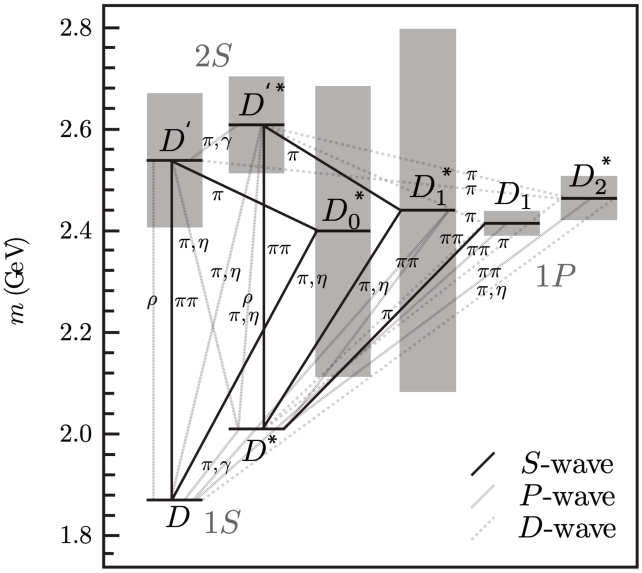


FIGURE 1.8: Strong decays of D^{**} mesons ($2S$ and $1P$ states). The bands correspond to the measured widths. Picture from Ref. [64].

Chapter 2

The Belle and Belle II experiments

In this chapter we describe the main features of the Belle experiment and of the accelerator KEKB that delivered the dataset used for the present work. From 2019, the new experiment Belle II at the upgraded accelerator, SuperKEKB, aims to collect a 50 times larger dataset in order to give a significant contribution to flavor physics. The key elements behind such upgrades will also be outlined, in order to prepare the reader to a discussion on the prospects of our analysis, mainly discussed in chapter 5.

2.1 The KEKB and SuperKEKB accelerators

The KEKB B-Factory was a double-ring e^+e^- collider at KEK, Japan, in operation from December 1998 to June 2010. It consisted of a low-energy ring (LER) for the 3.5 GeV-positrons, a high-energy electron ring (HER) for the 8.0 GeV-electrons, and an injector Linac, as sketched in Fig. 2.1. The two beams collided at interaction point (IP) around which the Belle detector was installed. In the same period (from 1999 to 2008) the *BABAR* experiment took place at the PEP-II accelerator (SLAC), providing a vital complementarity and competition with Belle in the common search for *CP* violation in B mesons.

The luminosity of $2.10 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, reached by KEKB in June 2009, has been the highest ever recorded until 2021, when the upgraded machine SuperKEKB obtained the same result. A new record has been reached by SuperKEKB on 22 June 2022: $4.71 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The luminosity of a collider is approximated by the following formula:

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

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. [66]). $\sigma^*_{x(y)}$ 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 on the beam focusing and varies with the position s along the ring due to the betatron oscillations [67]. 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

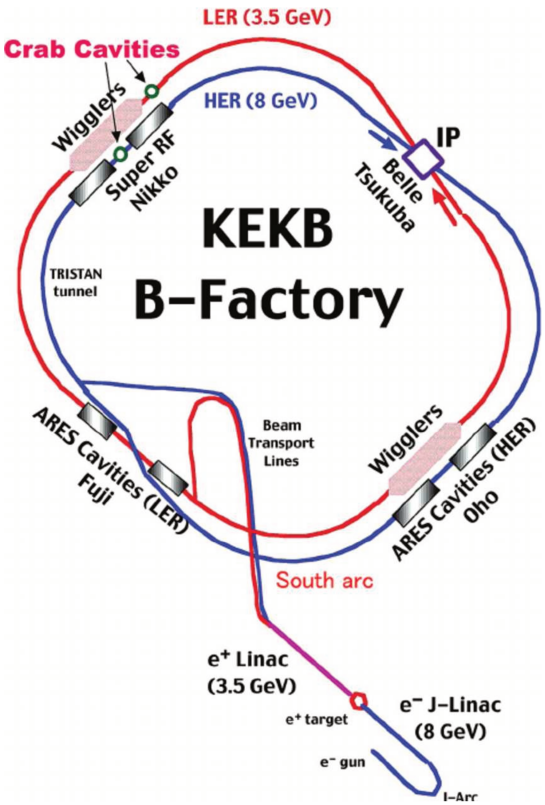


FIGURE 2.1: Pictorial representation of the KEKB facility complex, mainly composed of a linear accelerator (Linac) and the two main storage rings [65].

at the IP. The luminosity can be increased by rising the I and lowering the beam sizes. Indeed, at SuperKEKB the design currents are increased from 1.6A/1.2 A to 3.6/2.6 A, while the vertical size of the bunch σ_y is reduced for 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 mm) is longer than β_y^*, the actual luminosity is usually smaller than the nominal one; this effect is known as the *hourglass effect*.

At KEKB, the achieved peak luminosity was twice as high as the design value, despite the LER beam current was lower than the design (the HER beam current was higher, though). The daily integrated luminosity was also twice as high as the design because of the continuous injection mode and the usage of two bunches per radio-frequency (RF) pulse at the Linac [68]. In addition, chromatic corrections of the IP β functions were applied by means of skew-sextupole magnets installed in the winter of 2009 and contributed to the great performance summarized in Tab. 2.1, in comparison to the expected values for SuperKEKB.

At an electron-positron collider it is possible to produce B mesons by tuning the center-of-mass energy to the $\Upsilon(4S)$, the lightest strong resonance - the third among the radial excitations of the Υ[1] - with a mass sufficient to decay in b-flavored mesons ($B^0\overline{B}^0$ and B^+B^- pairs). Indeed the beauty mesons' threshold is at $2m_B \simeq 10.56$ GeV and

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

	KEKB Achieved	**SuperKEKB Design**
Energy (GeV)	3.5/8.0	4.0/7.0
ξ_y	0.129/0.090	0.090/0.088
β_y^*(mm)	5.9/5.9	0.27/0.41
I (A)	1.64/1.19	3.60/2.62
Luminosity ($10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	2.11	80

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

$m_{\Upsilon(4S)} \simeq 10.58\,\mathrm{GeV}$; this also explains the fact that this resonance has larger width than those of lower mass (see Fig. 2.2).

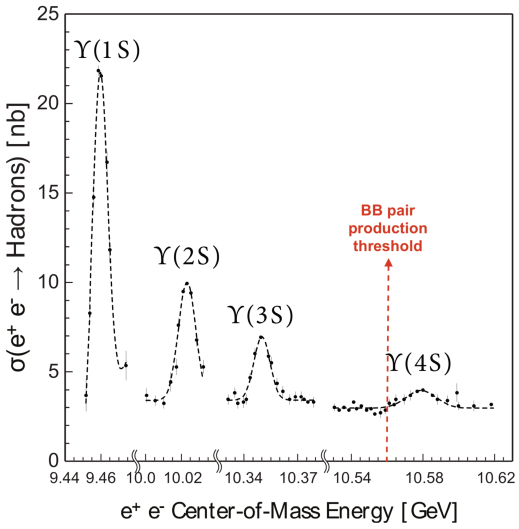


FIGURE 2.2: Hadronic cross section as a function of the center-of-mass energy in the region of the first four S-wave Υ resonances.

However, e^+e^- collisions at the $\Upsilon(4S)$ energy do not produce only $B\overline{B}$ pairs, but also a large amount of QED processes (Bhabha, di-gamma events), di-muon and di-tau events, as well as $q\overline{q}$ pairs. The cross sections of the main processes are listed in Tab. 2.2.

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

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 two storage rings must have different energies so that the $\Upsilon(4S)$, and therefore the pair of B's to which it decays, are produced with a significant relativistic boost $\beta\gamma$ along the beam axis, allowing precision measurements of lifetimes, mixing and decay-time-dependent CP violation parameters, the latter being the initial goal of B-factories. 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

$$\beta\gamma = \frac{|\mathbf{p}_H + \mathbf{p}_L|\,c}{\sqrt{s}} \simeq \frac{E_H - E_L}{\sqrt{4E_{e^-}E_{e^+}}} \simeq 0.42$$

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 of only $\sim 130\,\mu$m. The KEKB operations, visually summarized in Fig. 2.3, not only allowed to collect the largest $\Upsilon(4S)$ data sample to date, but also significantly large samples at others $b\bar{b}$ resonances like $\Upsilon(2S)$ and $\Upsilon(5S)$. The total recorded samples are listed in Tab. 2.3.

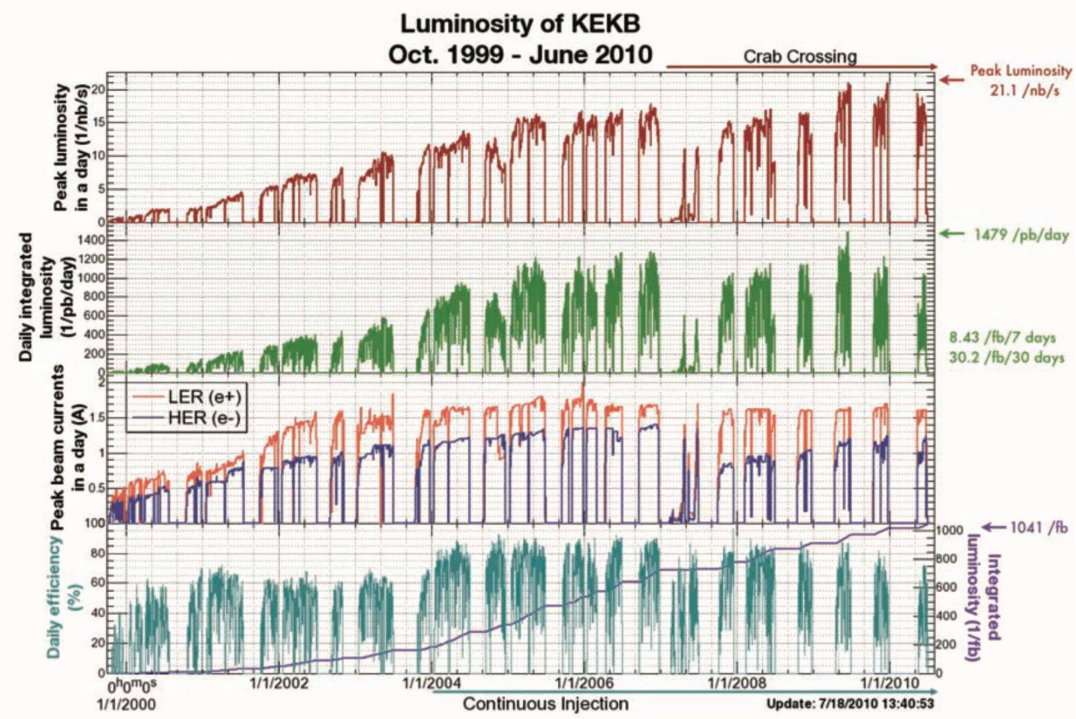


FIGURE 2.3: KEKB operation summarized in terms of beam currents, luminosity integrated per day, data acquisition efficiency and total integrated luminosity [68].

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

TABLE 2.3: Samples collected during Belle operation at different center-of-mass energies. From Ref. [69]

2.2 The Belle and Belle II detectors

Detectors at B-factories must provide a large acceptance and a high efficiency for B-decays to maximise the number of collected events, good momentum and energy resolutions to separate the small signal from the backgrounds, very good vertex position resolution, efficient and robust particle identification capabilities for hadrons. Since the average charged particle momentum is below $1\,\mathrm{GeV}/c$, the minimization of the amount of material producing multiple scattering is vital. The same necessity holds for the amount of uninstrumented material in front of the electromagnetic calorimeter, which has to detect showers with energy as low as 20 MeV. The detector has an asymmetric polar geometry reflecting the boost and comprises: silicon vertex detectors close to the beam pipe for secondary vertex measurement and to provide additional tracking points, drift chamber with helium-based gas for charged particle tracking and particle identification through ionisation and energy loss, Čerenkov detectors for particle identification and the CsI(Tl) crystal calorimeter for the measurement of neutral particles. The high light yield and small Molière radius of CsI crystals allow excellent energy and angular resolutions; moreover the high yield permits the use of silicon photodiodes, which can work in magnetic fields; all the above detectors are inside a 1.5 T cryogenic superconducting solenoid at a temperature of 4.5 K. 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 (shown in Fig. 2.4) reuses the structure, the solenoid, the CsI(Tl) crystals and part of the barrel RPCs from Belle, while most other components of the sub-detectors are new.

2.2.1 Coordinate system

The coordinate system is defined as follows:

(i) The x axis lies in the horizontal plan and points towards the outside part of the ring tunnel

(ii) The y axis is vertical and points upwards

(iii) The z axis coincides with the Belle 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).

In polar coordinates:

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

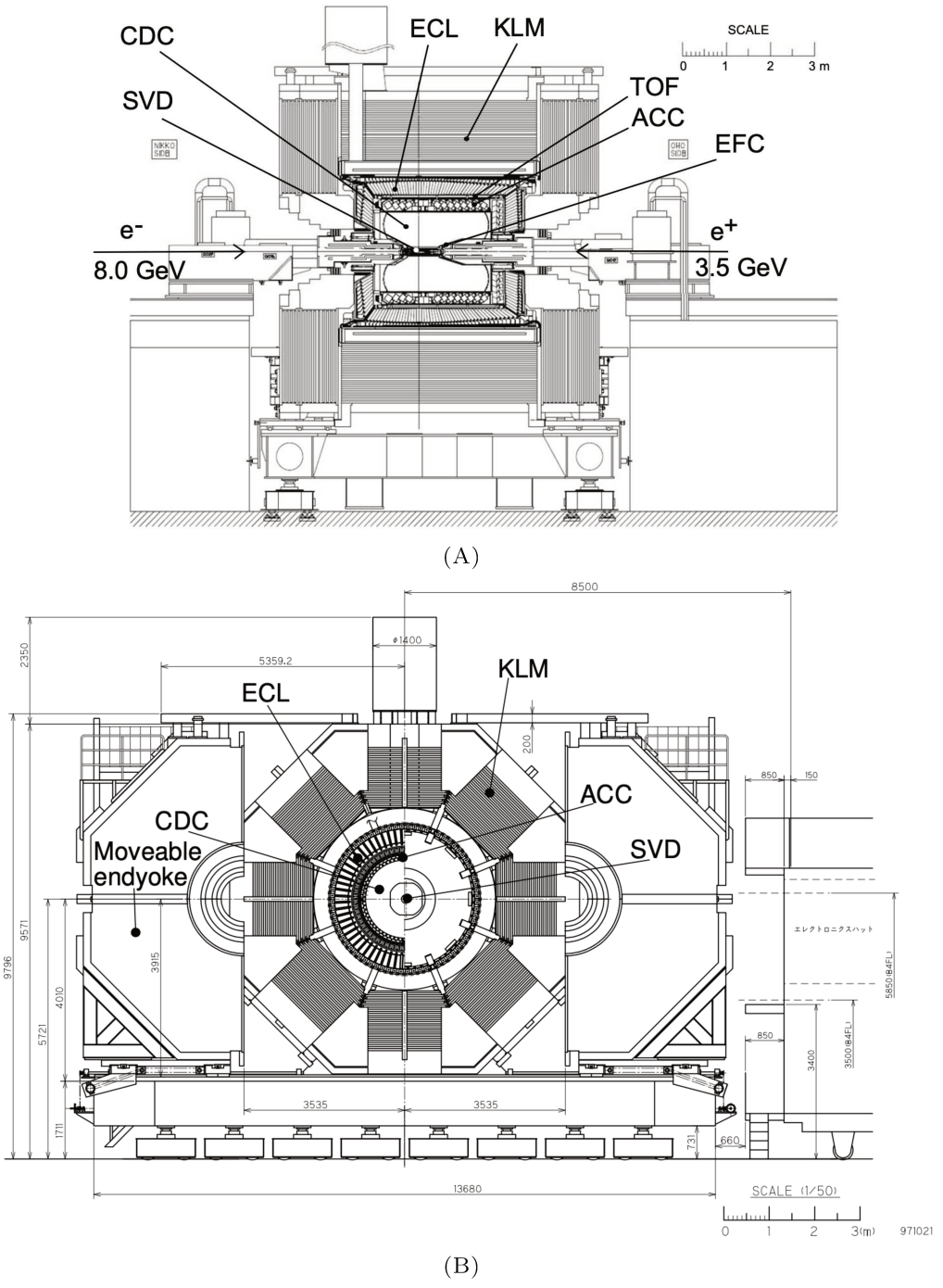


FIGURE 2.4: Belle detector: longitudinal (A) and side (B) views. From Ref. [70].

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

(ii) ϕ (azimuthal) angle $\in [0°, 360°]$, with $\phi = 0$ corresponding to the point with coordinates $(x, y, z) = (1, 0, 0)$

(iii) θ (zenith) angle $\in [-180°, 180°]$, with $\theta = 0$ corresponding to the point with coordinates $(x, y, z) = (0, 0, 1)$

We refer to *longitudinal* for the direction parallel to the z axis while *transverse* means lying in the $x - y$ (or $\rho - \phi$) plane. Furthermore, we will refer to *forward* as the direction towards positive z coordinates, *backward*, otherwise.

2.2.2 Tracking system

A superconducting solenoid provides the magnetic field of 1.5T 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 detector serves as the return path of magnetic flux and an absorber material for KLM. It also provides the overall support for the sub-detectors.

Silicon Vertex Detector

The main goal of the Belle vertex detector was to ensure a resolution on the B-decay lengths shorter than $200\,\mu$m. In fact many CPV studies rely on the measurement of time-dependent decay rate asymmetries of the two B mesons, where the decay time difference Δt is inferred from the decay length difference $\Delta z = \beta\gamma c\Delta t$. The first version of the vertex detector, called SVD1 and in operation from 1999 to 2003, consisted of three layers of double-sided silicon strip detectors (DSSDs), covering 86% of the full solid angle. The radii of the three layers were 30, 45.5 and 60.5 mm and each layer was constructed from independent mechanical units. The main limitations of this device were given by the poor radiation tolerance of the readout electronics, the tendency to suffer from the sensor damages and the dead time caused by the readout system. For this reason, in summer 2003 it was replaced by a four-layer, second generation silicon vertex detector, the SVD2. At the same time, the beam pipe was replaced with a narrower one (radius from 40 mm to 30 mm) to enable the radius of the innermost SVD2 layer to be reduced – as demonstrated by Fig. 2.5.

The main features of the upgraded detector were:

- Increased solid angle coverage ($17° < \theta < 150°$)[3]
- The innermost layer closer to the IP (2 cm against the 3 cm of SVD1) in order to achieve a better vertex resolution
- Improved charged particle tracking in the low momentum region and higher resolution for vertexing
- Improved radiation tolerance.

The impact parameter resolution in r-ϕ and r-z was measured to be $\sigma_r = 21.9 \oplus 35.5/p\,\mu$m and $\sigma_z = 27.8 \oplus 31.9/p\,\mu$m[4], respectively, where p represents the track momentum in GeV/c. Given that the outermost SVD layer and the innermost CDC layer were 88 mm and 110 mm, respectively, the reconstruction of low p_T tracks could be done by the CDC. Thus, the main purpose of the Belle SVD was to extrapolate the tracks reconstructed in the CDC to the decay vertices inside the beam pipe. The reconstruction of low p_T tracks with the CDC was efficient down to 70 MeV/c.

Upgraded components in Belle II: PXD and SVD

The Belle II vertex detector (VXD) consists of two devices, the silicon pixel detector (PXD) and a silicon vertex detector (SVD), with altogether six layers around a 10-mm radius Beryllium beam pipe. The first 2 layers have radii of respectively 14 mm and 22 mm. Such proximity to the IP ensures a better vertex resolution and higher K^0_S

[3] It corresponds to the total Belle solid-angle coverage, i.e. 91% of 4π.

[4] The $\oplus$ sign denotes summation in quadrature.

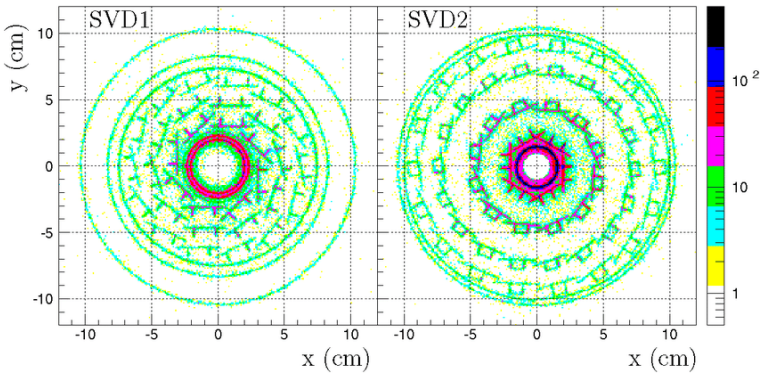


FIGURE 2.5: Distribution of the reconstructed pK_S^0 vertices in the Belle detector [71], allowing to distinguish and compare in the transverse plane the main differences between SVD1 and SVD2 configurations: reduction of the beam pipe diameter, addition of a fourth layer and larger radius of the outermost layer.

reconstruction efficiency[5], but also requires a higher radiation tolerance and acceptable occupancies, due to the fact that the background increases roughly with the inverse square of the distance from the IP. The DEPFET (DEPleted Field Effect Transistor) technology used for the PXD will provide precise vertex reconstruction in the harsh environment envisaged at the final stages of SuperKEKB. Furthermore, the 50 μm-thin sensors allow to minimize the multiple-Coulomb scattering and ensure a precise reconstruction of B-decay vertices.

At radii exceeding 30 mm, strip detectors are safe in terms of occupancy; for this reason, the SVD inherits from the Belle SVD2 the double-sided silicon strip sensor design and all the characteristics required at B-factories: low mass, high precision, immunity to background hits, radiation tolerance and long-term stability. The angular acceptance is the same as in Belle while the outer layer has larger radius (135 mm).

Central Drift Chamber

The central drift chamber (CDC) was the most crucial sub-detector, as it is required to:

- enable the reconstruction of charged particles trajectories - full 3D helix track via tracking in the magnetic field;
- provide information on the energy loss due to specific ionization in its volume ($\mathrm{d}E/\mathrm{d}x$) for the particle identification;
- participate to the global trigger decision (r-ϕ and z triggers).

The structure of the Belle CDC is shown in Fig. 2.6: it was an asymmetric volume in the z direction with an angular coverage of $17° < \theta < 150°$ and a maximum wire length of 2400 mm. The inner radius of the CDC lied at 110 mm and the outer radius was 880 mm. The chamber had 50 cylindrical layers with 8400 drift cells in total. Some modifications to the inner part of the CDC were necessary in 2003 to make room for the new, four-layer SVD layers. The volume was filled with He(50%):C_2H_6(50%) gas mixture which, because of the low Z nature of the gases, provided optimal momentum

[5]Due to the decrease of the Lorentz factor at Belle II, the smaller radii ensure the same resolution on Δz than in Belle.

resolution while retaining good energy loss resolution - important for minimizing the multiple scattering.

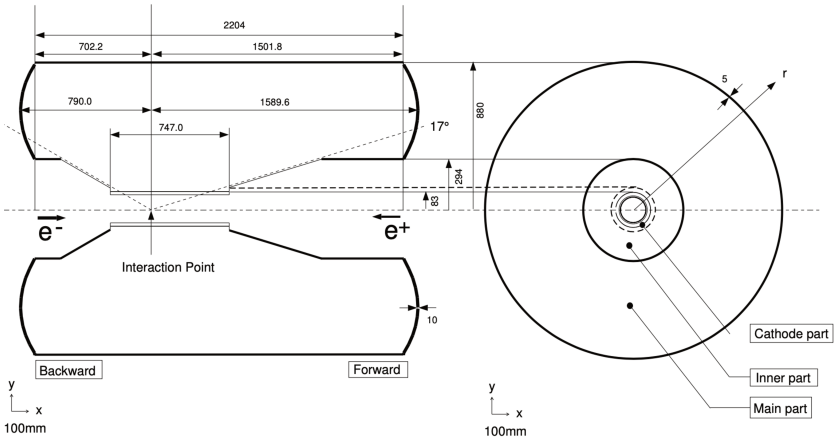


FIGURE 2.6: CDC lateral view. Picture from Ref. [70].

After detailed alignment and calibration, the overall spatial resolution was around 130 μm. The tracking system consisting of the SVD and CDC provided good momentum resolution, especially for low-momentum tracks thanks to the minimization of material inside the inner radius of the CDC: $\sigma(p_T)/p_T = 0.0019 p_T \oplus 0.0030/\beta$ (where p_T is expressed in GeV/c). The resolution on dE/dx, important for PID, was 7% for minimum-ionizing particles.

Upgraded components in Belle II: CDC

The Belle II CDC contains 14336 sense wires arranged in 56 layers; the main improvements are the longer lever arm (i.e. larger radius), which yields better momentum resolution, fast electronics and smaller drift cells. Even though the larger beam background in Belle II degrades the CDC tracking performance, the new software and the standalone tracking in the PXD and SVD together improve the overall reconstruction efficiency so that better performance with respect to Belle should be obtained.

2.2.3 Charged hadron identification

Particle identification for hadrons (in particular kaons and pions) was obtained by combining the information from three sub-detectors: the time-of-flight detector (TOF), the aerogel Čerenkov counter (ACC) and the CDC via the dE/dx information, as discussed in the previous section. Below are summarized the features of the other two.

Time of Flight Counters

The time-of-flight (TOF) system consisted of a barrel of 128 plastic scintillator counters that helped to distinguish between kaons and pions for tracks with momenta below 1.2 GeV/c. The system was designed to have a time resolution of 100 ps for muon tracks. The TOF system measured time of flight for charged tracks reconstructed by the CDC from the beam collision time of each event, t_{IP}. t_{IP} was determined by the RF clock signal used as a reference, and the time offset was calibrated offline on a run-by-run basis with high purity. The expected TOF for each charged particle was calculated taking into account the flight length measured by the CDC and tuning

the offset to give a zero deviation on average between the calculation and the TOF measurement for each photomultiplier.

Aerogel Čerenkov Counter

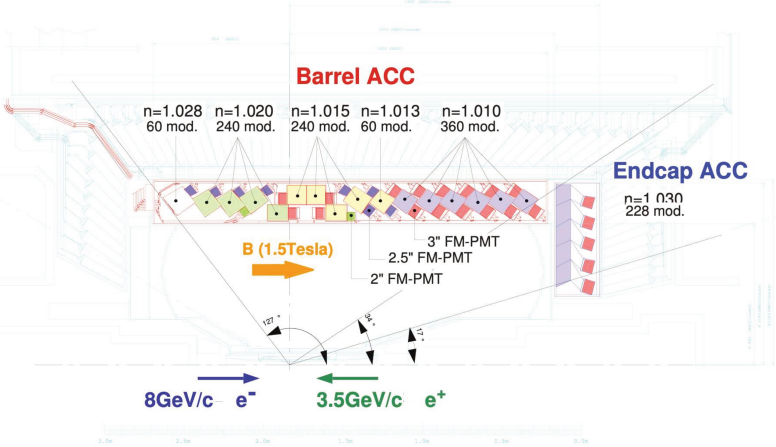


FIGURE 2.7: ACC lateral view. Picture from Ref. [70].

The Belle aerogel Čerenkov counter (Fig. 2.7) had an azimuthal angle coverage of $33.3° < \theta < 127.9°$ in the barrel and $13.6° < \theta < 33.4°$ in the forward end-cap. The detector was made of aerogel modules of ten types, varying in refractive index and in the number and size of fine-mesh photomultiplier tubes used to detect Čerenkov light in an axial magnetic field of 1.5 T, according to their position in polar angle.

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 [70]:

$$L(\alpha:\beta) = \frac{\mathcal{L}_\alpha^{\mathrm{CDC}}\mathcal{L}_\alpha^{\mathrm{TOF}}\mathcal{L}_\alpha^{\mathrm{ACC}}}{\mathcal{L}_\alpha^{\mathrm{CDC}}\mathcal{L}_\alpha^{\mathrm{TOF}}\mathcal{L}_\alpha^{\mathrm{ACC}} + \mathcal{L}_\beta^{\mathrm{CDC}}\mathcal{L}_\beta^{\mathrm{TOF}}\mathcal{L}_\beta^{\mathrm{ACC}}} \tag{2.2}$$

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 $L(K:\pi)$:

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

The performance obtained with the cut $K\mathrm{ID} > 0.6$ is shown in Fig 2.8: the efficiencies and the mis-identification rates for the kaon and pion identification were measured with the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ control sample for kaons in the barrel region. Discrepancies between data and MC were observed, especially in the mis-identification rate, but for both data and MC the kaon efficiency exceeded 80%, while the pion fake rate was kept below 10%.

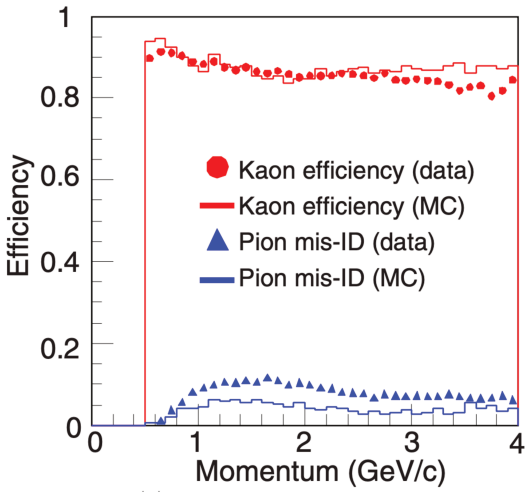


FIGURE 2.8: Kaon identification efficiency and mis-identification rate to pions as a function of the tracks' momentum. Data and MC are compared. Plot from Ref. [70]

Upgraded components in Belle II: TOP and ARICH

For the Belle II experiment, the particle-ID has gone through a substantial upgrade in both barrel and end-cap regions, in order to provide good performance over the full kinematic range of the experiment and minimize the material in front of the calorimeter. The detectors used for the hadron ID are: the time-of-propagation (TOP) counter, used in the barrel region, and the ARICH, for the forward end cap region only. The first is a special kind of Čerenkov detector in which two-dimensional information about a Čerenkov ring image is given by the time of arrival and impact position of Čerenkov photons at pixelated photo-detectors at one end of a 2.6 m long quartz bar.

In the forward end-cap, the proximity focusing Čerenkov ring imaging detector (ARICH) has been designed to separate charged kaons from pions over most of their momentum spectrum and to provide discrimination between pions, muons and electrons below 1 GeV/c. The ARICH is composed by an aerogel radiator, where Čerenkov photons are produced by charged particles, an expansion volume to allow the photons to form rings on the photon detector surface, an array of position-sensitive photon detectors, that is capable of detecting single photons in a high magnetic field with high efficiency and with good resolution in two dimensions, and a read-out system for the photon detectors.

2.2.4 Neutrals: Electromagnetic Calorimeter

Given that around 1/3 of B-decay products are π^0's or other neutral particles, producing photons in a wide energy range (20 MeV to 4 GeV), a high-resolution electromagnetic calorimeter was a fundamental part of the Belle experiment. CsI(Tl) was chosen as the scintillation crystal material because of its high light output, relatively short radiation length, good mechanical properties and moderate price. The main tasks of the calorimeter were:

- 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 overall configuration of the Belle calorimeter, the ECL, is shown in Fig 2.9. It consisted 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.

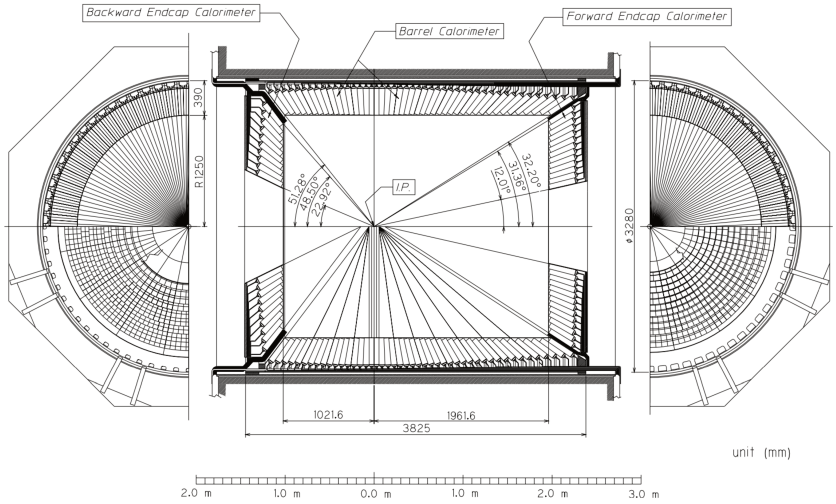


FIGURE 2.9: Exploded view of the ECL, showing the end-cap (forward and backward) and barrel regions with the corresponding polar acceptance. Picture from Ref. [72].

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 $4.5\,\mathrm{MeV}/c^2$.

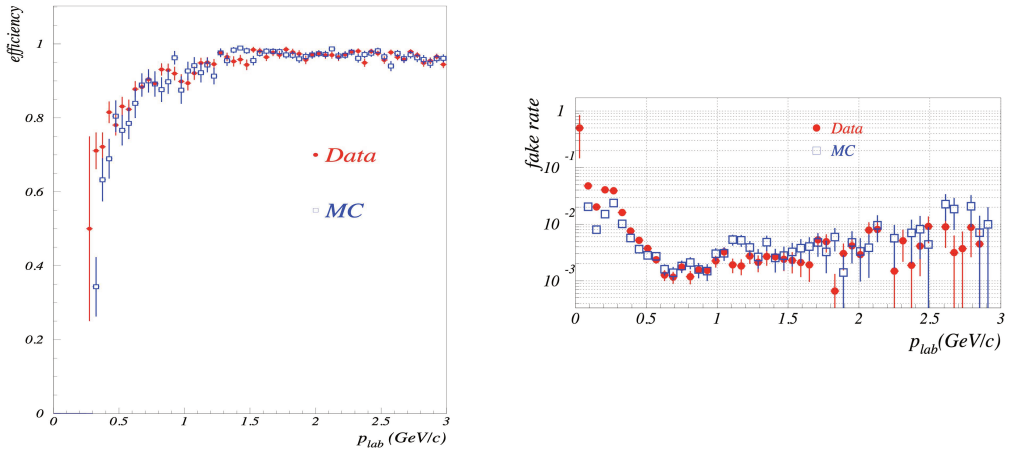
For electron identification, in addition to CDC ($\mathrm{d}E/\mathrm{d}x$) and ACC (light yield) information, the ECL was used to form the ratio of likelihoods. In fact, ECL provided the matching between the position of the track and the energy cluster, the E/p[6] and the information on the transverse shower shapes.

The variable used for electron identification was defined as

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

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

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 Fig. 2.10A. 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 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 Fig. 2.10B), obtained with $K^0_S \to \pi^+\pi^-$ events, was (0.25 ± 0.02)% for tracks' momenta between 1.0 and 3.0 GeV/c.



(A) (B)

FIGURE 2.10: eID> 0.5 performance. A: Efficiency in radiative Bhabha events as a function of the tracks' momentum. B: 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. Adapted from Ref. [73].

Upgraded components in Belle II: ECL

For Belle II, the CsI(Tl) crystals, preamplifiers and support structures have been reused, whereas the readout electronics and the reconstruction software have been upgraded. In absence of backgrounds, a very similar performance is obtained at Belle II. However, due to considerably higher background levels, the relatively long decay time of scintillations in CsI(Tl) crystals is expected to increase the overlapping of pulses from contiguous background events. To mitigate the resulting pile-up noise, the photo-sensors are equipped with waveform-sampling read-out electronics, allowing for the storage of ADC samples in FPGA internal buffers and the waveform fitting to discriminate signal from off-timing hits.

2.2.5 Muon and K_L^0 identification: KLM detector

The KLM was designed to identify K_L mesons and muons above 600 MeV/c with high efficiency. It consisted 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. 2.4).

The detector consisted of alternating layers of double-gap resistive plate counters (Fig. 2.11) and 4.7 cm thick iron plates. There were 15 detector layers and 14 iron layers in the octagonal barrel region and 14 detector layers and 14 iron layers in each end cap. The iron plates provided a total of 3.9 interaction lengths of material (in addition to the 0.8 X_0 in the ECL) for a hadron traveling normal to the detector planes. The hadronic shower from a K_L^0 interaction determined its direction (assuming an origin at the IP) but not its energy. The range and transverse deflection of a non-showering charged particle allowed to discriminate between muons and hadrons. The double-gap design resulted in a superlayer 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.

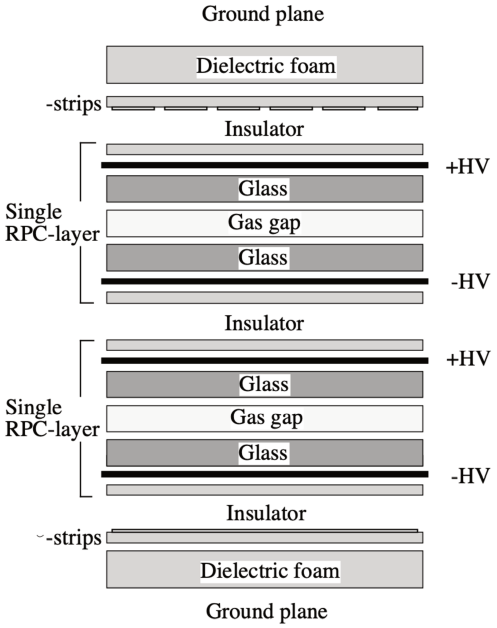


FIGURE 2.11: KLM's RPC structure. Picture from Ref. [70].

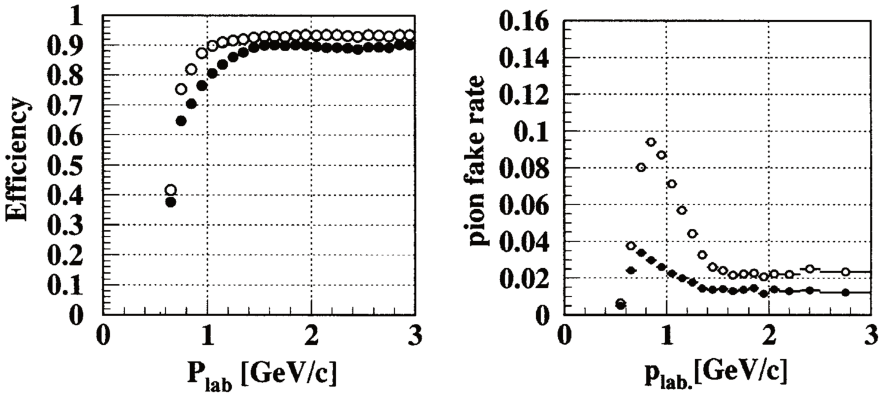
At Belle II, where larger backgrounds occur in both end-caps and the innermost layers in the barrel region due to neutrons, the RPCs have been replaced by layers of scintillator strips with wavelength shifting fibers, which are tolerant to higher rates. Such upgrade is expected to mitigate the long dead time of the RPCs due to the recovery of the electric field after a discharge, impacting the detection efficiency at high rates.

For muon identification, reconstructed hits in the KLM were compared to the extrapolation of the CDC track, 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, normalised 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. The kaon identification efficiency at $\mu\text{ID} > 0.9$ was larger than 80% for tracks' momenta above $1\,\text{GeV}/c$ while the fake rate for pions reached $(1.35 \pm 0.07)\%$ in the momentum region $1.0 - 3.0\,\text{GeV}/c$ (Fig. 2.12).



(A) (B)

FIGURE 2.12: μID> 0.9 performance measured with two-photon events $e^+e^- \to e^+e^-\mu^+\mu^-$. 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. Adapted from Ref. [74].

2.2.6 Trigger and Data Acquisition System

The Belle experiment operated with two levels of trigger; the first (hardware) called L1 and the second (software) called L3 (Fig. 2.13A). The L1 trigger combined the information from sub-triggers coming from CDC, ECL, TOF and KLM into a Global Decision Logic (GDL) (Fig. 2.13B). The L3 software trigger ran on an online computer farm and would receive from L1 the events with charged particles candidates. Other events, labelled by L1 as Bhabha, di-muon, two-photon or cosmic rays or having large energy deposits in the ECL would skip the L3 decision. The overall L1 efficiency for *hadronic* events was $> 99\%$. The final selection was performed offline by a so-called L4 trigger which would process the raw data and apply loose cuts as a first step for physics analyses.

The obtained trigger rate as a function of the operation time[7] is shown in Fig. 2.14: after experiencing an initial phase with very high rates, considering the low currents and luminosity, the GDL was adjusted to reduce the beam background effects. As a consequence, the normalised trigger rate (i.e. divided by the average luminosity) went down and was stable around 400 Hz until the end of operation, thanks also to better vacuum condition at the IP at higher currents. The total current, that reached its peak $\sim 3\,\text{A}$ around experiment 50, was drastically lowered by the introduction of the crab RF cavities from experiment 57[8].

[7] A run is a period of data-taking lasting typically from a few minutes to a few hours, while *experiment* indicates a period of data-taking typically lasting for several weeks, during which it is expected that there will be no major changes in Belle (II) or (Super)KEKB operating conditions. Each experiment contains many runs.

[8] Radio-frequency cavities deflecting the beam bunches giving them a rotational kick. This way, the bunches could go through head-on collision. This setup allowed to reach higher beam-beam parameter ξ_y and overall better beam-beam performance.

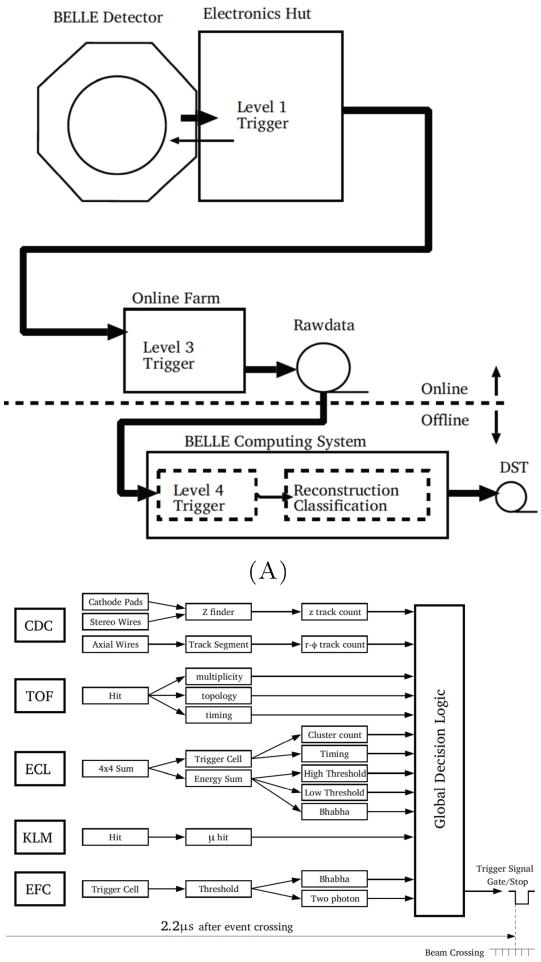


FIGURE 2.13: A: Belle Trigger scheme and B: Logic behind to global trigger decision of L1 stage. Figures from Ref. [70].

The Belle data acquisition system (DAQ) was required to read out event fragments from 8 detector subsystems with a total data size of 40 kbytes/event at a maximum rate of 500 Hz, and to record the data after event building and data reduction by real time processing. The initial read-out technology, unified for all the sub-detectors except for SVD, became very soon difficult to maintain; this and the increasing dead time due to the FASTBUS based readout system made necessary frequent upgrades of the DAQ system to cope with the luminosity increase.

The first change was made in 2001 to replace the event builder and the VME-based online computer farm that was used also for the L3 trigger stage. In 2003, the real time reconstruction farm (RFARM) was introduced. An improvement to the back-end system was made in 2005, when a second EFARM and RFARM were added in order to have sufficient bandwidth and processing power to cope with the expected increase in luminosity. For further reduction of the readout dead time, an upgrade of the FASTBUS readout system, to a pipelined version, was started. A new TDC was developed based on COPPER, a common pipeline readout module developed at KEK. The upgrade was performed detector by detector during the summer shut-down periods. The dead time was reduced to less than 1% as a consequence of upgrading 5 sub-detectors. Figure 2.15 shows the Belle DAQ configuration at the end of data-taking.

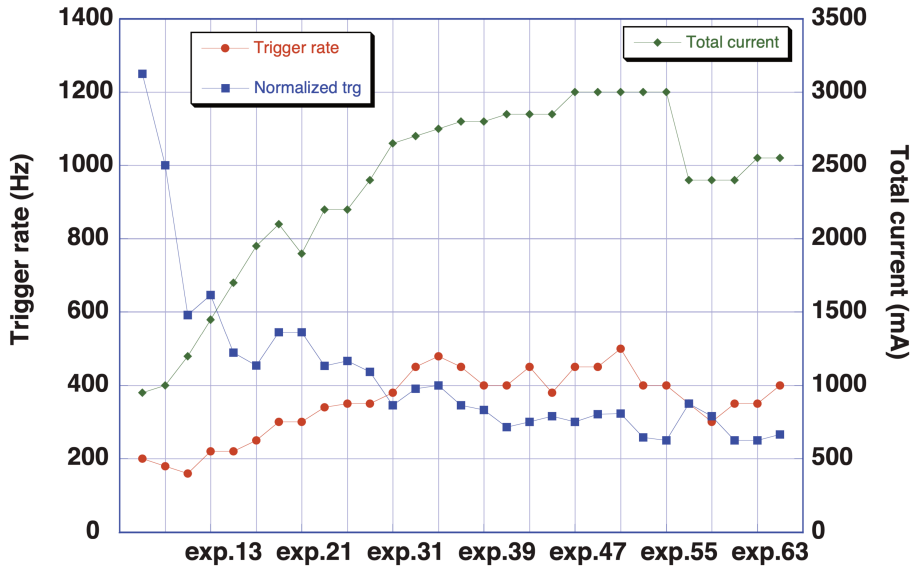


FIGURE 2.14: Hardware (L1) trigger rate as a function of the data-taking 'experiment' number. The trigger rate normalized to the luminosity is also shown, as well as the total average current from the colliding beams. From Ref. [70].

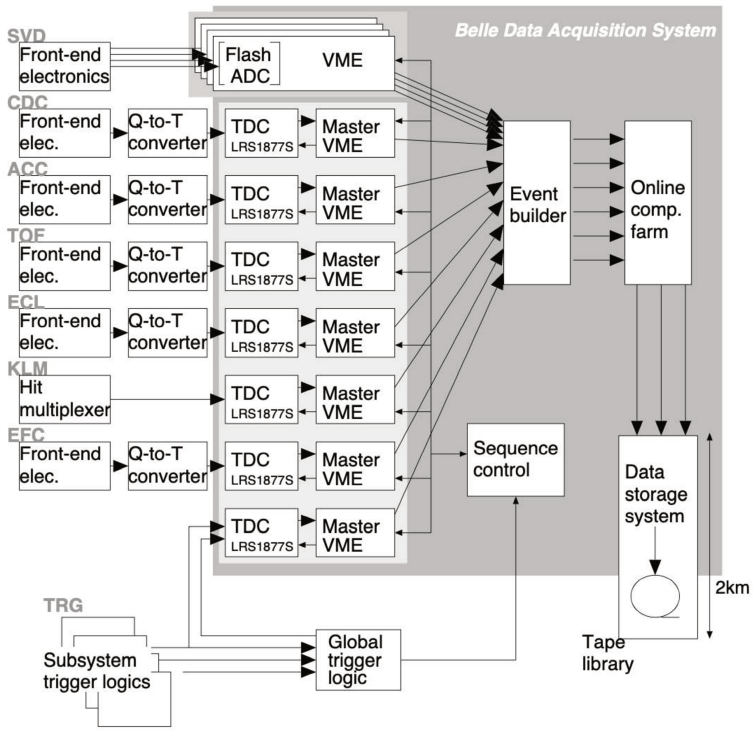


FIGURE 2.15: Data Acquisition System, final architecture. Picture from Ref. [72].

2.2.7 The High Level Trigger of Belle II

Two dedicated trigger systems are used at Belle II to reduce the data rate to a sustainable level: the hardware-based Level 1 (L1) trigger, which defines the events and performs a loose selection, and the software-based High-Level Trigger (HLT) for additional filtering [75].

In the current DAQ system (Fig. 2.16), the signals from the sub-detector front-ends are read out upon the Level 1 (L1) trigger decision. The data are transferred through several steps of data processing (like the event builder system) and delivered to the HLT farm. The task of HLT is to reconstruct the events in real-time using the regular Belle II software releases.

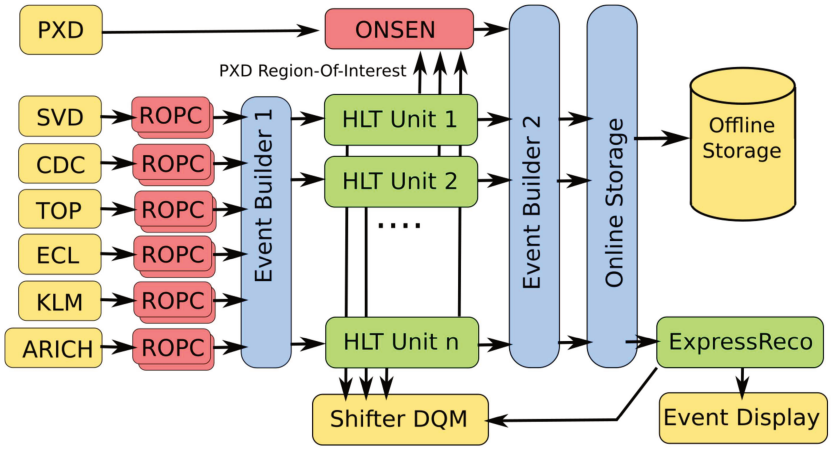


FIGURE 2.16: Scheme of the Belle II data flow, from the sub-detectors to the storage.

During the online reconstruction, the HLT decision is computed to:

1. Reduce the incoming rate by prescaling processes with high cross sections and discard background events;
2. Provide to the PXD the ROIs (Regions of interest) from the CDC and SVD-based tracks: they are used to select PXD hits combined in the second Event-Builder stage;
3. Flag the events compatible with interesting physics processes;
4. Provide information about data quality through the DQM.

The `SoftwareTrigger` module is the framework where the trigger results are calculated.
The two stages of HLT are called [76] *filter* and *skim*. The *filter* trigger, used to calculate the HLT decision, defines which events are discarded. On accepted events, an additional trigger decision called *skim* is calculated. It is used for the online tagging of the events; such information can be consulted during the processing of data.
The selection to be applied at the HLT *filter* stage was studied and refined during the early phases of Belle II data-taking, and the software trigger was switched on only for short periods. The IJCLab group contributed to the preparation for the start of the permanent *filtering* of data, which happened on March 2021 (or, from Experiment 17 in Belle II jargon, see Fig. 2.17). In this context, we developed a monitoring framework to get online feedback on the health of HLT and perform offline studies.

The effective cross section passing the HLT filter stage is, at the current stage of the experiment (Spring 2022), around 40 nb – roughly 1/5 of the incoming rate from the L1.

In the scheme adopted from the Belle II Data Production group, the HLT skim information is used to provide the samples for calibrations. In the past, such samples

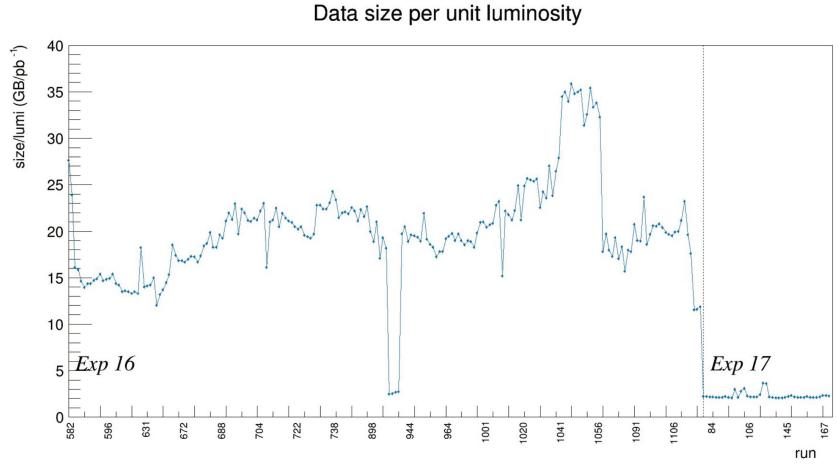


FIGURE 2.17: Normalized data size over time, expressed in terms of run and experiment number. The effect of allowing the HLT to discard events (*filter ON*) is visible at the edge between two data-taking periods (Experiment 16 and 17). A short period with *filter ON* is also present during experiment 16, around run 900.

were then used as input for the so-called *'offline skims'*, where more specific processes were selected through python scripts. Our task is to implement these selections directly into the HLT. We proceed as follows: we add to the software trigger the reconstruction needed to identify specific physics processes and assign the related *flags* to the event passing the selection criteria; consequently, the recorded data can be *skimmed*. A skim is a fraction of a given dataset where the purity of the desired signal channel is enhanced without a drastic loss of efficiency. Many flags are computed and allow to select processes from low-multiplicity (for example, di-muons, two-photons, or hadronic events) to specific B-decay channels [77]. Purity is a crucial performance for the skims, as they are used for calibrating the sub-detectors or monitoring their performance. High purity for a given process allows getting accurate constants out of the calibration algorithms or a better resolution for a given parameter when monitoring the data quality.

The effective cross sections of HLT flags have been constantly monitored during data-taking via Mirabelle, the data quality monitoring system for Belle II[9]. The framework was initially developed to check data quality after data (re)processing and identify the bad-quality runs to be discarded from calibration algorithms and physics analyses. Over time it has become a tool to check the health of data-taking through a list of selected summary plots where high-level variables are displayed. We have added many monitoring plots concerning the filter and skim stage performance; for example, the one showing the cross sections of selected HLT flags (shown in Fig. 2.18).

The idea is to have flags for single exclusive modes with $\mathcal{O}(1\%)$-level retention rates allowing fast calibration loops right after the data recording. In the last two years, we have implemented many selections for inclusive production channels that are useful for performance studies, for example:

- $J/\psi \to \ell^+\ell^-$. The selection does not use any particle-ID information and therefore the sample can be used for lepton-ID studies as it provides a clean sample of di-leptons.

[9]Website https://mirabelle.belle2.org/.

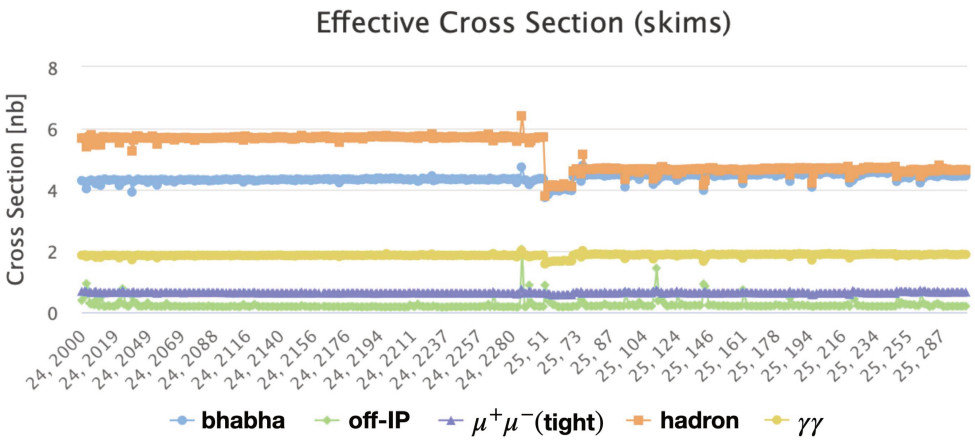


FIGURE 2.18: Effective cross section σ of physics processes selected by the Belle II HLT during data-taking, The plot, taken from the `Mirabelle` web page, shows the values for a selected data-taking period in between two experiments: *Experiment 24*, where data were collected at the $\Upsilon(4S)$ resonance (*on-resonance* data), *Experiment 25*, with data collected at $\sim$60 MeV below the $\Upsilon(4S)$ energy (*off-resonance* data). The effect of such energy shift can be seen in the σ_{hadron} because of the missing contribution coming from $B\overline{B}$ production.

- K^0_S. The inclusive K^0_S skim is the ideal source of pions for hadron-ID performance measurement over a low-momentum range spectrum. A clean peak of K^0_S gives also a valuable feedback during data-taking through the data-quality monitoring.
- Λ^0. For proton-ID.
- D^0. We reconstruct D^{*+}-tagged D^0 candidates in the channels $K\pi$, $K\pi\pi^0$, $K3\pi$ and $K^0_S\pi^+\pi^-$ as clean source of π^0 (π^0 resolution), soft pions from the D^* (tracking efficiency) and kaon for kaon-ID.

As a last implementation in the HLT, the reconstruction of the clean B-channels

$$B^0 \to D^{*+}(\overline{D}^0(\to K\pi, K\pi\pi^0, K\pi\pi\pi))\pi^+$$

$$B^0 \to D^{\mp}(K\pi\pi)\pi^{\pm}$$

$$B^+ \to D^0(K\pi, K\pi\pi^0, K\pi\pi\pi)$$

is performed to provide B-samples for the CM energy calibration in addition to $\mu^+\mu^-$ events. In the future, because of the increase in the instantaneous luminosity, the possibility of accessing tiny data samples with clean decay modes will be crucial for the efficient functioning of the experiment.

Chapter 3

$B^+ \to K^+\tau\ell$ search with Belle data and hadronic B-tagging

The present chapter describes the strategy in place for the search of the four LFV $B^+ \to K^+\tau\ell$ modes at Belle. We use the full dataset collected at the $\Upsilon(4S)$ resonance and a hadronic B-tagging approach to select the B^+B^- events. The workflow consists of many steps: the selection of signal events, the background characterization and suppression, the signal extraction and UL derivation in case of no observed signal. All these procedures are performed and optimised on simulated MC samples and are quite standard, aiming for a robust result, the first with the Belle data on these modes. For this reason, the performance obtained on MC (B-tagging efficiency, selection efficiency, fitting model etc.) is carefully checked and calibrated on data employing three control samples. In the last part, we show the results obtained on data and the extracted ULs on the BFs of the four modes, which are the most stringent to date. We conclude with a brief discussion of the current experimental scene. The achieved sensitivity is the indispensable baseline allowing us to develop on the improvement directions described in chapters 4 and 5.

3.1 MC samples

Signal MC is generated via the `EvtGen` [78] package[1] with a generic phase-space model (PHSP) to 3-bodies, i.e. the PHSP is uniformly populated, independently of whether it can be reached by an intermediate resonance (as illustrated in the Dalitz plot in Fig. 3.1A). An important kinematic variable is the q^2, which corresponds to the invariant mass of the di-lepton system:

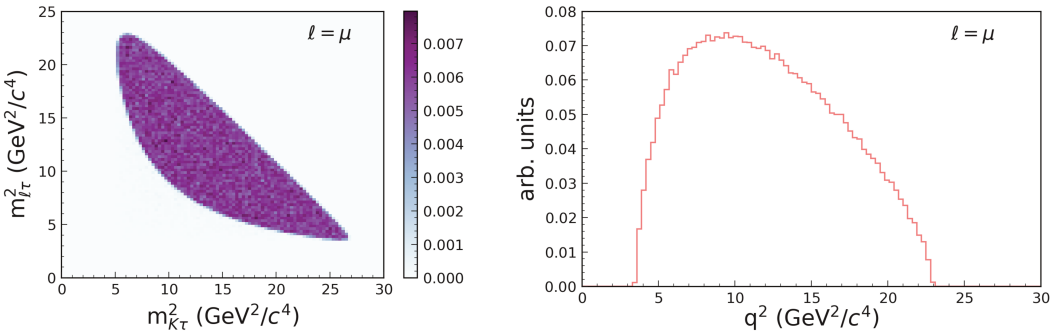
$$q^2 = (p_B - p_K)^2 = (p_\ell + p_\tau)^2 = m^2_{\ell\tau} \tag{3.1}$$

where p is the four-momentum of the particle. The q^2 has a span range given by the 4−momentum conservation $(m_\ell + m_\tau)^2 \leq q^2 \leq (m_B - m_K)^2$ and the shape obtained with the PHSP model is shown in Fig 3.1B.

The quark-level diagram showing the semileptonic $\bar{b} \to \bar{s}\tau\ell$ transition is represented in Fig. 3.2 for the two charge configurations; in both cases the kaon has the same charge of the parent B and the τ final-state, f, can be any of the allowed states, either *leptonic* ($f \equiv \ell\nu$) or *hadronic* ($f \equiv h$).

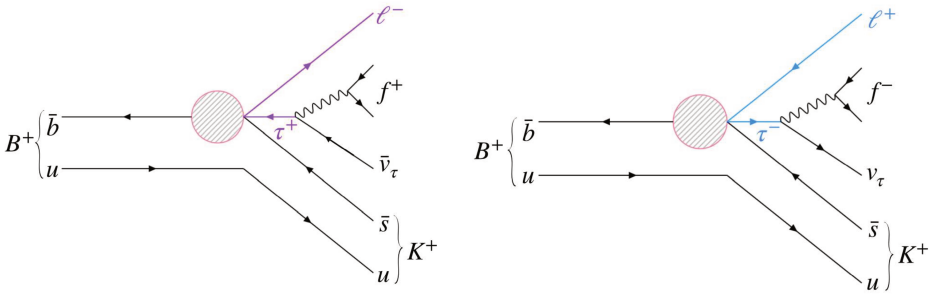
An equivalent variable to q^2 is $M(K^-X^+)$, defined as the invariant mass of the kaon and the oppositely-charged particle, that can be the prompt lepton or the particle from the τ (called t_τ) depending on the charge configuration - $\mathbf{OS}_\ell$ or $\mathbf{SS}_\ell$ (respectively

[1] The final state radiation (FSR) is simulated with the `PHOTOS` package [79].



(A) (B)

FIGURE 3.1: Kinematic distributions for $B \to K\tau\ell$ events generated with a PHSP model. A: Dalitz plot and B: $q^2 = m^2_{\ell\tau}$ distribution.



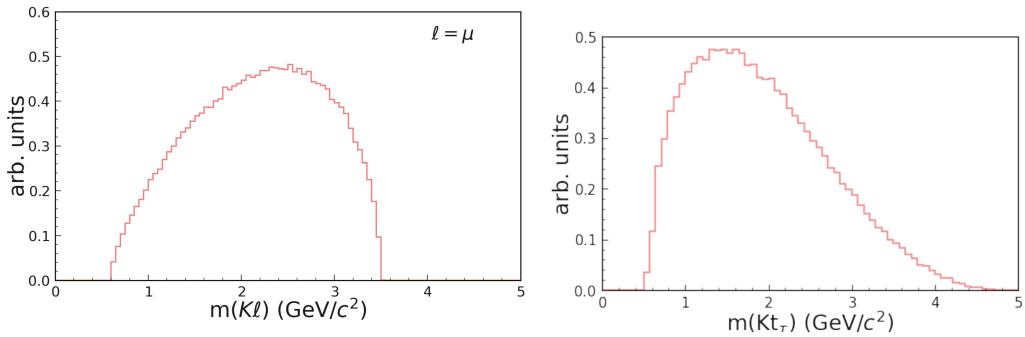
(A) (B)

FIGURE 3.2: Schematic diagram for $K\tau\ell$ decays. The BSM operator coupling to the four fermions is represented as a blob. A: OS and B: SS.

Fig. 3.3A and Fig. 3.3B):

$$X : \begin{cases} \ell^- & (\mathrm{OS}_\ell) \\ t_\tau^- & (\mathrm{SS}_\ell)\,. \end{cases} \tag{3.2}$$

The $M(K^-X^+)$ variable proves also to be crucial for the background suppression, as will be discussed in the next sections.



(A) (B)

FIGURE 3.3: $M(K^-X^+)$ distributions at generator level (PHSP). A: $X \equiv \ell$ (in this case $\ell \equiv \mu$) and B: $X \equiv t_\tau$, where t_τ is the charged particle from τ having the opposite charge of the signal kaon.

As anticipated in chapter 1, the PHSP is not a priori the most appropriate model to describe the physics process behind the $B \to K\tau\ell$ transitions and some variations

on the physics model should be tested in order to measure the impact on the reconstruction efficiency. At LHCb [22], different effective operators ($\mathcal{O}^{(\prime)}_{9,10}$ or $\mathcal{O}^{(\prime)}_{S,P}$) are introduced and the UL provided for each of them. A worse sensitivity is obtained with the scalar operators $\mathcal{O}^{(\prime)}_{S,P}$ (4.4×10^{-5}) with respect to the vector ones $\mathcal{O}^{(\prime)}_{9,10}$ (3.9×10^{-5}, the same as for the PHSP case) though, at the end, the PHSP is chosen as final result.
To perform a similar study at Belle two ways are possible:

(i) Generate the signal MC events via `EvtGen`. This is possible at present for $B \to K^* \ell\ell$ decays [80] but the corresponding `EvtGen` module has not yet been adapted to be used for $B \to K \ell\ell^{(\prime)}$ decays.

(ii) Re-weight the PHSP q^2 distribution according to the current set of best-fit (model-independent) parameters.

The second approach has been chosen because it did not require to write any new `EvtGen` model. The re-weighting has been performed according to the correct form factor [81] and the $\mathrm{d}\Gamma/\mathrm{d}q^2$ shown in Ref. [12]. More details are given in Appendix A.

To produce the signal samples, the `mcproduzh` package has been used. It allows to perform the generation (via `EvtGen`), the simulation of the detector response via `Geant3` [82] and the reconstruction of Belle data. The beam energy, IP profile, and detector configurations are experiment-dependent, but not run-dependent[2]. Also, the experiment-dependent background files are made of background events from all the runs of the experiment. The beam background, obtained from the random-triggered data, is overlaid to the simulated MC events.

For the background studies, the official *generic* samples with run-dependent beam background were used:

- $B\overline{B}$: `charged` (B^+B^-) and `mixed` ($B^0\overline{B}^0$);
- $q\overline{q}$: `uds` ($q = \{u, d, s\}$) and `charm` ($q = c$).

Multiple *streams* of these samples are available, each of them corresponding to the size of the Belle dataset. This is very useful to make more in-depth studies of the background nature but also to have independent samples of the same size to validate ML techniques. It is worth mentioning that the generic B samples only contain decay modes with $b \to c$ quark transitions. The suppressed *charmless* B-decays are simulated in specific MC samples, including the $b \to s/d$ (penguins) as well as the $b \to u\ell\nu$ transitions. Because of their small rates, these samples are generated with a much larger luminosity:

- $b \to u\ell\nu$: 20×;
- *rare* (electroweak and radiative penguins, charmless): 50×.

Two more MC datasets are used as control samples:

i. A privately-produced $B^+ \to D^-\pi^+\pi^+$ sample to assess the data-MC difference related to the $B\overline{B}$ background suppression. More details on the generation of such sample and the studies performed with it are provided in Sec. 3.10.1.

ii. A centrally-produced $B^+ \to J/\psi K^+$ sample for the checks related to the $q\overline{q}$ background suppression. It corresponds to 100 times the Belle dataset size.

[2]The same holds for the boost vector components. The definitions of 'experiment' and 'run' are given in Sec. 2.2.6.

BASF (Belle AnalysiS Framework) [83] 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. It blends the 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 for an easy access to different modules and their configuration for any tasks related to the experiment activities (generation of simulated data, unpacking of 'raw' data, reconstruction, calculation of high-level variables for the physics analyses).

In order to exploit the `basf2` software [84], [85] and use the Belle II hadronic B-tagging package, the workflow requires the usage of the `b2bii` framework [86], which converts the Belle `mdst` (mini-data summary tapes) files from the PANTHER [87] format (used in BASF) into ROOT-objects [88], used in `basf2`. After the conversion, all the Belle II analysis tools can be used and the same scripts can be used for studies on Belle II samples and possibly perform combined (Belle+Belle II) measurements. 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 `B2BIIConverMdst` module, responsible for the actual conversion. Afterwards, Belle II-type objects (Tracks, ECLClusters, PIDLikelihoods...) are available in the framework.

3.2 Data samples

The analysis is performed on the full Belle data sample collected at the $\Upsilon(4S)$ resonance between 1999 and 2010, corresponding to an integrated luminosity of $711\,\text{fb}^{-1}$ or 772×10^6 $B\overline{B}$ pairs [89]. Data were collected with two vertex detector configurations (called SVD1 and SVD2, see Sec. 2.2.2). Due to the fact that the instantaneous luminosity delivered by KEKB had increased after the SVD2 installation, the SVD1 run period corresponds to only 152×10^6 $B\overline{B}$ pairs ($\sim$20% of the total on-resonance data).

The reconstruction is performed on *skimmed* data, i.e. on a sub-set where minimal requirements are applied. In fact, at the same time of the `b2bii` conversion, the *HadronBJ* skim is applied [90]. 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 charged tracks and reconstructed photons) must be greater than 20% of $\sqrt{s}$. Further selection criteria remove the majority of beam-gas background and two-photon events; the first being reduced by requiring that the primary vertex position of the event to be close to the IP. Background events from QED and *tau*-pairs are suppressed by cutting on the total energy measured in the ECL and on 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 very efficient for inclusive ψ events. Therefore, the events with J/ψ and $\psi(2S)$ candidates are explicitly added to *HadronBJ*.

In summary, by using the pre-selection, or *skim*, described above we expect our MC data samples to be almost free from low-multiplicity processes (QED and $\tau^+\tau^-$). Even though the exclusive reconstruction of hadronic B-decays would naturally discard

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

those events, the skim prevents us from running the reconstruction on them which would have increased the processing time.

3.3 Recoil mass

As already mentioned in section 1.2.3, the strategy used at B-factories for the search of modes with missing energy consists in exploiting the full B_{tag} reconstruction and the 4-momentum conservation to extract the recoil mass M_{recoil}. In fact, if we consider the B_{sig}, the 4-momentum of the τ can be written as:

$$p_\tau = p_{B_{\text{sig}}} - p_K - p_\ell \tag{3.3}$$

where $p_{B_{\text{sig}}}$ is not known. In the frame where the $\Upsilon(4S)$ resonance is at rest (called center of mass - CM: $\mathbf{p}^*_{e^+e^-} = \mathbf{0}$), the two B mesons are back to back, hence[4]:

$$\mathbf{p}^*_{B_{\text{tag}}} = -\mathbf{p}^*_{B_{\text{sig}}}; \tag{3.4}$$

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

$$E^*_{B_{\text{tag}}} = E^*_{B_{\text{sig}}} = \frac{\sqrt{s}}{2} = E^*_{\text{beam}}\,. \tag{3.5}$$

In order to get a better resolution on the B meson energy, we can replace the $E^*_{B_{\text{tag}}}$ with E^*_{beam}, while preferring the reconstructed $p^*_{B_{\text{tag}}}$ over the average value $p^*_{\text{beam}} = \sqrt{E^{*2}_{\text{beam}}/c^2 - m_B^2 c^2}$. In fact the energy resolution of the detector is related to the uncertainties in the absolute values of the momenta of the decay products and is worse for the modes involving photons. On the contrary the E^*_{beam} depends on the spread in energy of the beams, typically $\sim 3\,\text{MeV}$. At Belle experiment the $\sqrt{s}$, and hence E^*_{beam} is calibrated with the kinematics of fully reconstructed hadronic B-decays. In particular, a deviation of $\sqrt{s}$ from the nominal value is inferred from the shift of the quantity $\sqrt{E^{*2}_{\text{beam}}/c^4 - \mathbf{p}^{*2}_B/c^2}$ from the nominal B mass. The resolutions on the reconstructed momentum and energy of the B are compared to those obtained by using the E^*_{beam} information in Fig. 3.4.

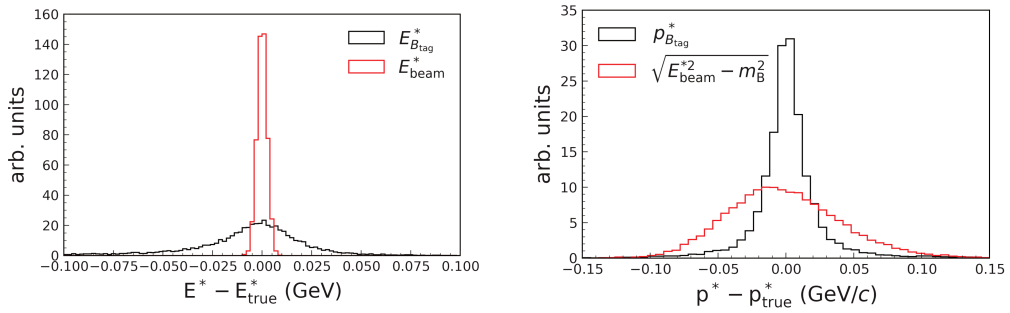


FIGURE 3.4: Comparison of the resolutions of $E^*_{B_{\text{tag}}}$ and E^*_{beam} (left) and of $p^*_{B_{\text{tag}}}$ and p^*_{beam} (right).

[4] All the observables denoted by an asterisk are measured in the CM frame.

Using the condition 3.4 and the substitution $E^*_B = E^*_{\text{beam}}$ in the equation 3.3 we get:

$$\begin{cases} \mathbf{p}^*_\tau &= -\mathbf{p}^*_{B_{\text{tag}}} - \mathbf{p}^*_K - \mathbf{p}^*_\ell \\ E_\tau &= E^*_{\text{beam}} - E^*_K - E^*_\ell \end{cases} \implies m^2_\tau = m^2_B + m^2_{K\ell} - 2(E^*_{\text{beam}} E^*_{K\ell}/c^4 + p^*_{B_{\text{tag}}} p^*_{K\ell} \cos\theta/c^2) \tag{3.6}$$

where θ is the angle between $\mathbf{p}^*_{B_{\text{tag}}}$ and $\mathbf{p}^*_{K\ell}$. The m_τ distribution of signal events is shown in Fig. 3.5 for the three recoil mass definitions, based on the three combinations of reconstructed and average E_B and p_B. The one corresponding to ($E^* = E^*_{\text{beam}}$, $p^* = p^*_{B_{\text{tag}}}$) leads to the highest resolution and will be used from now on throughout the entire manuscript with the two equivalent notations 'm_τ' or 'M_{recoil}'.

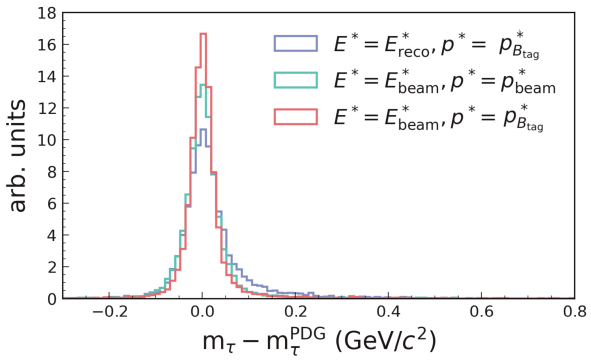


FIGURE 3.5: m_τ resolution with different configurations. The best resolution is obtained with $E^* = E^*_{\text{beam}}$ and $p^* = p^*_{B_{\text{tag}}}$.

3.4 B_{tag} reconstruction

The B_{tag} reconstruction is performed via the Full Event Interpretation (`FEI`) [8], a ML-based algorithm developed for Belle II. The `FEI` follows a hierarchical approach with six stages, as schematically rendered in Fig. 3.6. All the candidates available at each stage are combined to intermediate particle candidates in the subsequent stages, until the candidates for the B meson are created. Each intermediate particle has multiple possible decay-channels, which can be used to create valid candidates. The `FEI` reconstructs more than $\mathcal{O}(100)$ explicit decay-channels, leading to more than $\mathcal{O}(10^4)$ distinct decay-chains.

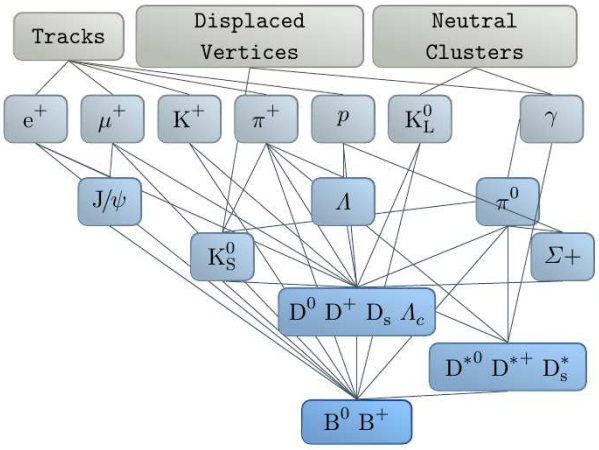


FIGURE 3.6: Hierarchical reconstruction of B candidates in the `FEI`. Figure from Ref. [8].

For each B meson, a value of the final multi-variate classifier output, called `SignalProbability` (or Σ, or $\mathcal{P}_{\text{FEI}}$), is assigned. This variable is distributed between zero and one, representing candidates identified as being background-like and signal-like, respectively.

For our study we configure the `FEI` in order to reconstruct only charged B hadronic decays not allowing for modes with K_L^0 as they are poorly measured. With such configuration, the four-momentum of the reconstructed B_{tag} is well-known and the tagged sample is pure. However, due to the fact that a typical hadronic B-decay has a BF of $\mathcal{O}(10^{-3})$ and that the intermediate states are explicitly reconstructed in a finite number of modes, only a very small fraction of events contains good B `FEI` tags. To quantify this, we recall the definition of three key-parameters used to establish the `FEI` performance:

- **tagging efficiency**: the fraction of $\Upsilon(4S)$ events that can be tagged
- **tag-side efficiency**: the fraction of $\Upsilon(4S)$ events with a correct tag
- **purity**: the fraction of tagged $\Upsilon(4S)$ events with a correct tag side.

In the case of hadronic B^+ `FEI`, the typical purity in MC is 10% for a 0.5% tag-side efficiency. The `FEI` performance is summarized in Fig. 3.7 and compared to that of the Belle exclusive B-tagging algorithm called Full Reconstruction (FR) [9]. The tag-side efficiency is higher in `FEI` compared to FR, but it drops very quickly as the purity grows. This behaviour seems to suggest that `FEI` is not very robust in background rejection.

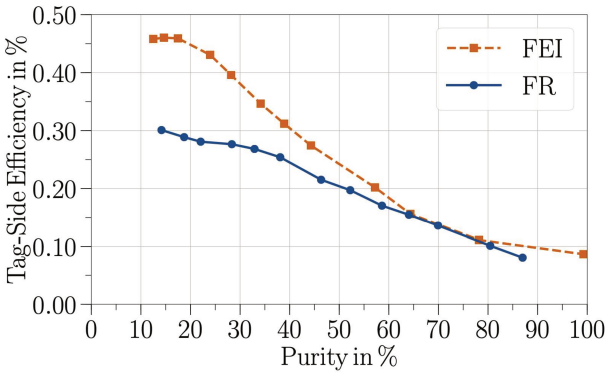


FIGURE 3.7: Tag-side efficiency as a function of the purity for B^+-tags. The *Full Reconstruction* (FR-Belle) performance is compared to the *Full Event Interpretation* (FEI-Belle II). Plot taken from Ref. [8].

Purity and tag-side efficiency are commonly measured by fitting the beam-constrained mass (M_{bc}), which is defined as:

$$M_{\text{bc}} = \sqrt{E^{*2}_{\text{beam}}/c^4 - p^{*2}_B/c^2}, \tag{3.7}$$

and is by definition independent of the mass hypotheses of the final state particles of the reconstructed B meson. The spread in M_{bc} is dominated by the spread in the beam energy.
The other common selection variable for the B_{tag} is the deviation from the beam-energy:

$$\Delta E = E^*_B - E^*_{\text{beam}}, \tag{3.8}$$

which instead is sensitive to mis-identifications. In this case, the spread on ΔE is dominated by the uncertainty of the reconstructed B meson energy. The ΔE and M_{bc} are shown in Fig. 3.8 while in Fig. 3.9 the achieved purity as a function of the cut on the `SignalProbability` is presented.

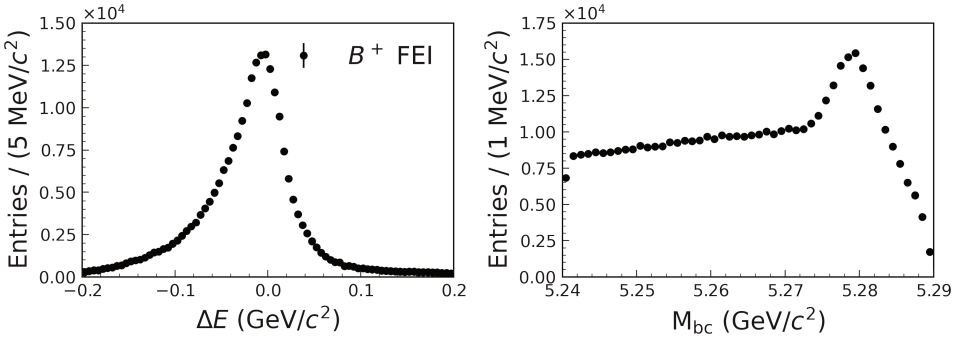


FIGURE 3.8: ΔE and M_{bc} distributions of B_{tag} candidates reconstructed by the `FEI` algorithm. The cuts $\mathcal{P}_{\text{FEI}} > 0.05$ and $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$ ($|\Delta E| < 0.1\,\text{GeV}$) are applied for ΔE (M_{bc}).

The total efficiency depends on two factors: the total BF (or coverage) corresponding to the exclusive modes included in the B-tagging algorithm; and the efficiency in

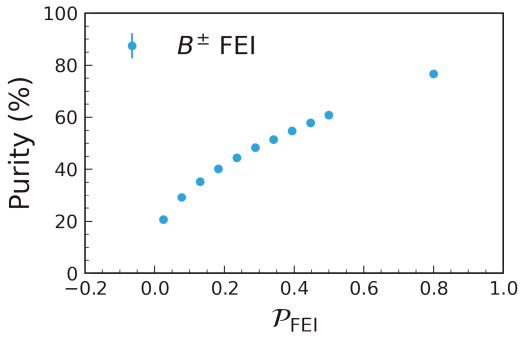


FIGURE 3.9: Tag-side purity of `FEI` candidates as a function of the threshold on the `SignalProbability`. The yields are measured with a fit to the M_{bc} distribution.

reconstructing all of the intermediate particles for each hadronic mode:

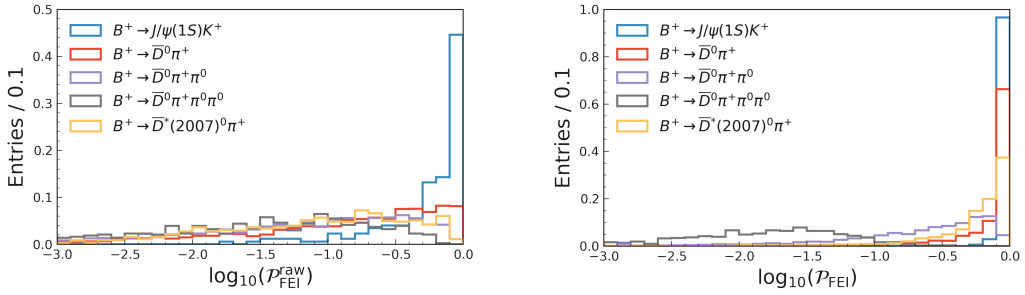
$$\varepsilon_{\text{tag}} = \sum_{i}^{N} \varepsilon_i \mathcal{B}_i \,. \tag{3.9}$$

To the ε_i terms contribute the geometrical acceptance (particles escaping the detection), tracking efficiency, particle-ID requirements, photon reconstruction efficiency etc. In the `FEI` algorithm, the coverage at B-level estimated from MC is, as will be detailed in the next chapter, section 4.1, around 25% and becomes $\sim$0.5% when considering the *coverages* of all the intermediate states and the efficiencies. For example, the D^0 (D_s^+) is reconstructed for roughly 40% (20%) of its total width; however, the final efficiency is half of it. On the other hand, the D^* mesons are reconstructed in all their possible decay modes (100% of coverage), but the effective efficiency is less than $\sim$ 10%, considering that, due to the small mass difference between D and D^*, the pion or the photon emitted have low momentum, making the reconstruction efficiency lower (outside the CDC or ECL p_t acceptance). The reconstruction of a D meson, despite being very powerful for background rejection because of the very good resolution on the reconstructed mass ($\sim 5\,\text{MeV}/c^2$), is very costly in terms of efficiency. For this reason, the B-decay modes with two D mesons (like in the *double charm* $\overline{D}^{(*)0} D_s^{(*)+}$ or $\overline{D}^{(*)0} D^{(*)0} K^{(*)}$ modes) have very small impact to the total tag-side efficiency. In chapter 4 this will be demonstrated with MC and data.

A different picture stands for the semileptonic tag, where a larger coverage ($\sim$ 18%) translates into a higher efficiency, around 2%. The advantages of this approach will be discussed in chapter 5.

As mentioned before, `FEI` employs multi-variate classifiers to assign to each candidate a probability $\mathcal{P}_{\text{FEI}}$ to discriminate B candidates against background and obtain purer samples. The features used for the training of the classifier are, for example, the kinematics and the vertex fitting information of the candidate and its daughters, as well as the `SignalProbability` of the daughters and their product (called $\mathcal{P}_{\texttt{FEI}}^{\text{raw}} = \Pi_{i=1}^{N_{\text{dau}}} \Sigma_i$). This means that the `SignalProbability` associated to the B candidate depends on the probabilities of each of its daughters; such feature can be visualised in Fig. 3.10, where the $\mathcal{P}_{\texttt{FEI}}$ of B^+ candidates (belonging to some selected decay modes) is compared to $\mathcal{P}_{\texttt{FEI}}^{\text{raw}}$. One can see that except for the clean modes like $J/\psi K^+$ or $\overline{D}^0 \pi^+$, where the probabilities associated to the candidates are close to one, for the others they span over different orders of magnitude. Furthermore, $\mathcal{P}_{\texttt{FEI}}^{\text{raw}}$ shows how the probability distribution looks like when using not only the information

of the B-daughters, i.e. before the training for the classifier used for the B-candidates. We will elaborate further upon this last point in the next chapter.



(A) (B)

FIGURE 3.10: `SignalProbability` of `FEI` candidates according to their reconstructed `decayModeID`. In (A) the probability is obtained as a product over the B daughters' Σ's, while in (B) the final `SignalProbability` is shown (output of the training).

A set of loose selection criteria is applied at each stage to reduce the number of candidates [91]; for each particle candidate, only the first n candidates are kept (the ranking is made according to a given variable – $|dM|$, $|dQ|$ etc.). After the training, a minimum `SignalProbability` cut is applied (see Tab. 3.1). The main rationale behind the choice of using loose cuts when building the B-candidates is to make the tag-side efficiency as high as possible. This choice is very convenient for the users as the `FEI` reconstruction can act as common ground for all the analyses targeting B-decays and the same sample with `FEI`-reconstructed B_{tag}'s can be used. The disadvantage is that the purity of the initial sample is very low, and the user must carefully study the selection that best fits their analysis and, consequently, calibrate the B-tagging performance.

The final state particles (FSPs) reconstructed by `FEI` are: $e^\pm, \mu^\pm, \pi^\pm, K^\pm, \overset{(-)}{p}, \gamma$. Impact parameter cuts are applied to the tracks in order to select only those originating close to the interaction point (IP). The two selection variables are:

i d_r (transverse impact parameter): the signed distance in the transverse plane between the origin of the coordinate system (nominal position of the IP) and the point of closest approach (POCA) of the helix describing the track. The sign convention is defined as follows: moving along the track into the direction of the particle's momentum, the sign is positive (negative) if the origin is to the right (left) of the track at the POCA, looking from the positive z-axis direction.

ii d_z (longitudinal impact parameter): the signed distance along the z-axis from the POCA to the helix.

Energy cuts are implemented for the selection of photon candidates: `goodBelleGamma`= 1 defined by their region of interaction in the ECL[5] while `clusterBelleQuality`=0 is a Belle legacy variable indicating a good cluster based on the ECL shower shape.
Intermediate state particles π^0, K_S^0, J/ψ, D^0, $D^\pm$, D^*, D_s and D_s^* are reconstructed using appropriate invariant mass windows. For the D^* mesons, these are replaced by a cut on the released energy Q of the decays. Other resonances (i.e. ρ, ω, K^* etc.) are not reconstructed. Finally, loose selections on the beam-constrained mass M_{bc}

[5]The energy threshold is $E > 50\,\mathrm{MeV}$ for the ECL barrel, $E > 150\,\mathrm{MeV}$ for the backward end-cap and $E > 100\,\mathrm{MeV}$ for the forward end-cap.

Particle List	FEI cuts	pre-cut	post-cut
$e^{\pm}$, $\mu^{\pm}$	$dr < 2\,\text{cm}, \lvert dz\rvert < 4\,\text{cm}$, $\mathcal{S} > 10^{-2}$	10 hi ℓID	5 hi $\mathcal{S}$, $\mathcal{S} > 10^{-2}$
$K^{\pm}$, $\pi^{\pm}$, $p^{\pm}$	$dr < 2\,\text{cm}, \lvert dz\rvert < 4\,\text{cm}$, $\mathcal{S} > 10^{-2}$	20 hi pID	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-2}$
γ	`goodBelleGamma`=1 `clusterBelleQuality`=0	40 hi E	20 hi $\mathcal{S}$, $\mathcal{S} > 10^{-2}$
π^0	$0.08 < M(\text{GeV}/c^2) < 0.18$, $\mathcal{S} > 10^{-2}$	20 lo $\lvert dM\rvert$	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-2}$
J/ψ	$2.6 < M(\text{GeV}/c^2) < 3.7$, $\mathcal{S} > 10^{-3}$	20 lo $\lvert dM\rvert$	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-3}$
K^0_S	$0.4 < M(\text{GeV}/c^2) < 0.6$, $\mathcal{S} > 10^{-2}$	10 lo $\lvert dM\rvert$	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-2}$
D^0, D^+	$1.7 < M(\text{GeV}/c^2) < 1.95$, $\mathcal{S} > 10^{-3}$	20 lo $\lvert dM\rvert$	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-3}$
D^+_s	$1.68 < M(\text{GeV}/c^2) < 2.1$, $\mathcal{S} > 10^{-3}$	10 lo $\lvert dM\rvert$	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-2}$
$D^{*0}, D^{*+}_{(s)}$	$0.0 < Q(\text{GeV}/c^2) < 0.3$, $\mathcal{S} > 10^{-2}$	20 lo $\lvert dQ\rvert$	10 hi $\mathcal{S}$, $\mathcal{S} > 10^{-3}$
B^+	$\lvert\Delta E\rvert < 0.5\,\text{GeV}$, $M_{\text{bc}} > 5.2\,\text{GeV}/c^2$		

TABLE 3.1: `FEI` selection, pre-cuts and post-cuts applied to each particle lists [8]. 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 `atcPID(4,3)` ($p^{\pm}$) – see Eq. 2.2. *'hi'* stands for *'highest'* and *'lo'* stands for *lowest'*.

and the energy difference ΔE are applied to the B_{tag} mesons to select the typical signal regions associated with B mesons. A total of 29 hadronic B^+ decay modes are reconstructed by the `FEI` package. They are listed in Table 3.2, along with all the reconstructed modes for the D mesons used to obtain the B^+ candidates (shown in Tab. 3.3).

The following intermediate states are also reconstructed:

1. $\pi^0 \to \gamma\gamma$;
2. $K^0_S \to \pi^+\pi^-, \pi^0\pi^0$;
3. $J/\psi \to \ell\ell\,(\ell = \{e, \mu\})$.

The `FEI` is trained on Belle MC events with a $\Upsilon(4S)$ decaying into charged or neutral B mesons (all the allowed decay channels). Hence, for the training no specific signal side is taken into account. The advantage is that the training can be done only once and be used for every analysis (i.e. every B_{sig}). It is worth mentioning that continuum events are not used for the training, as this background source is considered as easier to suppress by means of well-established techniques utilised at B-factories.

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.	$B^+ \to \overline{D}^0 D^+$
6.	$B^+ \to \overline{D}^0 D^+ K_S^0$
7.	$B^+ \to \overline{D}^{*0} D^+ K_S^0$
8.	$B^+ \to \overline{D}^0 D^{*+} K_S^0$
9.	$B^+ \to \overline{D}^{*0} D^{*+} K_S^0$
10.	$B^+ \to \overline{D}^0 D^0 K^+$
11.	$B^+ \to \overline{D}^{*0} D^0 K^+$
12.	$B^+ \to \overline{D}^0 D^{*0} K^+$
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.	$B^+ \to \overline{D}^0 K^+$
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.	$B^+ \to J/\psi K_S^0\pi^+$

TABLE 3.2: `FEI` B^+ decay modes.

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

TABLE 3.3: `FEI` D mesons decay modes.

The number of B_{tag} candidates per event[6] is reduced by taking only the candidate with the highest `SignalProbability`; such ranking is performed after having applied tighter cuts on M_{bc} and ΔE. The final selection of the B_{tag} candidates is:

- $M_{\text{bc}} \geq 5.27\,\text{GeV}/c^2$

[6] For $B\overline{B}$ events, when applying a minimal cut on `SignalProbability` of 10^{-5}, the average multiplicity is 4 candidates per event.

- $|\Delta E| \leq 0.1$ GeV
- $\mathcal{P}_{\text{FEI}} > 10^{-5}$
- Rank=1

where the ranking is performed according to the $\mathcal{P}_{\text{FEI}}$.

3.5 B_{sig} reconstruction

The full reconstruction of the B_{tag} provides the charge (i.e. the flavor) of one of the two B's, uniquely determining the charge of the other B, or B_{sig}, where we look for a $B^+_{\text{sig}} \to K^+_{\text{sig}}$ decay (or c.c.) with two more tracks: the *prompt* lepton ℓ_{sig} and the charged particle coming from the τ, t_τ, with opposite charge with respect to ℓ_{sig}. Because of the difference between the two (lepton) charge-conjugated modes, we reconstruct the combination $B^+ \to K^+\ell^- t^+_\tau$ for the $\mathbf{OS}_\ell$ modes and $B^+ \to K^+\ell^+ t^-_\tau$ for the $\mathbf{SS}_\ell$ modes. As it will be explained in the next section, we target 1-prong τ decays (Fig. 3.11) and therefore require the presence of 3 tracks in the signal side. To the 1-prong category belong not only the leptonic $\tau \to \ell\nu\bar{\nu}$ modes but also all the hadronic τ decays into a single charged track: $\pi\nu$, $\pi\pi^0\nu$ – which is mostly coming from ρ – and also $\pi 2\pi^0$ (mainy from a_1).

FIGURE 3.11: Visualisation of the τ decay BFs (%). The leptonic decays account for ∼35% of the total, the rest coming from hadronic decays, while ∼ 70% of the decays have only one charged particle (1 prong, either a lepton or a pion).

The selection criteria applied to the three signal side tracks are listed below.

Signal prompt lepton

- $|d_r| < 0.5\,\text{cm}, |d_z| < 5\,\text{cm}$
- $p_t > 0.2$ GeV/c
- ℓID $> 0.9\,(\ell = \{e, \mu\})$
- Bremsstrahlung recovery (only for the $e^\pm$): for all the photons with $E > 50\,\text{MeV}$ and inside a cone of radius 50 mrad centered around the $e^\pm$, their 4−momentum is added to that of the initial electron.

Signal prompt Kaon

- $|d_r| < 0.5\,\text{cm}, |d_z| < 5\,\text{cm}$
- $p_t > 0.2$ GeV/c
- kID > 0.6.

τ prong

- $|d_r| < 4\,\text{cm}, |d_z| < 20\,\text{cm}$

- $p > 0.5$ GeV/c (only for the π mode, to reduce the fake π rate)
- $p_t > 0.2$ GeV/c
- πID > 0.6 (π mode), ℓID > 0.9 ($\ell = \{e, \mu\}$).

The particle-ID variables used throughout this work have been defined in Sec. 2.2; we recall here that, for μID> 0.9 the average muon detection efficiency of 89%, with a pion mis-identification rate of 1.5%. For the $e^\pm$ candidates, the eID> 0.9 gives an average electron detection efficiency of 92% and a pion mis-identification rate below 1%. For the pion, the cut πID> 0.6 is used, which corresponds to an efficiency of 83% with a kaon mis-identification of 6%. The kaon is selected with the criterion kID> 0.6, which has an efficiency of 92% and a pion mis-identification rate of 5%. The cuts performed on the signal-side particles are summarized in Tab. 3.4.

Particle	Selection Cut
$K^\pm_{\text{sig}}$	$\lvert d_r\rvert < 0.5\,\text{cm}, \lvert d_z\rvert < 5\,\text{cm}$, $k\text{ID} > 0.6$
$e^\pm$	$\lvert d_r\rvert < 0.5\,\text{cm}, \lvert d_z\rvert < 5\,\text{cm}$, $e\text{ID} > 0.9$
	Bremss. recovery
$\mu^\pm$	$\lvert d_r\rvert < 0.5\,\text{cm}, \lvert d_z\rvert < 5\,\text{cm}$, $\mu\text{ID} > 0.9$
t_τ	$\lvert d_r\rvert < 4\,\text{cm}, \lvert d_z\rvert < 20\,\text{cm}$
All tracks	$p_t > 0.2$ GeV/c
π_τ	$p > 0.5$ GeV/c

TABLE 3.4: Signal side FSP selection criteria.

A vertex fit is performed with the two *prompt* tracks of the B_{sig} side: the information on the quality of the fit has double scope: perform a B_{sig} best candidate selection and also use the probability associated to the χ^2 of the fit as a training feature for the background suppression, as will be explained in section 3.7.

3.6 Event and best candidate selection

To each combination of B_{tag} and B_{sig} candidates (i.e. a $\Upsilon(4S)$ candidate) corresponds a Rest Of the Event (ROE), defined as the set of unused tracks and ECL clusters. The neutral ECL clusters which are allowed in the ROE have region-dependent threshold energies defined with the cut `goodBelleGamma`$= 1$.
The tracks in the ROE are selected within a region compatible with the IP: $\lvert d_r\rvert < 10.0\,\text{cm}$ and $\lvert d_z\rvert < 20.0\,\text{cm}$. With the above ROE definition, we reject events with ROE having more than zero tracks. No further cuts are applied on the neutral activity in the ROE, i.e. the number of clusters in the ECL or their energy; instead, these variables are later exploited as training features for the background suppression via BDT. The requirement of having no tracks in the ROE enhances the presence of 1-prong τ decays and helps reducing the multiplicity coming from the reconstruction of the t_τ. If more than one candidate is present for the t_τ the following priority is set: $\mu > e > \pi$. The priority is given to the leptons in order to avoid them, especially the muons, to be counted as pions. This cut is historical and was decided with the goal of exploiting the information coming from the different τ modes. In the end, we opted for not having a different selection for the three modes and treat all the event in an inclusive τ approach. This point could be a direction of improvement for a future measurement with Belle II data, as the $\tau \to \ell\nu\nu$ candidates have a different background nature. We will return to this point in Sec. 5.5.1.

B_{tag} \ B_{sig}	TM	Wrong t_τ	Wrong $K\ell$	FM	Total
TM	23.1	11.2	0.0	0.5	34.8
FM	36.8	25.9	0.2	2.3	65.2
Total	59.9	37.1	0.2	2.8	100.0

TABLE 3.5: Fraction (%) of signal events in different categories, according to the number of wrong particles in the event. The values refer to the reconstruction of the $\mathbf{OS}_e$ mode. Similar results are obtained for the other modes.

As shown in Eq. 3.6, the reconstruction of the B_{tag} side enters the formula only in the last term, where both p^*_B and $\cos\theta$ are used. In order to show the impact of the quality of the B_{tag} candidate on the m_τ resolution, we categorize them according to the number of wrong reconstructed B-daughters. Here a particle is considered as *wrong* when it does not originate from the most common mother, i.e., the B_{tag}. As we can see in Fig. 3.12, the M_{bc} of `FEI` candidates becomes background-like as soon as at least three particles are incorrectly combined[7], while in ΔE the wrong candidates still have a peaking shape; this is because ΔE is used as a training variable in the `FEI` algorithm and therefore `SignalProbability` ranking naturally selects candidates for which ΔE is closer to 0 GeV. In the m_τ dimension, almost all the candidates are peaking around the mass of the τ lepton and they will be considered signal candidates, as long as they belong to the m_τ signal window defined in Sec. 3.9. Clearly, the quality of B-candidates impacts the resolution of the reconstructed momentum and consequently on m_τ; we will explore in detail this point in chapter 5.

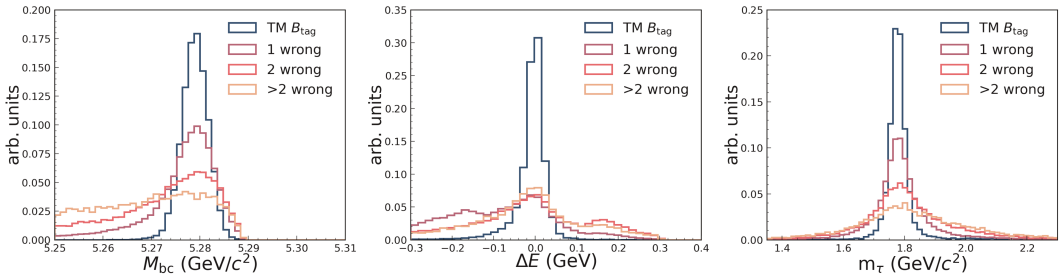


FIGURE 3.12: m_τ as a function of the number of wrongly reconstructed B_{tag} particles. The components are normalized to 1 and are obtained by selecting in a tighter window of the other two variables.

Fig. 3.13 shows the effects of wrongly reconstructing the t_τ, the $(K_{\text{sig}}, \ell_{\text{sig}})$ pair or the B_{tag}. In the first case the m_τ has the best attainable resolution, as long as the rest of the event is correctly formed; this is expected because the τ daughter does not enter the chosen recoil mass definition. The second case shows a combinatorial-like background; such component is almost negligible, after the best candidate selection and considering the narrow m_τ signal region (the fractions are provided in Tab. 3.5). The case where the B_{tag} candidate is not perfectly reconstructed has already been described before and is shown in the plot for comparison to the other case.

The information relative to all the generated particles is referred to as 'MC truth' and can be accessed by the analysts. During the analysis this information can be used

[7]Roughly 5% of the candidates, to be compared to the 35%, 35% and 25% for the TM, '1 wrong' and '2 wrong' candidates.

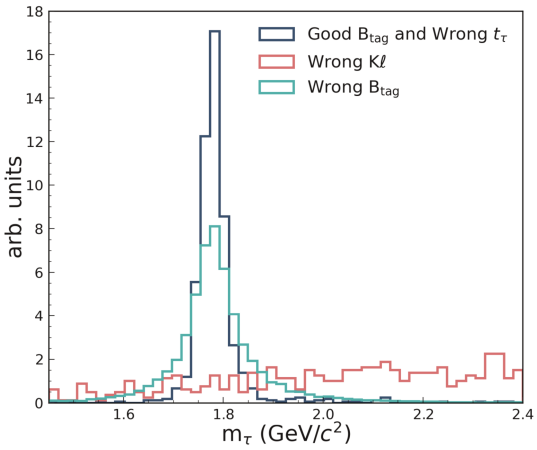


FIGURE 3.13: m_τ for candidates with partially wrong reconstruction. The presence of a wrong t_τ does not affect the resolution, as long as the B_{tag} is correctly reconstructed. On the other side, a wrongly reconstructed $K_{\text{sig}}\ell_{\text{sig}}$ pair irremediably leads to a background-like distribution. The different components are not in scale, the fractions can be found in Tab. 3.5.

to study the selection, or a particular decay chain, matching the reconstructed tracks with the MC truth (*truth-matching*). A `basf2` module performs the matching on reconstructed composite particles on a simulated sample. It maps, when applicable, reconstructed particles (`Particles`) to the generated ones (`MCParticles`). The MC matching module loops over all particles within the specified particle list and performs the matching on them as well as all daughter particles, searching for the first common mother (ancestor) of all daughter particles that form it and makes a `basf2` relation between `Particle` and `MCParticle`. We will therefore refer to *truth-matched* (TM) candidates whenever the `basf2` algorithm manages to match all the particles of its decay chain; if not, the candidates is labelled as *false-matched* (FM). With such definition, we detail the composition of candidates from signal MC events as the result of the reconstruction described in the last three paragraphs. Tab. 3.5 shows that for only a very small fraction of events the $K\ell$-pairs is mis-reconstructed ($\sim 0.1\%$), while on the tag-side, half of the `FEI`-candidates are not labelled as TM because of some failures in any of the particles composing them. Concerning the 'Wrong t_τ' part, the τ is TM when it is coming from $\tau^+ \to (\pi^+, e^+\nu_e, \mu^+\nu_\mu)\bar{\nu}_\tau$ decays but for the quoted $\sim$16%, in almost of the cases the track still corresponds to a pion but coming from a hadronic resonance, like ρ or a_1. Furthermore, when the B_{tag} is not properly reconstructed, swaps are also possible between the signal and the tag sides; in that case, the t_τ can come from B_{tag}. For all the possible $B_{\text{tag}}/B_{\text{sig}}$ combinations except the 'wrong' $K\ell$, the M_{recoil} peaks at m_τ and are therefore treated as signal. Nonetheless the quality of the B_{tag} affects the M_{recoil} resolution and produces longer tails.

3.7 Background

3.7.1 Sources of background

At the $\Upsilon(4S)$ resonance, the $e^+e^- \to$ hadron cross section σ_{had} is dominated by non-B events: of the total 4.8 nb, only $\sim$23% is related to the $\Upsilon(4S)$ production, which in turn decays most of the time into $B\overline{B}$ meson pairs (either charged or neutral). The rest is shared among the 4 possible light quark (u, d, s, c) pair modes. Despite the

larger cross section, the continuum background ($q\overline{q}$) is usually under control for analyses where a B is fully reconstructed. Besides, those events have distinct signatures opposed to $B\overline{B}$ events and therefore can be efficiently reduced. At the reconstruction level, the continuum events account for roughly 40% (20%) of the total background for $\mathbf{OS}_\ell$ ($\mathbf{SS}_\ell$) modes.

A possible way to classify the $B\overline{B}$ background sources consists in dividing the B-decays into two main categories: hadronic (`HAD`) and semileptonic (`SL`). They have different properties in terms of missing energy, topology and presence of high-momentum lepton. Under the approximation $\mathcal{B}_{\rm sl} = \mathcal{B}(B^+ \to X_c \ell^+ \nu_\ell) = 25\%$ (with ℓ being now any charged lepton) and $\mathcal{B}_{\rm had} = 1 - \mathcal{B}_{\rm sl} = 75\%$ one gets

- `SL`×`SL`: 6%
- `SL`×`HAD`: 38 %
- `HAD`×`HAD`: 56 %

i.e., the events where both B's decay hadronically are the most abundant at the generator level. The $B\overline{B}$ combinatorial background arises from a random combination of particles, or, in the case of B^+B^- events, a good $B_{\rm tag}$ candidate can be associated to other tracks that are compatible with our signal side (same final state). The main source of background in this last case is in the form of

- $B^+ \to \overline{D}^{(*)0}(\to K^+X^-)\ell^+\nu_\ell$ for $\mathbf{SS}_\ell$ modes;
- $B^+ \to \overline{D}^{(*)0}(\to K^+\ell^-\bar{\nu}_\ell)X^+$ for $\mathbf{OS}_\ell$ modes.

with the X^+ containing at least the t^-_τ (e^-, μ^- or π^-). One can already notice that while for the $\mathbf{SS}_\ell$ case the `SL`×`HAD` dominates merely because of the sign charges of the signal kaon and lepton, for the $\mathbf{OS}_\ell$ the X can either contain a lepton or hadrons and both combinations `SL`×`HAD` and `HAD`×`HAD` are possible. More precisely:

- For $\mathbf{SS}_\ell$ modes, the `SL`×`HAD` component reaches 90% and hadronic D decays are preferred.
- For $\mathbf{OS}_\ell$ modes, around 70% of $B\overline{B}$ background events is of the type `SL`×`HAD` while the rest consists of double-hadronic events[8]. Semileptonic $B_{\rm sig}$ decays are still favored because of the lower multiplicity and are compatible with the requirement of having two leptons on the signal side (for the $\tau \to \ell$ modes). The $\ell_{\rm sig}$ originates from a D meson for 70% of the time because of the charge signs.

These sources of background coming from charged B-decays are the most difficult to suppress because of their abundant production of kaons[9] and leptons. However, the different topology of the decay can be exploited. In fact, while for the signal candidates the kaon and the lepton are *prompt*, i.e. both coming directly from the B, for the background the kaon mostly comes from a D^0 and thus has larger impact parameter. In the $\mathbf{OS}_\ell$ configuration, the $K\ell$ pair is expected to come from the D and the two tracks make a good vertex but separated from IP because of the D-lifetime while for the $\mathbf{SS}_\ell$ case, only the $\ell_{\rm sig}$ is prompt and the vertex fit has a lower vertex fit χ^2.

[8] The case where the tag side decays semileptonically is almost negligible, due to the tight cuts on $M_{\rm bc}$ and ΔE.

[9] The inclusive BFs $\mathcal{B}(B^+ \to \overline{D}^0 X) = (79 \pm 4)\%$ and $\mathcal{B}(\overline{D}^0 \to K^+X) = (54.7 \pm 2.8)\%$ [13], while the anti-correlated charmed production is suppressed $\mathcal{B}(B^+ \to \overline{D}X) = (8.6 \pm 0.7)\%$ [53].

It is helpful to check the purity of each signal side particle to understand the background composition better and quantify how much of it is *irreducible*, i.e., provides the same final state to our signal. In Fig. 3.14 the μ_{sig} and K_{sig} of $\mathbf{OS}_\mu$ candidates are shown in terms of their true nature. The signal muon is quite pure ($\sim$80%) for $B\overline{B}$ background while being highly contaminated by pions (and kaons) for the $q\overline{q}$ events. Despite the low fake mis-identification rate for the chosen μID, it can be explained by the high multiplicity of pions in continuum events. Coming to the K_{sig}, for more than 80% of the times it corresponds to a true kaon for all backgrounds. In this context, it would be interesting to compare the performance of the $K^+\tau\ell$ to the $K^0_S\tau\ell$ modes, where the latter is expected to be even purer thanks to the powerful signature of the K^0_S. Indeed, the strange, neutral meson has a distinct displaced vertex due to its relatively long lifetime ($\tau \sim 90$ ps) and can be selected in a narrow region of $M(\pi^+\pi^-)$. However, different challenges related to this mode come from the additional cross-feed source due to the $B^0\overline{B}^0$ mixing and the lower B-tagging efficiency [8]. Nonetheless, we expect the neutral modes to provide similar sensitivity to the LFV $b \to s\tau\ell$ couplings.

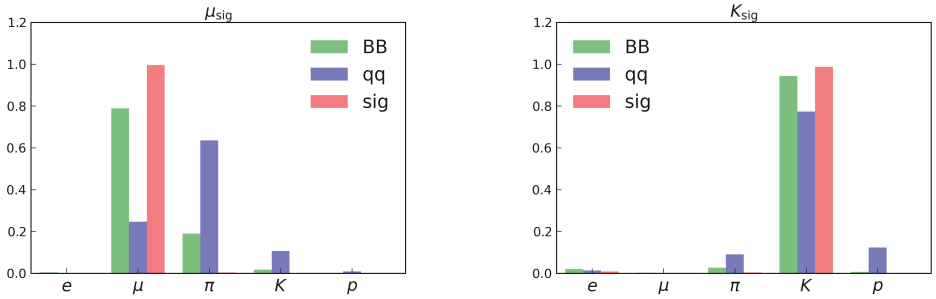


FIGURE 3.14: True nature of the particle identified as μ_{sig} (left) or K_{sig} (right). The $\mathbf{OS}_\mu$ case is here chosen for illustration.

The nature of t_τ is instead shown in Fig. 3.15 as a function of the assigned label to the τ prong:

- $t_\tau \equiv e$. The purity is $> 90\%$ for all the backgrounds;
- $t_\tau \equiv \mu$. As shown before for μ_{sig}, there is a large contamination from pions for the $q\overline{q}$ component;
- $t_\tau \equiv \pi$. The relatively high purity is obtained with the additional momentum cut.

3.7.2 $\overline{D}^0$ veto

In the dominant B^+B^- background, the K_{sig} is produced by a D^0, which preferably decays semileptonically for the $\mathbf{OS}_\ell$ case where a lepton with opposite charge is required. However, it can occur that the D^0 decays hadronically to $K^-\pi^+X$ and the pion is mis-identified as a muon (for $\ell_{\text{sig}} = e$, the mis-identification is much less frequent). In the $\mathbf{SS}_\ell$ case, where semileptonic B-decays are favored, the D^0 provides again the K_{sig} and the oppositely charged particle consistent with the t_τ (in case the τ prong is a pion). These two configurations are rendered in Fig. 3.16.

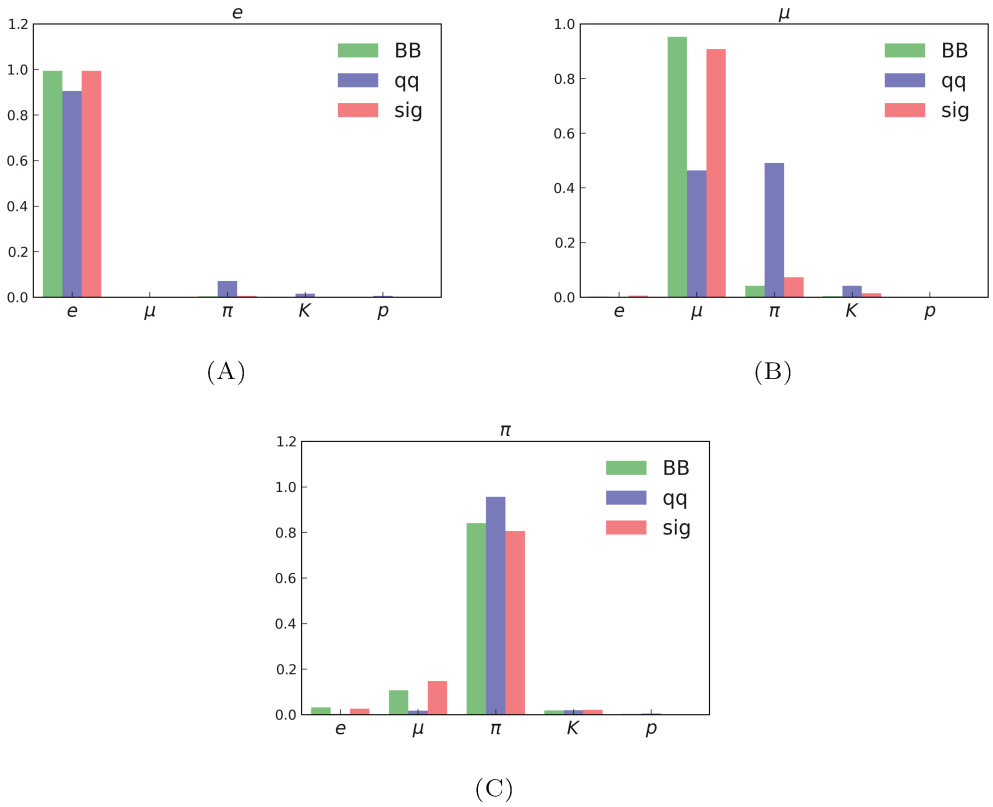


FIGURE 3.15: True nature of the particle identified as t_τ as a function of the assigned mode e_τ (A), μ_τ (B) and π_τ(C). The $\mathbf{OS}_\mu$ case is here chosen for illustration.

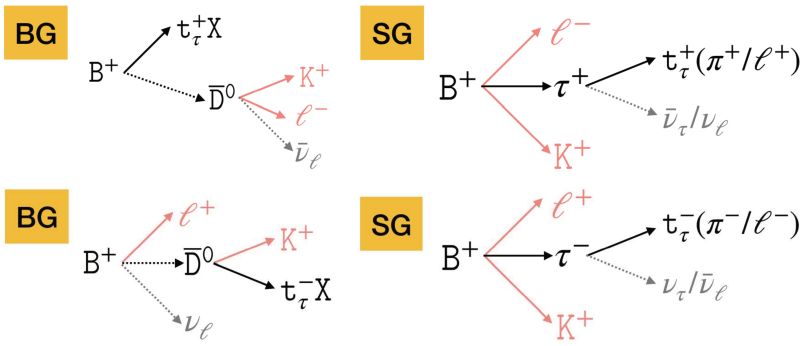
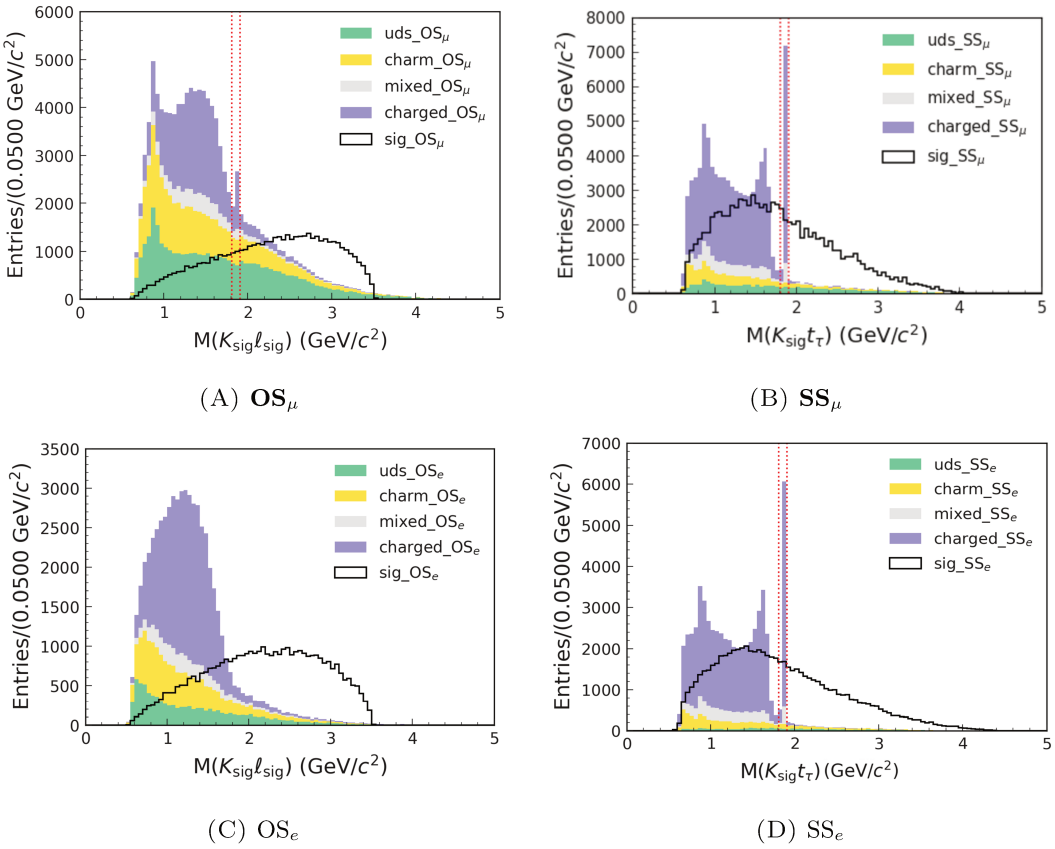


FIGURE 3.16: Background B^+ decay chains (left) on the signal side providing the same final state as $K\tau\ell$ (on the right). The common configuration for $\mathbf{OS}_\ell$ is displayed on the top figure while the $\mathbf{SS}_\ell$ is at the bottom.

In order to remove this charm component, the *BABAR* analysis [21] requires $M(K^-X^+) > 1.95\,\mathrm{GeV}/c^2$. The cut rejects between 97-99% (92-96%) of the background while retaining 32-37% (62-63%) of the signal for the $\mathbf{SS}_\ell$ ($\mathbf{OS}_\ell$) modes[10]. For the same purpose, a cut $M(K^-\mu^+) > 1.80\,\mathrm{GeV}/c^2$ is applied in the LHC*b* analysis [22].

We decide to remove only the trivial $D^0 \to K^-\pi^+$ background by vetoing the candidates with $M(K^-X^+) \in (1.81, 1.91)\,\mathrm{GeV}/c^2$ (as shown in Figure 3.17) for all the modes except the $\mathbf{OS}_e$ (as the fake rate of pions for the electron channel is very

[10]The signal loss is larger in the $\mathbf{OS}_\ell$ modes given that the τ prong has typically lower momentum than the ℓ_{sig}.



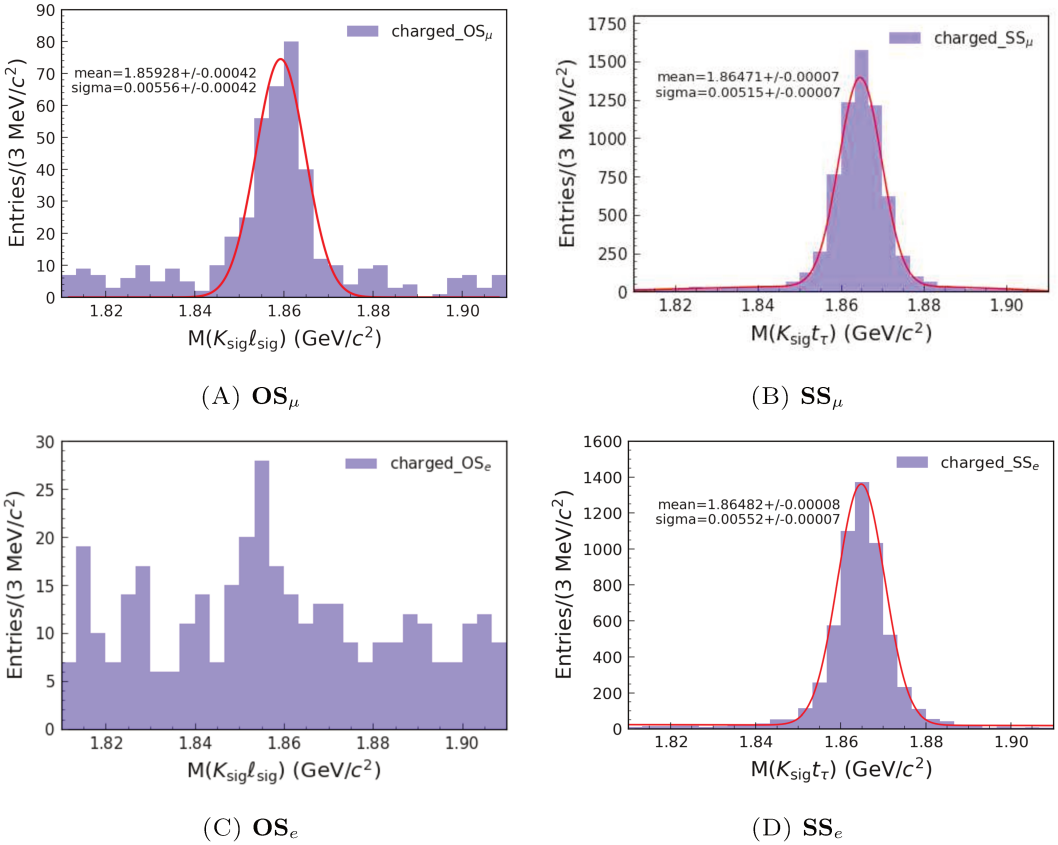
(A) $\mathbf{OS}_\mu$ (B) $\mathbf{SS}_\mu$

(C) OS_e (D) SS_e

FIGURE 3.17: $M(KX)$ distributions for background sources and signal (shape). The red dashed lines indicate the vetoed region, which corresponds to $D^0 \to K^-\pi^+$ decays. For the $\mathbf{OS}_e$ case no veto is applied.

small). This mass window extends over 10 times the measured resolution on both sides with respect to the D^0 nominal mass [92] - see Fig. 3.18. The signal loss due to the D^0 veto is less than 4% for the 3 modes involved, while the background rejection is, for the $B\overline{B}$ component, around 8% (20%) for the $\mathbf{OS}_\mu$ ($\mathbf{SS}_\ell$) mode.

After the veto, the $M(K^-X^+)$ variable is also exploited for the training of the BDT classifiers, in order to further reduce the backgrounds. As shown in Sec. 3.8, $M(K^-X^+)$ is the most important feature and a cut on the BDT output will indirectly favor higher values of its distribution.



(A) $\mathbf{OS}_\mu$ (B) $\mathbf{SS}_\mu$

(C) $\mathbf{OS}_e$ (D) $\mathbf{SS}_e$

FIGURE 3.18: Fit to the D^0 mass to determine the region to be vetoed.

3.8 Background suppression

3.8.1 BDT procedure

In order to increase the sensitivity on our channels, we perform a multivariate analysis (MVA) via Machine Learning (ML) instead of applying a sequence of one-dimensional cuts. ML consists of building algorithms to optimise a performance criterion using example data or past experience. We use a type of ML which is defined as *supervised* learning: MC simulated events with known label y (signal events with $y = 1$ and background events with $y = 0$) are used to build the statistical model. The main steps of such algorithms are:

- Pre-processing: The features needed for the learning are selected, the data samples (training/test) are chosen with the correct labels.
- Learning: The multivariate model is adjusted by tuning its parameters (called *hyper-parameters*). During this phase the fitting is performed. It is crucial to have a model which is not too simple (too complex) as this entails under-fitting (over-fitting): a scarce usage of the available information is obtained, hence a non-optimal inference on the data sample.
- Evaluation: The final model is evaluated on a test sample.
- Prediction on new data: The fitted classifier is applied to new data- points with unknown labels and new output labels are produced.

Some algorithms based on supervised learning are the decision trees (DTs), neural networks (NN) and random forests. We will focus on DTs.

The decision tree is a type of multivariate classifier that can make a prediction $\mathcal{O}$ depending on an ensemble of input observables w:

$$\mathcal{O} = f(w)$$

A DT performs a classification using a series of consecutive cuts (see Fig. 3.19). The maximum number of consecutive cuts is a hyper-parameter and is called the *depth* of the tree. The cuts are determined during the fitting phase using a training sample with known labels. At each node only training data-points which passed the preceding cuts are considered. For each feature at each node a cumulative probability histogram for signal and background is calculated, respectively. The histograms are used to determine the separation gain for a cut at each position in these histograms. The feature and cut-position (or bin) with the highest separation gain are used as the cut for the node. Hence each cut locally maximizes the separation gain between signal and background on the given training sample. The predictions of a deep DT is often dominated by statistical fluctuations in the training data-points. In consequence, the classifier is over-fitted and performs poorly on new data-points.

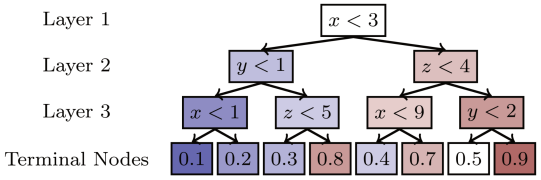


FIGURE 3.19: Decision tree. From Ref. [93].

A Boosted Decision Tree (BDT) constructs a more robust classification model by sequentially adding *shallow* DTs during the fitting phase. The DTs are fitted so that the expectation value of a negative binomial log-likelihood loss-function is minimized. The depth of the individual DT is strongly limited to avoid over-fitting. Therefore a single DT separates signal and background only roughly and is a so-called *weak* learner. The number of trees N coincides with the number of boosting steps and is anti-correlated to the shrinkage – or learning rate – $\eta \in (0, 1)$ which weights down the contribution of each learner to the ensemble to slow the learning process down and make the classifier more accurate. The two parameters, N and η are anti-correlated , i.e. decreasing the value of η increases the best value for N.

BDTs are widely used in the HEP community because they require short time for the training, they can simultaneously deal with continuous and discrete variables and don't need to transform inputs. Furthermore they are resistant to irrelevant variables and show good performance even with many training variables.

A classifier can be boosted by choosing the sequence of the trees in such a way that the error of the first tree is minimized. In the gradient boosted DTs (GBDT) the $(i+1)^{\text{th}}$ tree is chosen by moving towards the direction that minimizes the loss function for the model built using i trees. The loss-function is a metric for the classifier when to compared to the true labels y.

A GBDT is called *stochastic* when only a randomly drawn sub-set of the training sample is used for each DT and this increases the robustness against over-fitting (i.e. against statistical fluctuations). The sampling rate α determines the fraction of data-points that can be used at each step of the boosting.

FastBDT [93], the default MVA algorithm in Belle II `mva` software, implements a uniform stochastic gradient boosting algorithm. It was chosen because of its overall performance: the fitting is faster than other algorithms and showed to be reliable when used by non-experts in an automated environment. FastBDT uses an equal-frequency binning on the input data, which allows to map the continuous input features to integers. In fact, DTs only use the order statistic of the features; hence the algorithm only compares the values to one another and does not use the values themselves. As a consequence of the equal-frequency binning is that the best-cut found at each decision tree branch is only an approximation compared to a search performed on the original dataset.

A way to measure the performance of the model is via the Receiver Operating Characteristic (ROC), which shows the rate of True Positive events (true signal) against the rate of False Positive (false signal) as a function of the threshold on the output of the classifier (Fig. 3.20). The diagonal divides the ROC space. Points above the diagonal represent good classification results (better than random). A summary statistics that is commonly associated to the ROC is the 'area under the ROC curve', or AUC. The ideal model corresponds to AUC equal to 1.

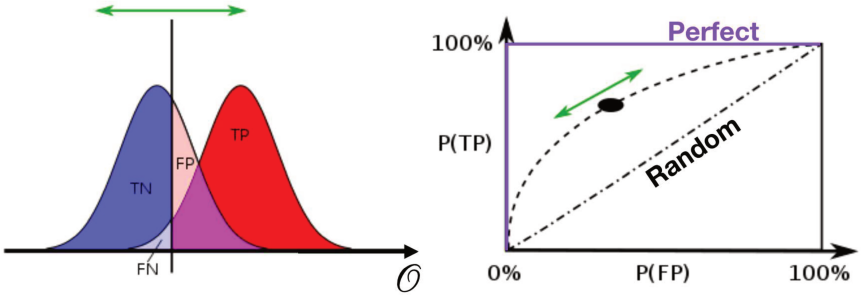


FIGURE 3.20: Pictorial representation of where signal and background events lie in the classifier's output $\mathcal{O}$. By changing the cut $\mathcal{O} > \bar{\mathcal{O}}$, the ROC curve on the right is obtained. Adapted from Wikipedia.

The multi-variate approach usually provides a higher signal efficiency (and hence sensitivity) because it takes into account the correlations between the input features. The available dataset is therefore exploited more efficiently and the statistical uncertainty of the final physics result is reduced. Despite providing clear advantages in analyses with dominating statistical uncertainty, as widely demonstrated in HEP, MVA needs careful validation in order to avoid mis-modeling effects and over-fitting problems. In order to avoid the last effect, we carefully choose the values of the hyper-parameters involved in the BDT training:

- NTrees (N): the number of trees used for the training;
- NTreeLayers: the maximum *depth* for each DT;
- NCutLevel: the number N of points within the range of each variable used to find the optimal cut to be used in each node of each tree (2^N bins are used for each feature);
- RandRatio (α): the fraction of events that is randomly chosen to train each decision tree;
- Shrinkage (η): the level of learning rate for the method of gradient boosting.

We also check the performance of other BDT packages than FastBDT supported in `basf2`. For example, in Fig. 3.21 the ROC curves obtained with FastBDT and `XGBoost` [94], obtained using the same training features and hyper-parameters setting, are compared: as expected, no other difference than training time is observed for the two options.

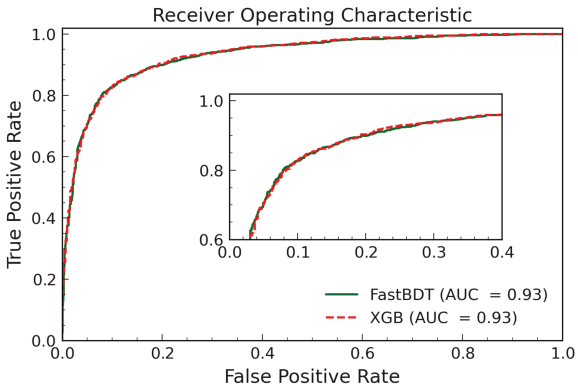


FIGURE 3.21: BDT training performance with FastBDT and XGBoost.

The BDT cuts are optimised using the figure of merit (FOM), defined in Ref. [10] as:

$$\mathcal{F}(a,t) = \frac{\varepsilon_{\text{sig}}(t)}{\frac{a}{2} + \sqrt{N_{\text{bg}}(t)}} \tag{3.10}$$

where

- t is chosen threshold
- $\varepsilon_{\text{sig}}(t) = N_{\text{sig}}(t)/N_0$ is the signal efficiency corresponding to the given selection.
- $N_{\text{bg}}(t)$ denotes the number of background events as a function of the selection.
- a is the number of sigmas corresponding to a one-sided Gaussian test at significance α. The value $a = 3$ is chosen for our search, although the optimization point does not strongly depend on the value for a.

This FOM is more suitable than other because the cross section of the searched process is unknown.

3.8.2 $B\overline{B}$ background

The training is performed in a narrower region $m_\tau \in (1.0, 3.5)\,\text{GeV}/c^2$. Due to the different reconstruction efficiency and background composition, for each mode a BDT classifier is trained and only the 10 best-scored features are kept. The rank and the set of the features slightly changes for each mode but they can be grouped in four categories:

1. $M(KX)$. The variable is described in section 3.7.2.
2. ROE-related. For well-reconstructed signal events, the ROE should have zero or small ECL activity, which can be related to background events or the neutral remnants from the τ (in case the reconstructed signal event has a $\tau \to \pi n\pi^0\nu$

decays). Therefore the number of ECL clusters (n_{ROE}(ECL)) and the related energy (E_{ROE}(ECL)) are expected to be smaller than for the background. For the same reason, the squared missing mass of the event, defined as

$$mm^2 \equiv |p_{\Upsilon(4S)} - \sum p_{\text{visible}}|^2$$

should be peaking at 0 for signal events when only the ν_τ escapes the reconstruction while being larger for the leptonic τ modes. Larger tails are observed for the background.

3. Vertex-related variables. The B_{tag} vertex is reconstructed from all the tracks in the tag side, while the B_{sig} vertex is fitted using the primary tracks: K_{sig} and ℓ_{sig}. As anticipated in Sec. 3.7.1 the origin of K^- in the combinatorial background is mainly from a D^0 decay, hence the distance of the B_{sig} vertex from the IP position (denoted as dr(B_{sig})) tends to have a longer tail. The distance of closest approach between the two primary tracks, dist(K_{sig}, ℓ_{sig}), is also used for the training. Other vertex-related variables are: dist(K_{sig},t_τ), dist(B_{sig} t_τ). All these variables are computed with the `basf2 DistanceCalculator` module.

4. KSFW-related variables [95]. KSFW is a Fisher discriminant developed at Belle for continuum suppression. The building blocks are the Fox-Wolfram (FW) moments, used to describe the distribution of momentum and energy flow in an event: for a collection of N particles with momenta p_i the k-th order FW moment H_k is defined as $H_k = \sum_{i,j}^{n} |\vec{p}_i||\vec{p}_j| P_k(\cos\theta_{ij})$ where θ_{ij} is the angle between $\vec{p}_i$ and $\vec{p}_j$, and P_k is the k-th order Legendre polynomial. The *modified* FW moments are divided into three parts depending on the origin of the particle: H_k^{ss}, H_k^{oo}, H_k^{so} if the sum runs over B_{tag} particles (label s), ROE particles (label o) or both (label so). The KFSW is a linear combination of the modified FW moments and further divides particles into charged, neutral and missing components.

Some of the training features can be seen in Fig. 3.22 for the $\mathbf{OS}_e$ mode, where the two background components ($B\overline{B}$ and $q\overline{q}$) are separated in order to show that, although targeting $B\overline{B}$ background suppression, the BDT classifier will also be able to reduce the $q\overline{q}$ components as many of the used variables show a good separation with the signal. The variables used for each mode and a short definition for each of them (as well as the aliases we will use throughout the manuscript) are listed in Table 3.6. Fig. 3.23 shows the linear correlations among the ten training variables, in addition to the m_τ: on the upper (lower) side of the diagonal the values refer to the background (signal) events. For background events the correlation with m_τ is not larger than 20% and is expected to be related to the $M(K\ell)$ variable as it enters the very definition of m_τ. This point will be further developed in chapter 5.

As mentioned before, the hyper-parameters for the BDT training are tuned in order to get a similar performance for the training and the test sample (as shown in Fig. 3.24).

For the optimal point attained, we check the ranking of the used variables in terms of their importance for the training (see Fig. 3.25). The importance is computed by the FastBDT method itself and uses the information gain of each applied cut by summing up the separation gain of each feature by looping over all trees and nodes. Except for the most important features, such metrics gives only a rough estimate because of the many correlations between the variables. Other importance types, evaluated in a method-agnostic way, are more accurate but require longer computing time. As

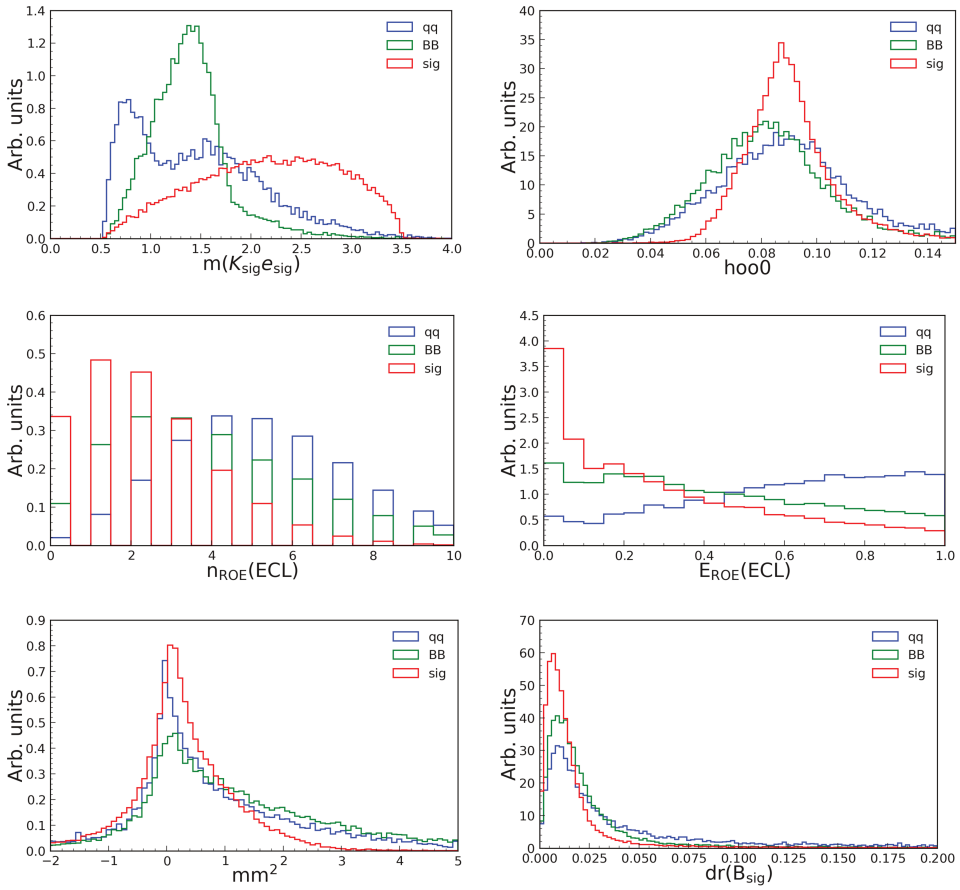


FIGURE 3.22: Training features for the BDT($B\overline{B}$) (**OS**$_e$ mode).

expected, the $M(K^-X^+)$ variable is the most powerful and the combination with the other nine variables only helps in obtaining a more efficient background suppression, as the MVA takes into account of the correlations between the features.

Fig. 3.26, representing the output of the BDT classifier trained to remove $B\overline{B}$ background, confirms what mentioned before: a good separation is obtained not only for signal against $B\overline{B}$, but also $q\overline{q}$. A further suppression step, this time dedicated on the residual $q\overline{q}$ component, will be described in the next section.

Finally, Fig. 3.27 (left plot) illustrates the performance of our training: for a cut on the BDT score of 0.9, the background rejection rate is $> 95\%$, at the price of loosing half of the signal events. However, as demonstrated by the right plot of the same figure, high BDT scores correspond to higher discovery sensitivity, as defined for the FOM of Eq. 3.10.

The M_{recoil} distribution of background and signal events (re-scaled to a yield equivalent to a BF of 5×10^{-5}) after the $B\overline{B}$ suppression is shown in Fig. 3.28 for the **OS**$_e$ mode. At this stage of the cut-flow, the $q\overline{q}$ background becomes as large as the $B\overline{B}$ ($\sim 50\%$ in the signal window), and thereby it is worth training a second classifier to reduce the continuum component further.

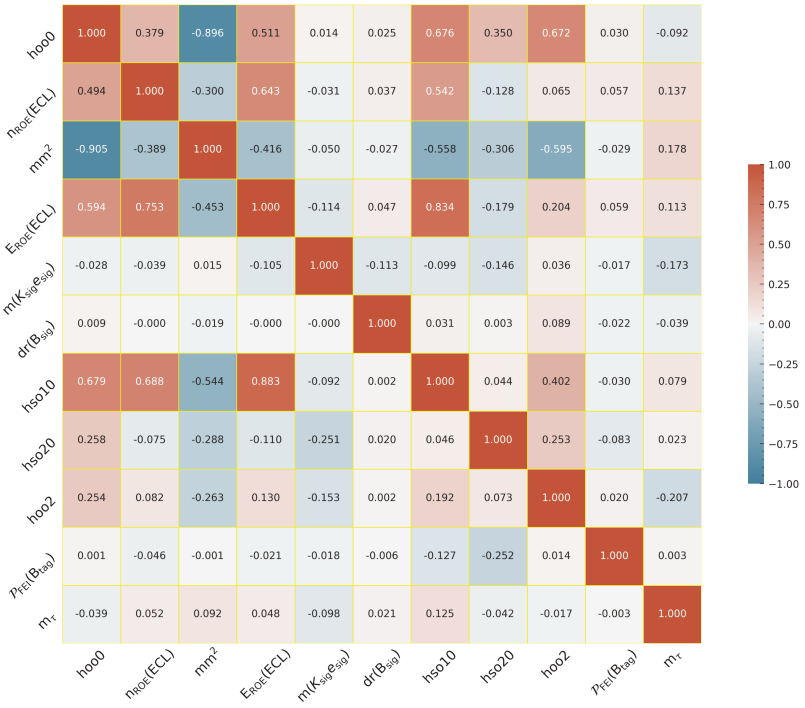


FIGURE 3.23: Correlation matrix of the variables used for the training for the $\mathbf{OS}_e$ mode, plus the m_τ. On the upper side of the diagonal, the values refer to background events while on the lower side signal events are shown.

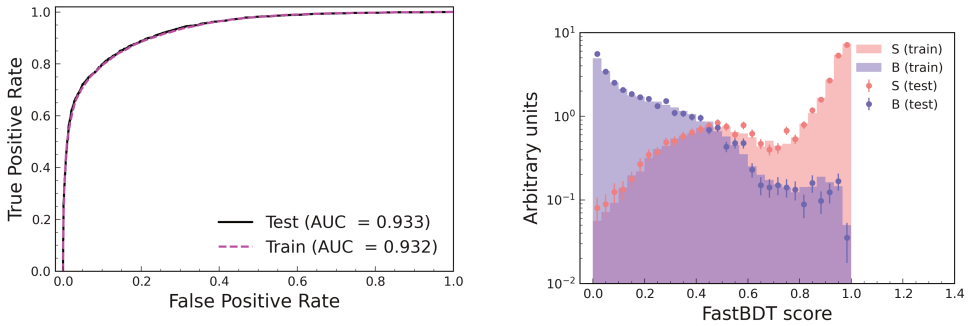


FIGURE 3.24: Performance in terms of ROC and BDT output of test and training samples for the supervised BDT training are compared. MC signal and background events for the $\mathbf{OS}_e$ mode are used.

Var. name	Definition	$\mathbf{OS}_\mu$	$\mathbf{SS}_\mu$	$\mathbf{OS}_e$	$\mathbf{SS}_e$
$\mathcal{P}_{\text{FEI}}$	`SignalProbability` associated to the B_{tag} candidate	✓		✓	
$M(K_{\text{sig}}t_\tau)$	Invariant mass of the prompt kaon and τ prong		✓		✓
n_{ROE}(ECL)	Number of ECL clusters associated to the ROE	✓	✓	✓	✓
E_{ROE}(ECL)	Energy of the ECL clusters associated to the ROE	✓	✓	✓	✓
mm2	Missing mass squared	✓	✓	✓	✓
$\mathcal{P}_{\chi^2}(V_{Kt_\tau})$	Probability associated to the $K\ell$ vertex fit		✓		✓
hso10	KSFW moment	✓	✓	✓	✓
$M(K_{\text{sig}}\ell_{\text{sig}})$	Invariant mass of the prompt kaon and prompt lepton	✓		✓	
$\text{dr}(\text{B}_{\text{sig}})$	Transverse impact parameter of the $(K_{\text{sig}}\ell_{\text{sig}})$ vertex	✓		✓	
hoo2	KSFW moment	✓		✓	
hso20	KSFW moment	✓		✓	✓
hoo0	KSFW moment	✓	✓	✓	✓
$\text{dist}(K_{\text{sig}}, \ell_{\text{sig}})$	Distance between the prompt kaon and prompt lepton		✓		✓
$\text{dist}(B_{\text{sig}}, K_{\text{sig}})$	Distance between the B_{sig} and the prompt kaon		✓		✓
$\text{dist}(K_{\text{sig}}, t_\tau)$	Distance between the prompt kaon and the τ prong		✓		

TABLE 3.6: List of the variables used for the BDT($B\overline{B}$) training and their definitions.

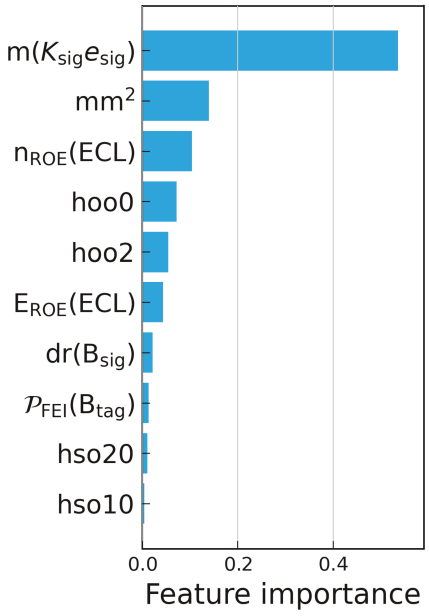


FIGURE 3.25: Ranking of the training features for BDT($B\overline{B}$) ($\mathbf{OS}_e$ mode).

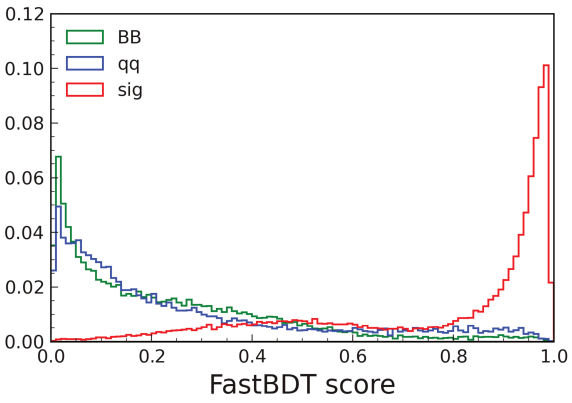


FIGURE 3.26: BDT($B\overline{B}$) score for background and signal events ($\mathbf{OS}_e$ mode).

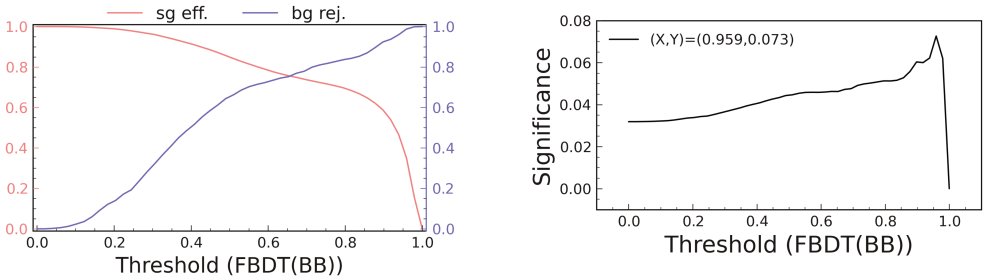


FIGURE 3.27: Left: Signal efficiency ε_{sig} and background rejection as a function of the cut on the BDT($B\overline{B}$) output. Right: FOM defined in Eq. 3.10. The $\mathbf{OS}_e$ mode is shown.

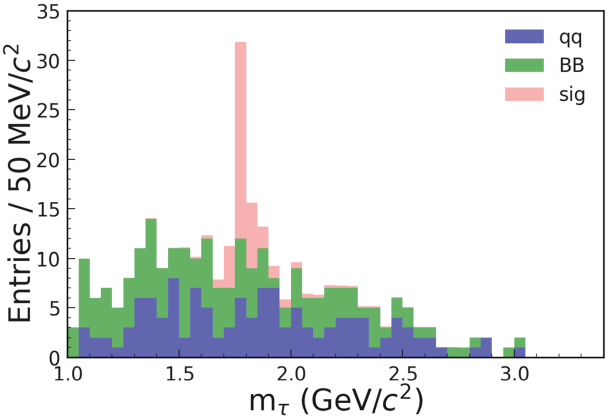


FIGURE 3.28: m_τ distribution for $\mathbf{OS}_e$ candidates after $B\overline{B}$ background suppression. The assumption $\mathcal{B}(B^+ \to K^+\tau^+e^-) = 5.0 \times 10^{-5}$.

3.8.3 $q\overline{q}$ background

The BDT classifier optimised for $B\overline{B}$ background suppression also removes a large amount of continuum background. However, a non-negligible $q\overline{q}$ component still remains, especially in the $\tau \to \pi\nu$ modes. For this reason, a second BDT classifier is trained on variables that are commonly used for continuum background suppression, based on the event topology.
In a $B\overline{B}$ event, both B mesons are produced almost at rest in the $\Upsilon(4S)$ frame, as the $m_{\Upsilon(4S)}$ is just above the $B\overline{B}$ production threshold; therefore the B-decay products are distributed isotropically, as opposed to $q\overline{q}$ events, where the quarks are produced with higher momentum, providing a back-to-back fragmentation into two jets of light hadrons. In this context, the usage of the thrust axis becomes crucial (it is defined as the direction that maximizes the sum of the longitudinal momenta of the considered particles). Furthermore, in $B\overline{B}$ events the angular distributions of the decay are uncorrelated, while for continuum there is a sizeable correlation because the particles from each B candidate tend to align with the direction of the jets.

The features used for the training are listed below, along with a short description (which is also summarized in Tab. 3.7):

- θ_T: the angle between the thrust axis obtained from final-state particles for the B_{tag} and for the rest of the event in the CM frame. For signal events, the $\cos\theta_T$ is flat, whereas for continuum events it peaks towards 1.

- R_2: Defined as the ratio of the 2-nd to the 0-th order FW moments. In the limit of vanishing particle masses, $H_0 = 1$ and the normalized ratio R_2 has values close to one for events with strongly collimated jets.

- Sphericity S: Scalar, event-based variable defined as the linear combination of the sphericity tensor eigenvalues λ_i: $S = (3/2)(\lambda_2 + \lambda_3)$. The distributions for background and signal events after $B\overline{B}$ background suppression of this variable and the previous two are shown in Fig. 3.29.

- CLEO Cones [96]: A set of nine variables corresponding to the momentum flow around the thrust axis of the B candidate, binned in nine cones of 10° around the thrust axis.

Var. name	**Definition**
R_2	Reduced Fox-Wolfram $R_2 = H_2/H_0$
$\cos\theta_T$	Cosine of angle between thrust axis of the B_{tag} and thrust axis of the B_{sig}
sphericity	Event sphericity
cleoConeThrust_i	i-th Cleo cone calculated with respect to the thrust axis

TABLE 3.7: List of the variables used for the BDT($q\overline{q}$) training and their definitions.

Figure 3.30 shows that $\cos\theta_T$ and the sphericity of the event are the most effective variables for the continuum suppression, while the cleoCones do not provide significant information gain.

As one can see from Fig. 3.31, this second classifier only separates the $q\overline{q}$ background as the topology of signal and $B\overline{B}$ events is very similar.

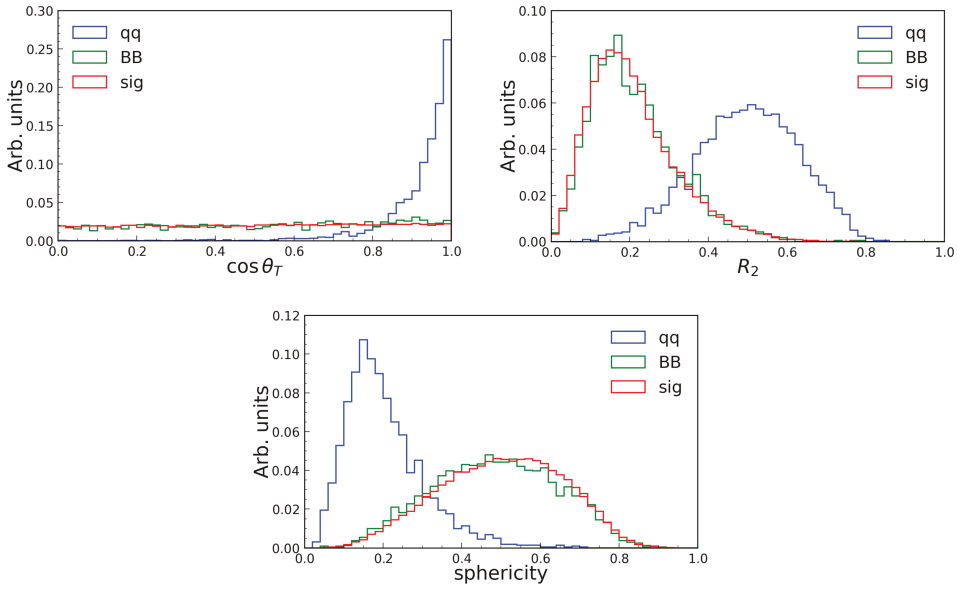


FIGURE 3.29: $\cos\theta_T$, R_2 and sphericity distributions after $B\overline{B}$ background suppression: they are the most important variables for the $q\overline{q}$ training. The $\mathbf{OS}_e$ mode is shown.

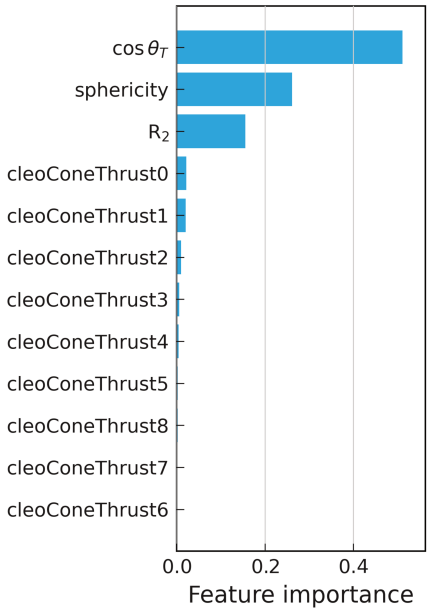


FIGURE 3.30: Ranking of the training features for BDT($q\overline{q}$) ($\mathbf{OS}_e$ mode).

Fig. 3.32 shows the signal efficiency and the background rejection as a function of the cut on the FBDT output for the $q\overline{q}$ classifier, as well as the relative FOM.

The cut values chosen for the $\mathbf{OS}_\ell$ and $\mathbf{SS}_\ell$ modes are summarized in Tab. 3.8, as well as the final reconstruction signal efficiencies computed in the signal region. Slightly looser cut values with respect to those maximizing the FOM are set in order to have an acceptable agreement with the performance of the training on the control sample – as will be explained in the next Sec. 3.10.

A summary of the entire cut-flow for the $\mathbf{OS}_e$ mode is presented in Tab. 3.9. The signal yields are, as usual, scaled to the tentative BF of 5×10^{-5}, while for the backgrounds, the number of events correspond to one *stream*, i.e., Belle dataset size.

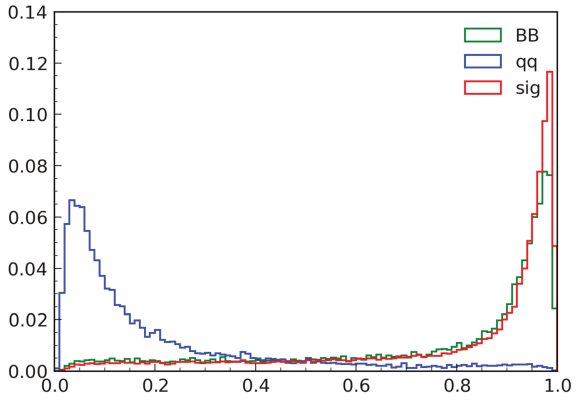


FIGURE 3.31: BDT($q\bar{q}$) score for background and signal events ($\mathbf{OS}_e$ mode).

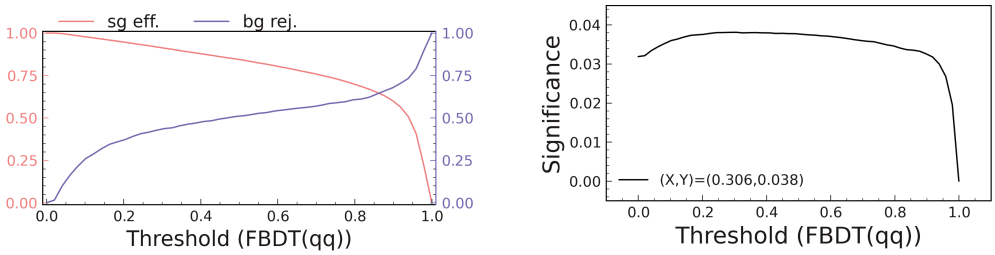


FIGURE 3.32: Left: Signal efficiency ε_{sig} and background rejection as a function of the cut on the BDT($q\bar{q}$) output. Right: FOM defined in Eq. 3.10. The $\mathbf{OS}_e$ mode is shown.

Cut value	$\mathbf{OS}_\mu$	$\mathbf{SS}_\mu$	$\mathbf{OS}_e$	$\mathbf{SS}_e$
BDT($B\overline{B}$)	0.90	0.92	0.90	0.90
BDT($q\bar{q}$)	0.50	0.66	0.20	0.40
ε_{sig} (%)	0.081	0.099	0.091	0.076

TABLE 3.8: BDT score thresholds and final signal efficiencies.

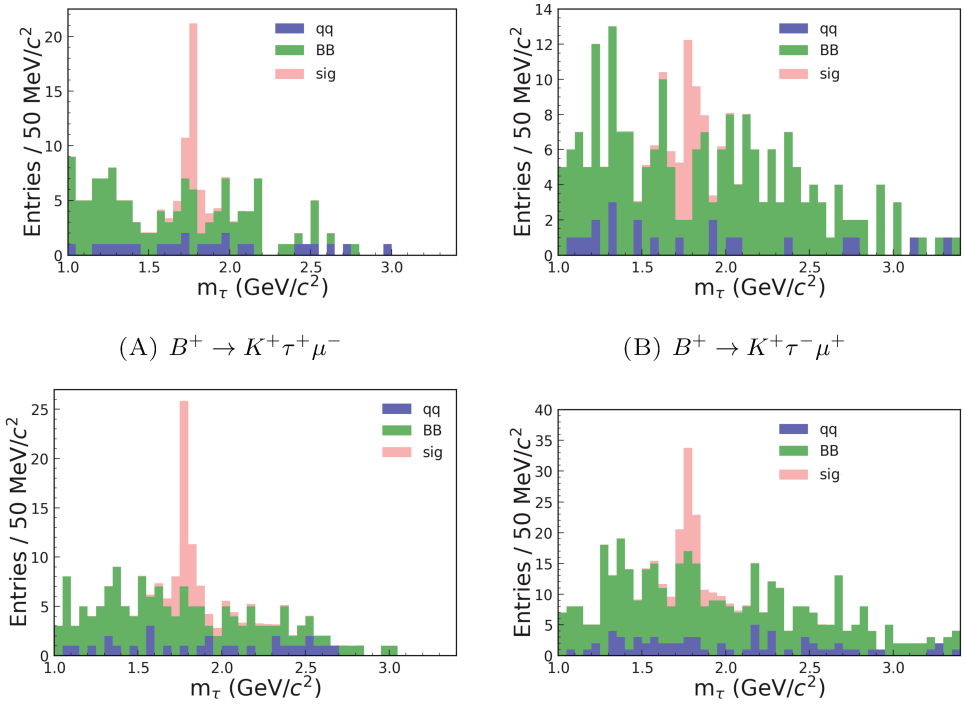
In the last column the significance is computed according to Eq. 3.10, demonstrating that the three selection steps increase the sensitivity to signal. It is also worth stressing that the efficiencies measured in MC will be corrected to take into account the known discrepancies with data, for example in the `FEI` reconstruction efficiency. The yields listed in Tab. 3.9 are thus only indicative as the fits for the signals extraction will not depend from them.

Even after the two BDT cuts, a residual background is present; this is mainly due to B^+B^- events that are, according to the MC simulation, mostly `HAD`×`HAD` for the $\mathbf{OS}_\ell$ modes (60%) and `HAD`×`SL` for the $\mathbf{SS}_\ell$ modes (70%). In the first case, hadronic B_{sig}'s prevail as they provide prompt kaons for K_{sig}. In this case, the ℓ_{sig} can either come from charmonium resonances or from D mesons (for example, in $D_s^{(*)}\pi K$ or $D^{(*)}K$ decays). For the $\mathbf{OS}_\ell$ modes, the $\overline{D}^{(*/**)0}\ell^+\nu_\ell$ background described previously is dominant: the ℓ_{sig} is assigned to the prompt lepton while the K_{sig} and the t_τ are associated to tracks coming from the $\overline{D}^0$. It should be stressed that in none of the configurations described, the resulting M_{recoil} would show a peaking structure in the signal region. The distributions of the final background events in m_τ for the four LFV

Cut	Signal	$\mathbf{B\overline{B}}$	$\mathbf{q\overline{q}}$	$\mathbf{B\overline{B}} + \mathbf{q\overline{q}}$
Reco	70.0 ± 0.4	555 ± 6	353 ± 4	910 ± 10
BDT($B\overline{B}$)	32.3 ± 0.3 (-53.9%)	19.0 ± 1.2 (-96.6%)	17.0 ± 1.1 (-95.2%)	36.1 ± 2.3 (-96.0%)
BDT($q\overline{q}$)	30.7 ± 0.3 (-5.0%)	18.6 ± 1.6 (-2.2%)	1.0 ± 0.1 (-94.2%)	19.6 ± 1.7 (-45.6%)

TABLE 3.9: Summary of the cuts for the $\mathbf{OS}_e$ reconstruction. The assumption $\mathcal{B}(B^+ \to K^+\tau^+e^-) = 5.0 \times 10^{-5}$ is made for the signal while for the background components the yields correspond to 1 stream of generic MC. The yields refer to the signal window $m_\tau \in (1.68, 1.87)\,\mathrm{GeV}/c^2$.

modes are illustrated in Fig. 3.33, where the re-scaled signal events are added.



(A) $B^+ \to K^+\tau^+\mu^-$ (B) $B^+ \to K^+\tau^-\mu^+$

(C) $B^+ \to K^+\tau^+e^-$ (D) $B^+ \to K^+\tau^-e^+$

FIGURE 3.33: m_τ distribution as a result of the full selection and background suppression procedure. The assumption $\mathcal{B}(B^+ \to K^+\tau\ell) = 5.0 \times 10^{-5}$ is made for each of the four modes.

3.9 Signal extraction

In order to measure the signal yield of the $B \to K\tau\ell$ modes the following strategy is adopted. The signal MC distribution is fitted to determine all the probability density function (PDF) parameters except the yield. The background shape is also modeled from the MC but the values of PDF parameters are not fixed. A toy MC study follows to assess the fit strategy and verify its stability and the absence of biases.
All the performed fits consist of unbinned extended maximum likelihood fits computed with the `RooFit` package [97]. In an extended fit the likelihood has a term proportional to the observed total number of events, which is the parameter we aim to measure, as well as its uncertainty.

A PDF is a function (normalized to one) of a set of observables $\mathbf{x}$ and parameters $\mathbf{p}$ $f(\mathbf{x};\mathbf{p})$. The *likelihood* of the PDF with respect to an observed data point $\mathbf{x}_0$ is defined as $f(\mathbf{x};\mathbf{p})$ and is a function of the parameters $\mathbf{p}$.

The likelihood can be extended to a set of data points by combining the single likelihoods as follows:

$$\mathcal{L}(\mathbf{p}) = \prod_i^n f(\mathbf{x}_i;\mathbf{p})$$

The best values for the $\mathbf{p}$ parameters are estimated by maximising the likelihood function or, equivalently, minimizing the function:

$$-\ln\mathcal{L}(\mathbf{p}) = \sum_{i=1}^{n} \ln f(\mathbf{x}_i;\mathbf{p})$$

If the number of observations n itself is a random variable with mean ν – n $\sim$Pois$(n;\nu)$ – then the likelihood is called *extended* and writes as:

$$\mathcal{L}(\mathbf{p}) = \frac{\nu^n e^{-\nu}}{n!} \prod_i^n f(x_i;\mathbf{p})$$

where in general $\nu = \nu(\mathbf{p})$. Furthermore we opt for an unbinned maximum likelihood (UML) fit because it is more robust for small sample sizes. UML fittings is obtained via minimization of the calculated likelihood function by the ROOT implementation of MINUIT. The fits are configured to execute the MIGRAD, HESSE and MINOS function in succession: the first finds the function minimum by iterative calculations of the gradient and following the local minimum, the second calculates the error matrix from the 2*nd*-order derivatives at the minimum. This method provides symmetric uncertainties, which is exact under the assumption that the likelihood is locally parabolic. On the contrary, MINOS calculates the uncertainties by explicitly finding the contours for which where $\Delta \log L = 0.5$ and they can thus be asymmetric.

3.9.1 Signal PDF

The signal m_τ distribution, defined in Eq. 3.6, is modeled with a PDF consisting of two components: a Crystal Ball PDF (CB), intended to fit the TM candidates and the right-hand tail of the distribution, and a Gaussian. As described in Sec. 3.6, we consider (self) cross-feed events as signal events because of their peaking nature. However, the resolution of partially mis-reconstructed candidates entails a larger resolution, that is encoded in the Gaussian component. The means of the CB and the

Gaussian are fixed to be equal to the current PDG value of the τ mass[11] while the rest of the parameters (n, α, σ_1 for the CB and σ_2 for the Gaussian), as well as the normalization of one component with respect to the other, are kept floated. The PDF describing the signal can be written as:

$$S(m|n,\alpha,\sigma_1,\sigma_2,f) = f \cdot \mathrm{CB}(m|n,\alpha,\sigma_1) + (1-f)\cdot \mathrm{G}(m|\sigma_2) \tag{3.11}$$

where the definition of the Crystal Ball is:

$$\mathrm{CB}(m|\mu,\sigma,\alpha,n) = \frac{1}{N}\begin{cases} \exp\left[-(m-\mu)^2/(2\sigma^2)\right] & , \quad m > \mu - \alpha\sigma \\ \frac{(n/\alpha)^n \exp(-\alpha^2/2)}{[(\mu-m)/\sigma + n/\alpha - \alpha]^n} & , \quad m \leq \mu - \alpha\sigma\,. \end{cases} \tag{3.12}$$

The parameters (μ and σ) are respectively the mean and the standard deviation of the Gaussian component of the CB, α is the parameter encoding the location of the switch from the Gaussian to the power-tail and n is the exponent of the power-tail.

The result of the fit to the signal m_τ distribution is shown in Fig. 3.34 and the values of the fitted parameters (as well as their uncertainties) are summarized in Tab. 3.10. The width associated with the CB component is used to define the signal window in which the cuts are optimised: for a $\sigma_1 \sim 0.3\,\mathrm{GeV}/c^2$, the 3-$\sigma$ range is $(1.68, 1.87)\,\mathrm{GeV}/c^2$.

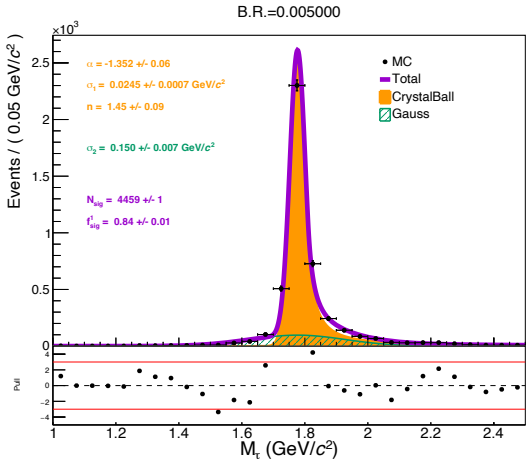


FIGURE 3.34: Fit to the m_τ for $B^+ \to K^+\tau^+e^-$ events passing the selection.

Parameter	Fit result
f_{CB}	0.84 ± 0.01
σ_{CB} (MeV/c^2)	24.5 ± 0.7
α	-1.352 ± 0.06
n	1.45 ± 0.09
σ_{Gauss} (MeV/c^2)	150 ± 7

TABLE 3.10: Fit parameters for the $\mathbf{OS}_e$ signal candidates.

3.9.2 Background PDF

A second-order polynomial PDF is used to fit the total background ($B\overline{B}$ and $q\overline{q}$ combined) surviving the whole selection process. The Chebychev polynomials are chosen as they are usually more stable due to milder correlations between the coefficients. In `Roofit` the Chebychev polynomials are defined as:

$$T(m|c_1,\ldots c_k) = \frac{1}{N}\cdot\left(T_0(m) + \sum_{i=1}^{k} c_i T_i(m)\right). \tag{3.13}$$

[11] PDG τ mass value: $(1776.86 \pm 0.12)\,\mathrm{MeV}/c^2$ [92].

Such polynomials are naturally defined and normalised in the domain $[-1, 1]$ and the range defined for the fit is mapped onto this range. An example of fit to background events is presented in Fig. 3.35, with the corresponding fit values in Tab. 3.11.

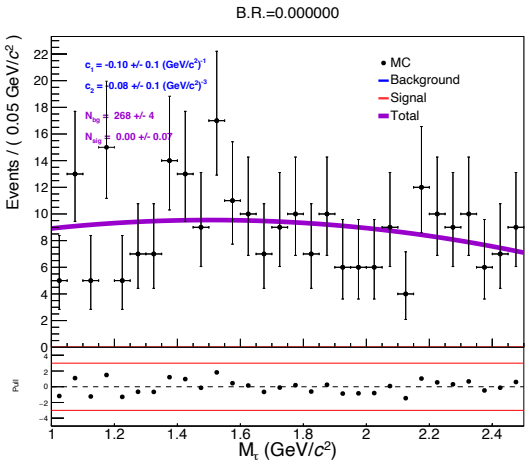


FIGURE 3.35: Fit to the m_τ for background events passing the selection.

Parameter	Fit result
c_1 (GeV/c^2)	-0.10 ± 0.1
c_2 (GeV/c^2)	0.08 ± 0.1

TABLE 3.11: Fit parameters for the $\mathbf{OS}_e$ background candidates.

3.9.3 Toy MC studies of fit stability

The stability of the fit is checked via two MC studies performed with `Roofit`. The first consists in generating 10^4 pseudo-data experiments (with a Poisson fluctuation on the total number of events) according to the following PDF:

$$F(m|N_{\rm sig}, N_{\rm bg}, c_1, c_2) = N_{\rm sig}S(m) + N_{\rm bg}B(m|c_1, c_2)\,, \tag{3.14}$$

and obtaining the pull distributions for the four fitted parameters $(N_{\rm sig}, N_{\rm bg}, c_1, c_2)$ obtained using the same PDF for the generation. A correct fit should lead to four normal distributions, i,.e. centered in zero (no biases) and with a standard deviation compatible with unity (correct uncertainty estimation). The distributions are shown in Fig. 3.36 for the $\mathbf{OS}_e$ mode, while an example of fit to the pseudo-data is given in Fig. 3.37.

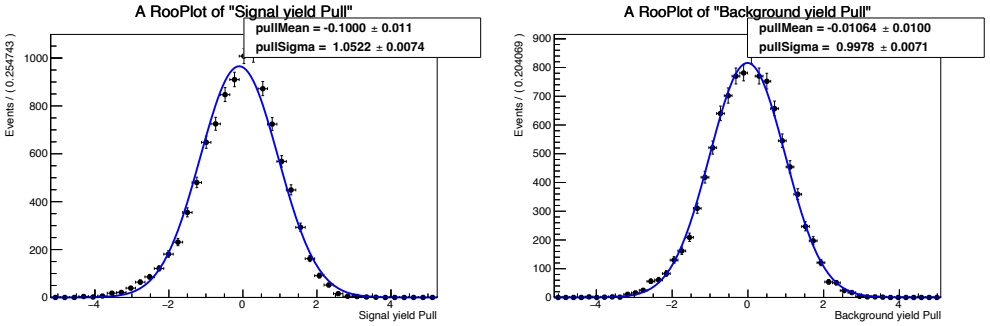
The BF is calculated using the formula:

$$\mathcal{B}^{(\rm UL)} = \frac{N_{\rm sig}^{(\rm UL)}}{\varepsilon \times N_{\rm B\bar{B}} \times f^{\pm/00}} \tag{3.15}$$

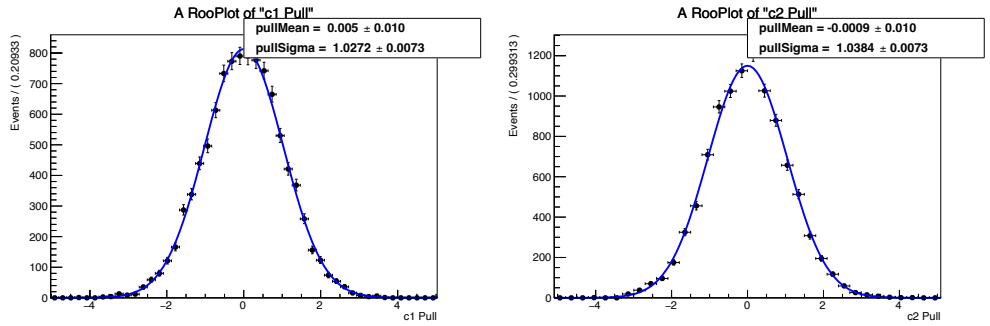
where

- $N_{B\bar{B}}$ is the number of $B\overline{B}$ pairs produced at KEKB: $(772 \pm 11) \times 10^6$ [98];
- ε is the final efficiency of signal MC;
- $f^{\pm/00} = f^{\pm}/f^{00}$ is the ratio of BFs between charged and neutral B pairs in $\Upsilon(4S)$ decays[12].

[12]This ratio is equal to 1.058 ± 0.024 [13] and is the result of the combination of $f^{\pm}/f^{00}$ measurements, obtained considering the isospin invariance in B^0 and B^+ decays, and of the direct measurements of f^{00} (assuming $f^{00} + f^{\pm} = 1$). $\mathcal{B}(\Upsilon(4S) \to B\overline{B}) = (51.4 \pm 0.6)\%$, $\mathcal{B}(\Upsilon(4S) \to B^0\overline{B}^0) = (48.6 \pm 0.6)\%$.



(A) N_{sig} (B) N_{bg}



(C) c_1 (D) c_2

FIGURE 3.36: Pull distributions of the fitted parameters for pseudo data with an assumed BF on the $K^+\tau^+e^-$ signal equal to 5×10^{-5}.

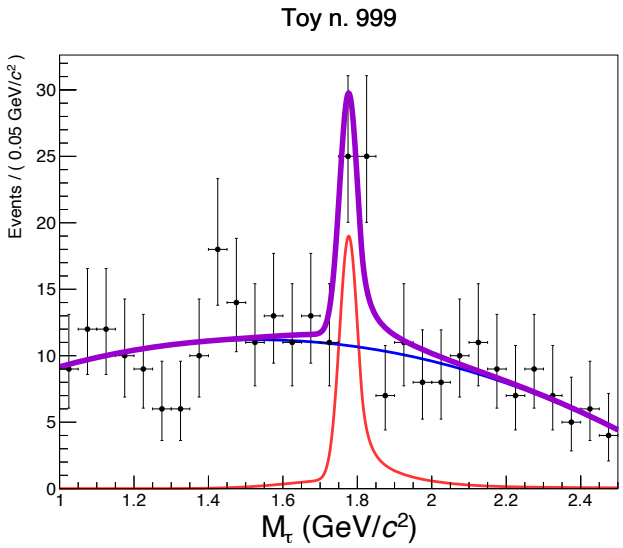
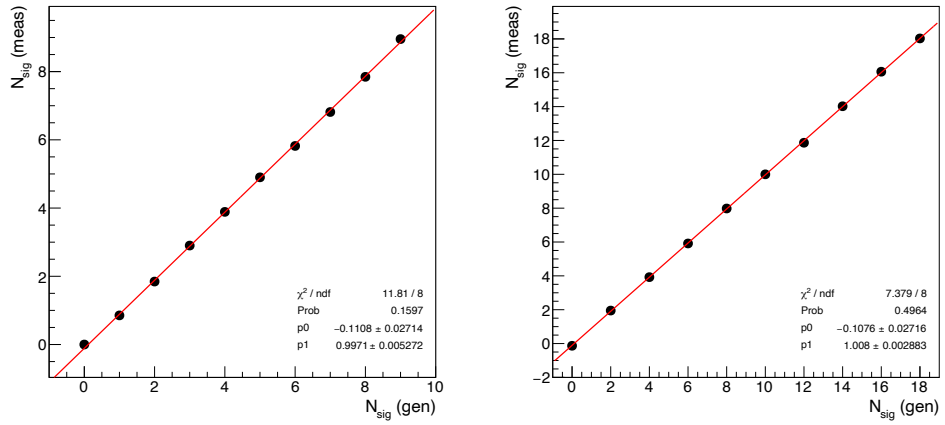


FIGURE 3.37: Fitted pseudo-experiment where the number of background events is compatible with the Belle data size and the injected BF for the signal $B^+ \to K^+\tau^+e^-$ is 5×10^{-5}.

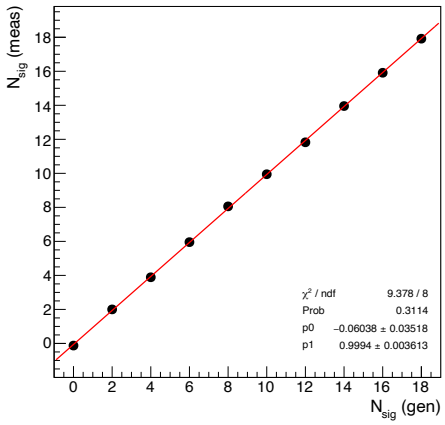
The second check consists in measuring a possible bias in the BF estimation. A 10-point scan in the signal yield $N_{\text{sig}}(\text{gen})$ with a mode-dependent range is performed, and the related measured yields ($N_{\text{sig}}(\text{meas})$) are shown in Fig. 3.38, with the linear fit proving the good agreement between the two sets. When significant, the differences between measured and injected N_{sig} are taken into account for the evaluation of the systematic uncertainty, otherwise only the statistical uncertainties on the linear fits

are taken; this point will be resumed in Sec. 3.11.

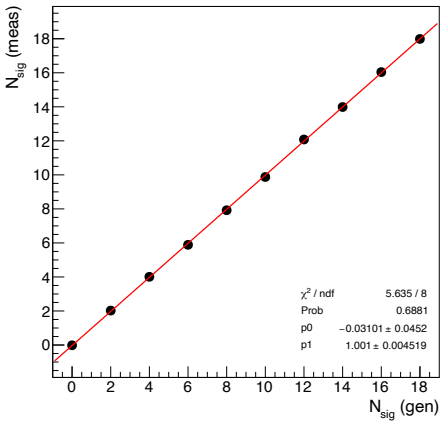


(A) $\mathbf{OS}_\mu$

(B) $\mathbf{SS}_\mu$



(C) $\mathbf{OS}_e$



(D) $\mathbf{SS}_e$

FIGURE 3.38: Linearity test: measured BF as a function of the BF assumption of the signal.

3.10 Control samples

Before looking at data, one must demonstrate that the developed analysis is mature enough and that a strategy is in place to measure the possible discrepancies between the MC simulation and the data. For this purpose, one needs to find one or more *control* samples, i.e., channels with similar properties to the signal, to perform the needed comparison between data and MC. We use the $B^+ \to D^-\pi^+\pi^+$ and $B^+ \to J/\psi K^+$ to measure on data the cut efficiencies related to the BDT selections. We also want to calibrate the signal PDF, which is fixed from MC simulation, to take into account possible differences with data and evaluate the related systematic effects. In this case, the so-called *inclusive* $B^+ \to \overline{D}^{(*)0}\pi^+$ sample is exploited.

The three calibration steps ($B\overline{B}$ and $q\overline{q}$ background suppression and signal PDF) are described in the next paragraphs.

3.10.1 $B^+ \to D^-\pi^+\pi^+$

The $B\overline{B}$ background suppression procedure might not perform on data the same way as on MC as a result of possible differences in the distributions of the input variables and the correlations among them. In order to take this effect into account, a control channel with similar kinematics to $B \to K\tau\ell$ with sufficient data is needed. With this in mind, we choose the mode $B^+ \to D^-(\to K^+\pi^-\pi^-)\pi^+\pi^+$: a three-body decay allowing to study the recoil mass in the same region as m_τ[13].

The total BF of this mode has been measured at Belle [99] and *BABAR* [100] and is equal to $(1.07 \pm 0.05) \times 10^{-3}$; in both measurements the resonant nature of the decay was observed coming from D^{**} decays (see Tab. 1.3) were measured. As will be discussed in sections 4.1 and 4.3, the $D\pi^+\pi^-$ modeling in Belle MC is not reflecting the current experimental indications. In order to have a more realistic model for a better agreement between data and MC and also to have a high statistic sample, we generate a $D\pi^+\pi^-$ sample following the Dalitz analysis performed by *BABAR* with the contributions below:

- $B^+ \to \overline{D}_0^{*0}(\to D^-\pi^+)\pi^+$: 65%
- $B^+ \to \overline{D}_2^{*0}(\to D^-\pi^+)\pi^+$: 35%
- $B^+ \to D^-\pi^+\pi^+$ non-resonant: 0%,

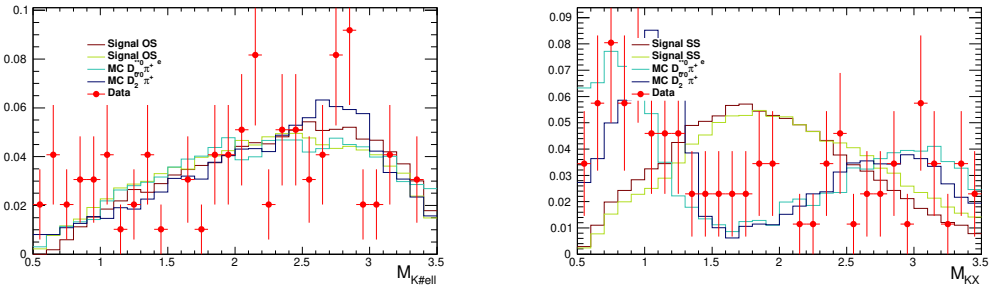
while the three-body $D^- \to K^+\pi^-\pi^-$ decay is generated with a Dalitz model taking into account of the measured $K^0\pi^-$ amplitude components [78].

The reconstruction procedure starts with the B_{tag} selection as described in section 3.4, with a veto on $B^+ \to D^-\pi^+\pi^+$ `FEI` mode to avoid the contamination from the signal side. Subsequently, the tracks that are not related with B_{tag} are combined to form the $B^+ \to D^-(\to K^+\pi^-\pi^-)\pi^+\pi^+$ with the cut $M_{K\pi\pi} \in (1.81, 1.95)\,\text{GeV}/c^2$. If more than one B_{sig} candidate passing the selection $M_{\text{bc}} \geq 5.24\,\text{GeV}/c^2$ and $|\Delta E| \leq 0.1\,\text{GeV}$ is obtained, the one with lowest ΔE is selected. Finally the request of having no tracks in the ROE is applied.

When forming the variables for the BDT training, the two prompt tracks from the B_{sig} are used as the primary kaon and lepton, while one of the pions from the D has the role of the τ prong, t_τ. The other two tracks from the D are considered as missing energy. It should be stressed that the $D\pi\pi$ system is far from having a PHSP-like kinematics: in general the prompt pion (π_1) has higher momentum than the one

[13] The PDG average of the D^- meson mass is $(1869.66 \pm 0.05)\,\text{MeV}/c^2$ while the τ mass is $(1776.86 \pm 0.12)\,\text{MeV}/c^2$ [13].

coming from the D^{**} decay (π_2). As a consequence, the $M(Kt_\tau)$ distribution, which is very important for background suppression, tends to have lower (higher) values when the assignment $K \equiv \pi_2$ ($K \equiv \pi_1$) is made. Such bias would entail a different BDT response with respect to $K\tau\ell$ for $\mathbf{SS}_\ell$ modes, where the $M(Kt_\tau)$. To avoid that, the $K_{\rm sig}$ is randomly assigned to π_1 or π_2. This problem does not arise for the $\mathbf{OS}_\ell$ modes, where the $M(K\ell)$ variable does not depend on the chosen assignment for the two prompt pions. In Fig. 3.39 we show the $M(K^-X^+)$ distributions for the $B^+ \to D^-\pi^+\pi^+$ control samples in both data and MC and compare them with the $K\tau\ell$ distributions. As we can see, the random assignment for the $\mathbf{SS}_\ell$ mode (Fig. 3.39B) provides a acceptable agreement between $D\pi^+\pi^-$ and $K\tau\ell$; furthermore, the used D^{**} model leads to good consistency with data for both $\mathbf{OS}_\ell$ and $\mathbf{SS}_\ell$ modes.



(A) (B)

FIGURE 3.39: $M(K^-X^+)$ distributions for data (red dots) compared to $B \to D^-\pi^+\pi^+$ MC sub-components and signal MC. $M(K_{\rm sig}\ell_{\rm sig})$ for $\mathbf{OS}_\ell$ (A) and $M(K_{\rm sig}t_\tau)$ for $\mathbf{SS}_\ell$ (B).

The response of the MC-trained classifier on the control sample thus composed is displayed in Figure 3.40. The obtained agreement is satisfactory and allows us to calibrate the efficiency of the $B\overline{B}$ rejection step.

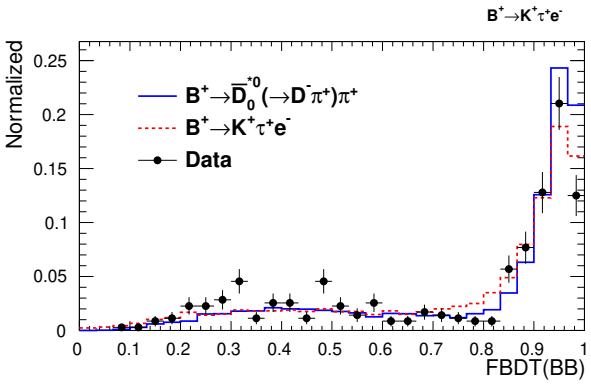


FIGURE 3.40: Data and MC comparison for BDT($B\overline{B}$) response for the $B^+ \to D^-\pi^+\pi^+$ sample obtained for the $\mathbf{OS}_e$ mode.

3.10.2 $B^+ \to J/\psi K^+$

Unlike the calibration of BDT($B\overline{B}$), for $q\overline{q}$ we do not need the control sample to be kinematically close to $B \to K\tau\ell$. The process $B^+ \to J/\psi(\to \ell\ell)K^+$ is good enough to establish the agreement between data and MC for the classifier based on variables describing the event shape and thrust information based on the ROE of the $B_{\rm tag}$

candidates.
The reconstruction is performed as follows:

1. B_{tag} candidate are reconstructed as before;
2. Kaon and leptons are identified as detailed in Sec. 3.5;
3. $J/\psi(\ell\ell)$ candidates must satisfy $M_{\ell\ell} \in (3.0, 3.2)\,\text{GeV}/c^2$;
4. B_{sig} candidates should have $|\Delta E| < 100\,\text{MeV}$.

As it will be described in detail in the next chapter, the $B^+ \to J/\psi(\to \ell\ell)K^+$ channels provides a very high purity because of the presence of the two high-momentum leptons with the constraint on their invariant mass to be compatible with a narrow resonance. The purity is even enhanced by the full reconstruction of the second B but at cost of very low final reconstruction efficiency: the total available yield on data is around 400 events. Conversely, for the MC study a much larger sample is available (100 times the size of the Belle on-resonance dataset). Figure 3.41 shows the agreement between data and MC of the BDT response for the $\mathbf{OS}_e$ mode, which is taken here for illustration.

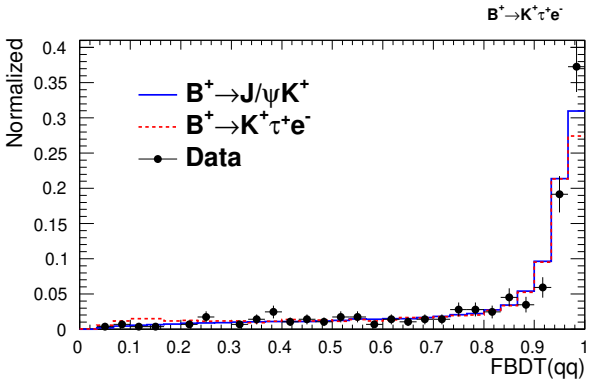


FIGURE 3.41: Data and MC comparison for BDT($q\bar{q}$) response, $\mathbf{OS}_e$ mode.

3.10.3 $B^+ \to \overline{D}^{(*)0}\pi^+$

A third control sample given by $B^+ \to \overline{D}^{(*)}\pi^+$ events is used to calibrate the signal PDF in data. It is preferred over the $B \to D\pi\pi$ described above because of the substantial gain in sample size: $\sim 36k$ against ~ 100 events. Such gain is due to the fact that the $\overline{D}^{(*)}$ mesons are not reconstructed and also to the relatively high BF of $B^+ \to \overline{D}^{(*)}\pi^+$ (roughly 0.5% for each modes).

Specifically, the M_{recoil} of $B^-_{\text{tag}} + \pi^+$ system shows clear $\overline{D}^{(*)}$ peaks that can be easily fitted as they appear at the lower side of the allowed PHSP region where the - exponentially growing - combinatorial background is still under control. The signal events in the MC sample are fitted with the same model described in Eq. 3.11 and all the parameters are the same for the D and the D^* events, except for the two means, which are fixed to the PDG average values. Fig. 3.42 shows the fit to MC $\overline{D}^0\pi^+$ events and Tab. 3.43 summarizes the fitted values for all the floated parameters.

The data and generic MC are fitted using the sum of the two signal PDFs for the D and D^* components, and the exponential function for the background:

$$M(m|N_{\text{sig}}, N_{\text{bg}}, \mu_{\text{shift}}, c_\sigma) = N_{\text{sig}}(f\mathcal{S}_D(m|\mu_{\text{shift}}, c_\sigma) + (1-f)\mathcal{S}_{D^*}(m|\mu_{\text{shift,c}_\sigma})) + N_{\text{bg}}e^{Am}\,.$$

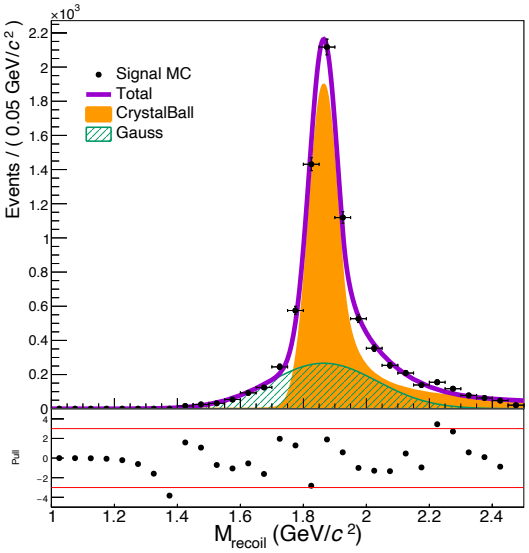


FIGURE 3.42: Fit to M_{recoil} corresponding to $B^+ \to \overline{D}^0\pi^+$ MC events.

Parameter	Fit result
f_{CB}	0.71±0.02
σ_{CB} (MeV/c^2)	45.7±1.4
α	−1.33±0.07
n	0.97±0.09
σ_{Gauss} (MeV/c^2)	170±5

FIGURE 3.43: Fit parameters for the M_{recoil} of $B \to D\pi$ events.

All the signal shape parameters are fixed but a shift on the mean (μ_{shift}) and a correction on the width of the CB (c_σ) are introduced. The exponential shape parameter is also left floated, as well as the total signal and background yields (while the D/D^* yield ratio is fixed).
The fits to MC and data show that the two μ_{shift} values are consistent within $\pm 1\,\text{MeV}/c^2$, which is used to evaluate the associated systematics. Concerning the width, the correction factor is obtained with the ratio $c_\sigma^{\text{data}}/c_\sigma^{\text{MC}} = (1.06 \pm 0.02)$ and its uncertainty is taken to estimate the systematic on the signal PDF width. Concerning the parameter f_{CB}, describing the weight of the Crystal Ball in the total signal PDF, we decide to assign as systematic a 10% variation. This choice is conservative considering that f_{CB} is large ($\sim$80%) and that the fit to $\overline{D}^{(*)0}\pi^+$ data shows a good agreement. Furthermore, the chosen sample does not allow to simply extract the correction on f_{CB} because of the more difficult modeling of the tails due to the high background level. As will be demonstrated in Sec. 3.12 and Tab. 3.12, the effect on the estimated BRs due to the PDF corrections are small compared to other systematics sources.
It is worth stressing that same results are obtained with different cut selections on the B_{tag} candidates; two configurations, called *Loose* and *Tight* and corresponding to the cuts

- $|\Delta E| < 100\,\text{MeV}$ and $\mathcal{P}_{\text{FEI}} > 10^{-3}$
- $|\Delta E| < 50\,\text{MeV}$ and $\mathcal{P}_{\text{FEI}} > 10^{-2}$

are shown in Fig. 3.44.

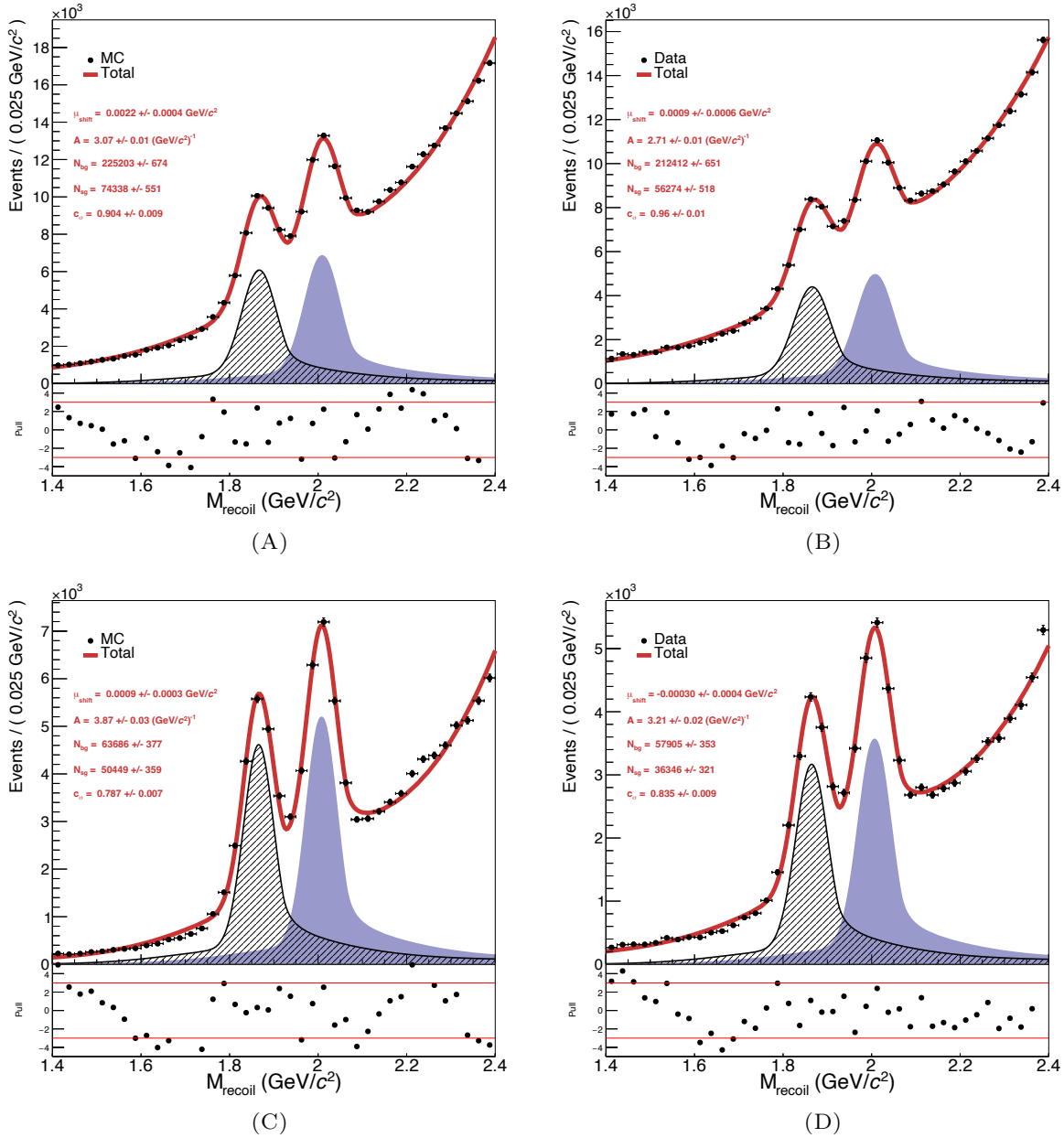


FIGURE 3.44: Fit to the recoil mass of the $B_{\text{tag}}+\pi$ system for MC (left column) and data (right column). Two cuts on `SignalProbability` of B_{tag} candidates are considered: the *Loose* cut (A-B) and the *Tight* one (C-D).

3.11 Systematic uncertainties

The systematic uncertainties can be divided into two categories: to the first belong the uncertainties related to the $N_{\rm sig}^{\rm UL}$ and they are therefore called 'additive'. They affect the fit yield and thus the significance of the measurement. In the second category all the effects that can give a bias to the terms at the denominator of Eq. 3.15 are regrouped and collectively called 'multiplicative'. They affect the central value of the measured BF but not its significance.

We provide below a review of the *multiplicative* systematic uncertainties:

FEI Calibration. As mentioned already, the tag-side efficiency for the hadronic `FEI` algorithm is rather different in simulated events and 'real' data. This is due to mis-modelings in our MC related to the poor knowledge of hadronic B-decays which has a non-trivial effect in the selection process of `FEI`, especially when training of the classifiers. To measure such discrepancy between data and MC with an acceptable accuracy, the full $\Upsilon(4S)$ dataset recorded at Belle is used and five semileptonic B^+ decays are reconstructed, in addition to the $B_{\rm tag}$, which is reconstructed via `FEI` [11]. In fact the semileptonic $B^+ \to \overline{D}^{(*)0}\ell\nu$ channels have relatively high BFs, which are also known with good precision from independent measurements. When comparing the total yields between simulation and data, all the known differences in terms of reconstruction efficiency are taken into account and the remaining disagreement is assumed to be caused by the different tag-side efficiency. Therefore, the ratio $\varepsilon_{\rm data}/\varepsilon_{\rm MC}$ is the calibration factor that must be considered as correction. The global factor used for this analysis is $0.863 \pm 0.014 \pm 0.050$, where the second uncertainty has a systematic nature and arises from the tracking, PID and BF corrections on the signal side, but most notably from the modeling of background for signal extraction. The above number is the result of an average of all the $B_{\rm tag}$ modes and is related to a specific channel on the signal side. Among the five chains studied in Ref. [11] (with different combinations of D/D^* hadronic decay modes), we consider $B^- \to D^0(K^-\pi^+)\ell^-\bar{\nu}_\ell$ because it has the same track multiplicity of $K\tau\ell$[14]. Furthermore, the tag-side selection performed for the calibration is very close to ours: no minimum cut on `SignalProbability` is applied but the candidate with highest `SignalProbability` is chosen as best one, while the selection on the $M_{\rm bc}$ and ΔE variables depends on the $D^{(*)}$ mode. The study in Ref. [11] also provides the correction factors for each `FEI` modes ε_i: our average correction factor is the result of the linear combination:

$$\sum_i f_i \varepsilon_i$$

where f_i is the fraction of signal events labelled as belonging to the i-th `FEI` mode. We obtain a calibration factor equal to $(85 \pm 5)\%$[15] and the uncertainty is taken as a systematic uncertainty.

Tracking. The systematic uncertainty due to the charged track reconstruction is evaluated using $D^{*+} \to D^0\pi^+$ with $D^0 \to K_S^0\pi^+\pi^-$, resulting in an uncertainty of 0.35% for each signal-side track with $p_T \geq 200\,{\rm MeV}/c$.

[14] The `FEI` performance can depend from the signal-side multiplicity as it introduces different possible combinatorial backgrounds.

[15] Statistical and systematic uncertainty are summed in quadrature

Lepton ID. The uncertainty due to the lepton identification is evaluated using a tag-and-probe method on $J/\psi \to \ell^+\ell^-$ events [101], resulting in an uncertainty of 0.3% for muons and 0.4% for electrons. The lepton identification efficiency does not only involve the $\ell_{\rm sig}$, but also t_τ.

K/π identification. The K/π separation can also be a source of systematic uncertainty. In fact the efficiency related to the kaon selection for $K_{\rm sig}$ (kID> 0.6) and pion selection for t_τ (πID> 0.6) do differ between data and MC. The study presented in Ref. [102] provides the efficiency corrections in bins of momentum and $\cos\theta$ of the track. The uncertainty on the efficiency factors is used to estimate the amount of systematic uncertainty on the final BF measurement.

BDT. The BDT to suppress background is trained with MC events. Differences in the input variable distributions between MC and experimental data might introduce a bias in the calculation of the signal efficiency. To estimate the associated uncertainty, the BDT output is calculated for the appropriate control modes. For the $B\overline{B}$ suppression, the $D\pi\pi$ sample described in Sec. 3.10.1 is used. The total signal yields before and after $B\overline{B}$ background suppression are obtained with a fit to the $M_{\rm bc}$ distributions and the ratios between the two provide the correction factors. The uncertainties on the said factors, obtained from the $M_{\rm bc}$ fit, are the source of systematic uncertainty. Similarly, the corrections to the efficiencies related to the $q\bar{q}$ suppression are estimated from the $M_{\rm bc}$ fitting – this time using the $B \to J/\psi K$ data, see Sec. 3.10.2.

MC statistics The signal efficiency $\varepsilon_{\rm sig}$ is computed from a MC signal sample with a large but finite number of events $N_{\rm tot}$. The uncertainty on $\varepsilon_{\rm sig}$ is $\sigma(\varepsilon_{\rm sig}) = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{\rm tot}}}$ and is considered as a source of systematic uncertainty.

$\mathbf{N_{B\overline{B}}}$ and $\mathbf{f^{\pm}/f^{00}}$. The total number of available B^+B^- at Belle is obtained by subtracting the off-resonance hadronic contributions retained by the *HadronBJ* skim, from the total number of on-resonance hadronic events [70]. The uncertainties on $N_{B\bar{B}}$ and also the one on f^{+-}/f^{00} are propagated to estimate the effect on the measured BF.

The *additive* systematic uncertainties are related to the possible bias induced by the model used to fit the signal component and differences in MC with respect to data. The effects we take into account for are related to the signal PDF parameters $\mu, \sigma_{\rm CB}, f_{\rm CB}$ (Eq. 3.11). Uncertainties in the shape of the PDFs used for the signal are evaluated by varying the first two parameters above by the uncertainty on the correction factors obtained from the $B^+ \to \overline{D}^{(*)0}\pi^+$ control samples (using $M_{\rm recoil}$ procedure with a pion and the $B_{\rm tag}$ – Sec. 3.10.3). Concerning $f_{\rm CB}$, it is varied in a $\pm 10\%$ range. The resulting change in the signal yield is taken as the systematic uncertainty. Another possible source of bias related to the fitting process has already been examined in Sec. 3.9.3: the linearity between measured and generated signal yields is checked and the uncertainties on the fits shown in Fig. 3.38 are taken into account as a systematic source. The systematic uncertainties described above are evaluated for the four $B^+ \to K^+\tau\ell$ mode and collected in Tab. 3.12, expressing them in terms of number (fraction) of signal events for the additive (multiplicative) terms.

A special remark needs to be made for the choice of the model for the $B \to K\tau\ell$ MC generation. As presented in Sec. 3.1, the analysis is performed and optimised with signal events generated according to the PHSP model. However, LFV in $b \to s\ell\ell$ decays can in principle be associated to BSM mediators which can be encoded in

the $\mathcal{H}_{\text{eff}}$ with either scalar or vector operators (or a combination of the two). These operators would entail very different q^2 distributions and, consequently, different signal efficiencies. In case LFV is produced by scalar operators only, the q^2 is as far from PHSP as possible; for this reason, weight factors assuming the extreme scenario with scalar-mediator only are applied to the PHSP-q^2 distributions in order to obtain a lower bound on ε_{sig}, which correspond to the most conservative BF. The numerical details of how the re-weighting is performed are given in Appendix A.

3.12 Results on data

The analysis on collision data consists in performing all the steps described in sections 3.4-3.8: reconstruction, best candidate selection and background suppression by applying the BDT classifiers trained on MC.

Fig. 3.45 shows the the BDT outputs and M_{recoil} distributions – after best candidate selection but before any cuts – for $\mathbf{OS}_e$ candidates. Black dots represent data and the MC distributions (colored histograms) are stacked and overlaid in the same plot.

The M_{recoil} is then fitted with the PDF defined in Eq. 3.14 to extract number of signal events N_{sig} and calculate the BF. Figure 3.46 collects the four plots with the fitted distributions of the events surviving all the cuts. The fitted N_{sig}, corrected for all the calibration factors described in Sec. 3.11, are listed in Tab. 3.13.

As we find no significant excess in the signal window for any of the four studied modes[16], we proceed with the derivation of the ULs on their BFs.

The last needed ingredient for the UL estimation is the systematic uncertainty. The total additive systematic uncertainty, expressed in number of events, and multiplicative one (shown as a fraction in percent) are obtained by summing in quadrature all systematic uncertainties of the same type and are summarized in Tab. 3.12. One can see that the largest contributions to the systematic uncertainty are given by the BDT background suppression efficiency corrections (between 6 and 13%). This is due to the small data size available for both control samples ($D\pi\pi$ and $J/\psi K$); however one should bear in mind that this does not alter the sensitivity of our search, which is anyway statistically limited. It is worth mentioning that while the contributions for the BDT selection could be reduced by choosing different samples with higher BRs, the one corresponding to the B_{tag} correction (6%) is limited by the systematic uncertainty on the `FEI` calibration factor. In chapter 4 we propose a new sample to obtain the calibration with a smaller systematic uncertainty.

Given the measured N_{sig}, the ULs at 90% C.L. are calculated following a frequentist method. We generate sets of pseudo-events, with each set being statistically equivalent to our data sample of $711\,\text{fb}^{-1}$. Signal and background events are generated according to their respective PDFs obtained from data. The number of generated background events is picked for each toy from a Poisson distribution with mean equal to the N_{bg} measured in data. The number of input N_{sig} is varied, and for each value we generate an ensemble of 50k data sets. We fit these data sets and calculate the fraction $\hat{f}$ with fitted signal yield less than that obtained on data. The 90% C.L. UL on the number of signal events ($N_{\text{sig}}^{\text{UL}}$) is the number of input signal events corresponding to $\hat{f} = 0.1$.

We include the systematic in our procedure by correcting the N_{sig} obtained from the MC fits with the fractional systematic uncertainty and adjusting it by a value sampled from a Gaussian distribution centered on zero and having a width equal to the total additive uncertainty.

The scan in N_{sig} used to derive $N_{\text{sig}}^{\text{UL}}$ is shown in the four plots of Fig. 3.47.

Table 3.13 summarizes the obtained signal yields N_{sig} and the corresponding $\mathcal{B}^{\text{UL}}$ (90% C.L.), computed according the Eq. 3.15. The expected limits for zero signal are also reported for comparison.

Fig. 3.48 combines the results shown in the table 3.13, adding the result from LHCb (B_{s2}^{*0}-tagged analysis with $9\,\text{fb}^{-1}$) for the $\mathbf{OS}_\mu$ mode, as well as the expected

[16] It was established that, in case of a signal yield significantly larger than zero, the significance would be given by $-2\ln(\mathcal{L}_0/\mathcal{L}_{\max})$. Here $\mathcal{L}_{\max(0)}$ denotes the value of the likelihood corresponding to the fit to the M_{recoil} when the signal yield is allowed to vary (fixed at zero).

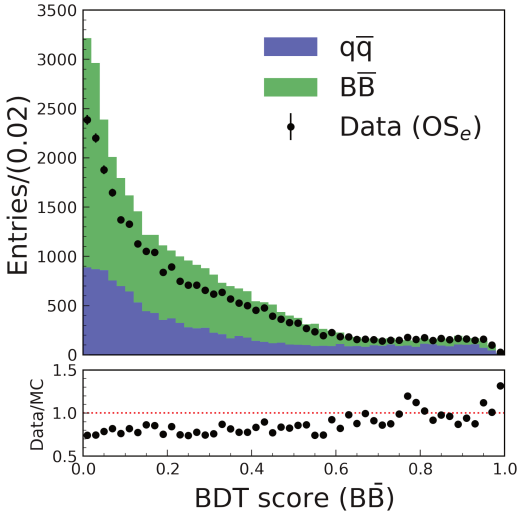
Source	$K^+\tau^+\mu^-$	$K^+\tau^+e^-$	$K^+\tau^-\mu^+$	$K^+\tau^-e^+$
PDF shape (mean)	0.09	0.01	0.08	0.08
PDF shape (width)	0.02	0.08	0.04	0.07
PDF shape ($f_{\rm sig}$)	0.28	0.16	0.11	0.16
Linearity	0.03	0.04	0.02	0.04
Total (evts)	**0.30**	**0.18**	**0.14**	**0.20**
$B_{\rm tag}$	6.1	6.1	6.1	6.4
Track reconstruction	1.1	1.1	1.1	1.1
Kaon identification	1.3	1.4	1.3	1.3
Lepton identification	0.3	0.4	0.3	0.4
h_τ identification	0.7	0.7	0.6	0.6
MC statistics	1.0	1.5	1.2	1.0
Number of $B\bar{B}$ pairs	1.4	1.4	1.4	1.4
BDT $B\bar{B}$ selection	10.6	10.0	12.7	12.6
BDT $q\bar{q}$ selection	8.8	8.6	9.2	6.6
f^{+-}	1.2	1.2	1.2	1.2
Total (%)	**15.4**	**14.9**	**17.1**	**15.8**

TABLE 3.12: Contributions to the systematic uncertainties of the measurements.

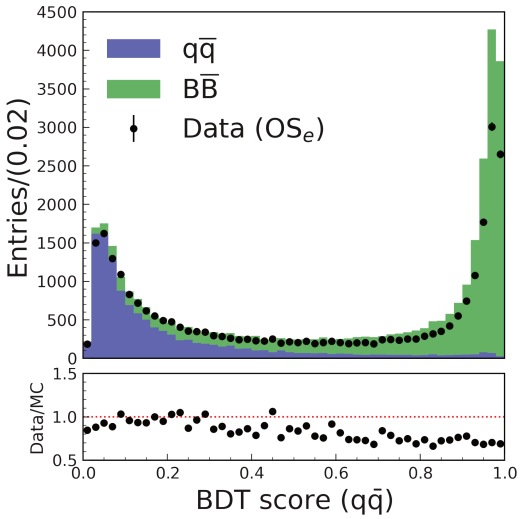
Mode	$\varepsilon_{\rm PHSP}(\%)$	$\varepsilon_{\rm BSM}(\%)$	$N_{\rm sig}$	$\mathcal{B}^{\rm UL} \times 10^{-5}$	$\mathcal{B}^{\rm UL}_{\rm exp} \times 10^{-5}$	PDG$\times 10^5$
$K^+\tau^+\mu^-$	0.064	0.058	-2.1 ± 2.9	0.65	1.18	2.8
$K^+\tau^-\mu^+$	0.046	0.038	2.3 ± 4.1	2.97	1.81	4.5
$K^+\tau^+e^-$	0.084	0.074	1.5 ± 5.5	1.71	1.34	1.5
$K^+\tau^-e^+$	0.079	0.058	-1.1 ± 7.4	2.08	2.29	4.3

TABLE 3.13: ULs for the four LFV $B^+ \to K^+\tau\ell$ modes. The results quoted by PDG [13] refer to the *BABAR* results.

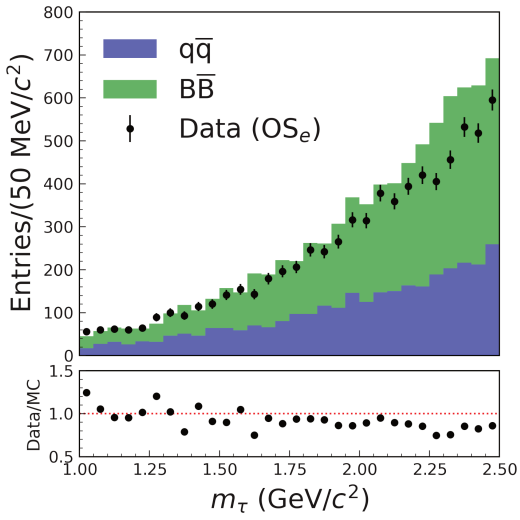
sensitivities at Belle and LHC*b*. As can be seen, the expected limits are significantly better than the previous results and so are the measured ones, despite the positive fluctuations for the $\mathbf{SS}_\mu$ and $\mathbf{OS}_e$ modes. Except for $\mathbf{OS}_e$, the obtained ULs are the most stringent to date.



(A)

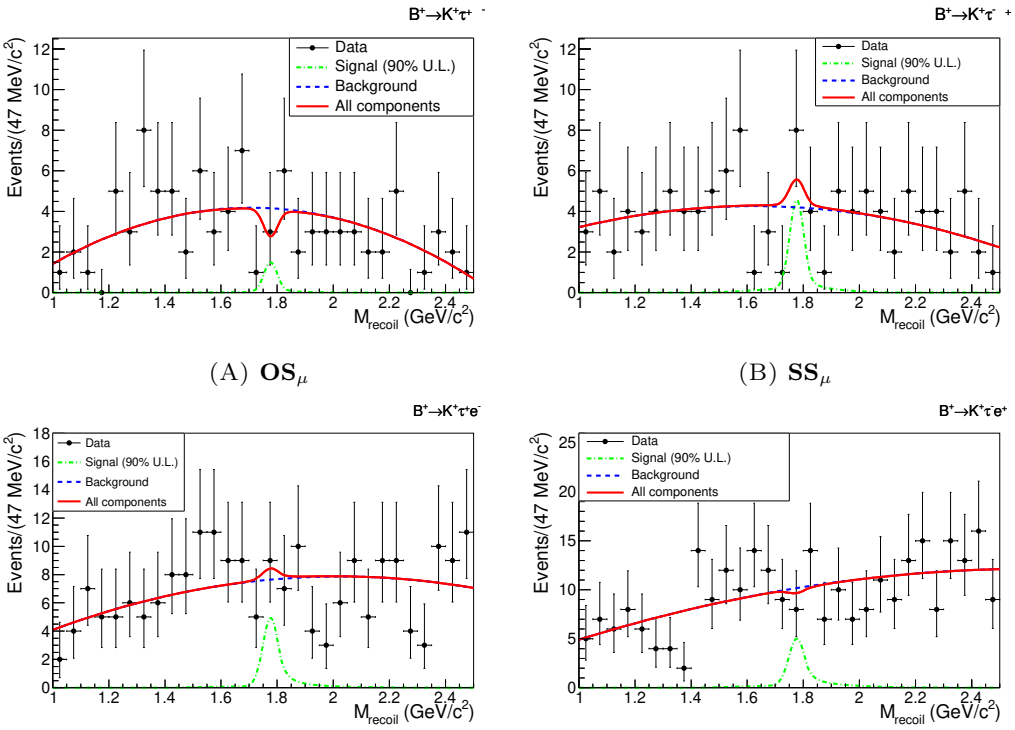


(B)



(C)

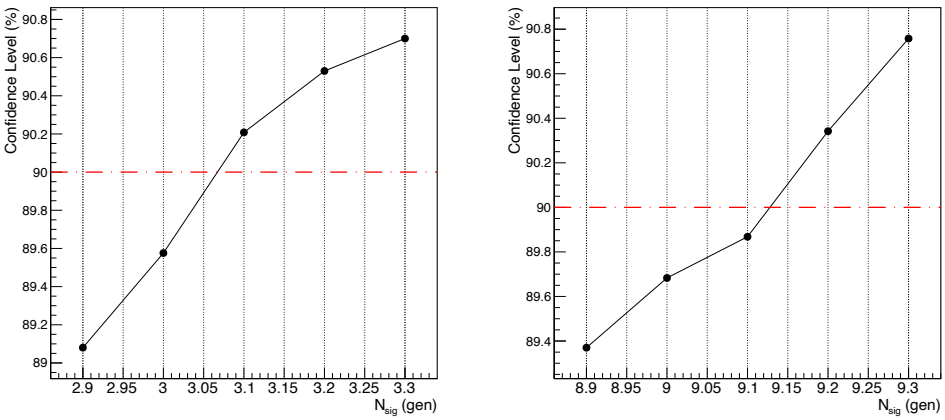
FIGURE 3.45: BDT responses (A-B) and m_τ (C) for $\mathbf{OS}_e$ data overlaid to MC simulation. Efficiency corrections due to data-MC discrepancies are not applied here.



(A) $\mathbf{OS}_\mu$ (B) $\mathbf{SS}_\mu$

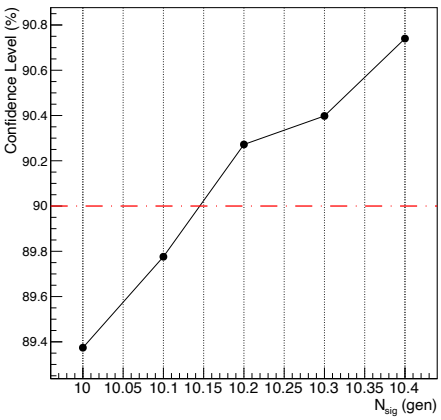
(C) OS_e (D) SS_e

FIGURE 3.46: Observed M_{recoil} distributions for the four LFV $B \to K\tau\ell$ modes and fit projections. The black dots represent the data points while the dashed blue and solid red curves show respectively the background components and the overall fit results. The dash-dotted green curves represent the signal PDFs with yields corresponding to the obtained limits on the BFs at the 90% C.L. .

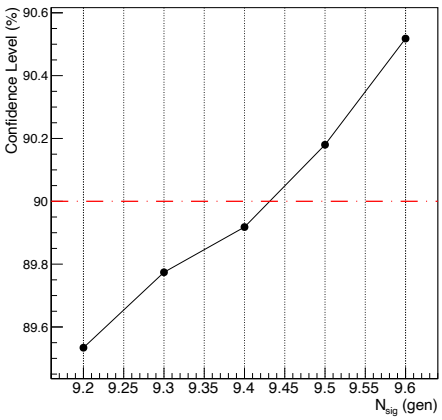


(A) $\mathbf{OS}_{\mu}$

(B) $\mathbf{SS}_{\mu}$



(C) $\mathbf{OS}_{e}$



(D) $\mathbf{SS}_{e}$

FIGURE 3.47: Scan in the injected N_{sig} as the way to estimate, with a frequentist approach, the ULs on the BFs. The 90% threshold is indicated with the red dashed line and provides the $N_{\text{sig}}^{\text{UL}}$ used for the BF calculation (Eq.3.15).

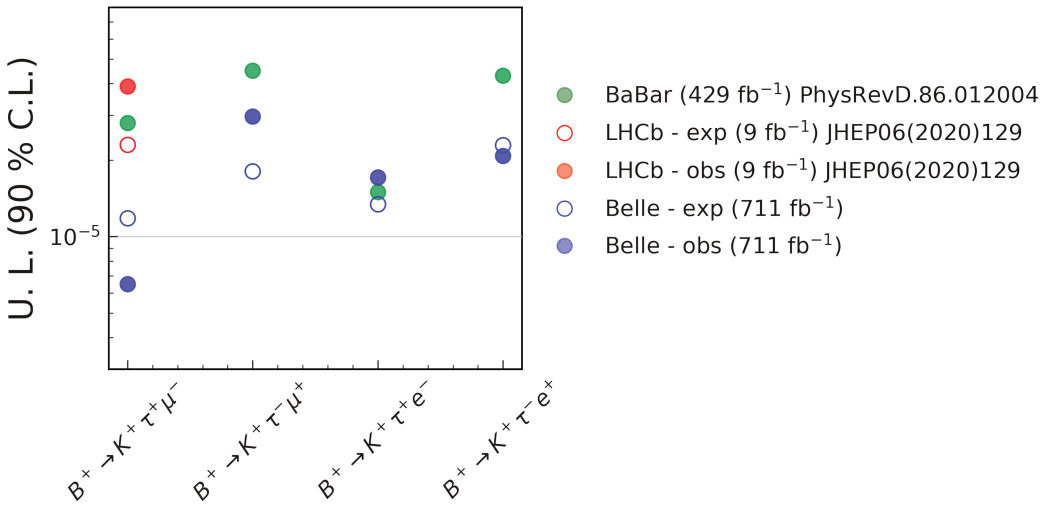


FIGURE 3.48: Summary of the current experimental status for the LFV search through the modes $B^+ \to K^+\tau\ell$. The previous limits set at *BABAR* and LHC*b* are shown, compared to the new results obtained at Belle. The expected limits are also added, when available.

Chapter 4

Improving the hadronic B-tagging

In this chapter, we examine the performance of the FEI algorithm in the context of the $B^+ \to K^+\tau\ell$ search and explore its possible improvements. Besides ours, many other analyses performed at Belle and Belle II [103] involving missing-energy modes would significantly benefit from in-depth studies of B-tagging and a better understanding of B-decays in general. Compared to other algorithms (SER at *BABAR* [70] and FR at Belle [9]), the FEI provides a higher efficiency because of the addition of more exclusive decay chains and loose selection cuts (see section 3.4). However, the tag-side efficiency has to be associated with its purity, which is a function of the SignalProbability. As shown in the previous chapter, we observe a quick drop in efficiency for high purity. Another crucial element regards the very different FEI efficiency between data and MC. To take this discrepancy into account, a global calibration factor has to be derived using a control sample (for example $D\ell\nu$) and applied to the final signal efficiency – as we do for the $B \to K\tau\ell$ BF ULs extraction in Sec. 3.11. The overall calibration factor for B^+ FEI at Belle is 0.85, but a closer look shows that the calibration strongly depends on the FEI mode: some modes have correction factors as low as $\sim$0.5, which is uncomfortable. The two aspects above can be explained by the fact that many B^+-modes have a wrong or incomplete description in our simulation. As FEI is based on BDT classifiers trained on MC, making sure that the MC is well modeled is essential for optimal performance (for any ML-based B-tagging).

In the first part of the chapter, we study the decay chains contributing the most to the total FEI efficiency and propose the necessary fixes to our MC. In order to isolate the FEI B_{tag} candidates and study their properties in MC and data, we use a $\sim$100%-pure sample of $B^+ \to J/\psi K^+$. This strategy, combined with the in-depth literature reading, improves the agreement with data not only regarding the FEI mode-dependent efficiency but also the kinematic distributions. Secondly, we suggest some improvements inspired by the research in the chapter's first part. The FEI could better exploit the knowledge of the intermediate states (for example, the presence of narrow intermediate resonances), producing a given final state; furthermore, a few B-modes encountered in our readings and the recovery of partially-reconstructed tags, could easily increase the total efficiency.
Most of the results presented in this chapter have been documented in an internal Belle II note [104].

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

Unlike semileptonic B-decays, where the MC modeling is easier due to the smaller total BF $\mathcal{B}(B^+ \to X_c\ell^+\nu_\ell) \sim 25\%$ and the higher BF of the single modes $\mathcal{O}(10^{-2})$, for hadronic decays the task is much more complex, with vast number of possible decay

modes of typical rates of $\mathcal{O}(10^{-3})$ or less. Even if we consider the $b \to c$ transition, or better the $B^+ \to \bar{D}^{(*)0} n\pi$ modes, the description becomes highly non-trivial as soon as more than two pions are involved. This is due to the increasing number of possible intermediate states and the fact that theoretical predictions become scarce (in many cases, the factorization does not hold anymore), and the experimental results are lacking.

Even interpreting or correcting our decay table[1] is not easy, given that measurements are often old and with large uncertainties, and PDG listings require some careful interpretation from the reader. For example, BFs of multi-body decays are often reported for the resonant sub-modes and can be mistaken for total ones. Regarding the maintenance of the decay table, the history of the changes applied is also hard to track.

The poor modeling also comes from the fact that `PYTHIA` produces more than half of the hadronic decay, requiring special care when tuning its parameters. B-decays are generated by `EvtGen`, the framework collecting modules describing the decay amplitudes to simulate the entire decay chains with all the angular properties. It has an interface to `PYTHIA`, which takes over not only for the modeling of $q\bar{q}$ events but also the fragmentation into all the possible B-final states which are not explicitly mentioned in the decay table. The `PYTHIA` algorithm checks whether the obtained final state coincides with any of those explicitly listed (having a known BR), in which case, it performs the fragmentation again to produce an alternative one. Except for tuning some global properties of the samples, like the known inclusive productions of notable mesons, this generation process via `PYTHIA` is basically blind as the final states are produced with uncontrolled BF.

As an illustration, the line below shows how `EvtGen` invokes `PYTHIA` for the generation of B^+ decays with a given quark content (in this case, $\bar{b} \to \bar{c}ud$ current and u is the *spectator* quark), BF and physics model described via the `modeID`:

```
0.31234    u    anti-d    anti-c    u    PHOTOS PYTHIA  48.
```

Some of the `modeID`'s in Belle MC, corresponding to the `PYTHIA` version 6, are:

- **48**: $n \geq 3$-body weak decays with PHSP model[2],
- **13**: Decays of $n \geq 2$ bodies. The multiplicity is decided based on a Poisson distribution (turn partons into a random number of hadrons),
- **23, 24**: Color suppressed baryonic decays (PHSP model with fixed multiplicity).

As illustrated later, the BFs obtained with `PYTHIA` are difficult to tune, and `PYTHIA`'s interplay with `EvtGen` can easily lead to double counting when different intermediate modes can produce the same final state.

A representation of how the B^+ width is shared between hadronic and semileptonic decays, and between `EvtGen` and `PYTHIA` in the Belle official MC is shown in Fig. 4.1.

The MC description is much simpler if we look at other particles, for example the τ lepton or the D mesons. In the first case, almost 90% of the total width is reached by

[1] The decay tables used by `EvtGen` contains the list of known decay channels and their BFs for particles up to the $b\bar{b}$ resonances. It is regularly updated with the available experimental and some guidance from theory.

[2] According to the `PYTHIA` manual [105], the charmed meson decay products are distributed according to the V−A matrix element, whereas the rest of the system is assumed isotropic in its rest frame.

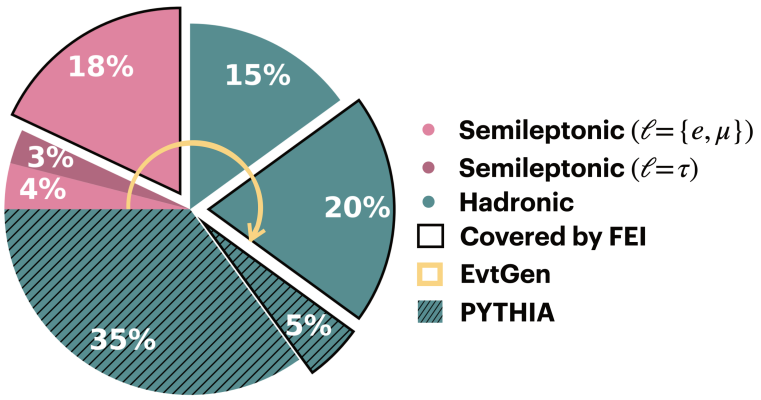


FIGURE 4.1: Composition of simulated B^+ events in Belle MC ($b \to c$ transitions). Pink color indicates the $B^+ \to X_c \ell^+ \nu_\ell$ processes (darker shade for semi-tauonic decays) while green stands for hadronic. The dashed area corresponds to the BF simulated by `PYTHIA`. Solid black line contours show the effective BF covered by the `FEI` algorithm.

just considering the leptonic decays and the hadronic decays into one, two and three pions (Fig. 4.2). The $\pi^-\pi^0$ component is almost entirely coming from the resonant $\rho^-\nu$ while the $a_1^-\nu$ is almost evenly shared between $\pi^-2\pi^0$ and $\pi^-\pi^+\pi^-$ (the a_1^- main channels are $\rho^-\pi^0, \rho^0\pi^-, f_0(600)\pi^-$).

As for the D-decays, tree-level hadronic and semileptonic decays occur through the $c \to sW^+$ current and thus have amplitudes governed by the CKM matrix element $|V_{cs}|$ (CF processes). They decay dominantly (84%) into hadronic final states and, since the charm-quark mass is quite sizable, the number of hadronic D decays is rather large with BFs of $\mathcal{O}(10^{-2})$. About 63% of the width consists of two-body decays (if intermediate resonances are considered as single particles). Depending on how the W^+ materializes, one gets the CF $c \to s\bar{d}u$ modes, the singly Cabibbo-suppressed ones like $c \to d\bar{d}u$ or $c \to s\bar{s}u$ and finally the doubly Cabibbo-suppressed modes via $c \to d\bar{s}u$.

Getting back to the B-decays, we compare the available predictions (using Eq. 1.10 in Sec. 1.3) with the composition of our MC B^+ samples. We see from Fig. 4.3 that the main contributions come from the $b \to c\bar{u}d$ transitions, accounting for 50% of the total BF, followed by $b \to c\bar{c}s$ decays (20%), responsible for double charm decays and charmonia production. These two contributions are consistent with the predictions [15], considering the large uncertainties on the theoretical values[3].

A breakdown of $b \to c\bar{u}d$ decays, illustrated in Fig. 4.4, shows that only one-third of them (corresponding to a BF of 14.4%) is described with exclusive modes while the rest (35.9%) is left to `PYTHIA`. This huge part of BF, generated via the quark fragmentation, corresponds to many wrong or unverified decay modes, among which some could be interesting for B-tagging.

A strategy to understand the MC simulation and the differences with data, restricting ourselves to the modes contributing to `FEI`, is presented in the next section.

4.2 Probing the B with $B^+ \to J/\psi K^+$ events

To study the properties of B-decays in data and MC with high purity and sufficient data size, we need to follow a sort of 'tag and probe' method, where we isolate a

[3] We remind the reader that the values used are $r_{c\bar{u}d} = 4.1 \pm 0.4$ and $r_{c\bar{c}s} = 1.95 \pm 0.50$.

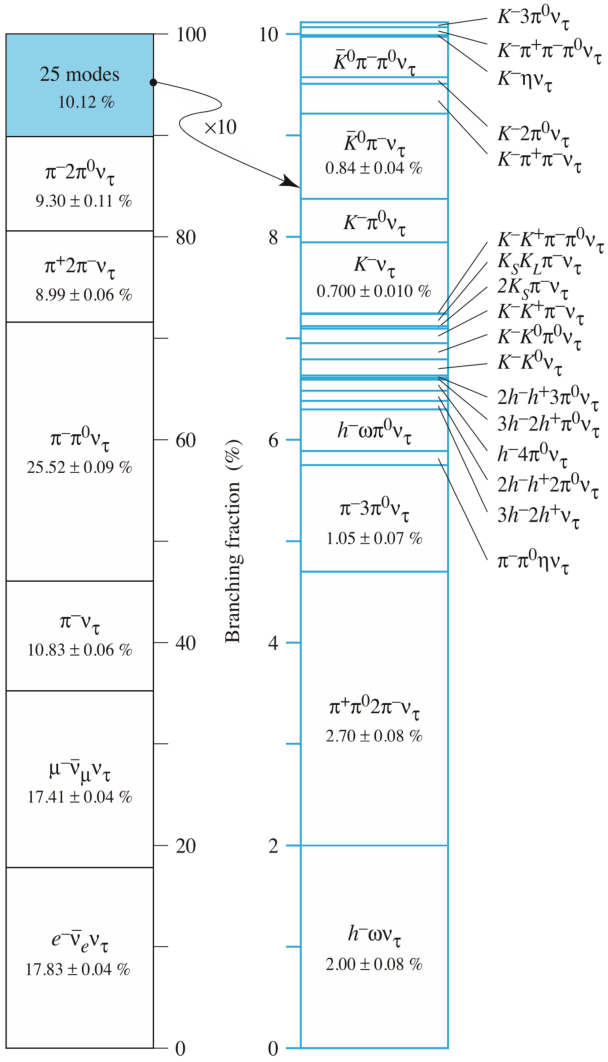


FIGURE 4.2: τ lepton decay modes [106]. Leptonic and hadronic decays with $n\pi n_0\pi^0$, ($n = \{1,3\}, n_0 = \{0,2\}$) cover ∼90% of the total width.

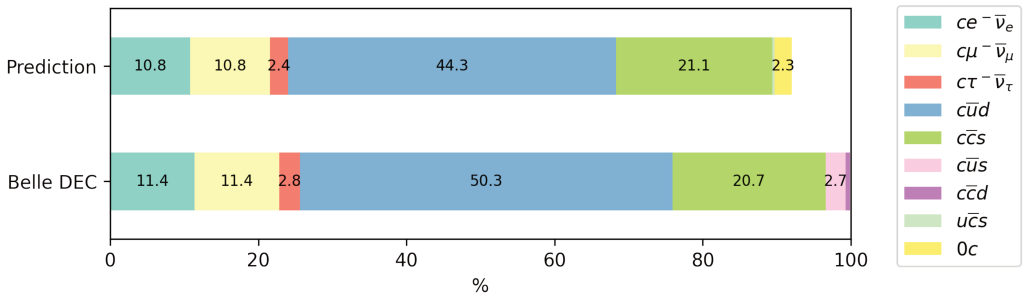


FIGURE 4.3: B^+ decays in MC compares with predictions obtained from Eq. 1.10 with the inputs provided by [15] – see also section 1.3.

single B meson by reconstructing the other via a clean, hadronic mode. Our choice falls on $B^+ \to J/\psi K^+$ because of its very high purity (∼ 99%) and relatively large data sample (∼ 35k events for the Belle dataset), allowing us to obtain a sample of

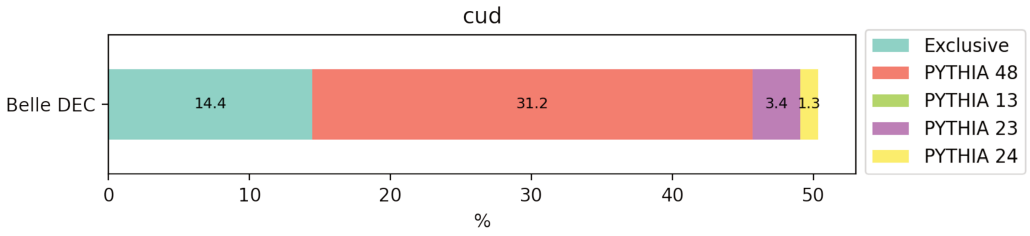


FIGURE 4.4: $b \to cud$ transitions in Belle MC: composition in terms of generator mode – `EvtGen` or `PYTHIA`– and model used.

B's with known flavor. Furthermore, as shown in chapter 3, we have at our disposal a huge MC sample (100 times the size of the Belle dataset) that allows us to observe small-BF modes with enough data.

The following selection criteria are adopted:

- Tracks: $|dr| < 1\,\text{cm}; |dz| < 3\,\text{cm}$
- Lepton candidates: eID$> 0.9\,(e^{\pm})$; μID$> 0.9\,(\mu^{\pm})$[4]
- Kaon candidates: $P_{K/\pi} > 0.6$
- J/ψ candidates: $2.947(2.847) < M_{\mu^+\mu^-(e^+e^-)} < 3.177\ \text{GeV}/c^2$

And then, we select the signal $B^{\pm} \to J/\psi\,(\to \ell^+\ell^-)K^{\pm}$ candidates with:

- $M_{\text{bc}} > 5.27\ \text{GeV}/c^2$
- $|\Delta E| < 0.1$ GeV.

This selection provides 35255 (3455039) events in data (MC). As it can be seen in Fig. 4.5, where data and MC (scaled to the Belle total integrated luminosity) are compared, the cut on M_{bc} is almost ineffective as the obtained sample is already very pure whereas the cut on ΔE discards the candidates arising from *reflections* related to $B \to J/\psi K^*$ events where a pion is missing.

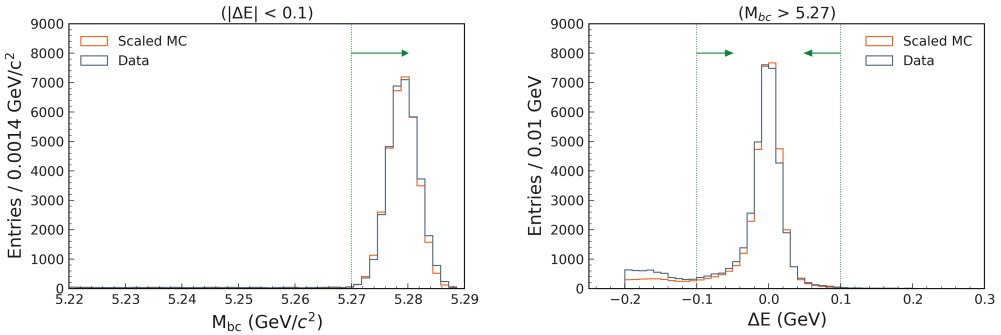


FIGURE 4.5: M_{bc} and ΔE pre-selection cuts on the $B^+ \to J/\psi\,K^+$ candidates.

In the ROE obtained after excluding the particle objects that are used to form the $J/\psi K$ candidate, we run the hadronic `FEI` reconstruction and, among the candidates passing the selection $M_{\text{bc}} > 5.24\ \text{GeV}/c^2$, $|\Delta E| < 0.1$ GeV and $\mathcal{P}_{\text{FEI}} > 10^{-5}$, we

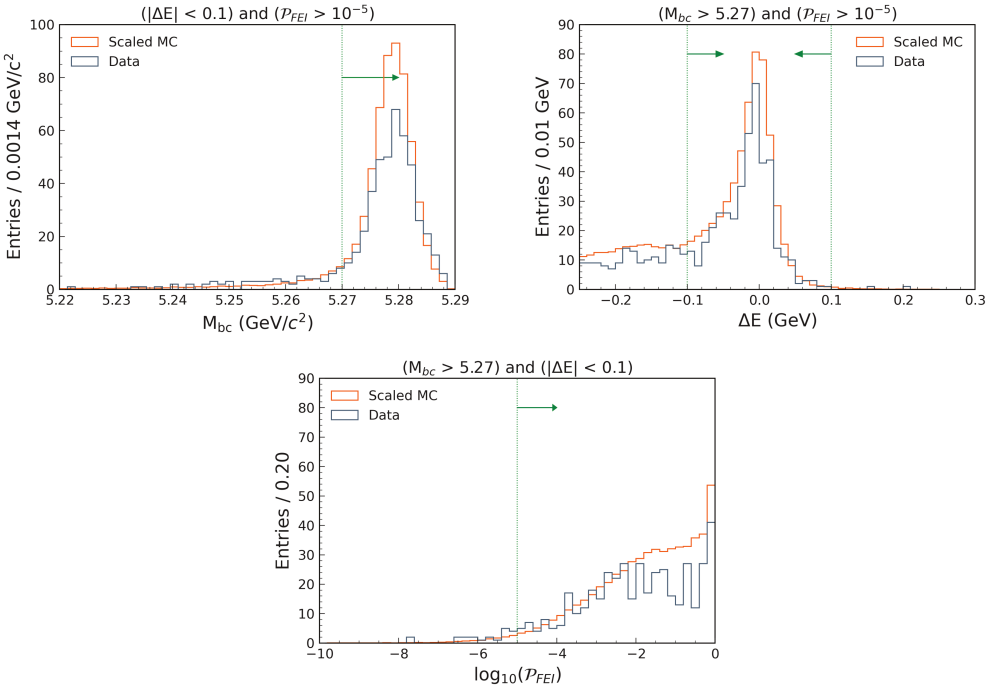


FIGURE 4.6: Pre-selection cuts on B_{tag} candidates. For each variable, the selection on the other two variables, shown with the green arrows, is applied.

consider the one with the highest `SignalProbability`. The applied cuts are shown in Fig. 4.6, where a clear difference in yields can be seen between data and MC.

After this selection, in the signal window, we retain 422 (55257) events in data (MC), distributed across the `FEI` modes as shown in Fig. 4.7 (upper plot). As mentioned earlier, most of the `FEI` efficiency (90%) comes from $D^{(*)}n\pi$[5] modes, which are therefore the target of our study. We also notice some large differences between MC and data, especially for the $B \to D^{(*)}\pi\pi\pi(\pi^0)$ modes. It is interesting to mention that for the modes from $b \to c\bar{c}s$ transitions, despite having very high BFs, the `FEI` efficiency is strongly affected by the exclusive reconstruction of two D mesons, as anticipated in Sec. 3.4. For example, the $\overline{D}^{(*)}D_s^+$ modes considered by `FEI` sum to a BF of $\sim$2.6% in our MC (Fig. 4.7, lower plot), 3.5 times smaller than the BF of $D^{*0}n\pi$ modes; however, when comparing the `FEI` efficiency of these two categories, the ratio goes to 14. The same effect is observed for the `FEI` $\overline{D}^{(*)}\overline{D}^{(*)}X$ modes which have a total BF of 3%.

The different efficiency between data and MC is measured and taken into account by calculating the calibration factors ε^i, defined as

$$N^i(\text{data})/N^i(\text{MC})$$

where i corresponds to the `FEI` mode ID. The standard way of measuring the yields consists of fitting the M_{bc} distribution with a PDF that combines the Crystal Ball and the Argus shapes (in Fig. 4.8 the M_{bc} distributions for $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$ candidates in data and MC are shown as an example). However, the procedure suffers from the fact that some of the modes have a large fraction of mis-reconstructed candidates, which

[4]For $e^{\pm}$ candidates, bremsstrahlung recovery is applied, see definition in Sec. 3.5.

[5]We combine $\overline{D}^{(*)0}$ and D^- modes.

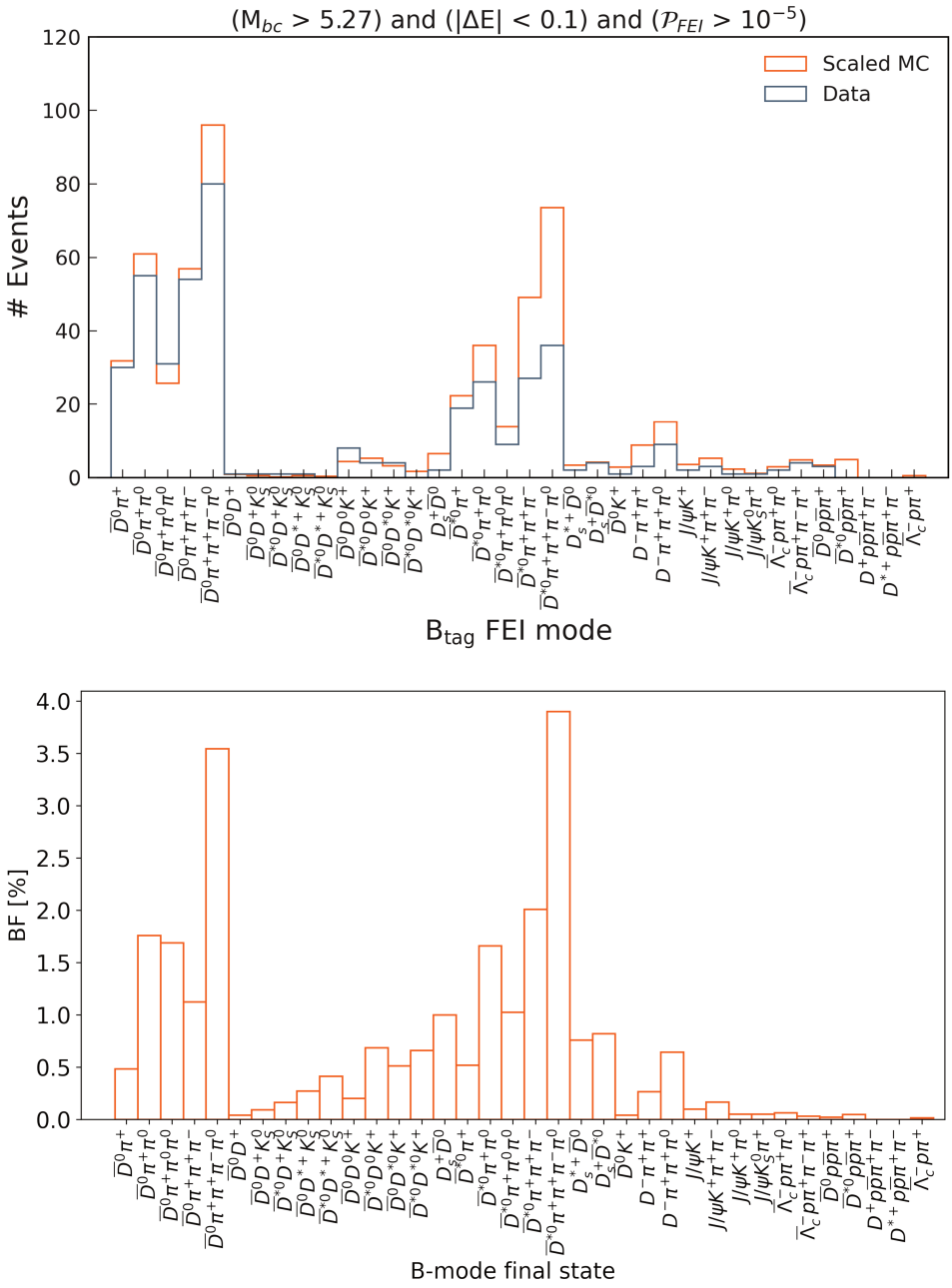


FIGURE 4.7: Top: the number of events per `FEI` mode in data and MC are compared. Bottom: the total BF contributing to each `FEI` mode.

still peak in $M_{\rm bc}$, making it difficult to disentangle them from the TM ones[6]. Furthermore, the fraction of mis-reconstructed candidates could differ between MC and data; hence, we cannot fix the PDF shape from MC and decide to consider as signal all the events corresponding to $M_{\rm bc} > 5.27$ GeV/c^2. The simple counting is justified by the 99%-purity of the $J/\psi K^+$ sample and the scarce sample size for collision data. As a result, we get an average ε[7] equal to 0.77 ± 0.04, which is compatible with the official

[6]We remind the reader that, as defined in Sec. 3.6, TM stands for truth-matched and indicates the candidates for which all of their constituent particles have been correctly identified; they are called false-matched (FM) otherwise.

[7]The average is over all the `FEI` modes.

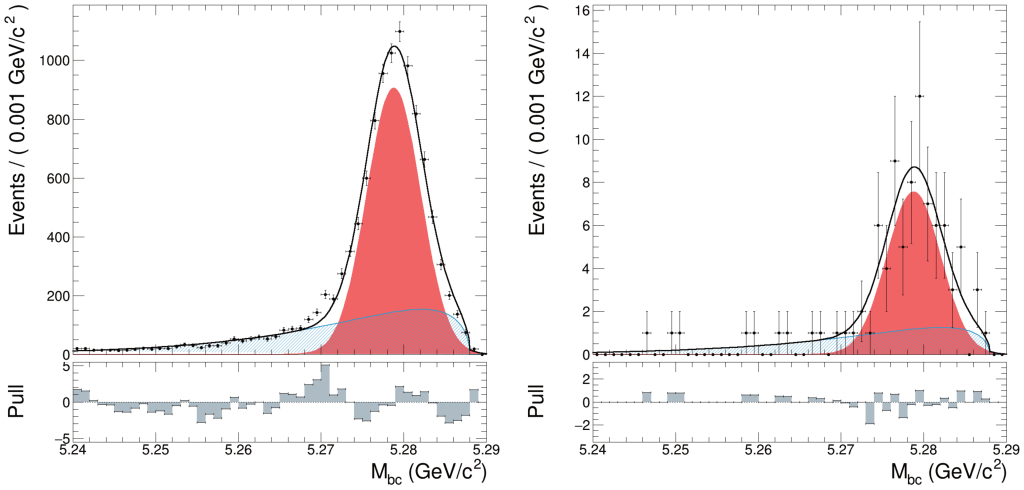


FIGURE 4.8: Example of fits to the $M_{\rm bc}$ distributions of FEI $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$ candidates for official MC (left) and data (right).

one 0.86 ± 0.05 [11] (discussed already in Ref. 3.11) within the uncertainties. The same strategy will be used when estimating the calibration factors for the updated MC, $\varepsilon_{\rm upd\,MC}$ obtained with the modifications to the modeling that are described later in the next section.

Fig. 4.9 shows the yields of tagged B belonging to the twelve modes. The official MC (in orange) is compared to data (black points) and to the official calibrations [11] (in blue), used for the data-MC correction in chapter 3.

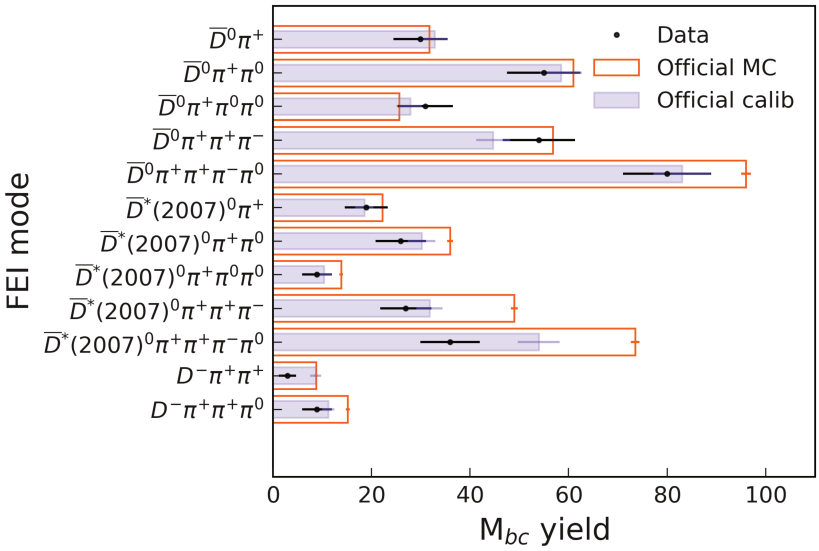


FIGURE 4.9: Yield per $B_{\rm tag}$ FEI mode obtained by counting the selected events with $M_{\rm bc} > 5.27\,{\rm GeV}/c^2$. Data and MC are compared and overlaid to the yields corresponding to the calibration factors from Ref. [11].

It is important to emphasize that the procedure just described is entirely new and paves the way for all the improvements shown in the following sections. Not only does the $B^+ \to J/\psi K^+$ sample make it possible to isolate the single B-chain with very high purity, but applying FEI on it reduces the number of decays to be studied drastically.

Our decay table contains about 300 hadronic exclusive modes[8] that can overwhelm whoever tries to get a hand on it. Instead, focusing only on the B-decays contributing to hadronic B-tagging not only makes the task relatively easier but also optimizes the effort on the main goal: improving B-tagging for our searches for channels with missing energy.

[8]Only for $b \to c$ transitions; to be compared with the 8 lines for the `PYTHIA` contributions and the $\mathcal{O}(30)$ lines for semileptonic decays.

4.3 Corrections to the Belle MC B^+ simulation

The two main concerns about the MC description of B^+ hadronic decays are the normalisation discrepancies (mode-dependent `FEI` yields) and the kinematic information resulting from the wrong simulation which is directly and indirectly used for the training of the `FEI` classifiers[9] giving as output the `SignalProbability`. While the first effect is usually corrected for in our analyses, we do not generally account for the second discrepancy.

The reasons above motivate an in-depth study of the main contributions, according to the official Belle simulation, to the `FEI` modes. A parallel study of the Belle II decay table has been performed, in order to track the main changes and the contributions of `PYTHIA` (which, over time, moved from v6 [107] to v8 [108]). The situation in Belle II was, surprisingly, not very much improved with respect to Belle: although some decay modes had been corrected, new errors and misinterpretations were introduced, resulting in an overall calibration factor that is further from 1: 0.65 ± 0.02 for a loose selection on the B_{tag} quality [109].

Despite the parallel work on both Belle and Belle II, we decide for clarity to only show the critical aspects found in Belle simulation and the proposed modifications. Only a few elements of the Belle II decay table will be highlighted when important.

The fundamental guidelines we follow are:

- The $Dn\pi$ final states reconstructed by `FEI` can be considered as the result of two separate currents ($\bar{u}d$ and bc) producing the X and Y systems (see Fig. 4.10). They include some 'narrow' states like the $D^0, D^{*0}, D_1, D_2^*, \omega, \eta$ and some broader resonances ($\rho^{(\prime)}, a_1^+, D_0^*, D_1^*$). Regardless of the intermediate states, the final products are of the form $Dn\pi n_0\pi^0$ with $D = \{D^0, D^{*0}, D^{**0}\}$, $n = \{1, 3\}$ and $n_0 = \{0, 1, 2\}$.

- When experimental results are available, update according to PDG and rely on a more critical interpretation of the PDG tables.

- If not, the few prescriptions provided by the theory (for example, what has been discussed in Sec. 1.3) can be used to adjust the inclusive productions. Further guidance is provided by the following principles:

$$D^0X : D^{*0}X : D^{**0}X \simeq 1 : 1 : 1\,, \tag{4.1}$$

$$Y\pi : Y\rho : Ya_1 \simeq 1 : 2.5 : 2.5\,. \tag{4.2}$$

  The first relation is based on the observation of $D^0\pi^- : D^{*0}\pi^- : D^{**0}\pi^-$ and $D^0\rho^- : D^{*0}\rho^-$ and confirmed by [14], while the second comes from the theoretical predictions and the parallel with $\tau^- \to h^-\nu_\tau$ decays [13].

- For D^{**} decays, use the Belle II model and $D_2X : D_1X \sim 0.4$ (Eq. 36 of [16]).

- A general principle is to prefer the resonant modes over the non-resonant (NR) ones. In this scheme modes like $D^{(*)}\rho\rho$ are considered as negligible as there is no known resonant production for them (and also there is no counterpart in the τ decays).

- No fine-tuning is performed at this stage for two reasons: the available data are not enough to draw conclusion and also because sometimes the measured BF

[9] The momenta of all the daughter particles as well as particle-ID and vertexing information for FSPs selection are exploited.

have very large uncertainties. In this regard, it would be important to update some measurements with Belle or Belle II data; we list them at the end of this chapter.

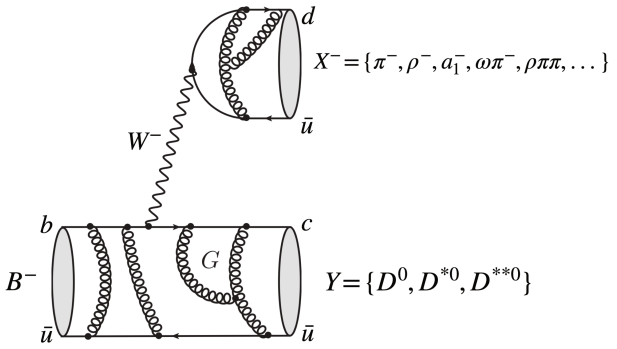


FIGURE 4.10: Factorization model where the products of the $b \to c$ transitions can be seen as the result of a D meson, called Y, and a system of n-pions (X). Adapted from Ref. [110].

The issues found on the current (*official*) Belle decay table and the proposed fixes are listed below, where each item corresponds to a given `FEI` mode, or 'final' state. To corroborate our findings, we compare the distributions obtained with the $J/\psi K$ sample to those corresponding to the tag side of the $B \to K\tau\ell$ samples (signal and background MC). This parallel allows us to discuss how the correction of some MC contributions might affect our sensitivity, for example, because of a less effective background rejection or a worse resolution of signal candidates. To better categorize the signal MC events, we use `topoAna` [111], a C++-based tool providing full access to the whole generated chain in MC from the generation-level information of the samples. Among the events with FM B_{tag}'s, we distinguish two sub-categories: the *self* Cross-feed (sCF), for which the B `FEI` mode is compatible with the true B_{tag} decay, and cross-feed (CF) for the rest. It is worth stressing that the current algorithm does not check the decay chains of the B daughters.

- $B^+ \to \overline{D}^{(*)0}\pi^+$
  These two modes are precisely measured and reliable, given the simpler two-body decay chain. The only needed corrections consist in updating the BFs to the latest PDG results [13]. As expected, `PYTHIA` does not contribute to the $B^+ \to \overline{D}^{(*)0}\pi^+$ final states. The contributions to the $\overline{D}^{(*)0}\pi^+$ `FEI` modes in the official MC simulation, as well as the proposed modifications are listed in Tab. 4.1.

B^+ `FEI` mode	Contribution	$\mathcal{B}^{\text{off}}(\%)$	$\mathcal{B}^{\text{upd}}(\%)$	References
$\overline{D}^0\pi^+$	$\overline{D}^0\pi^+$	0.48	0.47	$(0.468 \pm 0.013)\%$ [13]
		0.48	**0.47**	
$\overline{D}^{*0}\pi^+$	$\overline{D}^{*0}\pi^+$	0.52	0.49	$(0.490 \pm 0.017)\%$ [13]
		0.52	**0.49**	

TABLE 4.1: Update of BFs based on the studies summarized in chapter 4. The contributions to each of the $Dn\pi$ ($n = 1$) `FEI` modes are shown in the second column with the respective *official* (updated) rates in the third (fourth) column.

- $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0$
  The total $\mathcal{B}(B^+ \to \overline{D}^0\pi^+\pi^0)$ is, according to PDG [13], equal to 1.7% and it is known that its dominant contribution is the $\overline{D}^0\rho^+$ channel. With a data sample corresponding to 0.9 fb^{-1}, CLEO [112] measured the BF of $B^+ \to \overline{D}^0\rho^+$ to be $(1.34 \pm 0.18)\%$ and this result was correctly propagated to the Belle simulation. The rest of the $\overline{D}^0\pi^+\pi^0$ production is mostly due to $\overline{D}^{*0}\pi^+$ via $\overline{D}^{*0} \to \overline{D}^0\pi^0$ decay ($0.5\% \times 0.66 \sim 0.32\%$).
  With a larger dataset of 9 fb^{-1}, CLEO [113] updated in 2003 the measurement on $\mathcal{B}(B^+ \to \overline{D}^{*0}\rho^+)$, obtaining $(0.98 \pm 0.17)\%$. This latest result was not updated in Belle simulation and is therefore corrected now (see Tab. 4.2). For the same final state, there are modest contributions from the NR mode, and $B \to D^{**}\pi$ decays; we decide to set the first to zero and leave the rest as it is.

- $B^+ \to D^-\pi^+\pi^+(\pi^0)$
  Belle [99] and *BABAR* [100] have measured $\mathcal{B}(B^+ \to D^-\pi^+\pi^+) \sim 10^{-3}$, dominated by $B^+ \to \bar{D}_2^{*0}(\to D^-\pi^+)\pi^+$ (0.6×10^{-3}) and $B^+ \to \bar{D}_0^{*0}(\to D^-\pi^+)\pi^+$ (0.4×10^{-3}). Correspondingly, in Belle MC, the $D^-\pi^+\pi^+$ final state should be only attributed to the $D^{**}\pi$ component rather than the NR part. Except for the needed removal, no further correction is required until a clearer picture of the D^{**} contributions is obtained. Tab. 4.2 summarizes the status of official MC and the proposed modifications concerning the modes with two pions.

B^+ `FEI` mode	Contribution	$\mathcal{B}^{\text{off}}(\%)$	$\mathcal{B}^{\text{upd}}(\%)$	References
$\overline{D}^0\pi^+\pi^0$	$\overline{D}^0\rho^+$	1.34	1.34	$(1.34 \pm 0.18)\%$ [13]([112])
	$\overline{D}^{*0}\pi^+$	0.32	0.30	$(0.490 \pm 0.017)\%$ [13]$\times$0.67
	$\overline{D}^0\pi^+\pi^0$ (NR)	0.05	0.00	
	$\overline{D}_0^{*0}\pi^+$	0.03	0.03	
	$\overline{D}_2^{*0}\pi^+$	0.02	0.02	
		1.76	**1.69**	
$\overline{D}^{*0}\pi^+\pi^0$	$\overline{D}^{*0}\rho^+$	1.55	0.98	$(0.98 \pm 0.17)\%$ [113]
	$\overline{D}^{*0}\pi^+\pi^0$ (NR)	0.05	0.00	
	$\overline{D}_1^0\pi^+$	0.02	0.04	
	$\overline{D}_1'^0\pi^+$	0.02	0.02	
	$\overline{D}_2^{*0}\pi^+$	0.01	0.01	
		1.65	**1.05**	
$D^-\pi^+\pi^+$	$D^-\pi^+\pi^+$ (NR)	0.17	0.03	
	$\overline{D}_0^{*0}\pi^+$	0.06	0.06	6.8×10^{-4} [99], [100]
	$\overline{D}_2^{*0}\pi^+$	0.03	0.03	3.5×10^{-4} [99], [100]
		0.26	**0.12**	$(1.07 \pm 0.05) \times 10^{-3}$ [13]

TABLE 4.2: Update of BFs based on the studies summarized in chapter 4. The contributions to each of the $D^{(*)}n\pi$ ($n = 2$) `FEI` modes are shown in the second column with the respective *official* (updated) rates in the third (fourth) column.

Turning to the $D^-\pi^+\pi^+\pi^0$ final state (and hence to the modes with three pions, summarized in Tab. 4.3), the only input is given by the $\overline{D}^{**}\pi^+$ discussed above; instead, it is not clear how other components, like $D^-\rho^+\pi^+$ or $\overline{D}^{**}\rho^+$, were decided. The D^{**} candidates have been looked for through the invariant masses of $D^-\pi^+$ and $D^-\pi^+\pi^0$ combinations by choosing for each event the combination

giving the lowest mass (we expect the pion from $D^{**} \to D^*\pi$ transition to have low momentum) but the $J/\psi K$ method does not provide enough entries to draw any conclusions.

- $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$
  An early paper from CLEO [114] in 1992 using only 212 pb^{-1} data measures the four-body $\overline{D}^0\pi^+\pi^+\pi^-$ rate at $(1.15 \pm 0.29 \pm 0.21)\%$ and tries to disentangle the three sub-components $\overline{D}^0\pi^+\pi^+\pi^-$, $\overline{D}^0\rho^0\pi^+$ and the resonant $\overline{D}^0 a_1^+$, assigning them the fractions 48%, 34% and 18%, respectively The only other measurement is from an early dataset of LHC*b* [115], corresponding to 35 pb^{-1}, which provides the ratio of BFs $\mathcal{B}(B^+ \to \overline{D}^0\pi^-\pi^+\pi^-)/\mathcal{B}(B^+ \to \overline{D}^0\pi^+) = 1.27 \pm 0.06 \pm 0.11$. It contributes to the $\mathcal{B}(B^+ \to \overline{D}^0\pi^+\pi^+\pi^-) = (0.56 \pm 0.21)\%$ quoted in PDG [13] where the uncertainty is significantly scaled (by a factor 3.7 to cover the large difference between CLEO and LHC*b* measurements). The LHC*b* 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 to be from $\overline{D}^0_{1,2}\pi^+$ decays as described in Fig. 4.11 from LHC*b* data.

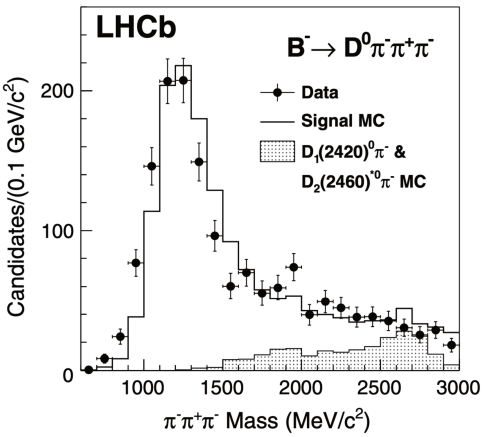


FIGURE 4.11: 3π-system invariant mass for $B^- \to D^0\pi^-\pi^-\pi^+$ candidates. Data points are overlaid with MC simulation (solid and shaded lines). Plot taken from Ref. [115].

Unfortunately, no quantitative information is provided, other than the total BF of $\overline{D}^0\pi^+\pi^+\pi^-$ (or more exactly its ratio to $D^0\pi^-$).

The Belle decay table, as shown in Tab. 4.3 is based on the old CLEO measurement – significantly overestimating the contributions of the two NR components – while including the more recent $\overline{D}^{**0}\pi^+$ measurements. Consequently, the total rate sums to 1.1% instead of the $\sim$0.6% indicated by LHC*b*. As mentioned above, there is an interesting contribution from $B^+ \to \overline{D}^{**0}(\to D^{*-}\pi^+)\pi^+$ decays which can be observed via the variable $\Delta M = M(\overline{D}^0\pi^-) - M(\overline{D}^0)$. This component is very precious because of the presence of a narrow resonance and would deserve to be selected separately, as opposed to the current `FEI` algorithm. A larger data sample is needed to measure the $\overline{D}^{**0}$ component and, if necessary, correct it in our MC.

We can verify these observations with our $J/\psi K$ sample where the other B is reconstructed by `FEI`. After vetoing the D^{*-} component, the $M(\pi^+\pi^+\pi^-)$ distribution (upper plot of Fig. 4.12) confirms the dominance of Da_1 in the data

while an exaggeratedly large contribution from the NR $D3\pi$ and $D\rho\pi$ components is present in the MC. In the right plot of Fig. 4.12 we show for comparison the background and $B \to K\tau\ell$ signal events for the $\overline{D}^0\pi^+\pi^+\pi^-$ modes. One can see that an a_1^+ selection in the $M(3\pi)$ would remove a significant fraction of background coming from continuum and $B^0\overline{B}^0$events and, more importantly, the combinatorial background from B^+B^- events.

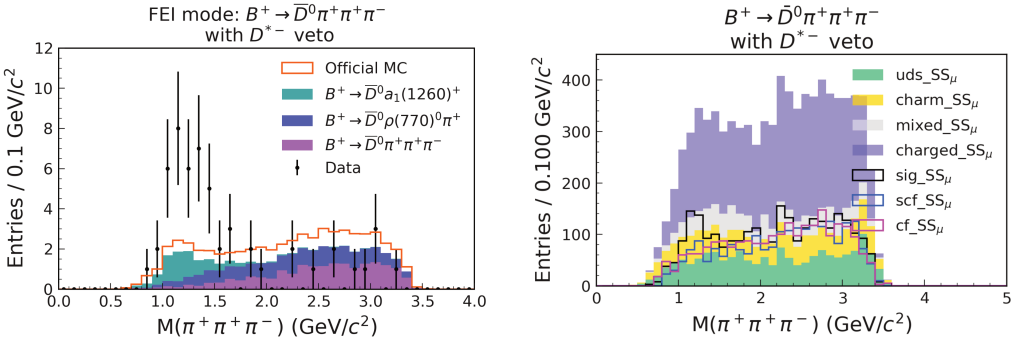


FIGURE 4.12: Left: Data-MC comparison of $M(\pi^+\pi^+\pi^-)$ with D^{*-} veto in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-$ `FEI` mode. Right: Background and signal distributions in the same variable for $B \to K\tau\ell$ studies.

The proposed version of Belle MC sets to zero the rates of the decays $\overline{D}^0\pi^+\pi^+\pi^-$ (PHSP), $\overline{D}^0\rho^0\pi^+$ and $D^{*-}\pi^+\pi^+$ with the purpose of preventing `PYTHIA` from generating them. The $\overline{D}^0 a_1^+$ BF is also corrected (see Tab. 4.3). We also notice that in Belle II a NR contribution from $B^+ \to D^{*-}\pi^+\pi^+$ (with a 0.09% BR) is present; this confirms that the corresponding decay table does incorporate the most up-to-date measurements, but sometimes introducing different mistakes which result in different calibration factors with respect to Belle.

- $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$
  In 1994, CLEO [112] measured $\mathcal{B}(B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-) = (0.94 \pm 0.20 \pm 0.17)\%$ with a sample of $0.9\,\text{fb}^{-1}$, noting that the three-pion mass is dominantly 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 NR (4-body) contribution. In 2004, Belle [61] 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 more than the total BF. We decide to set to zero the NR $\overline{D}^{*0}\pi^+\pi^+\pi^-$ contribution and also remove the NR `PYTHIA` contribution (see Table 4.3). It is worth mentioning that in Belle II decay table, the NR $\overline{D}^{*0}\pi^+\pi^+\pi^-$ contribution had been commented out with right intentions; however, as a result of not setting its BF to zero, `PYTHIA` generates it with a BF equal to 0.2%. In the context of the $K\tau\ell$ search, the $M(\pi^+\pi^+\pi^-)$ distribution of `FEI` candidates shows the large NR component extending up to $3\,\text{GeV}/c^2$ for both signal and background events (Fig. 4.13). One can also notice that this mode has less continuum background compared to the $\overline{D}^0$ counterpart because of the presence of the $\overline{D}^{*0}$, which reduces the combinatorial obtained when reconstructing a $\overline{D}^0$. We expect the suppression of the NR component to reduce the $q\overline{q}$ and $B\overline{B}$ background components even more.

- $B^+ \to \overline{D}^{(*)0}\pi^+\pi^0\pi^0$
  The $B^+ \to \overline{D}^0\pi^+\pi^0\pi^0$ final state is largely generated from the $B^+ \to \overline{D}^{*0}\rho^+$

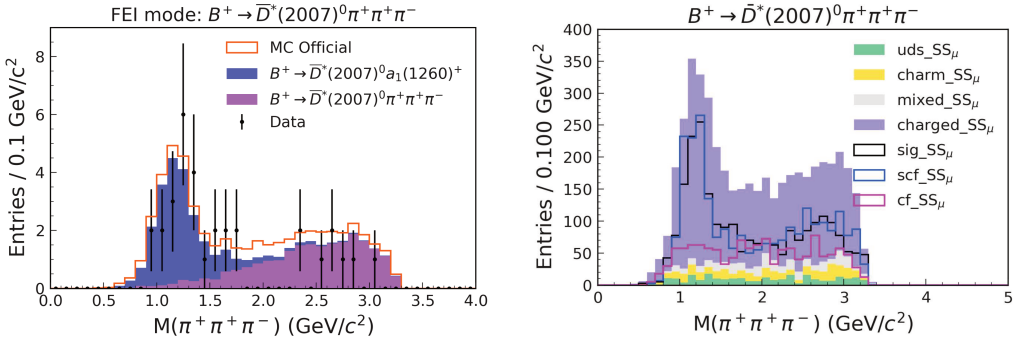


FIGURE 4.13: Left: Data-MC comparison of the 3π invariant mass in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-$ `FEI` mode. right: Background and signal distributions in the same variable for $B \to K\tau\ell$ studies.

decay, which was first measured by ARGUS [116] and later by CLEO [113], resulting in a PDG average of $\mathcal{B}(B^+ \to \overline{D}^{*0}\rho^+) = (0.98 \pm 0.17)\%$, as it has been discussed above. The second largest contribution comes from $B^+ \to \overline{D}^0 a_1^+$ due to a_1 modes into two neutral pions (around 42% of the total width). In addition to the above components, `PYTHIA` is significantly contributing with unexpectedly large BFs in $B^+ \to \overline{D}^0\rho^+\pi^0$ and $B^+ \to \overline{D}^0\pi^+\pi^0\pi^0$ decays, which we both set to zero. The major change proposed here corresponds to the correction of $\mathcal{B}(B^+ \to \overline{D}^0 a_1^+)$ discussed above.

Similarly to the previous mode, the main source of $\overline{D}^{*0}\pi^+\pi^0\pi^0$ is from $B \to \overline{D}^{*0}a_1^+$, measured by Belle [61] through the $a_1^+ \to \pi^+\pi^0\pi^0$ mode. As done so far, the BFs of $\overline{D}^{*0}\pi^+\pi^0\pi^0$ and $\overline{D}^{*0}\rho^+\pi^0$ are set to zero. Adjustments are made to make the $\overline{D}^{**0}\rho^+$ rates twice the $\overline{D}^{**0}\pi^+$ ones.

- $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$
  The $\overline{D}^{*0}n\pi$ $(n = 4)$ mode is the one with the largest effective BF in our MC – 3.9%, see Fig. 4.7. This final state has been studied by the CLEO collaboration with a sample of 9.0 fb^{-1}. The analysis described in Ref. [60] provides two measurements: the total BF, equal to $(1.80 \pm 0.24 \pm 0.27)\%$, and the sub-component $\mathcal{B}(B^+ \to \overline{D}^{*0}\omega\pi^+) = (0.45 \pm 0.10 \pm 0.07)\%$. As shown in Tab. 4.4, these two results were misinterpreted and the $\overline{D}^{*0}\omega\pi^+$ contribution was wrongly added on top of the total one. We therefore keep the right value for the $\overline{D}^{*0}\omega\pi^+$ while setting to zero the BF of the NR five-body mode, making the total BF closer to the measured value. The $\overline{D}^{*0}\omega\pi^+$ decays are searched for in the $M(\pi^+_{1,2}\pi^-\pi^0)$ spectrum represented in Fig. 4.17A[10]; in data, the ω peak is barely visible, probably because of the low statistics. In the same distribution, we can also check the $\overline{D}^{*0}\eta(\to \pi^+\pi^-\pi^0)\pi^+$ contribution; in parallel with the τ sector (Sec. 1.3), the disintegration of the W^{*+} into a $\eta\pi^+$ system is suppressed with respect to $\omega\pi^+$ because of G-parity conservation; therefore, we expect that the $\overline{D}^{*0}\eta\pi^+$ mode is reached only via D^{**} decays and consequently has a small BF. Instead, our MC generates it via `PYTHIA` with a quite large BR=0.6%[11], which is most likely overestimated, as it seems to be confirmed from data.

[10] Because of the low statics in data, we sum the two possible combinations $M(\pi_1^+\pi^-\pi^0)$ and $M(\pi_2^+\pi^-\pi^0)$.

[11] This BF has to be multiplied by $\mathcal{B}(\eta \to \pi^+\pi^-\pi^0) = 23\%$.

We also check the $M(4\pi)$ distribution, which, according to the CLEO measurement, corresponds to a ρ' component of a quasi-two-body $B^+ \to \overline{D}^{*0}\rho'^+$ decay[12], with $\rho'^+ \to \omega\pi^+$. Differently from data, where the distribution has a peak at lower values which could be compatible with the presence of the ρ', the MC seems to prefer a PHSP like distribution (as captured by Fig. 4.14, left plot). It is worth mentioning that, according to the `PYTHIA` model used for the fragmentation of $b \to c\bar{u}d$ decays (`modeID`=23), the products of the W^*, i.e. the $n\pi$ system, are generated with a PHSP model. This would result in a $M(4\pi)$ distribution that is closer to data, with respect to what one would obtain with a pure PHSP of all the decay products (D meson included). In the context of the $K\tau\ell$ search, the `FEI` training with the correct kinematic of the four-pion system should improve the background rejection (see Fig. 4.14, right plot).

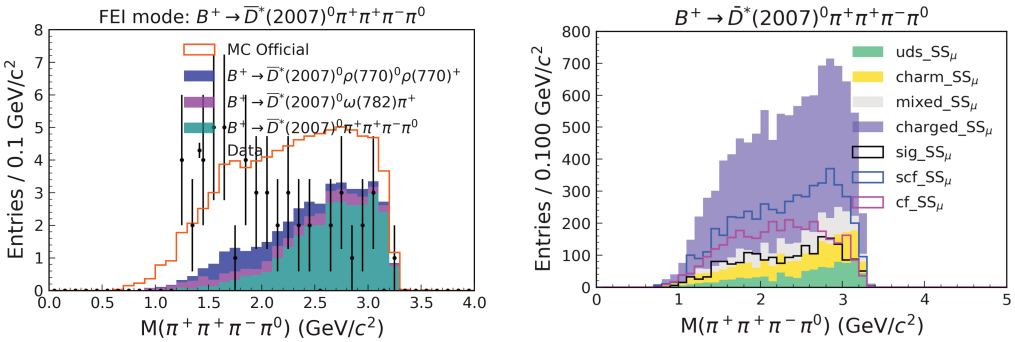


FIGURE 4.14: Left: Data-MC comparison of $\pi^+\pi^+\pi^-\pi^0$ invariant mass in $B^+ \to \overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$ `FEI` mode. Right: Background composition of the same in $K\tau\ell$ studies.

In the proposed MC model, in addition to fixing the mentioned double counting issue, we reduce the $\overline{D}^{*0}\eta\pi^+$ BF and removed the large `PYTHIA` $\overline{D}^{*0}\rho^0\rho^+$ component, in order to get a total BF consistent with the measurement.

- $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$

  This final state represents the second-highest contribution to the `FEI` with its BF equal to 3.5%. One of the possible contributions to this mode comes from the decay $B^+ \to D^{*-}\pi^0\pi^+\pi^+$ with a rate, according to the ARGUS measurement [116], equal to $(1.5 \pm 0.7)\%$ (to be multiplied by the BF $(D^{*-} \to \overline{D}^0\pi^-) = 0.67$). It has a very large uncertainty due to the small data sample available, which can be directly verified from the $M_{\rm bc}$ data points that are fitted to extract the signal, see Fig. 4.15. This measurement looks in contradiction with the left plot of Fig. 4.16, where the Belle MC D^{*-} yield is clearly overestimated in our MC. According to our model, the uncorrelated production of D^{*-} should be related to decays like $\overline{D}^{**0}(\to D^{*-}\pi^+)\rho^+$. We therefore set to zero the BF of the NR $B^+ \to D^{*-}\pi^+\pi^+\pi^0$ decay.

  A second experimental input comes from the same CLEO study encountered before [60] which provides $\mathcal{B}(B^+ \to \overline{D}^0\omega\pi^+) = (0.41 \pm 0.07 \pm 0.06)\%$. Accounting for the BF of $\omega \to \pi^+\pi^-\pi^0$ at 89.2%, this corresponds to the contribution listed in Tab. 4.4. Like for the previous mode, we compare the $M(\pi^+_{1,2}\pi^-\pi^0)$

[12] This state, sometimes identified as $\rho(1450)$, is the first radial excitation of the ρ. According to Ref. [60] the most likely spin-parity values are $J^P = 1^-$, consistently with those of the $\omega\pi$ system measured in τ decays (though the low m_τ mass distorts the mass spectrum significantly compared to B's).

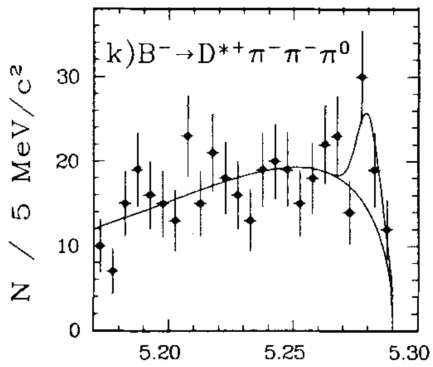


FIGURE 4.15: $M_{\rm bc}$ distribution of $B^- \to D^{*+}\pi^-\pi^-\pi^0$. Plot taken from the CLEO paper, Ref. [116].

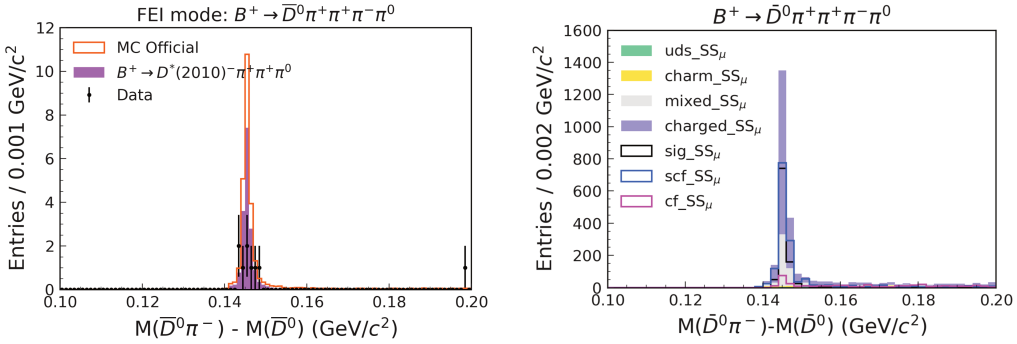
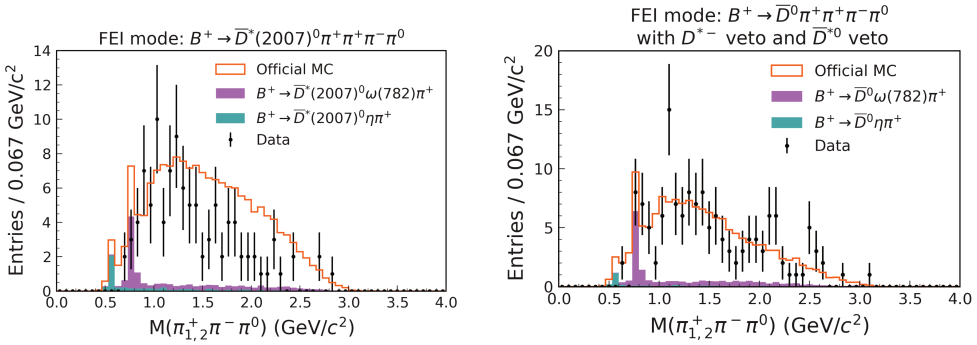


FIGURE 4.16: Left: Data-MC comparison of $\Delta M = M(\overline{D}^{*0}) - M(\overline{D}^0)$ in $B^+ \to \overline{D}^0\pi^+\pi^+\pi^-\pi^0$ `FEI` mode. Right: Background and signal distributions in the same variable for $B \to K\tau\ell$ studies.

distribution of data and MC, this time after vetoing the (small) $D^{*-}\pi^+\pi^0$ and $\overline{D}^{*0}\pi^+\pi^+\pi^-$ contributions to put in evidence the resonant structure of the three-pion neutral system. The ω peak looks consistent with data whereas the $\overline{D}^0\eta\pi^+$ mode looks again over-modeled in MC. However, `PYTHIA` generates it with a BF equal to 0.2% and we decide to keep it for now. It would be important to measure the $\overline{D}^{(*)0}\eta\pi^+$ modes at Belle II to have more reliable BFs or, at least, get ULs on their rates.

Table 4.4 summarizes all the MC contributions and the relevant updates to the $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$ final state. In the new proposed version, the main contributions are given by $\overline{D}^{*0}a_1^+$[13], $\overline{D}^0\omega\pi^+$ – both with unchanged BF – and then some $\overline{D}^0\rho\pi\pi$ decays generated by `PYTHIA`. We again set to zero the BF of the modes $\overline{D}^0\rho^0\rho^+$ (consistently with what has been done for $\overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$) and $B^+ \to D^{*-}\rho^+\pi^+$ (as it produces the same final state as $D^{*-}\pi^+\pi^+\pi^0$).

[13]The NR $D^{*0}\pi^+\pi^+\pi^+$ has been set to zero before.



(A) (B)

FIGURE 4.17: A (B): Data-MC comparison of $M(\pi^+\pi^-\pi^0)$ in $B^+ \to \overline{D}^{*0}(\overline{D}^0)\pi^+\pi^+\pi^-\pi^0$ `FEI` mode. The two possible combinations, due to the presence of two π^+ in the final state, are summed.

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{off}}(\%)$	$\mathcal{B}^{\text{upd}}(\%)$	References
$D^-\pi^+\pi^+\pi^0$	$D^-\pi^+\pi^+\pi^0$ (NR)	0.20	0.00	
	$D^-\rho^+\pi^+$	0.20	0.00	
	$\overline{D}^{**0}\rho^+$	0.09	0.18	
	$\overline{D}^{**0}\pi^+$	0.04	0.06	[99]
	$\overline{D}^{**0}\pi^+\pi^0$	0.11	0.00	
		0.64	**0.24**	
$\overline{D}^0\pi^+\pi^-\pi^+$	$\overline{D}^0\pi^+\pi^-\pi^+$ (NR)	0.46	0.00	
	$\overline{D}^0\rho^0\pi^+$	0.39	0.00	
	$\overline{D}^0 a_1^+$	0.18	0.58	$(1.24 \pm 0.13)\% \times 0.48$ [115]
	$\overline{D}_1^0\pi^+$	0.04	0.08	[99], [115]
	$\overline{D}_1'^0\pi^+$	0.03	0.03	[99], [115]
	$\overline{D}_2^{*0}\pi^+$	0.02	0.02	[99], [115]
	$\overline{D}^0\omega\pi^+$	0.01	0.01	$(0.41 \pm 0.09)\% \times 0.015$ [60]
		1.11	**0.71**	$(5.6 \pm 2.1) \times 10^{-3}$ [13]
$\overline{D}^{*0}\pi^+\pi^-\pi^+$	$\overline{D}^{*0}\pi^+\pi^-\pi^+$ (NR)	1.03	0.00	[112], [61]
	$\overline{D}^{*0} a_1^+$	0.91	0.91	$(1.9 \pm 0.5)\% \times 0.48$ [112]
	$\overline{D}^{*0}\omega\pi^+$	0.01	0.01	$(0.45 \pm 0.12)\% \times 0.015$ [60]
	$\overline{D}^{*0} f_0\pi^+$	0.07	0.00	
		2.01	**0.92**	$(1.03 \pm 0.12)\%$ [13]
$\overline{D}^0\pi^+\pi^0\pi^0$	$\overline{D}^{*0}\rho^+$	0.96	0.61	$(0.98 \pm 0.17)\% \times 0.67$ [116], [113]
	$\overline{D}^0 a_1^+$	0.15	0.50	$(1.24 \pm 0.13)\% \times 0.42$ [13, 115]
	$\overline{D}^{*0}\pi^+\pi^0$	0.03	0.00	
	$\overline{D}^0\rho^+\pi^0$	0.30	0.00	
	$\overline{D}^0\pi^+\pi^0\pi^0$ (NR)	0.10	0.00	
	$\overline{D}^{**0}\rho^+$	0.04	0.09	
	$\overline{D}^{**0}\pi^+$	0.02	0.02	[99]
	$\overline{D}^{**0}\pi^+\pi^0$	0.05	0.00	
		1.68	**1.26**	
$\overline{D}^{*0}\pi^+\pi^0\pi^0$	$\overline{D}^{*0} a_1^+$	0.79	0.79	$(1.9 \pm 0.5)\% \times 0.42$ [112]
	$\overline{D}^{*0}\rho^+\pi^0$	0.05	0.00	
	$\overline{D}^{*0}\pi^+\pi^0\pi^0$ (NR)	0.05	0.00	
	$\overline{D}^{**0}\rho^+$	0.05	0.14	
	$\overline{D}^{**0}\pi^+\pi^0$	0.04	0.00	
	$\overline{D}^{*0} f_0\pi^+$	0.03	0.00	
		1.02	**0.93**	

TABLE 4.3: Update of BFs based on the studies summarized in chapter 4. The contributions to each of the $D^{(*)}n\pi$ ($n = 3$) FEI modes are shown in the second column with the respective *official* (updated) rates in the third (fourth) column. Shaded rows correspond to contributions generated via PYTHIA in the official MC. Decays with $\overline{D}^{**0}$ denote the sum of the possible $\{D_1, D_0, D_1', D_2\}$ decays. Contributions to the effective BF smaller than 10^{-4} are omitted.

B^+ FEI mode	Contribution	$\mathcal{B}^{\text{off}}(\%)$	$\mathcal{B}^{\text{upd}}(\%)$	References
$\overline{D}^0\pi^+\pi^-\pi^+\pi^0$	$D^{*-}\pi^+\pi^+\pi^0$	1.02	0.00	[116]
	$\overline{D}^{*0}\pi^+\pi^-\pi^+$	0.64	0.00	
	$\overline{D}^{*0}a_1^+$	0.56	0.56	$(1.9 \pm 0.5)\% \times 0.48 \times 0.67$ [112]
	$\overline{D}^0\omega\pi$	0.37	0.37	$(0.41 \pm 0.09)\% \times 0.89$ [60]
	$D^{*-}\rho^+\pi^+$	0.14	0.00	
	$\overline{D}^{*0}\omega\pi$	0.00	0.00	
	$\overline{D}^0\rho^0\rho^+$	0.20	0.00	
	$\overline{D}^0\eta\pi^+$	0.05	0.05	
	$\overline{D}^0\omega\rho^+$	0.00	0.00	
	$\overline{D}^0\rho^+\pi^+\ \pi^-$	0.20	0.20	
	$\overline{D}^0\omega\pi^+\pi^0$	0.00	0.00	
	$\overline{D}^0\rho^-\pi^+\pi^+$	0.10	0.10	
	$\overline{D}^0\rho^0\pi^+\pi^0$	0.10	0.15	
	$\overline{D}_2^{*0}\rho^0\pi^+$	0.02	0.02	
	$\overline{D}_0^{*0}\omega\pi^+$	0.00	0.00	
	$\overline{D}_0^{*0}\rho^0\pi^+$	0.03	0.03	
	$\overline{D}_0^{\prime 0}\pi^+\pi^0$	0.05	0.00	
	$\overline{D}_2^{*0}\omega\pi^+$	0.00	0.00	
	$\overline{D}_2^{*0}\pi^+\pi^0$	0.02	0.00	
	$\overline{D}_2^{*0}f_0\pi^+$	0.04	0.00	
		3.53	**1.49**	
$\overline{D}^{*0}\pi^+\pi^-\pi^+\pi^0$	$\overline{D}^{*0}\pi^+\pi^-\pi^+\pi^0$ (NR)	1.80	0.00	
	$\overline{D}^{*0}\omega\pi$	0.41	0.41	$(0.45 \pm 0.12)\% \times 0.89$ [60]
	$\overline{D}^{*0}\eta\pi^+$	0.14	0.01	
	$\overline{D}^{*0}\rho^0\rho^+$	0.49	0.00	
	$\overline{D}^{*0}\omega\rho^+$	0.01	0.01	
	$\overline{D}^{*0}\rho^0\pi^+\pi^0$	0.40	0.44	
	$\overline{D}^{*0}\rho^+\pi^-\pi^-$	0.40	0.56	
	$\overline{D}^{*0}\omega\pi^-\pi^0$	0.00	0.01	
	$\overline{D}^{*0}\rho^-\pi^+\pi^+$	0.20	0.26	
	$\overline{D}_2^{*0}\rho^0\pi^+$	0.01	0.01	
	$\overline{D}_2^{*0}\omega\pi^+$	0.00	0.00	
	$\overline{D}_1^{\prime 0}\omega\pi^+$	0.00	0.00	
	$\overline{D}_1^{\prime 0}\rho^0\pi^+$	0.03	0.03	
		3.89	**1.73**	$(1.8 \pm 0.4)\%$ [60]

TABLE 4.4: Update of BFs based on the studies summarized in chapter 4. The contributions to each of the $D^{(*)}n\pi$ ($n = 4$) FEI modes are shown in the second column with the respective *official* (updated) rates in the third (fourth) column. Shaded rows correspond to contributions generated via PYTHIA in the official MC. Contributions to the effective BF smaller than 10^{-4} are omitted.

To summarize, we have removed the double-counted components, which mainly consist of NR decays, and explicitly set to zero in the decay table some of the modes obtained with PYTHIA to ensure they will no longer be generated. As a result, the total exclusive BF has shrunk from 13.8% to 8.5%, and the remaining fraction has been put back to the PYTHIA line responsible for the $Dn\pi$ decays, which goes from 31% (as previously shown in Fig. 4.4) to $\sim$36%. This increase is not ideal because we do not control how the additional 5% BF will be spread over some modes. Nonetheless, we can forbid PYTHIA to generate those decays that we consider prohibited or highly suppressed in our model. The changes in the BF contributing to each FEI mode is displayed in Fig. 4.18. A better agreement with data is observed, as shown in Fig. 4.19 and summarized in Table 4.5, where the calibration factors obtained with the $\overline{D}^0(\to K^+\pi^-)\ell^+\nu_\ell$ sample are compared to the $J/\psi K$ sample, before ($\varepsilon_{\text{off MC}}$) and after ($\varepsilon_{\text{upd MC}}$) our corrections. The overall calibration factor goes from 0.77 ± 0.5 to 0.99 ± 0.5 (statistical uncertainty only. As we have shown throughout this chapter, the low average $\bar{\varepsilon}$ is primarily due to the highly wrong description of $\overline{D}^{(*)0}\pi^+\pi^+\pi^-$ and $\overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$ modes and their large weights on the total FEI efficiency. By only fixing those components, the agreement with data significantly improves. We should stress, though, that even in our updated model, some FEI modes still show some deviations from 1; one possible reason is that the central values taken from PDG often have very large uncertainties, which our MC does not take into account for. For example, the $\overline{D}^{*0}\rho^+$ mode, affecting not only $\overline{D}^{*0}\pi^+\pi^0$ but also $\overline{D}^0\pi^+\pi^0\pi^0$, has a 17% uncertainty on its measured BF; similarly, the BFs of $\overline{D}^{(*)0}\omega\pi^+$ modes have an uncertainty of 22(7)%. This effect should be encoded in the systematic uncertainty, the estimation of which is beyond the scope of this work. Other systematic uncertainties are expected to be small, given the high purity of the $B^+ \to J/\psi K^+$ sample and the fact that $\varepsilon_{\text{off MC}}$ is consistent with $\bar{\varepsilon}$ obtained with the $\overline{D}^0(K^+\pi^-)\ell^+\nu_\ell$ sample.

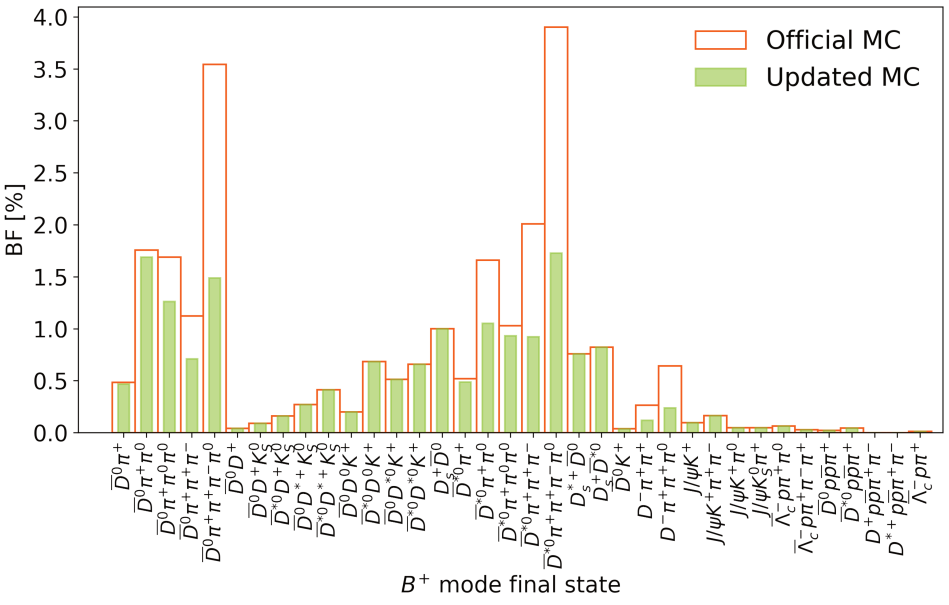


FIGURE 4.18: Total BF contributing to each FEI mode: the official and updated MC samples are compared.

Another important aspect is the agreement in the kinematic shapes; this is a crucial element at the stage of the FEI training. Improvements are clearly obtained, as shown with the four plots in Fig. 4.20. Concerning the $\overline{D}^0\pi^+\pi^+\pi^-$ mode (Fig. 4.20A),

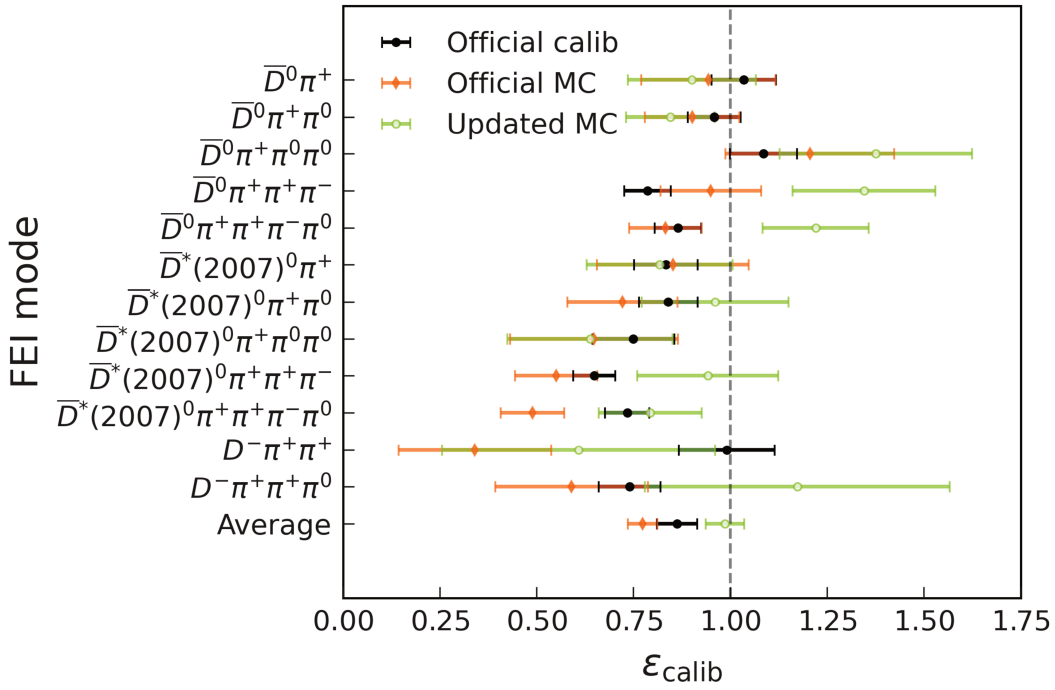


FIGURE 4.19: Calibration factors $\varepsilon_{\text{calib}} = N(\text{data})/N(\text{data})$. 'Official calibration' (black points) refer to the values obtained in the study [11] with the $B^+ \to \overline{D}^0(K^-\pi^+)\ell^+\nu_\ell$ sample and Belle official MC. The 'Official MC' (orange) points are obtained with the $J/\psi K^+$ sample while the 'Updated MC' points (green) are obtained with the corrections described in Sec 4.3.

the NR 4-body and the $\overline{D}^0\rho^0\pi^+$ have been removed whereas the $\overline{D}^0 a_1^+$ BF has been enhanced based on the LHC*b* measurement [115]. It is worth mentioning that due to the BF change from 1.11% to 0.71%, the calibration factor for the this mode looks worse than before; in this case, the good agreement with data obtained with the official MC (as shown in Fig. 4.9) can be considered as accidental and due to the large NR contributions. However, the predominance of a_1^+ is confirmed by measurements and from the $M(\pi^+\pi^+\pi^-)$ distribution in data. Moving to the $\overline{D}^{*0}\pi^+\pi^+\pi^-$ mode (Fig. 4.20B), the only needed correction is the removal of the NR component, making not only the total BF closer to the measured one (see Tab. 4.3), but also the $M(3\pi)$ shape closer to data. For the $\overline{D}^0\pi^+\pi^+\pi^-\pi^0$ mode (Fig. 4.20C), the main change consists in setting to zero the NR 4-body $D^{*-}\pi^+\pi^+\pi^0$, clearly overestimated by ARGUS, as well as the exclusive $D^{*-}\rho^+\pi^+$ contribution; in the updated MC version, the D^{*-} is produced only via the D^{**} modes. In conclusion the $\overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$ (Fig. 4.20D), for which we show again the pion-system invariant mass distribution: the updated MC (obtained removing components like the NR 5-body decay and the $\overline{D}^{*0}\rho^0\rho^+$) is in better agreement with the data. Furthermore, the total normalization is now consistent with the measurement.

Despite of its undeniable power, the $J/\psi K^+$ method suffers from low statistical power, and a better channel is needed to validate the proposed modifications and further tune the BF for the channels for which no experimental nor theoretical input is available. The chosen new sample is briefly described in the next paragraph.

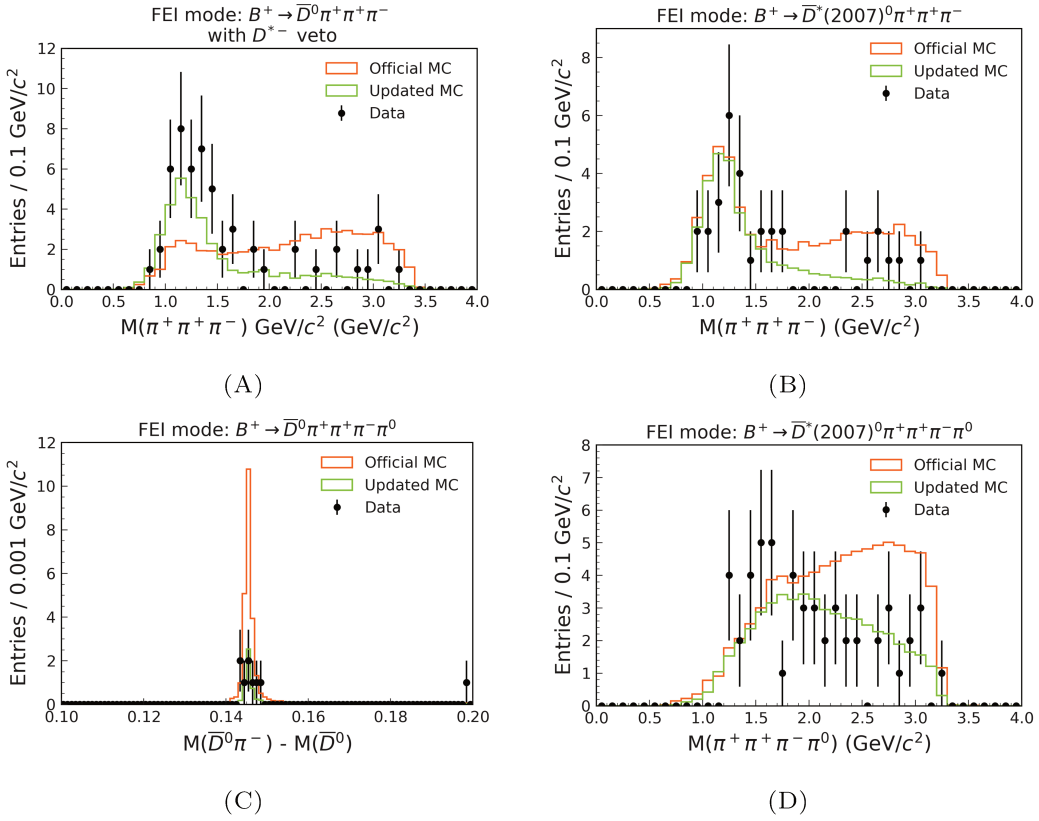


FIGURE 4.20: Invariant masses distributions capturing some of the significant modifications applied to the B^+ Belle MC for the $\overline{D}^{(*)0}\pi^+\pi^+\pi^-$ modes on top and $\overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$ modes at the bottom. Data points are overlaid with MC (official and updated versions).

4.3.1 An improvement to the $J/\psi K$ method: $D^{(*)}\pi$ and M_{recoil}

The $J/\psi K^+$ method allows to achieve very high purity but suffers from a low BF, $\mathcal{B}(B^+ \to J/\psi(\to \ell\ell)K^+) \sim 10^{-4}$; as a result, most of the comparisons between data and MC made in terms of yields and kinematic spectra need confirmation. A novel approach is necessary for comparable purity and significantly higher yield. We propose using the inclusive[14] $D^{(*)}\pi$ reconstruction, already encountered in Sec. 4.3.1 and the fact that, if one reconstructs one B via hadronic tagging and associates a pion to it, the recoil mass shows clear peaks from $B^+ \to \overline{D}^0\pi^+$,$\overline{D}^{*0}\pi^+$ and $\overline{D}^{**0}\pi^+$ events. Such technique has been exploited by *BABAR* to perform the measurement of the inclusive BFs of $\overline{D}^{(*,**)0}\pi^+$ decays [117] (where D^{**} stands for the ensemble of many resonant states), where after background subtraction, the yields are obtained by fitting the excess in data, and the three BFs are inferred. The same technique has been exploited for the $B^0 \to \tau\ell$ search at Belle [118], where the $D^{(*)-}\pi^+$ modes serve as control channels to assess the signal shape parameters and for the data-MC efficiency corrections related to the hadronic B_{tag} reconstruction with the `FR` algorithm.
This method allows to probe the hadronic B-decays relevant for `FEI` and measure with higher precision the correction factors for each mode (data-MC normalisation). In fact, the available sample for $D^{(*)}\pi^+$ events is around $32k$ ($16k$ for each mode), to be compared with the $\sim$400 events obtained with $J/\psi K^+$. As an example, the three-pion invariant mass distribution obtained for the $\overline{D}^{*0}\pi^+\pi^+\pi^-$ candidates is presented

[14] Inclusive because the D mesons are not reconstructed.

FEI Mode	$\bar{\varepsilon}$	$\varepsilon_{\text{off MC}}$	$\epsilon_{\text{upd MC}}$
$\overline{D}^0\pi^+$	1.03 ± 0.08	0.94 ± 0.17	0.90 ± 0.17
$\overline{D}^0\pi^+\pi^0$	0.96 ± 0.07	0.90 ± 0.12	0.85 ± 0.11
$\overline{D}^0\pi^+\pi^0\pi^0$	1.09 ± 0.09	1.21 ± 0.22	1.38 ± 0.25
$\overline{D}^0\pi^+\pi^+\pi^-$	0.79 ± 0.06	0.95 ± 0.13	1.35 ± 0.18
$\overline{D}^0\pi^+\pi^+\pi^-\pi^0$	0.87 ± 0.06	0.83 ± 0.09	1.22 ± 0.14
$\overline{D}^{*0}\pi^+$	0.83 ± 0.08	0.85 ± 0.20	0.82 ± 0.19
$\overline{D}^{*0}\pi^+\pi^0$	0.84 ± 0.08	0.72 ± 0.14	0.96 ± 0.19
$\overline{D}^{*0}\pi^+\pi^0\pi^0$	0.75 ± 0.11	0.65 ± 0.22	0.64 ± 0.21
$\overline{D}^{*0}\pi^+\pi^+\pi^-$	0.65 ± 0.05	0.55 ± 0.11	0.94 ± 0.18
$\overline{D}^{*0}\pi^+\pi^+\pi^-\pi^0$	0.73 ± 0.06	0.49 ± 0.08	0.79 ± 0.13
$D^-\pi^+\pi^+$	0.99 ± 0.12	0.34 ± 0.20	0.61 ± 0.35
$D^-\pi^+\pi^+\pi^0$	0.74 ± 0.08	0.59 ± 0.20	1.17 ± 0.39
Average	0.86 ± 0.05	0.77 ± 0.04	0.99 ± 0.05

TABLE 4.5: Calibration factors per mode in Belle MC. The *official* factors from Ref. [11] (statistical and systematic uncertainties are summed here in quadrature) are compared to those obtained with the *official* and *updated* MC, for which uncertainties are statistical only.

in Fig. 4.21 for the $\overline{D}^0\pi^+$ events[15] and shows the dramatic improvement of the new MC with respect to the official one, confirming what seen in Fig. 4.20 but with much higher precision.

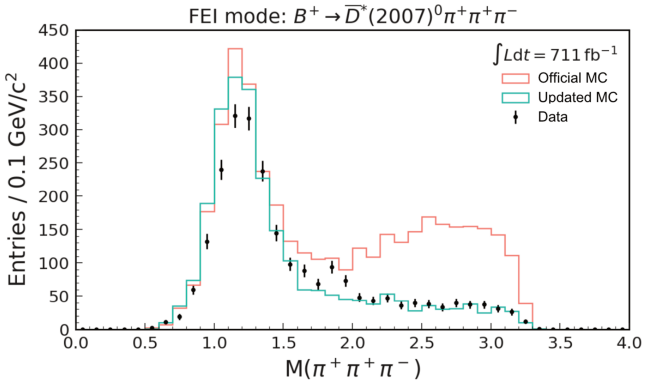


FIGURE 4.21: $M(\pi^+\pi^+\pi^-)$ distribution for $\overline{D}^{*0}\pi^+\pi^+\pi^-$ FEI candidates. The official MC is compared to the updated one and data. The plot is obtained with the $\overline{D}^{(*)}\pi^+$ method and taken from Ref. [104].

The same checks are also being performed on Belle II MC samples and will provide valuable information for many ongoing *tagged* analyses. Currently, the available dataset at Belle II collected at the $\Upsilon(4S)$ resonance is around $363\,\text{fb}^{-1}$ – half of Belle, but enough to perform the first round of checks and correct the obvious mistakes. Another advantage of the $D^{(*)}\pi$ sample is related to systematic uncertainties associated with the calibration procedure. Until now, it has been performed using high-BF processes on the signal side like inclusive $X_c\ell^+\nu_\ell$ [109] or $\overline{D}^0\ell^+\nu_\ell$ [119] and extracting the yields with binned fit to the lepton spectrum p^*_ℓ or the squared missing

[15] Corresponding to the M_{recoil} region $(1.80, 1.92)\,\text{GeV}/c^2$.

mass $M^2_{\text{miss}} = m^2_\nu = \left(p_{B_{\text{sig}}} - p_\ell - p_{D^0}\right)^2$, respectively. Due to the large semileptonic BFs, these signal samples provide copious data, and the systematic uncertainty of the measurement, related to the low purity obtained and the consequent difficulty in parametrizing the background[16], becomes predominant (it is around 5%, see Ref. [119]). The $\overline{D}^{(*)0}\pi^+$ approach could pave the way to measuring the calibration factors with hadronic B-modes, despite the smaller sample size. In fact, as these modes have a smooth background shape that is easier to model and a larger signal-to-noise ratio, they should lead to better precision.

4.4 Additional improvements

While in the previous section we focused on the dependence of `FEI` from MC modeling, we want to discuss here some easier modifications to the `FEI` package. Although we expect that the greatest improvement will be achieved by re-training, it is worth elaborating more on some aspects already mentioned before that could lead to easy gains in purity and efficiency. For example, exploiting the knowledge of the intermediate states and applying dedicated cuts, adding new B-modes and also recover 'broken' B_{tag}'s which are localized in kinematics.

Intermediate states

A notable example of an intermediate state which is not explicitly reconstructed by `FEI` is D^{*-}, expected to be produced mainly in D^{**} decays. Including the $\Delta M = M(D\pi_s) - M(D)$ criterion for existing `FEI` modes would improve purity, considerably restricting the amount of combinatorial background. The effectiveness is enhanced by the excellent experimental resolution in ΔM ($\sim 1\,\text{MeV}/c^2$) arising from the cancellation of experimental uncertainty in the determination of the momenta of particles comprising the D. The usage of ΔM is relatively less powerful when D^* decays into a π^0, because of the larger combinatorial background which reduces the purity. This can be seen in Fig. 4.22, where $\overline{D}^{*0}$ and D^{*-} events are shown in the ΔM dimension from the backgrounds of the $K\tau\ell$ analysis. The $\overline{D}^{*0}$ candidates for $\overline{D}^{*0}\pi^+\pi^0$ show a worse resolution with respect to the D^{*-} ones, obtained combining the $\overline{D}^0$ and the π^- in $\overline{D}^0\pi^+\pi^+\pi^-$`FEI` candidates. This effect can be explained if we consider that ΔM is used in `FEI` selection and therefore wrong $\overline{D}^{*0}$ candidates with ΔM closer to $\sim 142\,\text{MeV}/c^2$ are preferred.

Additional modes for `FEI`

As expressed in Eq. 3.15, the sensitivity reach on the $K\tau\ell$ modes depends on the final reconstruction efficiency, which is below the percent-level for hadronic `FEI`. A way to obtain a higher efficiency consists of adding new exclusive modes with sizeable BFs and acceptable purities to the `FEI` package. Looking for new exclusive modes is complicated: as extensively demonstrated, the knowledge of hadronic B^+-decays is not yet satisfactory. However, the studies needed to update the decay table made us encounter a few possible additions that we believe are worth exploring, for example the two-body $\overline{D}^{**}X$ decays. The model described in Sec. 4.3 can provide us a direction to find new exclusive modes for `FEI` with $\overline{D}^{**0}$. Experimental results seem to confirm the relation in Eq. 4.1 for the π^+ and from the Eq. 4.2 we expect the modes with ρ^+ and a_1^+ to be enhanced with respect to the π^+ by a factor 2.5. The total $\overline{D}^{**0}a_1^+$

[16] Especially the $B \to D^{**}\ell\nu$ and $B \to D^{(*)}n\pi\ell\nu$ components, which are not well known.

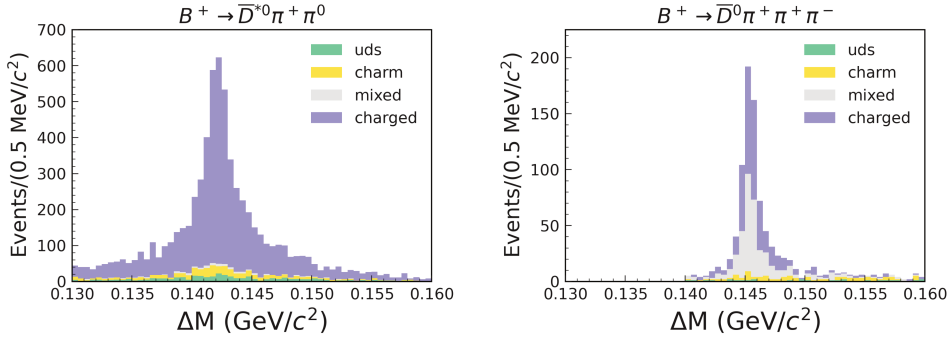


FIGURE 4.22: $\overline{D}^{*0}$ (D^{*-}) candidates represented in the ΔM dimension and obtained in the $K\tau\ell$ backgrounds from $\overline{D}^{*0}\pi^+\pi^0$ ($\overline{D}^0\pi^+\pi^+\pi^-$) FEI candidates. The $\overline{D}^{*0}$ is reconstructed by FEI, while the D^{*-} is searched for by combining the $\overline{D}^0$ candidates with the π^-.

could thus be as large as 1% and, focusing on the $\overline{D}^{**0}$ resonances decaying into D^{*-} (from Tab. 1.3) one can estimate the $D^{*-}(4\pi)^{++}$ production rate associated to D^{**}. Assuming 48% for the BF of a_1^+ into three charged pions, one gets a rate of roughly 0.1%. Actually the BF of the $B^+ \to D^{*-}(4\pi)^{++}$ final state has been measured by the Belle collaboration with a sample of 140 fb^{-1} [61]. A significant signal yield of 446 ± 32 events is found, corresponding to $\mathcal{B}(B^+ \to D^{*-}(4\pi)^{++}) = (2.56 \pm 0.26 \pm 0.33) \times 10^{-3}$. This BF is not irrelevant, considering that it corresponds to roughly one-fourth of the yield of $\overline{D}^{*0}\pi^+\pi^+\pi^-$ (1728 ± 53 events) – one of the highest-ranked FEI modes and responsible for 6% of the total FEI efficiency in data. One can therefore expect a 'clean' 1-2% of efficiency gain by adding the $B^+ \to D^{*-}(4\pi)^{++}$ to the FEI, considering the presence of a D^{*-} and absence of π^0's which, as we have seen before, can bring a larger combinatorial. The good signal-to-noise ratio of this mode is demonstrated by the plot from Ref. [61] (left column of Fig. 4.23). The $D^{*-}(4\pi)^{++}$ mode is not present in the Belle decay table (it is generated only through small $D^{**}\rho(\omega)\pi$ contributions via PYTHIA) while Belle II uses the correct measured BF from Ref. [61]. However, PHSP is used in the generation, while in the corresponding Belle internal note [120], the search of quasi-two-body decays confirms that roughly half of the signal is from $\overline{D}^{*0}_{1,2}a_1^+$ decays.

In the same Belle study [61] a significant signal yield for $B^+ \to \overline{D}^{*0}(5\pi)^+$ was observed: 341 ± 41 events corresponding to $\mathcal{B}(B^+ \to \overline{D}^{*0}(5\pi)^+) = (5.67 \pm 0.91 \pm 0.85) \times 10^{-3}$ (Fig. 4.23, right plot). Possible contributions to this final state come from $\bar{b} \to \bar{c}c\bar{s}$ transitions like $\overline{D}^{*0}D_s^+$ via the five-body D_s^+ decay[17] – currently not covered by FEI. Other than that, no other resonant production is known. The $B^+ \to \overline{D}^{*0}(5\pi)^+$ might not be a good candidate for FEI, because of the low BF, the lower reconstruction efficiency due to the high multiplicity and the worse purity (see right plot of Fig. 4.23).

Other modes that can provide interesting signatures for B-tagging belong to the family $B^+ \to \overline{D}^{(*)0}K^+\overline{K}^{(*)0}$. These states can be produced either via quasi two-body decays, where a light unflavored meson (like a_1 or ρ') decays in the couple of light strange mesons, or with a NR three-body decay, as a result of the production of an $s\bar{s}$ pair from the $W^+ \to u\bar{d}$ current (as illustrated in Fig. 4.24).

For this family of modes, the only available experimental result comes from Belle [121], which found significant signal yields for the $B^+ \to \overline{D}^{(*)0}K^+\overline{K}^{*0}$ and $\overline{D}^0K^+\overline{K}^0$ and set a UL to the BF of $\overline{D}^{*0}K^+\overline{K}^0$. The sum of the measured BFs amounts to the

[17] The BF being (0.79 ± 0.08)%.

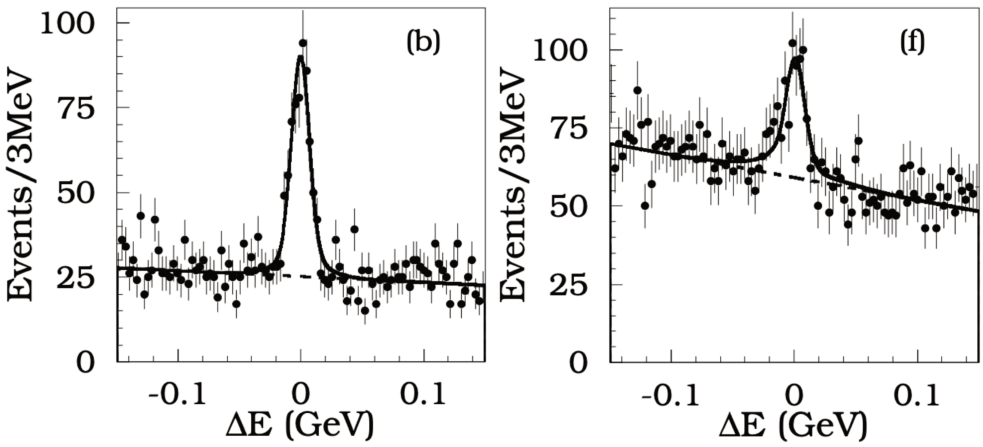


FIGURE 4.23: ΔE distributions from Belle [61] for $D^{*-}(4\pi)^{++}$ (left) and $\bar{D}^{*0}(5\pi)^{++}$ (right) obtained with 140 fb^{-1}.

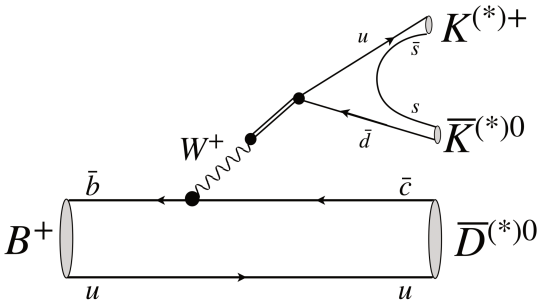


FIGURE 4.24: Diagram for DKK production. The double line indicate possible intermediate resonances producing the KK-system with the allowed isospin 1.

0.32%[18], roughly 0.7 times the $\overline{D}^0\pi^+$ BF, which brings 7% efficiency to `FEI`. Under the approximation of the same reconstruction efficiency for $\overline{D}^0\pi^+$ and $\overline{D}^0K^+$ and using 50% efficiency for the $K^{*0}/K^0_S \to \pi\pi$ reconstruction, the four DKK modes could bring an overall efficiency to `FEI` equal to $\sim 2\%$ (similarly to the $D^-\pi^+\pi^+\pi^0$ mode). Like for the $D^{*-}(4\pi)^{++}$, these additional modes would have the advantage of being pure, in this case, because of the presence of a kaon and a narrow kaon resonance. It is essential to clarify the properties of DKK with the full Belle dataset and Belle II to have a proper MC simulation and to exploit them for B-tagging.

Shifted ΔE bin

Another possible direction to obtain a higher tag-side efficiency is to recover partially reconstructed B_{tag}'s. As we have learned from chapter 3, wrong B-candidates still often provide, for $K\tau\ell$ events, a recoil mass within the signal region around m_τ (see for example Fig. 3.12). We want to verify if $B^+ \to \overline{D}^{*0}n\pi$ events reconstructed as $B^+ \to \overline{D}^0 n\pi$ (because of the missed reconstruction of a photon, or a soft π^0, from $\overline{D}^{*0} \to \overline{D}^0\gamma/\pi^0$ decays) can be used to increase the signal efficiency. With the current ΔE selection $|\Delta E| < 100$ MeV, this is not possible because these events would lie at ~ -200 MeV below our threshold. Even enlarging the ΔE window, we would not see these candidates belonging to the *shifted* ΔE bin because of the correlation between `SignalProbability` and ΔE. In fact, as anticipated in Sec. 3.6, the BDT output

[18] Only the K^0_S component is considered

$\mathcal{P}_{\text{FEI}}$ depends on the ΔE (Fig. 4.25) because this variable is part of the current set of variables used to train the `FEI` classifiers.

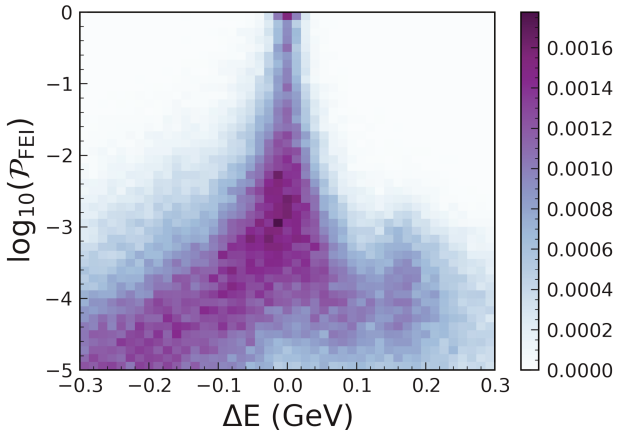


FIGURE 4.25: $\mathcal{P}_{\text{FEI}}$ (in logarithmic scale) of B_{tag} candidates as a function of the ΔE for $B \to K\tau\ell$ events.

To evaluate this effect without performing a new training, we replace $\mathcal{P}_{\text{FEI}}$ by $\mathcal{P}_{\text{FEI}}^{\text{raw}}$ (defined in Sec. 3.4) when ranking the signal $K\tau\ell$ candidates. This approach is justified by the fact that $\mathcal{P}_{\text{FEI}}^{\text{raw}}$ is used as a training variable for the B^+-classifier and therefore the two variables show a good linearity – as illustrated in Fig. 4.26.

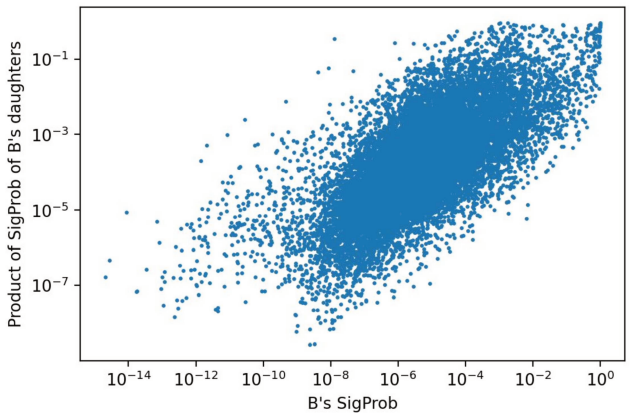


FIGURE 4.26: $\mathcal{P}_{\text{FEI}}^{\text{raw}}$ versus $\mathcal{P}_{\text{FEI}}$ for `FEI` candidates. The former can be used as alternative ranking variable to avoid biases in the ΔE dimension.

We then compare the ΔE distributions for $\overline{D}^0\pi^+\pi^+\pi^-$ and $\overline{D}^{*0}\pi^+\pi^+\pi^-$ `FEI` candidates obtained with the two different ranking approaches: from Fig. 4.27 (upper part) one can see that for the first mode the two methods provide similar peaks at $\Delta E \sim 0$, to which is added a component in negative ΔE in the case of $\mathcal{P}_{\text{FEI}}^{\text{raw}}$. This new peak consists of $\overline{D}^{*0}\pi^+\pi^+\pi^-$ component with a missed photon or π^0. In the bottom part of the same plot, the event distributions in M_{recoil} are compared; for the first case ($\mathcal{P}_{\text{FEI}}$), the $\overline{D}^{(*)0}\pi^+\pi^+\pi^-$ correspond to a $|\Delta E| < 50\,\text{MeV}$, in the second ($\mathcal{P}_{\text{FEI}}^{\text{raw}}$), the range $|\Delta E - 0.2\,\text{GeV}| < 0.1\,\text{GeV}$ is added, in order to recover the $\overline{D}^{*0}$ candidates. Even though the total number of events in the signal window (specified by the dashed lines) increases by $\sim$40%, the extra part does not correspond to high resolution but instead tends to populate the tails of the distribution. One could apply

further constraints to improve the resolution of those events by assuming $\Delta E \equiv 0$. Strictly speaking, the proper way to estimate the gain should consist of selecting first the candidates by the usual $\mathcal{P}_{\text{FEI}}$ ranking and within the narrow ΔE candidates (to favor the fully reconstructed $\overline{D}^{*0}$ modes) and only after adding those events that do not fall in our usual window, recovered with the usage of $\mathcal{P}^{\text{raw}}_{\text{FEI}}$. In this sense, the shifted bin brings a net gain. One needs to make sure that a too high background level does not cancel out the gain in the number of events, i.e., estimate a FOM. Similar procedure can be applied to other D^* modes, like $\overline{D}^{*0}\pi^+$, which can have different background levels.

Although this needs to be verified, we know that the additional bin is not very large and that the background does not grow rapidly in departing from zero.

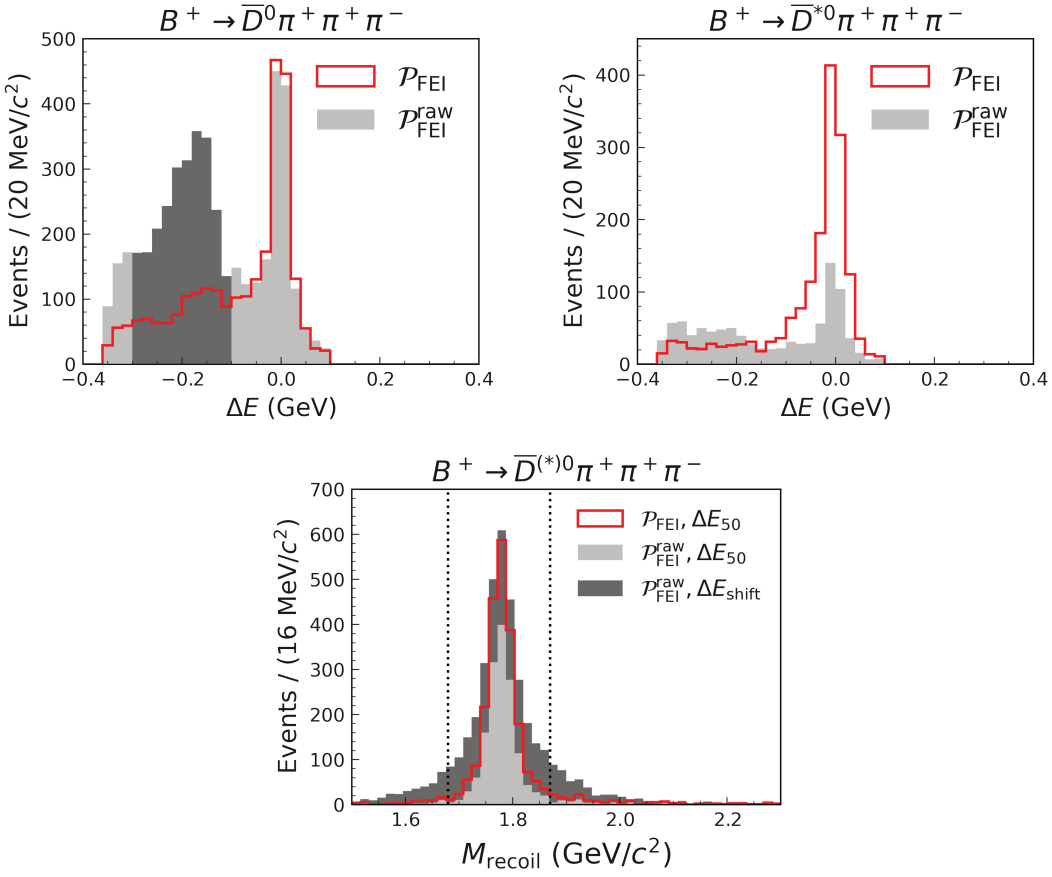


FIGURE 4.27: Distribution for $K\tau\ell$ candidates obtained by applying different best B_{tag} selection, based either on $\mathcal{P}_{\text{FEI}}$ (red) or $\mathcal{P}^{\text{raw}}_{\text{FEI}}$ (gray). On the upper part, the ΔE for $D^{(*)0}\pi^+\pi^+\pi^-$ candidates is shown, on the bottom, the resulting distribution in M_{recoil} obtained after selection appropriate ΔE cuts for the two approaches (see text).

In summary, we want to stress the importance of measuring new B^+ decay rates – or updating them if too old. Right now, Belle II has the unique opportunity, missed in most cases by Belle and *BABAR*, to perform many measurements before entering the stage when priority will be given to the *golden* modes. With new experimental inputs, we can not only further improve the MC modeling but also feed `FEI`, or any other B-tagging strategy. The effort in this direction has already started, and a few analyses are targeting BF measurements. Concerning the BF updates, for example, the $B^+ \to \overline{D}^{(*)0}\rho^+$ modes, which have rather large uncertainties (around 20%) and

represent necessary control samples for many analyses. In addition, the DKK modes will soon be measured to update the Belle result and clarify the structure of the KK^* system. As mentioned before, the cleanest DKK exclusive modes could be added to `FEI`, as they have non-negligible BFs and require the reconstruction of only one D meson. In the context of more inclusive B-tagging approaches, the DKK decays are of great interest, considering that we predict a large inclusive production $\mathcal{B}(B^+ \to DKKX) \simeq 6\%$[19]. The possibility of reconstructing a D and a pair of strange mesons with given flavors might be constraining enough to tag an event as being $B\overline{B}$. Moving to the cases where no experimental result is available, the $B \to D^{(*)}\eta\pi$ are currently under study. This is relevant example of how the measurement would help correct `PYTHIA`, in case any signal is observed, or tune it to match the ULs. The ultimate goal would be to reduce as much as possible the `PYTHIA` contribution and enrich the exclusive fraction of the chart of Fig. 4.1.

The same strategy described in this chapter is being applied to neutral B-decays: a thorough study of B^0 MC simulation and available experimental and theoretical information, as well as the usage of $D^{(*)-}\pi^+$ sample, to verify the final states' kinematics and the goodness of the applied corrections. There is room for improvement for B^0's as well, as the measured calibration factors are 0.85 ± 0.06 [11] at Belle and 0.83 ± 0.03 at Belle II [109].

[19] Where $D = D^{\pm,0,*}$, $K = K^{\pm,0,*}$.

Chapter 5

$B^+ \to K^+\tau\ell$ search with semileptonic B-tagging

Many analyses of Belle and *BABAR* experiments have exploited a semileptonic B-tagging approach to obtain higher efficiencies. In fact, for channels like $B \to D^{(*)}\tau\nu$ (see the $R(D)$-$R(D^*)$ measurement [31]) or $B^+ \to \tau^+\nu_\tau$ the signal yield is measured from the distribution of the ECL energy associated with the rest of the event (ROE). This variable would peak at zero for signal events independently of the reconstructed B_{tag} decay modes (hadronic or semileptonic). The BF of $B^+ \to \tau^+\nu_\tau$ was measured with the full Belle dataset using the hadronic B-tagging, leading to a $3.0\,\sigma$ significance [122]; a few years later, the measurement has been done with an improved strategy and the semileptonic B-tagging, providing a better significance ($3.8\,\sigma$) [123]. As the two samples are independent, Ref. [123] also provides a combined significance for the two results, obtaining 4.6σ. This example shows the importance of the B-tagging strategy and the availability of independent samples when it comes to observing rare decays.

Other modes with specific topologies allow for even fewer constraints from the tag side and provide very competitive results with an *un-tagged* approach. In this category we can mention the 2-body $B \to \mu\nu$ decay searched at Belle [124] or the $B^+ \to K^+\nu\bar{\nu}$ at Belle II [125]; in both cases, the B_{sig} is characterised by the presence of a single track. The inclusive approach has also been attempted for the $B \to K\tau\ell$ search [126], where only the $\tau \to \pi$ mode is used to exploit the additional kinematic constraint given by the 2-body decay. The analysis was not optimised and did not lead to a competitive result.

In this chapter, we apply for the first time the semileptonic (SL) B-tagging approach to the $B^+ \to K^+\tau\ell$ search (in particular, we focus on the $\mathbf{OS}_\mu$ mode). The main problem one needs to tackle is the degraded resolution on the M_{recoil} due to the presence of an additional neutrino in the event, this time from the B_{tag}. On the other hand, we expect the semileptonic `FEI` B-tagging to be significantly more efficient than the hadronic (ε is roughly four times higher, see [8]), and this could compensate for the impact due to the worse resolution. The scope of the first part of this chapter is to obtain from MC a realistic background estimation for the $B^+ \to K^+\tau^+\mu^-$ mode, as it cannot simply be predicted based on the analysis performed with the hadronic B-tagging. In the second part, we put in place a strategy to improve the resolution on M_{recoil}, detailing all the constraints that can be exploited.

5.1 Recoil mass properties

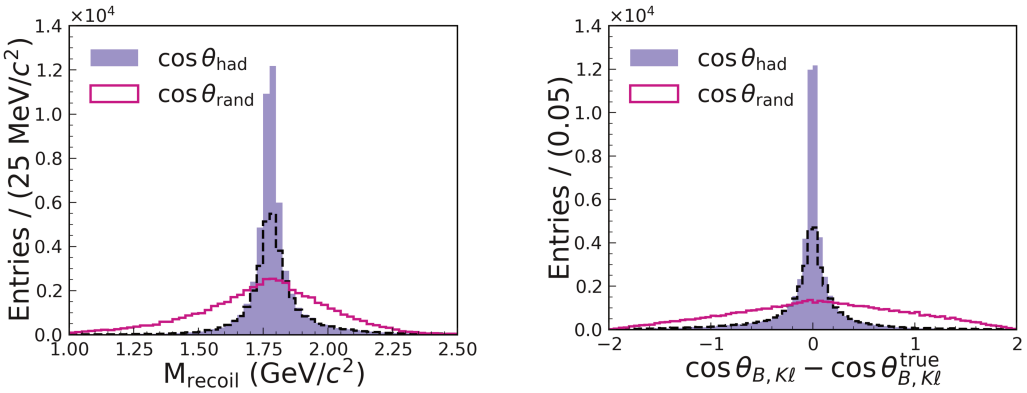
The peculiarity of our mode (compared with those conserving the leptonic flavor) lies in the fact that the only source of missing energy of the B_{sig} comes from the τ, which allows – in the case of hadronic B-tagging – to use the M_{recoil} to extract the signal. In

the case of semileptonic B-tagging, the momentum of the B_{tag}, $\mathbf{p}_{\text{Btag}}$ is unknown due to the presence of the undetected neutrino. As a consequence, the expression of Eq. 3.6 must be revised: $p^*_{B_{\text{tag}}}$ can be replaced with the average momentum obtained with the relation $p^*_{\text{beam}} = \sqrt{E^{*2}_{\text{beam}}/c^2 - m_B^2 c^2}$ while the $\cos\theta$[1] term is tentatively replaced with a random number picked from a uniform distribution in the range $(-1, 1)$.

Hence, the new fit variable is

$$M^2_{\text{recoil}} = m_B^2 + m^2_{K\ell} - 2(E^*_{\text{beam}} E^*_{K\ell}/c^4 + p^*_{\text{beam}} p^*_{K\ell} \cos\theta_{\text{rand}}/c^2). \tag{5.1}$$

Because of these two substitutions, the resolution on M_{recoil} is much worse, as shown in Fig. 5.1A in comparison with the distribution obtained with hadronic B_{tag}'s. The resolution on the $\cos\theta$ term is also shown (Fig. 5.1B), while the effect of using p^*_{beam} instead of $p^*_{B_{\text{tag}}}$ has already been presented in section 3.3 and Figure 3.4.



(A) (B)

FIGURE 5.1: M_{recoil} distribution (A) and resolution on the $\cos\theta$ term of Eq. 3.6 (B) for signal MC events, comparing hadronic B_{tag}'s ($\cos\theta_{\text{had}}$, filled blue) and the semileptonic ones ($\cos\theta_{\text{rand}}$, solid purple line). The sub-component of false-matched hadronic candidates is highlighted with the dashed black line.

The fit to the M_{recoil} variable defined in Eq. 5.1 is performed using a combination of Crystal Ball (CB) and Gaussian PDFs (see Eq. 3.11). Unlike for the hadronic case, the mean of the Gaussian μ_{Gauss} is floated, in order to reflect the left-side tail; this can be seen in Fig. 5.2 and Tab. 5.1. It is worth emphasizing the different resolution with respect to the hadronic (HAD) B-tagging: $\sigma^{\text{SL}}_{\text{CB}} \sim 160\,\text{MeV}/c^2$ against $\sigma^{\text{HAD}}_{\text{CB}} \sim 25\,\text{MeV}/c^2$ and that the n parameter (see Tab. 5.1) is compatible with zero, i.e. we could replace the CB with a simple Gaussian. However, we keep the CB in order to use the same PDF when comparing the performance of semileptonic tag with different resolutions and ultimately with the hadronic B_{tag}'s (see the next sections). The signal region is defined as the M_{recoil} window centered on m_τ^{PDG} and extending on each side for roughly three times the width of the CB, resulting in the range $(1.33, 2.21)\,\text{GeV}/c^2$. This window will be used to decide the cut value on the BDT output for the background reduction. In Sec. 5.5, we will discuss how to improve the resolution on the M_{recoil} by using the kinematic constraints and the vertexing information.

[1] θ is the angle between $\mathbf{p}_{\text{Btag}}$ and $\mathbf{p}_{K\mu}$.

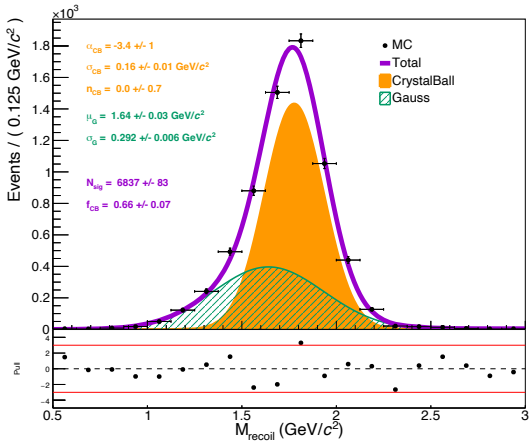


FIGURE 5.2: Fit to M_{recoil} corresponding to $B^+ \to K^+\tau^+\mu^-$ MC events tagged through semileptonic decay modes.

Parameter	Fit result
f_{CB}	0.66 ± 0.07
σ_{CB} (MeV/c^2)	160 ± 10
α	-3.4 ± 1.2
n	0.0 ± 0.7
μ_{Gauss} (GeV/c^2)	1.64 ± 0.03
σ_{Gauss} (MeV/c^2)	292 ± 6

TABLE 5.1: Fit parameters for the M_{recoil} of $B^+ \to K^+\tau^+\mu^-$ events.

Before that, we describe the semileptonic `FEI` reconstruction and the selection to obtain a manageable background level and estimate the expected UL on the $B^+ \to K^+\tau^+\mu^-$ BF in case of no signal observed.

5.2 Reconstruction

The B_{tag} is reconstructed using the same BDT-based hierarchical algorithm presented in chapter 3; the reconstruction of the decay chain starts with the identification of a charged lepton and is followed by the reconstruction of a D meson (most commonly a $\overline{D}^0$ for B^+ decays) via the same channels listed in Tab. 3.3 for the `FEI` hadronic B-tagging. The D meson can be combined with a soft pion to form a D^* which is compatible with the chosen $\Delta M = M(D^*) - M(D)$ window. The full list of B-modes considered in the semileptonic version of the `FEI` algorithm is shown in Table 5.2: in addition to the $\overline{D}^{(*)0}\ell^+\nu_\ell$ modes, which account for $\sim 70\%$ of $X_c\ell^+\nu_\ell$ decays and $\sim 90\%$ of the total tagging efficiency, a small contribution is given by the $D^{(*)-}\pi^+\ell^+\nu_\ell$ modes.

ID	**Decay Mode**
0.	$B^+ \to \overline{D}^0 e^+$
1.	$B^+ \to \overline{D}^0 \mu^+$
2.	$B^+ \to \overline{D}^{*0} e^+$
3.	$B^+ \to \overline{D}^{*0} \mu^+$
4.	$B^+ \to D^- \pi^+ e^+$
5.	$B^+ \to D^- \pi^+ \mu^+$
6.	$B^+ \to D^{*-} \pi^+ e^+$
7.	$B^+ \to D^{*-} \pi^+ \mu^+$

TABLE 5.2: Semileptonic `FEI` B^+ decay modes.

In semileptonic $B \to D^{(*)}(n\pi)\ell\nu$ decays, the 4-momentum carried by the neutrino:

$$p^*_\nu = p^*_B - p^*_{D^{(*)}(n\pi)} - p^*_\ell \tag{5.2}$$

is unknown and consequently the direction of the parent B, $\mathbf{p}^*_B$. However, in the $\Upsilon(4S)$ rest frame, p^*_B is on average 0.34 GeV/c and can be neglected, if compared to the momenta of the lepton or the $D^{(*)}$ meson. Under this assumption, the Eq. 5.2 is simplified and so is the expression:

$$m^2_\nu = p^{*2}_\nu = \left(E^*_{\text{beam}} - E^*_{D^{(*)}(n\pi)} - E^*_\ell\right)^2 c^{-4} - \left(\mathbf{p}^*_{D^{(*)}(n\pi)} + \mathbf{p}^*_\ell\right)^2 c^{-2}, \tag{5.3}$$

representing the approximate missing mass squared related to the candidate B, where $E^*_B \equiv E^*_{\text{beam}}$ and $\mathbf{p}^*_B \equiv \mathbf{0}$. For correctly-reconstructed candidates, where only the neutrino is escaping the reconstruction, the m^2_ν distribution has a strong peak at zero and a spread of roughly $0.850\,\text{GeV}^2/c^4$. For background events, the m^2_ν values are distributed in a wider range. From the squared version of Eq. 5.2 one can derive a second relation for the cosine of the angle between the direction of the B and that of the 'visible' system:

$$\cos\theta_{\text{BY}} = \frac{2E^*_B E^*_Y - m^2_B c^4 - m^2_Y c^4}{2|\mathbf{p}^*_B||\mathbf{p}^*_Y|c^2} \tag{5.4}$$

with the notation $Y \equiv D^{(*)}(\pi)\ell$ and having imposed $p^2_\nu = m^2_\nu = 0$. The $\cos\theta_{\text{BY}}$ variable must lie in the interval $(-1, 1)$ for signal events, which geometrically means that the direction of the B is constrained to lie on a cone centered on the $D^{(*)}(n\pi)\ell$ momentum direction and of aperture given by Eq. 5.4. This property is illustrated in Fig. 5.3.

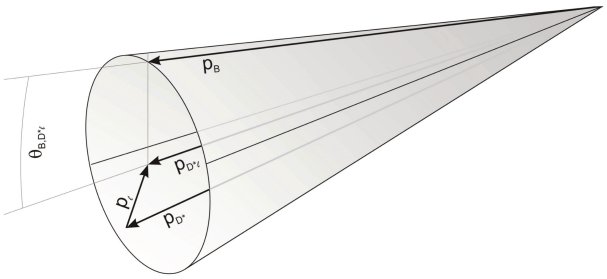


FIGURE 5.3: Direction of the B meson is constrained to lie on the cone of aperture $\theta_{B,D^{(*)}\ell}$ and centered on $\mathbf{p}_{D^{(*)}\ell}$, which is the visible system reconstructed by `FEI`. Figure adapted from Ref. [127].

The `FEI` B_{tag} candidates are selected with the two following criteria:

- $|\cos\theta_{\text{BY}}| < 4$ - as a pre-selection cut,
- $\mathcal{P}_{\text{FEI}} > 10^{-3}$ - to enrich the sample of B meson-like candidates.

Among the candidates surviving the above selection – in average 5 per event – only the highest $\mathcal{P}_{\text{FEI}}$-ranked per event is taken, corresponding to a 70% efficiency in selecting the correct (TM) one. The distributions of $\mathcal{P}_{\text{FEI}}$ and $\cos\theta_{\text{BY}}$ for TM and FM signal candidates are shown in the left column of Fig 5.4. The signal and background shapes after the cuts are given in the right column of the same plot. One can notice that even for TM B_{tag} candidates, the $\log_{10}(\mathcal{P}_{\text{FEI}})$ does not peak at zero like for the hadronic B_{tag}'s; this is a known feature of semileptonic `FEI`, but the reason is yet to be found.

At this stage, we do not apply any cuts on the momentum of the lepton candidate, which we expect to have larger values for semileptonic B-decays as opposed to D-decays or leptons coming from continuum events. We will exploit this information at a later stage of the selection. Regarding the signal-side reconstruction and the final

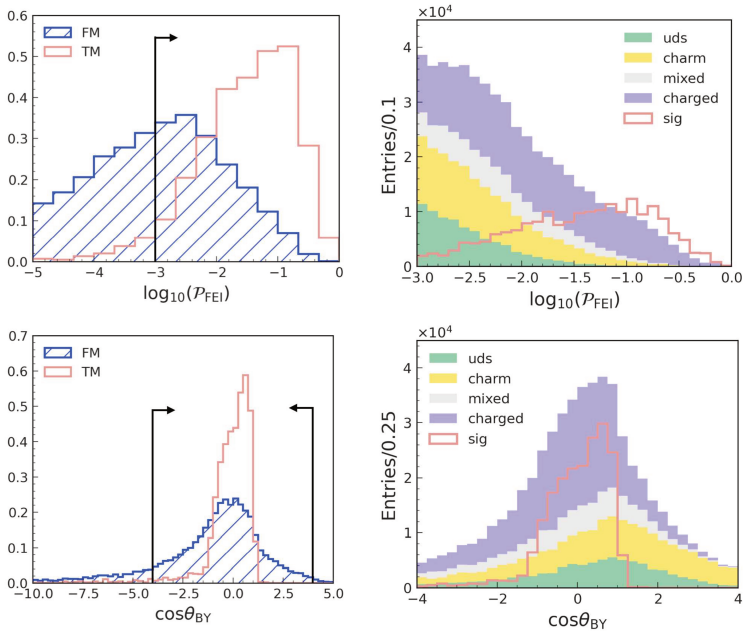


FIGURE 5.4: Top: Distribution of $\log(\mathcal{P}_{\rm FEI})$ for signal events (TM and FM, normalized to the same area) and distribution of the same variable for background and signal events (with arbitrary normalization) after the cut $\mathcal{P}_{\rm FEI} > 10^{-3}$. Bottom: Distribution of $\cos\theta_{\rm BY}$ (defined in Eq. 5.4) for signal events and distribution of the same variable for background and signal events after the cut $|\cos\theta_{\rm BY}| < 4$.

event selection, the same criteria described in sections 3.5-3.6 are applied here. The τ is reconstructed as a single track (1-*prong* decays) and is labelled in three possible ways e, μ, π, the latter being highly contaminated by the $\tau \to \rho\nu$ component. The possibility of explicitly reconstructing the $\tau \to \rho\nu$ decay and performing a more refined classification of the τ modes will be discussed later. As shown in Tab. 5.3, $\sim$70% of $B_{\rm tag}$ candidates are FM because of errors in the reconstruction of the final states – mainly the D meson and for the $D\ell\nu$ modes, where there is not the extra constraint given by the ΔM cut. Nonetheless, from the MC information, we know that 85% of the time the event has a semileptonic B-decay in the $B_{\rm tag}$ with the candidate lepton being correctly assigned. Similarly to what we have encountered in chapter 3, when the $(K^+_{\rm sig}, \mu^-_{\rm sig})$ pairs are correctly reconstructed and the $B_{\rm tag}$'s are TM, the wrong t_τ's (category 'Wrong t_τ') are mostly pions coming from $\tau \to \rho$ decays. Conversely, when the $B_{\rm tag}$ is classified as FM, it is possible that a track from the tag-side is picked as t_τ candidate (it happens in $\sim$30% of the cases). It should be stressed that, besides the cases where the $(K^+_{\rm sig}, \mu^-_{\rm sig})$ pairs are wrong, all other categories have an $M_{\rm recoil}$ distribution peaking at m_τ and with the same resolution. In fact, no term in Eq. 5.1 depends from the reconstructed $B_{\rm tag}$, nor from t_τ.

5.2.1 Resonance vetoes

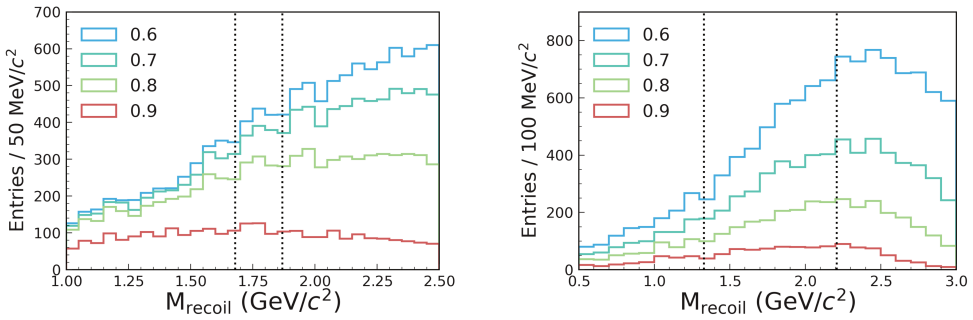
Before moving to the background suppression via BDT, we look for convenient cuts that can be applied on the narrow resonances producing our final state. For example, $D^0 \to K^-\pi^+$ events can be seen in the $(K^-_{\rm sig}, \mu^+_{\rm sig})$ variable, due to mis-identification of the pion to a muon. In general, the whole $M(K_{\rm sig}\mu_{\rm sig}) \lesssim 1.91\,{\rm GeV}/c^2$ spectrum can be related, for $B\overline{B}$ background, to D^0 meson (mainly semileptonic) decays. In contrast to Sec. 3.7.2, where we apply a narrow D^0-veto and use the $M(K_{\rm sig}\mu_{\rm sig})$ for

B_{tag} \ B_{sig}	TM	Wrong t_τ	Wrong $K\ell$	Wrong	Total
TM	19.6	11.5	0.0	0.4	31.5
FM	29.7	31.1	2.0	5.7	68.5
Total	49.3	42.6	2.0	6.1	100.0

TABLE 5.3: Fraction (%) of signal MC events in different categories, according to the number of wrong particles in the event. The values refer to the reconstruction of the $\mathbf{OS}_\mu$ mode with semileptonic B-tagging.

the training, here we decide to remove the whole region compatible with a charm decay. This is justified by two considerations:

- For $\mathbf{OS}_\mu$, the cut is not very demanding efficiency-wise (75% signal efficiency against 60% background rejection);
- As shown in Eq. 5.1, the M_{recoil} linearly depends on $M(K_{\text{sig}}\mu_{\text{sig}})$. As $M(K_{\text{sig}}\mu_{\text{sig}})$ is also the most important feature for the BDT, we risk sculpting the M_{recoil} background distribution when applying a cut on the BDT output. In Fig. 5.5 we compare the effect of the BDT cut on background events for the hadronic-tagged analysis of chapter 3 with those obtained with the selection described in the previous section. In the hadronic case, such dependence is not an issue because of the excellent resolution of signal events ($\sim 30\,\text{MeV}/c^2$). As we have demonstrated in Fig. 5.2, the resolution is much worse in the semileptonic case; consequently, the signal and fit regions are larger, making it more delicate to use the $M(K_{\text{sig}}\mu_{\text{sig}})$ variable for the BDT training.



(A) (B)

FIGURE 5.5: Effect of the cut on the BDT output at different thresholds on the background M_{recoil} shape, for (A) hadronic and (B) semileptonic-tagged analyses. The dashed, black lines indicate the signal regions.

Turning to the sources of $\ell^+\ell^-$ pairs, an obvious one consists of $B^+ \to K^+ X_{c\bar{c}}(\to \ell^+\ell^-)$ events. In the context of $B^+ \to K^+\tau^+\mu^-$ search with hadronic B-tagging, the only possible di-lepton combination is given by $\mu^\pm_{\text{sig}}\mu^\mp_\tau$ for $\tau \to \mu\nu\nu$ candidates. With the semileptonic B-tagging, an additional lepton is added to selection (ℓ_{tag}) and therefore more $\ell^+\ell^-$ combinations are possible, i.e., $\mu^\pm_{\text{tag}}\mu^\mp_\tau$ ($e^\pm_{\text{tag}}e^\mp_\tau$) for $\tau \to \mu\nu\nu$ ($\tau \to e\nu\nu$) modes. For this reason we apply J/ψ and $\psi(2S)$ vetoes in the variables $\ell_{\text{tag}}\ell_\tau$ and $\mu_{\text{sig}}\mu_\tau$. It is worth mentioning that, despite those components might not look significant from Fig. 5.6, they would provide a non-negligible fraction of background events at the end of the cut-flow.

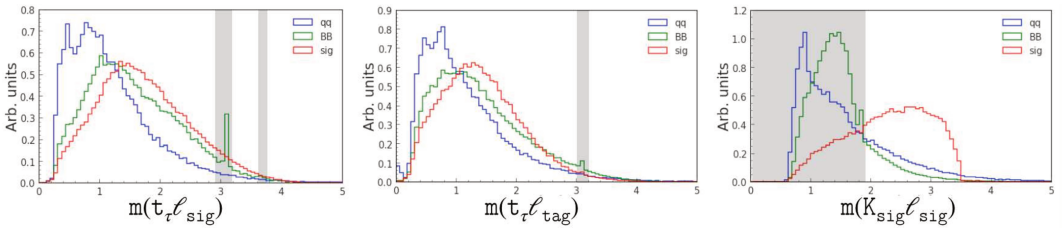


FIGURE 5.6: Invariant masses of pairs of particles with opposite charges ($\mathbf{OS}_\mu$ mode). The shaded areas correspond to the vetoed regions that do not enter the BDT training and the signal extraction.

Another source of di-leptons is given by photon conversions $\gamma \to e^+e^-$, where one lepton is selected as t_τ and the other as ℓ_{tag}. These events correspond to a peak in $M(t_\tau \ell_{\text{tag}}) < 0.1\,\text{GeV}/c^2$ but we decide not to apply a veto, as we notice that such component is mostly present in $q\overline{q}$ events, which are easily suppressed by the BDT classifier.

In Tab. 5.4 we summarize all the vetoes explored in the present paragraph with the corresponding vetoed regions, highlighting with the blue color the contributions that are specific to the semileptonic B-tagging.

		$c\bar{c} \to \ell^+\ell^-$	$\bar{D}^0 \to K^+\pi^-$	$\gamma \to e^+e^-$
$\mathbf{OS}_\mu$	e	$e^+_\tau e^-_{\text{tag}}$		$e^+_\tau e^-_{\text{tag}}$
	μ	$\mu^+_\tau \mu^-_{\text{sig}} / \mu^+_\tau \mu^-_{\text{tag}}$	$K^+_{\text{sig}} \pi^-_{\text{sig}}$	
	π			
Inv. Mass Veto		$(2.847, 3.177)\,\text{GeV}/c^2$ $(3.60, 3.75)\,\text{GeV}/c^2$	$(0, 1.91)\,\text{GeV}/c^2$	$(0, 0.1)\,\text{GeV}/c^{2\dagger}$ † not applied

TABLE 5.4: Summary of the narrow resonances compatible with the signal final state particles (τ-mode-dependent) and the applied vetoes. $c\bar{c}$ refers to the J/ψ and $\psi(2S)$ resonances. In blue are the resonances specific to the $\mathbf{OS}_\mu$ channel and semileptonic B-tagging.

5.3 Background studies

In order to describe the background composition at the different stages of the selection we utilise the same classification of chapter 3, i.e. dividing the events in three groups depending on the number of prompt leptons (0, 1, 2). At the initial stage of the selection, $\sim 2/3$ of the total background is from $B\overline{B}$ events (see Tab. 5.5) and its composition depends on the nature of the t_τ. In fact, in case of $t_\tau = \ell$ the SL×SL component prevails ($\sim 70\%$ of the total) because of the charge configuration of the $\mathbf{OS}_\mu$ mode, the t_τ track has the same charge of the $B_{\rm sig}$ candidate and thus the lepton ℓ from $B^+ \to \overline{D}^{(*)0}\ell^+\nu_\ell$ decays can be taken as τ prong. For the $\tau \to \pi$ mode, the $B_{\rm sig}$ components is preferably decaying into a hadronic channel (SL×HAD at $\sim 75\%$). For both cases, the $K^\pm_{\rm sig}$ and the $\mu^\mp_{\rm sig}$ candidates are predominantly coming from the D^0 mesons.

Only after the cut $M(K_{\rm sig}\mu_{\rm sig}) > 1.91\,{\rm GeV}/c^2$, the $B\overline{B}$ background is reduced for the reasons explained above and the $q\overline{q}$ component becomes comparable to the $B\overline{B}$.

5.3.1 Background suppression via BDT

Similarly to what was done for the hadronic-tagged analysis, a BDT classifier is trained to get an efficient background suppression. Unlike the previous study, however, we decide to apply a single-step training and simultaneously reduce $q\overline{q}$ and $B\overline{B}$ backgrounds, motivated by the fact that a large part of the 'easy' $B\overline{B}$ background is removed with the $M(K_{\rm sig}\mu_{\rm sig})$ cut. The features used for the training can be divided into four main categories:

- **Event shape/continuum suppression**
  We use variables related to the event shape (sphericity and $\cos\theta_T$) and those built from the ROE of the $B_{\rm tag}$ candidates, like the KSFW moments[2]. As shown in chapter 3, they provide some separation power not only against $q\overline{q}$ but also $B\overline{B}$ events.

- **ROE**
  The energy deposits in the ECL related to the ROE – $\rm n_{ROE}$(ECL) and $\rm E_{ROE}$(ECL).

- **Invariant masses**
  For the $\mathbf{OS}_\mu$ configuration, $\ell_{\rm tag}$ and $\ell_{\rm sig}$ have the same sign; the following variables are used in the training after having applied the vetoes described in the previous section: $M(t_\tau\ell_{\rm sig})$, $M(t_\tau\ell_{\rm tag})$, $M(K_{\rm sig}\ell_{\rm tag})$.

- **$\mathbf{B}_{\rm tag}$**

  - m^2_ν as defined in Eq. 5.3. The m^2_ν distribution would peak at zero for true semileptonic decays and tend to have larger tails for background events.
  - $\cos\theta_{\rm BY}$ as defined in Eq. 5.4. In the `basf2` version used for the analysis, the semileptonic `FEI` uses the same training variables as the hadronic version. Some of these variables relate to the angles and the visible products of the B, which are related to the $\cos\theta_{\rm BY}$. As a result, the semileptonic-`FEI` classifiers learn the signal shape of the $\cos\theta_{\rm BY}$ and sculpt the background shape. In future versions of the training, this effect will be fixed and most likely the $\cos\theta_{\rm BY}$ will provide a higher separation power.

[2]The used ones are hso10, hoo1, hso02, hoo2, hso20, hso01.

- As mentioned in Sec. 5.2, we do not perform cuts on the lepton momentum $p^*(\ell_{\rm tag})$ at the $B_{\rm tag}$ selection stage; hence we exploit it for our BDT training along with the momentum of the charmed meson, denotes as $p^*(D_{\rm tag})$, which shows some separation for the $q\overline{q}$.
- `SignalProbability`.

A subset of the variables used for the BDT training are shown in Fig. 5.7, where the shapes for signal events are compared to those for $q\overline{q}$ and $B\overline{B}$ events. Some of the features have already been shown for the analysis in chapter 3 and are omitted here. The training features are ranked according their importance in Fig. 5.8; one can see that the ROE information, the continuum suppression variables and those related to the $B_{\rm tag}$ properties have high score rank. The features are also represented in the correlation matrix of Fig. 5.9, where below (above) the diagonal the values refer to signal (background) events. The m_τ variable is added to show the linear correlations with the training variables – all lower than ±20%.

As it can be seen in Fig. 5.10 a better separation of the $q\overline{q}$ component with signal is obtained; as a result, after the BDT selection (next paragraph), only the $B\overline{B}$ survive while the continuum becomes negligible.

Fig. 5.11A provides the background rejection and the signal efficiency as a function of the BDT score and the same scan is performed computing the FOM defined in Eq. 3.10. The maximum significance is found for a BDT score of $\simeq 0.9$, as shown in Fig. 5.11B, corresponding to a final efficiency $\varepsilon_{\rm sig} = 0.131\%$ – roughly 60% larger than the one obtained for the $\mathbf{OS}_\mu$ mode with the hadronic B-tagging[3]. Such increase is mainly due to the higher tag-side efficiency obtained with semileptonic decays. We stress that the adopted cut-flow, summarized in Tab. 5.5, might not be optimal but is sufficient to get an UL estimation and demonstrate the capabilities of the semileptonic B-tagging in this search.

Cut	**Signal**	$\mathbf{B\overline{B}}$	$\mathbf{q\overline{q}}$	$\mathbf{B\overline{B} + q\overline{q}}$
Reco	217.7 ± 1.3	11850 ± 50	6877 ± 27	18730 ± 70
$M(K\ell)$ cut	163.5 ± 1.1 (-24.9%)	3822 ± 27 (-67.8%)	3310 ± 24 (-51.9%)	7130 ± 50 (-61.9%)
$c\bar{c}$ vetoes	159 ± 1.1 (-2.7%)	3383 ± 25 (-11.5%)	3286 ± 24 (-0.7%)	6670 ± 50 (-6.5%)
BDT	43.4 ± 0.6 (-72.7%)	107 ± 7 (-96.8%)	2.2 ± 0.1 (-99.9%)	110 ± 7 (-98.4%)

TABLE 5.5: Summary of the cuts for the $\mathbf{OS}_\mu$ reconstruction. The assumption $\mathcal{B}(B^+ \to K^+\tau^+\mu^-) = 5.0 \times 10^{-5}$ is made for the signal. For the background components the yields correspond to 1 stream of generic MC. All the yields are estimated in the signal window $M_{\rm recoil} \in (1.33, 2.21)\,{\rm GeV}/c^2$.

The background surviving the BDT cut is composed at 30% of $B^0\bar{B}^0$ events, while the rest (B^+B^-) can be divided as follows:

- $D_s^{(*)-}\pi^+(\pi^0)K^+$: 30%
- $\bar{D}^{(*(*))0}X^+_{\rm had}, X^+_{\rm had} = \{n\pi K^{(*)+}, D_s^{(*)+}, \overline{K}^{(*)0}K^{(*)+}\}$: 40%
- $\bar{D}^{(*(*))0}\ell^+\nu_\ell$: 20%

[3] 0.0813% without data-MC corrections.

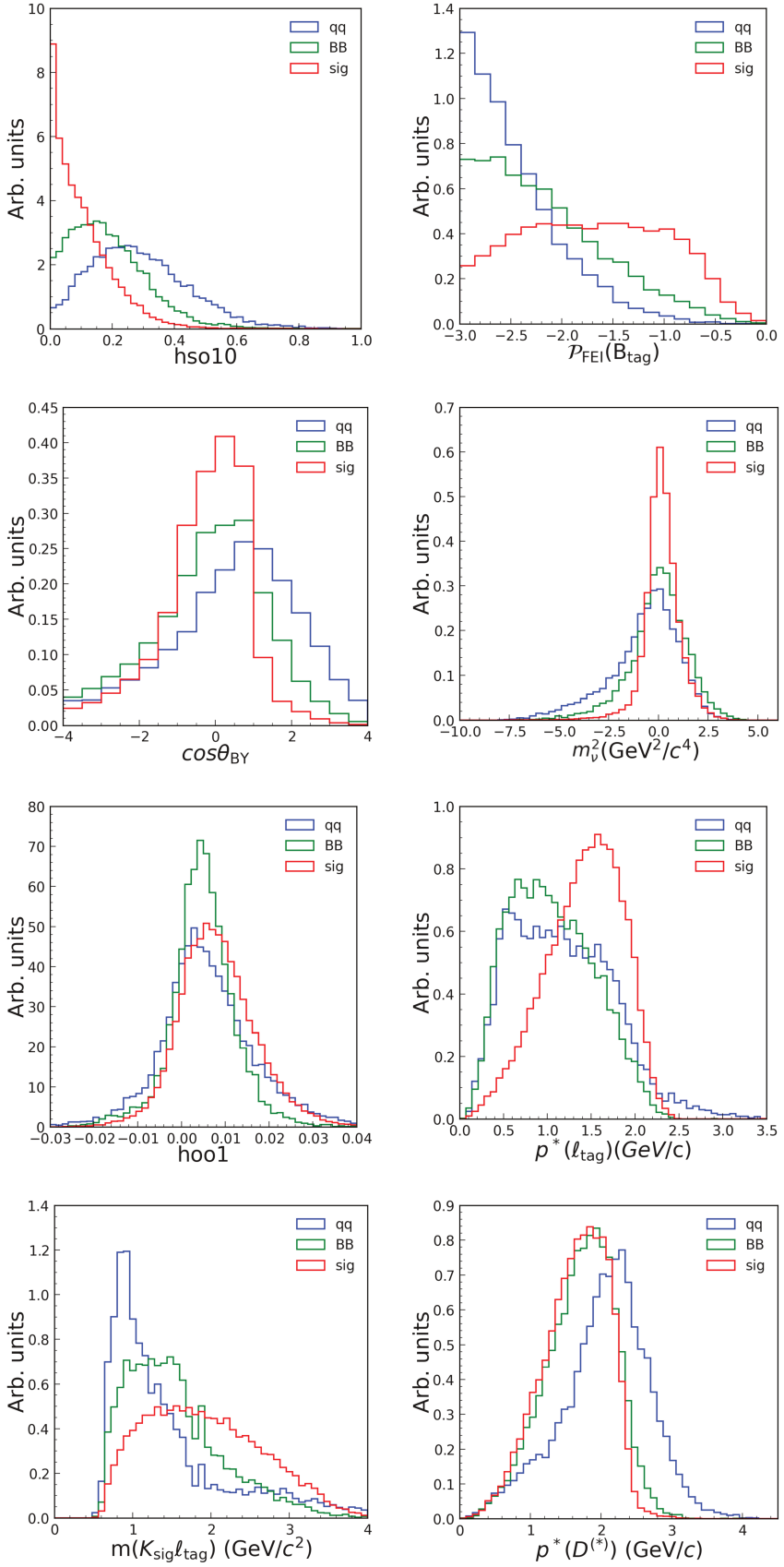


FIGURE 5.7: Selected training features for the FastBDT classifier ($\mathbf{OS}_\mu$ mode).

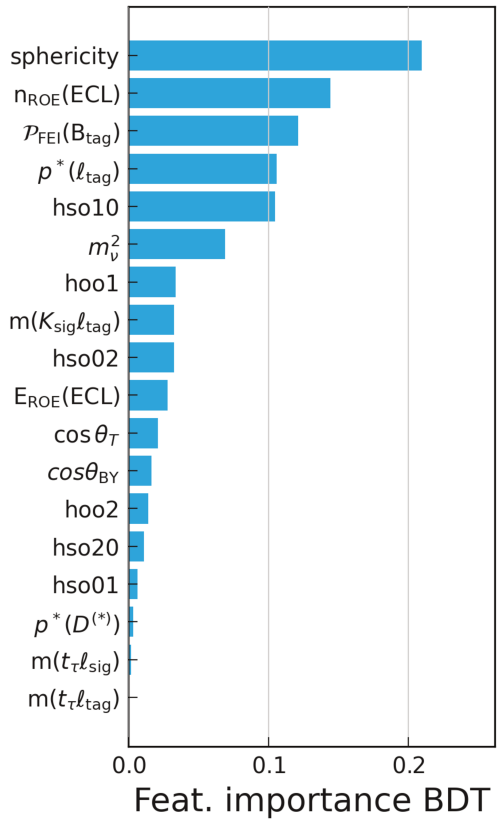


FIGURE 5.8: Importance ranking of the features used for training of the BDT classifier.

- $(c\bar{c})K^+$: 10%

A large fraction of the B^+B^- events has a prompt kaon ($\sim 50\%$) while the D_s^+ meson decays leptonically ($D_s^+ \to \tau^+/\mu^+\nu_{\tau/\mu}$) providing the μ_{tag} candidate. The latter case, although it can lead to the exact final state as the signal, cannot provide a peak at the m_τ in the M_{recoil} distribution. Due to the $M(K_{\text{sig}}\mu_{\text{sig}})$ cut, the K_{sig} and the μ_{sig} cannot come from the same D meson, leading to two possible cases[4]:

- If μ^-_{sig} is produced by a $\overline{D}^0$, the K^+_{sig} is forced to be *prompt* and the t^+_τ can either come from the B^+ or a D;

- when μ^-_{sig} is from other sources (e.g., $D_s^-, J/\psi$ or B^- in case it swaps with the ℓ^-_{tag}), the signal kaon can either come from the $\overline{D}^0$ or the B^+ and the t_τ privileges the B^+ or J/ψ. There is indeed a surviving component of $J/\psi \to (\mu_{\text{sig}}\mu_\tau)$ for candidates outside the veto region.

We consider these events as our ultimate background that we can fit to get a sensitivity estimate. A more refined selection based, for example, on different training variables for the BDT classifier could be implemented, probably without changing our conclusions significantly.

[4]For clarity we fix here all the charges: $B^-_{\text{tag}}B^+_{\text{sig}} \to K^+\tau^+\mu^-$.

FIGURE 5.9: Correlation matrix of the variables used for the training, plus the m_τ. On the upper side of the diagonal, the values refer to background events while on the lower side signal events are shown.

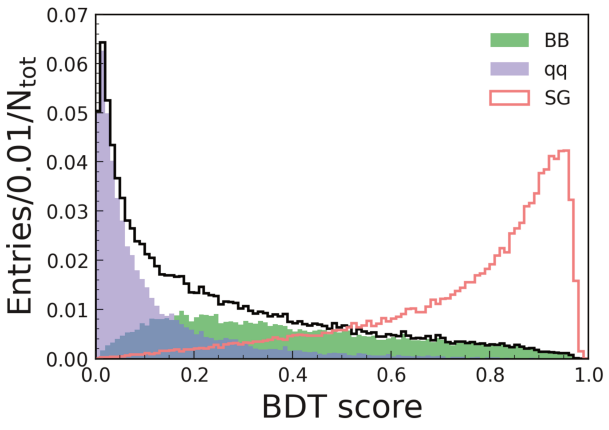
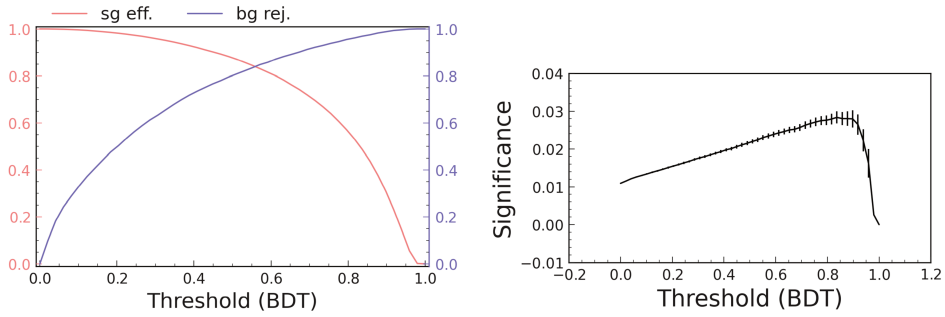


FIGURE 5.10: BDT response as a result of the training with the features described in Sec. 5.3.1.



(A) (B)

FIGURE 5.11: BDT classifier performance: Fig. A shows the survived signal and the rejected background as a function of the threshold on the BDT output, while in Fig. B the significance defined in Eq. 3.10.

5.4 Signal and UL extraction

The signal extraction follows the approach described in Sec. 3.9.3: the model defined in Eq. 3.14 with the fitted values of Tab. 5.1 is used to extract the signal yield, $N_{\text{sig}}^{\text{meas}}$. The unbinned ML fit to MC for the usual hypothesis $\mathcal{B}(B^+ \to K^+\tau^+\mu^-) = 5 \times 10^{-5}$ is shown in Fig. 5.12.

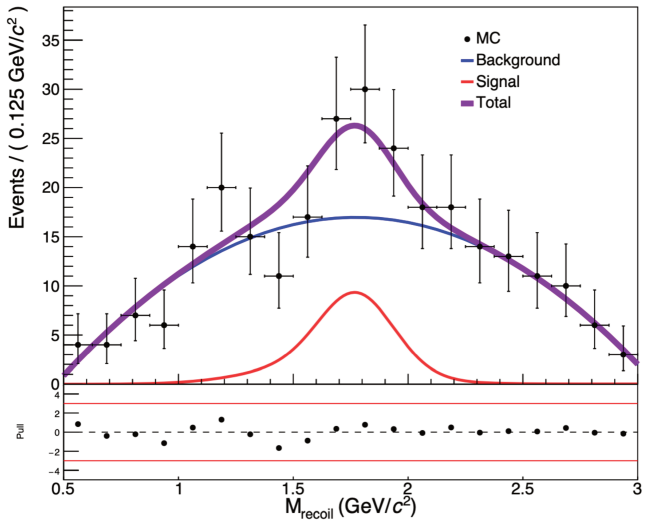
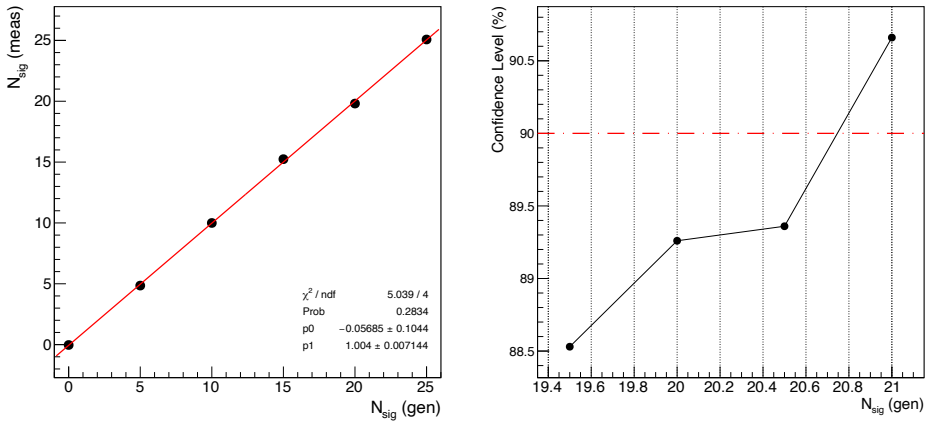


FIGURE 5.12: Fit to MC data surviving the BDT selection (see text). The signal events, re-scaled to the BR= 5×10^{-5}, are fitted with a fixed shape with parameters taken from Tab. 5.1. The background is modeled with a $2nd$-order polynomial.

The last step consists in deriving the UL with the same frequentist approach used in chapter 3: we produce 10k pseudo-experiments for a set of injected N_{sig} values. $N_{\text{sig}}^{\text{UL}}$ is the N_{sig} for which 90% of the pseudo-experiments have a $N_{\text{sig}}^{\text{meas}} \geq 0$. The obtained value is then converted into a BF by using the relation 3.15, yielding to $\mathcal{B}^{\text{UL}}(OS_\mu) = 2.0 \times 10^{-5}$. The linearity of the fit model and the scan in $N_{\text{sig}}(\text{gen})$ for the upper limit derivation are shown in Fig. 5.13.

The achieved UL is worse than what obtained with the hadronic B-tagging: 1.2×10^{-5}, considering the expected UL for the zero signal case. This result seems to suggest that, despite the higher tag-side efficiency leading to higher ε_{sig}, the higher background level combined combined with the degraded resolution on M_{recoil}, produces a worse



(A) (B)

FIGURE 5.13: Scan in $N_{\rm sig}$(gen) to check the linearity of the fitting procedure (A) and obtain the upper limit at 90% C.L. for the zero signal case (B).

sensitivity. Regardless, this result is extremely interesting because of the novelty of the approach. In fact, not only was the semileptonic B-tagging never tried before for these LFV modes but also being able to set an UL with the very same procedure (i.e., employing a fit to the $M_{\rm recoil}$) was not apparent by any means. It should be mentioned that, like for other measurements at Belle involving modes with missing energy ($B \to \tau\nu, B \to D^{(*)}\tau\nu$), we could use the $E_{\rm ROE}$(ECL) as the variable of signal extraction. In that case, the $E_{\rm ROE}$(ECL) should not be used for the background suppression, while $M_{\rm recoil}$ could be exploited as a selection variable.

In the next section, we explore the possibility of improving the resolution of $M_{\rm recoil}$ and try to estimate the ultimate potential of the semileptonic tag for our search.

5.5 The resolution path: MAOS

M_2 belongs to a family of variables that have been extensively used for searches of new particles at LHC. The initial versions, called M_{T2} [128,129], or transverse masses, were developed to find a lower bound on the masses of supersymmetric particles in processes where two BSM invisible particles would decay into visible (SM) and invisible (BSM) products. The M_{T2} variables use the transverse momenta of invisible particles as the parameters of multidimensional constrained minimization for a bump hunt, while M_2 constitutes a Lorentz-invariant extension exploiting both transverse and longitudinal momenta and making use of *on-shell* mass relations [130–132]. This reconstruction method is also called M_2-Assisted On-Shell, or MAOS [133,134]. Recently, the usage of M_2 at B-factories has been proposed [17] in the context of searches of LFV τ decays into a lepton and invisible particles. The concepts exploited in the study are here adapted and extended for the B decays, where similar kinematic constraints hold.

Since there is generally no analytic form for M_2, the calculation requires constrained numerical minimization. In the study presented here and documented in Ref. [18], we use the C++-based library `YAM2` [135] which allows us to derive the M_2 variables and apply them to our physics case. The package utilizes the sequential quadratic programming (SQP) method implemented in the `NLopt` [136] library.

We target $B\overline{B}$ events that can be expressed as:

$$e^+e^- \to B_1B_2 \to V_1(p_1)\chi_1(k_1) + V_2(p_2)\chi_2(k_2), \tag{5.5}$$

where V_i represents the visible particles with total momentum p_i and χ_i those escaping the detection (with momentum k_i). The event topology is rendered in Fig. 5.14, where $B_{\text{sig}} \equiv B_1$ and $B_{\text{tag}} \equiv B_2$.

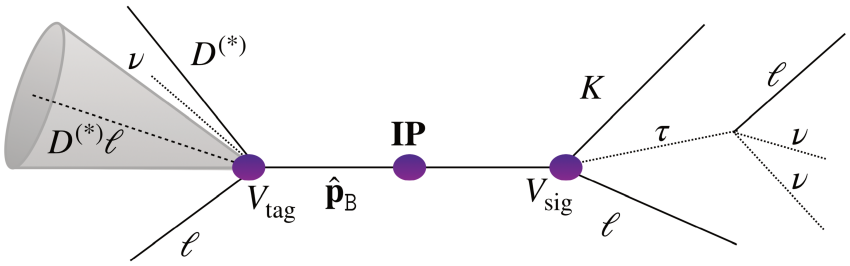


FIGURE 5.14: Schematic view of the $B\overline{B}$ process under study. One of the two B mesons (B_{tag}) decays semileptonically and the other (B_{sig}) in the LFV $K\tau\ell$ channel.

Given that the initial 4-momentum is known, different M_2 variables can be built out of the process of Eq. 5.5 depending on the constraints we decide to introduce for the minimization. A general definition for M_2 is:

$$\begin{aligned} M_{2\mathcal{C}} =& \min_{\mathbf{k}_1,\,\mathbf{k}_2\in\mathbb{R}^3}\Big[\max\Big\{M(p_1,\,k_1),\,M(p_2,\,k_2)\Big\}\Big] \\ &\text{subject to } \begin{cases} \mathbf{k}_1 + \mathbf{k}_2 = \mathbf{P}^{\text{miss}}, \\ \text{more constraints} \to \text{fix } \mathcal{C}\ , \end{cases} \end{aligned} \tag{5.6}$$

where M is the invariant mass constructed from the sum of the momenta. The first constraint in Eq. 5.6 consists in requiring that the total invisible 3-momentum should coincide with the $\mathbf{P}^{\text{miss}} = \mathbf{p}_{e^+e^-} - \mathbf{p}_1 - \mathbf{p}_2$. This condition reduces to three the degrees of freedom (d.o.f.) over which the minimization is performed. Additional constraints (globally described with $\mathcal{C}$) will characterize the corresponding result of the minimization, called $M_{2\mathcal{C}}$. In the next pages two sets of constraints, collectively denoted as M_{2sB} and M_{2V}, are discussed.

The $\mathbf{M_{2sB}}$ solution

A first constraint derives from the knowledge of the total collision energy $\sqrt{s}$:

$$(p_1 + k_1 + p_2 + k_2)^2 = s \tag{5.7}$$

and an additional one using the fact that the parent particle on each 'branch' of the $\Upsilon(4S)$ decay is a B meson:

$$(p_1 + k_1)^2 = (p_2 + k_2)^2 = m_B^2. \tag{5.8}$$

The two constraints together reduce to zero the number of d.o.f in the M_2 minimization. In this situation M_2 is not anymore a distribution with an endpoint (the decaying parent mass) but a solver of the event kinematics with a unique root. We call M_{2sB} the solution obtained combining the Eq. 5.6 with the constraints of Eqs. 5.7-5.8. The

last constraint has never been used before, reasonably due to the fact that the $M_{(T)2}$ variables were created for searches of particles with unknown mass.

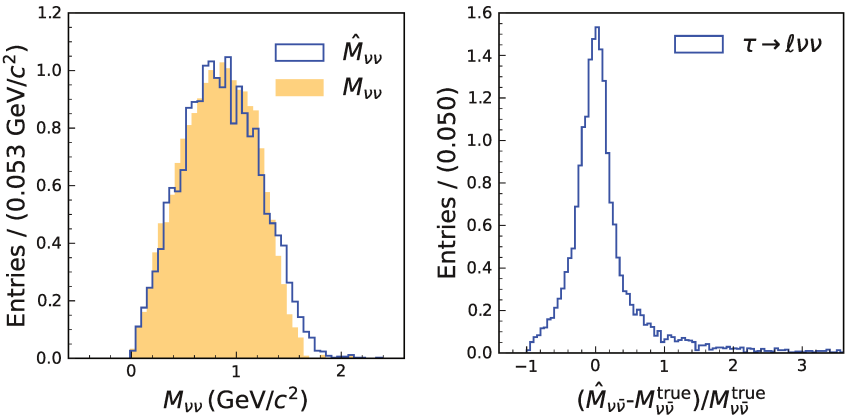


FIGURE 5.15: Distribution for $M_{\nu\nu}^{\text{approx}}$ as defined in the text, and comparisons with the truth-level distribution.

The minimization depends upon the knowledge of the invariant masses of the invisible systems on both signal (k_1^2) and tag side (k_2^2). While $k_2^2 = 0$, k_1^2 is unknown for $\tau^+ \to \ell^+\nu_\ell\bar{\nu}_\tau$ decays as in this case $k_1^2 = M^2(\nu_\ell\bar{\nu}_\tau) = M^2_{\nu\bar{\nu}}$. The simple ansatz $k_1^2 = 0$, often adopted in the literature, is not realistic, as the $M^2_{\nu\bar{\nu}}$ distribution peaks around 1 GeV/c^2 (see left plot of Fig. 5.15).

We then propose an improved ansatz for k_1^2 as follows. We consider the B meson in its rest frame, and boost it with the total $\Upsilon(4S)$ momentum: then k_1 can be obtained by subtracting the sum of the visible final-state 4-momenta from the approximated B 4-momentum. This solution neglects the back-to-back momentum of the $B - \bar{B}$ pair in the center-of-mass frame, which is however small with respect to the boost induced by the beams asymmetry. As a result of our ansatz, the k_1^2 distribution is close to the truth-level one for about 86% of the events, and yields negative values for the remaining 14%. For these events, we change the k_1^2 sign. The comparison between the true $M_{\nu\bar{\nu}}$ and our ansatz (denoted as $\hat{M}_{\nu\bar{\nu}}$) is shown in Fig. 5.15 (left) while the 'pull' of the two is on right plot.

We apply the $\hat{M}_{\nu\bar{\nu}}$ in $M_{2s(B)}$ and test the effect on M_{recoil}, in comparison with the two cases where either the naive hypothesis $M_{\nu\bar{\nu}} = 0$ or the true distribution are used. In Fig. 5.16 we show these M_2 definitions, in addition to the random-cos θ distribution as a reference, for $\tau \to \ell\nu\bar{\nu}$ events. Here the true kinematics of the B-daughters is used, in order to focus only on the effect on the $M_{\nu\bar{\nu}}$ hypotheses. One can see that the improvement in resolution due to $\hat{M}_{\nu\bar{\nu}}$ with respect to zero is very slight for M_{2sB} but corrects the bias appearing for the $M_{\nu\bar{\nu}} = 0$ ansatz. In brief, M_{2sB} displays the largest sensitivity to the accuracy of the $M_{\nu\bar{\nu}}$ ansatz. This sensitivity seems to be mainly due to the 'B' constraint because the same effect is not observed for the M_{2s} case. Nonetheless, the M_{2sB} is also the variable showing the largest improvement in resolution when comes to hadronic τ decays, for which no approximation on $M_{\nu\bar{\nu}}$ is needed as $k_1^2 = 0$.

When moving to the reconstructed information to take into account the detector smearing affecting the $\boldsymbol{p}_i$ and $\boldsymbol{P}^{\text{miss}}$ inputs (see Eq. 5.5) – not only the true momenta are replaced with those measured by Belle tracking system, but also the τ decay mode is assigned based on the particle-ID criteria (see Sec. 3.5), resulting in self-cross feed due to fake rates. Furthermore, in the current reconstruction an inclusive $\tau \to 1$ prong selection is adopted and the $\rho \to \pi^+\pi^0$ is not reconstructed, making impossible to test

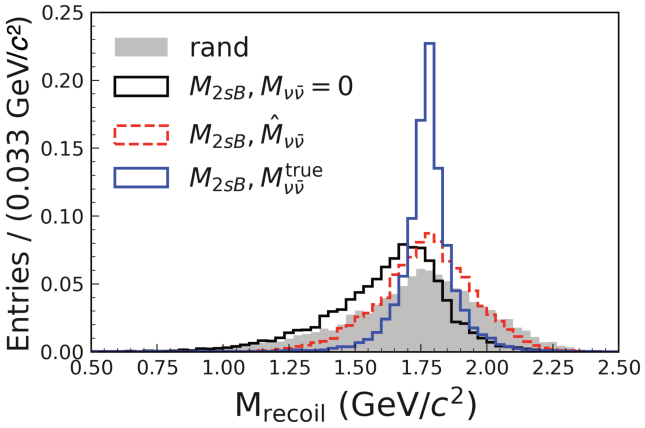


FIGURE 5.16: M_{recoil} distributions for M_{2sB} as a function of $M_{\nu\bar{\nu}} = \{M^{\text{true}}_{\nu\bar{\nu}}, \hat{M}_{\nu\bar{\nu}}, 0\}$. The random-$\cos\theta$ distribution is also shown as reference.

the M_{2sB} constraints. Hence we decide to keep using the true information for the τ-prong in order not to discard the $\tau \to \rho\nu$ events.

In the plot of Fig. 5.17 we show the fit to the M_{recoil} of signal events obtained with the M_{2sB} constraints. As can be seen by eye and from the fit values listed in Tab. 5.6, the resolution obtained under the assumptions explained above is a factor ~ 2 better than the $M^{\text{rand}}_{\text{recoil}}$[5]. This shows that the exclusive reconstruction of the ρ and the proper separation of leptonic and hadronic modes are crucial to improve our sensitivity. For the hadronically tagged analysis the same kinematic constraints hold and one could use the additional information on the τ-decay modes at some stage of the selection and not necessarily as a constraint to improve the M_{recoil} resolution.

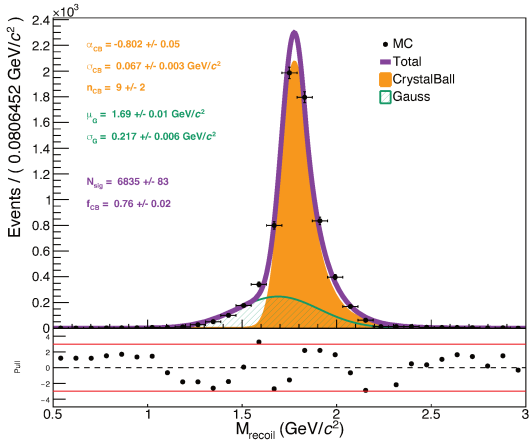


FIGURE 5.17: Fit to M_{recoil} (M_{2sB}) corresponding to $B^+ \to K^+\tau^+\mu^-$ MC events tagged through semileptonic decay modes.

Parameter	Fit result
f_{CB}	0.76 ± 0.02
σ_{CB} (MeV/c^2)	67 ± 3
α	-0.067 ± 0.003
n	9 ± 2
μ_{Gauss} (GeV/c^2)	1.69 ± 0.01
σ_{Gauss} (MeV/c^2)	287 ± 6

TABLE 5.6: Fit parameters for the M_{recoil} (M_{2sB}) of $B^+ \to K^+\tau^+\mu^-$ events.

The M_{2V} solution

In addition to the 'classic' kinematic constraints, we try to exploit the vertex information of the B-parents to infer their 3-momenta[6].

[5] Obtained with the random $\cos\theta$ term, $\cos\theta_{\text{rand}}$.

[6] In the CM frame, $\boldsymbol{p}^*_{B_{\text{tag}}} = -\boldsymbol{p}^*_{B_{\text{sig}}}$ and the boost from the lab to the CM frames is known; hence only one of the two $\mathbf{p}_B$ is needed.

The idea of using the vertex information for decays with missing energy is not completely new to B-factories: at *BABAR*, for example, the B^0 lifetime is obtained by tagging $B^0 \to D^{*-}\ell^+\nu_\ell$ decays and then looking at $\overline{B}^0$'s in the same channel on the signal side. Both B-decay vertices are measured with a fit to the prompt lepton and the *soft* pion coming from the D^* [137]. At Belle [138] the τ lifetime has been measured using $\tau^+\tau^-$ decays where both leptons decay into $3\pi\nu_\tau$. The direction $\mathbf{p}_\tau$ is not known *a priori* because of the undetected neutrinos but there are a few constraints that can reduce the degrees of freedom:

(i) The direction of the τ in the CM frame is constrained to lie on the cone of aperture $\cos\theta^* = \frac{2E^*_\tau E^*_X - m^2_\tau c^4 - m^2_X c^4}{2p^*_X \sqrt{E^{*2}_\tau - m^2_\tau c^2}}$ $(X \equiv 3\pi)$.

(ii) The two τ leptons are back-to-back in the CM frame.

(iii) Conditions (i) and (ii) provide two solutions for $\vec{\mathbf{p}}^*_\tau$; the average of the two is taken as estimate of the τ direction.

(iv) The X system provides a vertex and the direction of the τ (boosted into the lab frame) is constrained to pass through it.

(v) The two directions in the lab frame should point to a common IP position, but because of resolution, they might not intersect: the distance of closest approach provides the production vertex positions that are then used to compute the flight distance of the τ's.

The same condition (i) holds for our B_{tag} side (see Eq. 5.4) and in principle also for the B_{sig} side, in case we replace the m_ν with the m_τ constraint. However, using the condition (i) for the B_{sig} would prevent us from using the m_τ as the variable to extract the signal. Like in the point (iv), both B_{tag} and B_{sig} can provide decay vertices, the first from a fit to the $D^{(*)}(\pi)\ell$ system and the second from the $(K_{\text{sig}}, \mu_{\text{sig}})$ pair. An additional piece of information is given by the knowledge of the IP position, represented as a blob in Fig. 5.14 together with those for the two B-decay vertices.

The B_{sig} flight direction can be determined as $\hat{\boldsymbol{v}}_{\text{sig}} = (\boldsymbol{r}_{\text{sig}} - \boldsymbol{r}_0)/|\boldsymbol{r}_{\text{sig}} - \boldsymbol{r}_0|$, where $\boldsymbol{r}_0$ is the production vertex of the B and $\boldsymbol{r}_{\text{sig}}$ is the B_{sig}-decay vertex. As both $\boldsymbol{r}_{\text{sig}}$ and $\boldsymbol{r}_0$ are measured with a finite resolution, the $\hat{\boldsymbol{v}}_{\text{sig}}$ constraint should be implemented as:

$$\arccos(\hat{\boldsymbol{p}}_{B_{\text{sig}}} \cdot \hat{\boldsymbol{v}}_{\text{sig}}) \leq \delta\varphi_{\text{sig}} \ . \tag{5.9}$$

Here $\hat{\boldsymbol{p}}_{B_{\text{sig}}}$ denotes the unit vector corresponding to $\boldsymbol{p}_{B_{\text{sig}}} = \boldsymbol{p}_{\text{sig}} + \mathbf{k}_{\text{sig}}$, with $\mathbf{k}_{\text{sig}}$ estimated through M_2. This constraint expresses the fact that the B_{sig} direction forms a cone of maximal aperture $2\delta\varphi_{\text{sig}}$ with $\hat{\boldsymbol{v}}_{\text{sig}}$. As the uncertainty on $\boldsymbol{x}_{IP}$ tends to zero, so does $\delta\varphi_{\text{sig}}$. The constraint in Eq. 5.9 under the form of inequality can be implemented with the SQP method in the `YAM2` software. To simplify, we assume that the secondary vertex $\boldsymbol{r}_{\text{sig}}$ is determined accurately (resolution on the vertex coordinates are of the order of 45 μm at Belle and 30 μm at Belle II – see Fig. 5.18), and so are $(\boldsymbol{r}_0)_{x,y}$ because of their small size especially at Belle II.

At Belle, the beamspot[7] spread in the z-direction is $\sigma_z^{\mathbf{r}_0} \sim 5$ mm, which makes the $\hat{\boldsymbol{v}}_{\text{sig}}$ constraint ineffective, as the B-decay length is more than 10 times smaller than $\sigma_z^{\mathbf{r}_0}$. However, the current $\sigma_z^{\mathbf{r}_0}$ at Belle II is $\sim$350 μm and is expected to further decrease to $\sim$150 μm at the final stages of the operations [66]. The simulated beamspot size at Belle II can be seen in Fig. 5.19.

[7] It is defined as the probability distribution function of the primary interaction vertices, or the beams' overlap 'luminous' region at the IP.

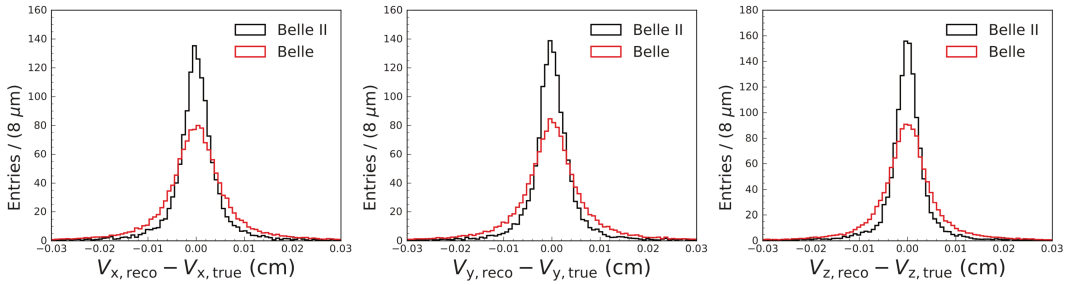


FIGURE 5.18: B_{sig} vertex resolution on the x, y, z coordinates according to official Belle and Belle II $B \to K\tau\ell$ MC.

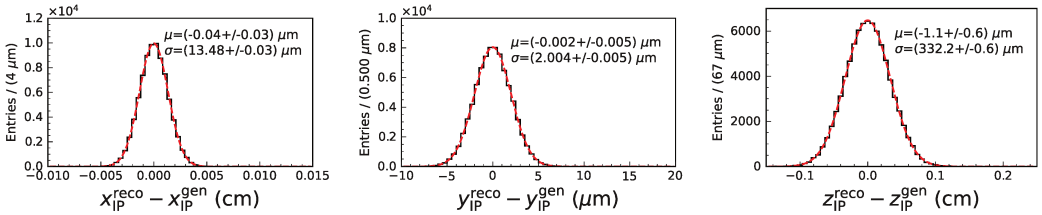


FIGURE 5.19: IP resolution on the x, y, z coordinates according to Belle II $B \to K\tau\ell$ MC.

The effective longitudinal size of the interaction region is the result of the fact that collisions at Belle(II) are not head-on but happen with a non-zero crossing angle $2\phi_x$ (see Tab. 5.7). For this reason, the size of the IP along z, denoted as σ_z^*, is not given by the bunch size σ_z but by the relation:

$$\sigma_z^* = \frac{\sigma_x^*}{\sin 2\phi_x} \simeq \frac{\sqrt{\epsilon_{x,H}\beta_{x,H}^* + \epsilon_{x,L}\beta_{x,L}^*}}{2\phi_x} , \tag{5.10}$$

which can be explained by the not-to-scale rendering given in Fig. 5.20. As usual the subscripts 'H' and 'L' denote the HER and LER rings (electron and positron beams, respectively) and the beam size σ_i^* along the i-axis is defined as $\sqrt{\varepsilon_i^* \times \beta_i^*}$, where ε is the emittance and β the amplitude function.

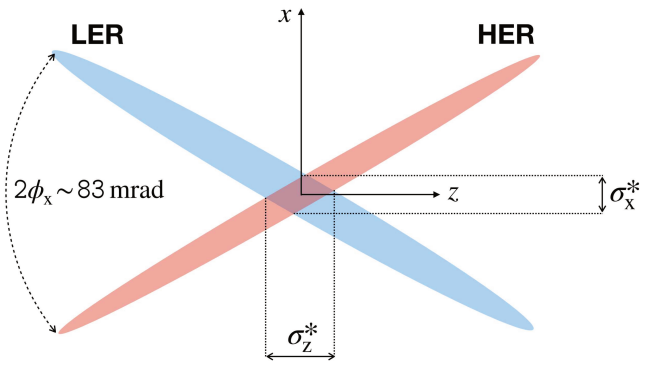


FIGURE 5.20: Pictorial representation and not to scale of the crossing of the two beams at SuperKEKB. The sizes along x and z axis of the overlap region - the *beamspot* - are shown. In particular, the σ_x^* is in relation with the crossing angle according to the Eq. 5.10.

The above considerations suggest that $\hat{\boldsymbol{v}}_{sig}$ may be simply imposed as an equality constraint, rather than an inequality, as in Eq. 5.9: detector effects will smear negligibly $\boldsymbol{r}_{sig}$ and $(\boldsymbol{r}_0)_{x,y}$ around their respective true values, whereas $(\boldsymbol{r}_0)_z$ can be

	KEKB		**SuperKEKB**	
	HER	LER	HER	LER
σ_y^*	2 μm		240(60) nm	
ε_x (nm)	18		4.6(4.6)	4.0(3.2)
β_x^* (mm)	330		60(25)	80(32)
$2\phi_x$(mrad)	22		83	
σ_z^*	5 mm		300(180) μm	
$\beta\gamma$	0.42		0.28	

TABLE 5.7: Design values at KEKB and the achieved (target) ones at SuperKEKB for the main accelerator parameters. From Ref. [68] and Ref. [66], respectively.

estimated as a normally distributed random number with a standard deviation equal to σ_z. We will denote the thus-obtained equality constraint on the $B_{\rm sig}$ flight direction as 'V'. The M_2 definition of Eq. 5.6, plus the equality V constraint just discussed, lead to the variable M_{2V}; to our knowledge, it has never been considered before. M_{2V} reduces the number of d.o.f. available in the M_2 minimization by two units as the constraint fixes one direction in the 3-dimensional space. Note that, if we further use the knowledge of s, we can fix not only the orientation, but also the magnitude of $\mathbf{p}_{B_{\rm sig}}$, and thereby 'close' the kinematics. In this case the minimization must necessarily land at $\mathbf{k}_1 = \mathbf{k}_1^{\rm true}$ and to the M_2 solution called M_{2sV}. On the other hand, if one does not use the s information, as in M_{2V}, s is an outcome of the M_2 algorithm.

We apply this procedure for M_{2sB} as well as M_{2sV} with $\sigma_z^{\mathbf{r}_0} = 150\,\mu$m and in all cases we use the $\hat{M}_{\nu\bar{\nu}}$ ansatz. We also try different resolution scenarios for the $B_{\rm sig}$ vertex and find that only for excellent vertex resolutions a similar performance can be attained with M_{2V} compared to M_{2sB}. A proof of this is provided in Fig. 5.21, where M_{2sB}, M_{2sV} and M_{2V} are presented. The three plots correspond to the τ hadronic, leptonic, and all decay modes, respectively. One can see that M_{2sB} provides the best resolution for $\tau \to h\nu$ modes while M_{2V} provides a better result for $\tau \to \ell\nu\bar{\nu}$ decays, with the result of having similar performance with the two modes combined (third plot). However, it should be stressed that $\sigma_{x,y,z}^{\mathbf{r}_{\rm sig}} \sim 5\,\mu$m is unrealistic and reflects the fact that extreme good resolution is needed to separate the production and decay B-vertices which are only $\sim 100\,\mu$m apart. A way to exploit even further the knowledge of the system is to use both B-vertices positions and the fact that in the CM frame the two B mesons are back to back, as was done at Belle for the τ lifetime. However, this could not be implemented due to time constraints.

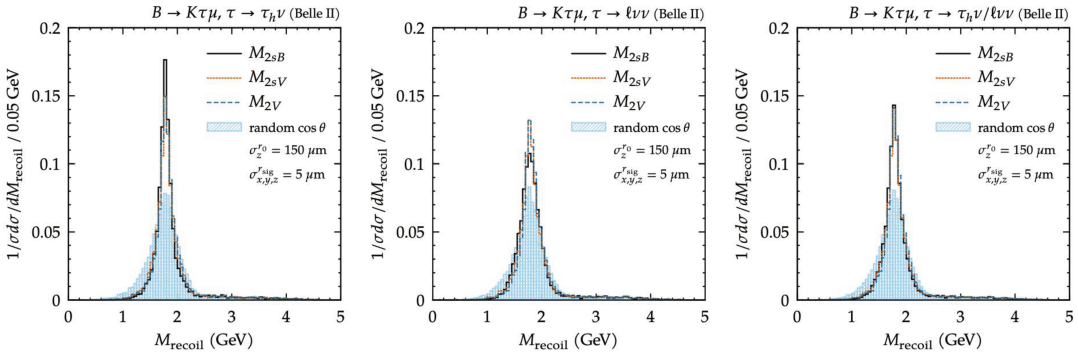


FIGURE 5.21: $M_{\rm recoil}$ distributions for Belle II with $\sigma_z = 150$ μm and for the different decay modes of the signal τ (leftmost to rightmost columns: hadronic, leptonic τ decays, or both).

5.5.1 Obtained sensitivities

The findings presented in the last two paragraphs show that M_{2sB} and M_{2sV} have distinct strengths and weaknesses:

- The former has a strong sensitivity to the $M_{\nu\bar{\nu}}$ hypothesis but shows great performance for hadronic τ decays and can be used at Belle.

- The second weakly depends on $M_{\nu\bar{\nu}}$ and allows using the vertexing information at Belle II, although the current vertex resolution and target beasmpot size do not improve the M_{recoil} resolution sufficiently to make it competitive with M_{2sB}.

These considerations advise to prefer M_{2sB} for Belle data and set M_{2sV} aside for now. So far we have applied constraints to M_2 of Eq. (5.6) in the form of equalities, in which case the total number of independent relations should not exceed the total number of kinematic d.o.f. available. One may exceed this number by implementing constraints as inequalities instead, thus 'overloading' M_2. This is useful when several complementary pieces of information are available, with each of them not accurate enough to dominate the algorithm outcome, so that they improve the algorithm outcome only when used concurrently. In this context, the M_{2V} approach could be revised to incorporate further information from the tag-side and exploit the properties of the Belle II environment. We leave this possibility to next iterations.

In this manuscript we focus only on the two following resolution scenarios:

- $\cos\theta_{\text{TRUE}}$: the true $\mathbf{p}_B^*$ is used. This certainly is an extreme condition but not very far for the hadronic B-tagging case: as shown in Fig. 5.1B the resolution on $\cos\theta$ is very good for hadronically tagged candidates.

- M_{2sB} as described in the previous paragraphs.

In order to set an UL, we check whether the minimization algorithm, fed with the background events where the M_{recoil} is the result of a random combination of particles, would return biased values or even spurious peaks. As shown in Fig. 5.22, this is not the case: the smooth distribution is preserved for different configurations and both before and after the background suppression; hence we feel confident we can use the background as defined in Eq. 5.1 while replacing the signal events with the MAOS values.

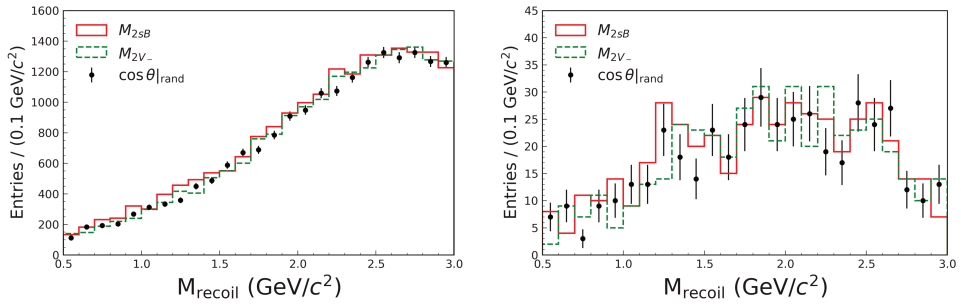


FIGURE 5.22: Output of MAOS from the M_{recoil} obtained from background events before (left) and after (right) the BDT cut at 0.9.

In Fig. 5.23 we show the M_{recoil} distribution for the same background events shown of Fig. 5.12 and the signal events related to the two configurations M_{2sB} (left) and

$\cos\theta_{\rm TRUE}$ (right). In fact we opt not to optimize the BDT cut for each plot but instead provide the resulting UL derived from the mere improvement in resolution.

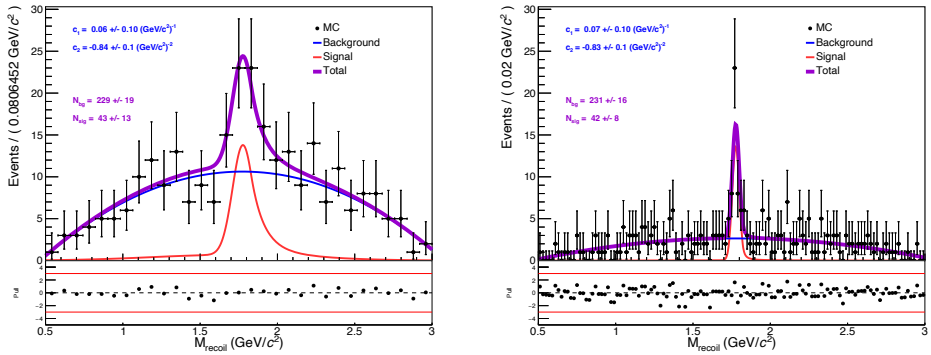


FIGURE 5.23: Fit to MC data surviving the BDT cut at 0.9. The signal events, re-scaled to the BR$= 5 \times 10^{-5}$, are obtained with the M_{2sB} constraints (left) and with the true $\mathbf{p}^*_B$ (right). In both cases the signal is fitted with fixed shapes and the background is modeled with a 2*nd*-order polynomial.

The UL at 90% C.L. for SL-tagged events in the $\mathbf{OS}_\mu$ mode are provided in Tab. 5.8 for $\cos\theta_{\rm TRUE}$ and $\cos\theta_{\rm RAND}$ (the two 'extreme' scenarios), and M_{2sB}, which gives an estimation of the improvement brought by the constraints discussed in this section. It is surprising, and also promising, to see that the limit obtained with M_{2sB} (1.2×10^{-5}) is the same than the one with the hadronic B-tagging for the zero signal case ($\cos\theta_{\rm had}$). The limits corresponding to $M^{\rm rand}_{\rm recoil}(\cos\theta_{\rm rand})$ and with $M^{\rm true}_{\rm recoil}$ ($\cos\theta_{\rm true}$) are respectively 2.0 and 0.6, both in units of 10^{-5}. The expected ULs on the BF of the $B^+ \to K^+\tau^+\mu^-$ mode are summarized in the plot of Fig. 5.24: the circle corresponds to the hadronic B-tagging and the triangles are for the semileptonic B-tagging (obtained using $\cos\theta_{\rm rand}$, M_{2sB} or $\theta_{\rm true}$).

Mode	$N^{\rm UL}_{\rm sig}$	$\mathcal{B}^{\rm UL}(\times 10^{-5})$
$\cos\theta_{\rm true}$	5.7	0.6
$\cos\theta_{\rm rand}$	20.8	2.0
M_{2sB}	12.9	1.2

TABLE 5.8: ULs for the SL-tagged $\mathbf{OS}_\mu$ mode. $\cos\theta_{\rm true}$ and $\cos\theta_{\rm rand}$ represent the two extreme scenarios, while M_{2sB} gives an estimation of the improvement brought by the constraints discussed in Sec. 5.5.

In this chapter we have provided the first estimation of the sensitivity for the $B^+ \to K^+\tau^+\mu^-$ mode with semileptonic B-tagging. The UL on the BF is obtained with the full Belle MC background and using the $M_{\rm recoil}$ definition of Eq. 5.1. This scenario with realistic background level and $M_{\rm recoil}$ resolution without any constraint can be seen as the most pessimistic result. To improve the resolution on $M_{\rm recoil}$, we have followed two paths: one exploits the kinematic information and leads to the M_{2sB} result – for now the most promising. The second path, based on the vertexing information, does not bring enough constraints and will need more work. The M_{2sB} scenario allows to obtain a better resolution which, combined with the higher efficiency, makes the semileptonic B-tagging competitive with the hadronic one. The fact that we can access an orthogonal sample to the hadronic one and obtain a similar sensitivity is very encouraging, the semileptonically tagged data can be used to get an independent measurement in case a signal is observed with the hadronic tags.

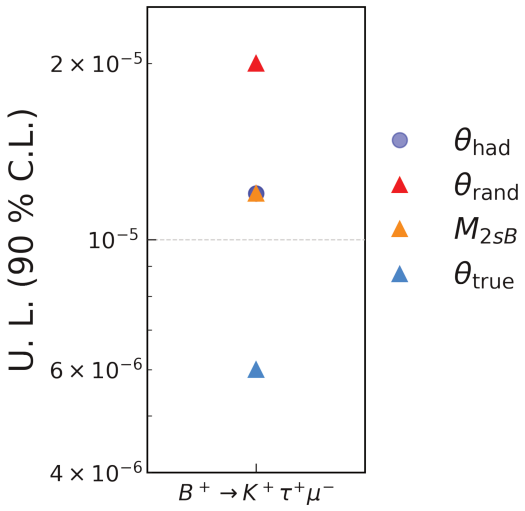


FIGURE 5.24: UL at 90% C.L. for the zero signal case. The limit obtained with hadronic$B$-tagging (blue circle) is compared with those obtained with semileptonic B-tagging (triangles), for which the three configurations are reported.

Chapter 6

Conclusion and outlook

This thesis has covered three main topics related to the search of the lepton-flavor-violating $B^+ \to K^+\tau\ell$ decays at Belle and Belle II. In the following sections we detail the obtained results and provide suggestions for improvements and directions for future developments.

$B^+ \to K^+\tau\ell$ search with Belle data and hadronic B-tagging

We have searched for the four $B^\pm \to K^\pm\tau^\pm\ell^\mp$ ($\ell = \{e, \mu\}$) decay modes with the full data sample collected by the Belle experiment at the $\Upsilon(4S)$ resonance and corresponding to 772 million $B\bar{B}$ pairs. This measurement has never been performed at Belle nor Belle II, while the *BABAR* and LHC*b* experiments previously set some ULs on the rates of such modes. At B-factories, the standard strategy for modes with missing energy exploits B-tagging, i.e., the full reconstruction of the other B in the event. This approach not only helps to reduce the background but – in the context of our search – also allows us to compute the recoil mass of the $B^\pm_{\text{tag}}K^\mp\ell$-system ($M_{\text{recoil}}$, or m_τ) which is the observable used to extract the signal yield. We use the hadronic B-meson reconstruction provided by the `FEI` algorithm, the official exclusive B-tagging package at Belle II, for which a conversion of the Belle data format is needed. The background nature depends on the charge configuration of the signal particles: 'Opposite sign' for $B^+ \to K^+\tau^+\ell^-$ modes and 'Same sign' for $B^+ \to K^+\tau^-\ell^+$ modes. In both cases, Cabibbo-favored $B^+ \to \overline{D}^0(\to K^+X^-)X^+$ decays are the dominant sources of $B\overline{B}$ background, which is reduced employing a multivariate analysis. We exploit kinematic variables, the information from the calorimeter and vertexing to suppress these background events. The $q\bar{q}$ component is also removed at a later stage, using the event shape properties.

As we find no evidence of signal for any of the four modes, we set the following ULs at 90% C.L. on their branching fractions:

$$\begin{aligned}
\mathcal{B}(B^+ \to K^+\tau^+\mu^-) &< 0.65 \times 10^{-5} \\
\mathcal{B}(B^+ \to K^+\tau^+e^-) &< 1.71 \times 10^{-5} \\
\mathcal{B}(B^+ \to K^+\tau^-\mu^+) &< 2.97 \times 10^{-5} \\
\mathcal{B}(B^+ \to K^+\tau^-e^+) &< 2.08 \times 10^{-5}
\end{aligned}$$

which incorporate the systematic uncertainties. These results, except for the $B^+ \to K^+\tau^+e^-$ mode, are the most stringent to date and provide new constraints to the BSM models related to the LFV and the B-anomalies – especially for the $B^+ \to K^+\tau^+\mu^-$ that has entered the 10^{-6} region. This analysis, described in detail in chapter 3, is currently at the final stage of internal review and aims for publication.

Recently the LHC*b* collaboration has shown a new result for the neutral modes $B^0 \to K^{*0}\tau^\pm\mu^\mp$ with the full dataset ($9\,\text{fb}^{-1}$), providing the first ULs around $1.0 \times$

10^{-5} [139]. Although it is encouraging to see that our results, obtained with a data sample collected more than 10 years ago, are still relevant and competitive with LHCb's, it is imperative to improve our strategy and find new, complementary channels that can contribute to probe the $b \to s\tau\ell$ transitions. For example, the promising $B^0 \to K^0_S\tau\ell$, where we can exploit the high purity of K^0_S and the experience gained with the hadronic tag.

The current limitation on our measurement comes from the efficiency, drastically reduced by the hadronic B-tagging. Therefore, we pursue an in-depth study of the `FEI` performance to exploit the available data better. This work is timely considering that the first long Belle II shutdown is ongoing: we have the opportunity to understand better our hadronic B-tagging algorithm before more data come.

Hadronic B-tagging improvements

The main result of chapter 4 is the improvement of the MC description of B^+ decays. A significant discrepancy with data in the `FEI` performance suggests important mismodelings in our simulation. A careful interpretation of the experimental results, combined with theory's prescriptions, allowed us to correct many of the $D^{(*)}n\pi$ modes, responsible for 90% of the `FEI` efficiency. Major corrections were needed for the $\overline{D}^{(*)0}\pi^+\pi^+\pi^-$ modes in terms of 3π-invariant mass shape and $\overline{D}^{(*)0}\pi^+\pi^+\pi^-\pi^0$ modes which were largely overestimated in the MC. The proposed updates to the simulation have been verified with a $B^+ \to J/\psi\, K^+$ sample, which allows to obtain a pure sample of B's where we can study their properties through the candidates passing the `FEI` selection. Besides reducing the MC/data discrepancy in terms of efficiency, this work on the MC is expected to be beneficial for the `FEI` purity. The `FEI` consists of cuts depending on BDT classifiers that are trained on MC. Since our simulation wrongly describes the B-modes contributing to the B-tagging, the background rejection is non-optimal. A new training of `FEI` based on the updated MC will soon show the expected improvements.

We also provide additional directions that are worth exploring for a more efficient `FEI`:

1. Revise the current selection and incorporate the physics knowledge into the algorithm (for example, exploiting the presence of intermediate resonances);
2. Revise the set of training variables that can introduce undesired dependencies (e.g. ΔE);
3. Increase the efficiency by explicitly adding new decay chains (e.g. $B^+ \to D^{*-}(4\pi)^{++}$, $B^+ \to \overline{D}^{(*)0}K^+K^{(*)0}$) or, for example, allowing for an optimized ΔE selection to recover partially mis-reconstructed B_{tag}'s.

The same work applied on the B^+ Belle MC is ongoing for the $B^0 \to D^{(*)-}n\pi$ `FEI` modes. In parallel, the Belle II simulation is also under study, as it developed independently from Belle's. The shared strategy employs the $B \to D^{(*)}\pi$ inclusive samples because of the larger statistics and can be used as a validation tool to compare the `FEI` performance for different versions of the software (and training) at Belle II. The obtained improvements on the simulation will be beneficial for the search of $B^0 \to K^0_S\tau\ell$ decays and many tagged analyses at Belle II.

$B^+ \to K^+\tau^+\mu^-$ search with Belle data and semileptonic B-tagging

We explore the possibility of using a semileptonic tag approach for our search. This method has never been tried for the $B \to K\tau\ell$ modes and has two main advantages.

The first is the higher reconstruction efficiency due to the large BFs of $B \to D^{(*)}\ell\nu$ decays. Secondly, only a few modes are used for the B-tagging and they are well known, making the MC description more reliable and the `FEI` performance closer to data. However, as opposed to the hadronic tag, the momentum of the B meson cannot be measured; hence, the resolution on the M_{recoil} is roughly a factor 5 worse. In chapter 5 we study how these two effects combined – higher efficiency and worse signal resolution – impact the final sensitivity. We have focused on the $B^+ \to K^+\tau^+\mu^-$ mode to get a realistic estimate of the background level from MC with a complete selection procedure similar to the one applied in chapter 3, and obtain an expected UL on the BF equal to 2.0×10^{-5}. We also try to apply additional constraints to improve the resolution on M_{recoil}. We demonstrate that the knowledge of the event kinematics improves the resolution of M_{recoil}, especially for hadronic two-body τ decays, leading to an UL of 1.2×10^{-5}. In this context, it would be beneficial to have a better control on the $\tau \to \rho\nu$ mode and perform a simultaneous fit to leptonic and hadronic τ events in order to incorporate the knowledge on the different background nature and resolution during the signal extraction.

Additional constraints can be imposed based on the vertexing information and different resolution scenarios related to Belle II and the upgraded collider SuperKEKB, exploiting the reduced beamspot size and the excellent vertexing performance due to PXD. However, in the current implementation of the constraints, this piece of information does not bring a significant gain. This study is documented in Ref. [18].

The results shown are encouraging because they suggest that it is possible to consider semileptonic B-tagging as a complementary method for the $B \to K\tau\ell$ search, allowing us to exploit an independent data sample in addition to the hadronic tags. This point would be crucial to confirm the presence of signal for any significance level.

Appendix A

q^2 parametrization for LFV scenarios

We study the differential decay rate for LFV $\bar{B} \to \bar{K}\ell_1^-\ell_2^+$ decays as described in Ref. [12]. We focus on the cases where one of the two leptons is a τ and the other is $\ell = e, \mu$. In this context, two groups of decays are possible: (i) $\mathbf{OS}_\ell$ where the kaon and the light lepton have opposite charge signs, i.e. $(K^\pm\ell^\mp\tau^\pm)$ and (ii) $\mathbf{SS}_\ell$ where the kaon and the light lepton have same charge signs, i.e. $(K^\pm\ell^\pm\tau^\mp)$. Simple expressions can be obtained if we use the fact that one of the lepton masses is much larger than the other:

$$\begin{aligned} SS:&\ m_2 = m_\tau \gg m_\ell = m_1\,, \\ OS:&\ m_1 = m_\tau \gg m_\ell = m_2\,. \end{aligned}$$

In this case, the phase-space functions assume a simpler form,

$$\begin{aligned} \varphi_9(q^2) = \varphi_{10}(q^2) = \varphi_V(q^2) &= \frac{1}{2}|f_0(q^2)|^2 m_\tau^2 \frac{(m_B^2 - m_K^2)^2}{q^2}\left(1 - \frac{m_\tau^2}{q^2}\right) \\ &\quad + \frac{1}{2}|f_+(q^2)|^2 \lambda(m_B, m_K, \sqrt{q^2})\left[1 - \frac{m_\tau^2}{q^2} - \frac{(q^2 - m_\tau^2)^2}{3q^4}\right], \\ \varphi_S(q^2) = \varphi_P(q^2) = \varphi_S(q^2) &= \frac{q^2|f_0(q^2)|^2}{2(m_b - m_s)^2}(m_B^2 - m_K^2)^2\left(1 - \frac{m_\tau^2}{q^2}\right), \\ \varphi_{10P}(q^2) = \pm\varphi_{9S}(q^2) = 2\varphi_I(q^2) &= \frac{|f_0(q^2)|^2}{m_b - m_s} m_\tau (m_B^2 - m_K^2)^2\left(1 - \frac{m_\tau^2}{q^2}\right). \end{aligned} \tag{A.1}$$

The hadronic form factors $f_{+,0}(q^2)$ (Fig. A.1) are related to the matrix element

$$\langle \bar{K}(k)|\bar{s}\gamma_\mu b|\bar{B}(p)\rangle = \left[(p+k)_\mu - \frac{m_B^2 - m_K^2}{q^2}q_\mu\right] f_+(q^2) + \frac{m_B^2 - m_K^2}{q^2} q_\mu f_0(q^2)\,. \tag{A.2}$$

It is interesting to note that $\varphi_I(q^2) = \varphi_S(q^2)\, q^2/[2m_\tau(m_b - m_s)]$ and that the interference term φ_{9S} changes sign depending on the charge configuration because of the presence of the linear term $(m_1 - m_2)$. Moreover, we have ignored the electromagnetic operator $\mathcal{O}_7$ as its contribution is negligible when considering the LF(U)V effects [4].

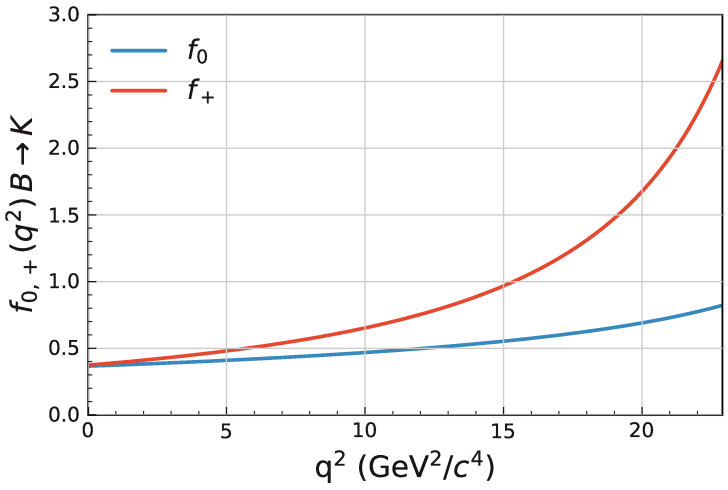


FIGURE A.1: Relevant $B^+ \to K^+$ form factors for LFV $B^+ \to K^+\tau\ell$ decays.

Under these assumptions, the differential decay rates can be written as:

$$\begin{aligned}\frac{\mathrm{d}\mathcal{B}}{\mathrm{d}q^2}(\bar{B} \to \bar{K}\ell_1^{\mp}\ell_2^{\pm}) = |\mathcal{N}_K(q^2)|^2 \times \Big\{ & \varphi_V(q^2)\left[|C_9 + C_9'|^2 + |C_{10} + C_{10}'|^2\right] \\ & + \varphi_S(q^2)\left[|C_S + C_S'|^2 + |C_P + C_P'|^2\right] \\ & + 2\varphi_I(q^2)\mathrm{Re}[(C_{10} + C_{10}')(C_P + C_P')^* \pm (C_9 + C_9')(C_S + C_S')^*]\Big\},\end{aligned} \tag{A.3}$$

where the normalization factor in Eq. (A.3) reads

$$|\mathcal{N}_K(q^2)|^2 = \tau_{B_d}\frac{\alpha^2 G_F^2 |V_{tb}V_{ts}^*|^2}{512\pi^5 m_B^3}\frac{\lambda^{1/2}(\sqrt{q^2}, m_1, m_2)}{q^2}\lambda^{1/2}(\sqrt{q^2}, m_B, m_K). \tag{A.4}$$

The $|\mathcal{N}_K(q^2)|^2$ function for $m_1 = m_\tau$ is shown in Fig. A.2.

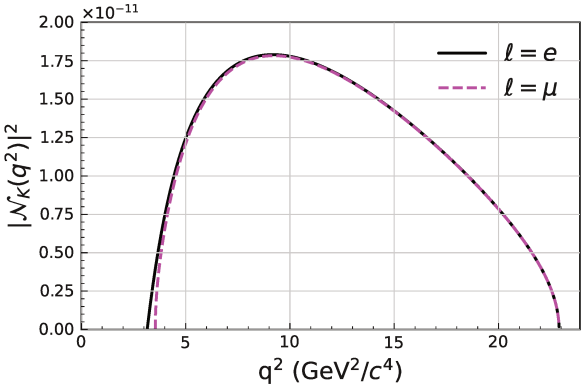


FIGURE A.2: Normalization function $|\mathcal{N}_K(q^2)|$ as defined in Eq. A.4.

We can re-define the linear combinations of the WCs in the form:

$$\begin{aligned}
\mathcal{K} &= |C_9 + C'_9|^2 + |C_{10} + C'_{10}|^2 + |C_S + C'_S|^2 + |C_P + C'_P|^2 \\
r_V &= \frac{|C_9 + C'_9|^2 + |C_{10} + C'_{10}|^2}{\mathcal{K}} \\
r_S &= \frac{|C_S + C'_S|^2 + |C_P + C'_P|^2}{\mathcal{K}} \\
r_I^{\pm} &= \frac{2Re\left[(C_{10} + C'_{10})(C_P + C'_P)^* \pm (C_9 + C'_9)(C_S + C'_S)^*\right]}{\mathcal{K}}
\end{aligned} \tag{A.5}$$

where the following conditions hold

$$\begin{cases} r_V + r_S = 1 \\ r_I^{\pm} \leq 1 \\ r_I^{\pm} \leftrightarrow OS(SS). \end{cases} \tag{A.6}$$

With the assumptions detailed above, the Eq. A.3 reads:

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{B}}{\mathrm{d}q^2}(\bar{B} \to \bar{K}\ell_1^-\ell_2^+) &= |\mathcal{N}_K(q^2)|^2 \cdot \mathcal{K} \times \left\{ r_V \varphi_V(q^2) + r_S \varphi_S(q^2) + r_I^{\pm} \varphi_I(q^2) \right\} \\
&= \mathcal{K} \times \left\{ r_V \phi_V(q^2) + r_S \phi_S(q^2) + r_I^{\pm} \phi_I(q^2) \right\},
\end{aligned} \tag{A.7}$$

where $\phi_i(q^2) = |\mathcal{N}_K(q^2)|^2 \varphi_i(q^2)$. The decay rate is therefore expressed as a linear combination of only three functions (shown in Fig. A.3).

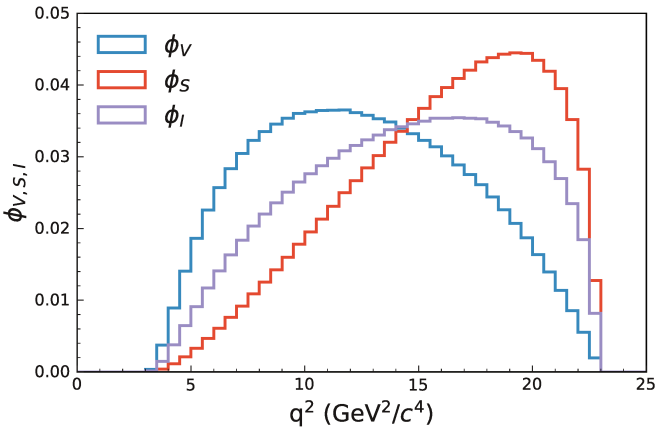


FIGURE A.3: PHSP functions (normalized to the unit area) describing the $B \to K\tau\ell$ transitions (the mass of the light lepton is neglected).

We summarize below a few well-motivated BSM scenarios (explaining the 2021 data) that could explain the discrepancies observed in $R_{K^{(*)}}$ and potentially contribute to the LFV transition $b \to s\ell_1\ell_2$ as well.

- $Z' \sim (\mathbf{1}, \mathbf{1}, 0)$: One of the simplest scenarios consists in extending the SM with a neutral vector boson that couples to quarks and leptons of second and third generations,

$$\mathcal{L}_{Z'} \supset g^L_{\ell_i\ell_j} \bar{\ell}_i \gamma^\mu P_L \ell_j Z'_\mu + g^L_{sb} \bar{s} \gamma^\mu P_L b Z'_\mu + (L \to R) \tag{A.8}$$

In this case, only (axial)vector coefficients are induced [12],

$$C_9^{(\prime)\ell_1\ell_2} = -\frac{\pi}{\sqrt{2}m_{Z'}^2}\frac{1}{\alpha G_F V_{tb}V_{ts}^*}g_{sb}^{L(R)}(g_{\ell_1\ell_2}^R + g_{\ell_1\ell_2}^L)\,, \tag{A.9}$$

$$C_{10}^{(\prime)\ell_1\ell_2} = -\frac{\pi}{\sqrt{2}m_{Z'}^2}\frac{1}{\alpha G_F V_{tb}V_{ts}^*}g_{sb}^{L(R)}(g_{\ell_1\ell_2}^R - g_{\ell_1\ell_2}^L)\,. \tag{A.10}$$

while the $C_{S(P)}^{(\prime)}$ coefficients are zero if the Z' couples only to left-handed fermions (as suggested by the B-anomalies).
It can be noticed that, if $g_{\ell_1\ell_2}^R = 0$, $C_9^{(\prime)\ell_1\ell_2} = -C_{10}^{(\prime)\ell_1\ell_2}$.
Furthermore if LFV is due to Z' only, the expected BF for the OS and SS mode are the same: $\mathcal{B}(B^\pm \to K^\pm\tau^+\mu^-) \equiv \mathcal{B}(B^\pm \to K^\pm\tau^-\mu^+)$. Concerning R_K, the Z' would couple to $\bar{s}b$ (pure LH) and $\mu^+\mu^-$ (LH or RH).
The four-fermion operators describing the interaction of the Z' are

i $2q2\ell$: contribute to both $b \to s\mu\mu$ and $b \to s\nu\nu$

ii $4q$: contributing to $B_s^0 - \overline{B}_s^0$ mixing

iii 4ℓ: contributes to $\nu_\mu N \to \nu_\mu N\mu^+\mu^-$.

In addition to flavor observables, direct searches at LHC ($pp \to Z'\ell\ell$) must also be taken into account when building models with the Z'. The way the experimental results are incorporated into the theory depends on the model itself, where precise assumptions on the coupling with fermions are made, as well as on the total BF for $Z' \to \ell\ell$.

- $S_3 \sim (\mathbf{3}, \mathbf{3}, 1/3)$ is a weak triplet of LQs and represents the only scalar boson compatible with the $R_{K^{(*)}}$ measurements at three level [140, 141].
We call y_L^{ij} the Yukawa couplings of the S_3 with a quark of index i and lepton of index j. The effective WCs are[1]

$$C_9^{\ell_1\ell_2} = -C_{10}^{\ell_1\ell_2} = \frac{\pi v^2}{V_{tb}V_{ts}^*\alpha_{\text{em}}}\frac{y_L^{b\ell_1}(y_L^{s\ell_2})^*}{m_{S_3}^2}$$

while $C_i^{(\prime)} = 0$ otherwise (left-handed couplings only).
Due to the nature of S_3 (its fermion number $F = 3B + L = -2$), it is expected that $\mathcal{B}(B^\pm \to K^\pm\tau^\pm\mu^\mp) \gg \mathcal{B}(B^\pm \to K^\pm\tau^\mp\mu^\pm)$ (i.e. the OS mode should be enhanced with respect to the SS mode).

- $U_1 \sim (\mathbf{3}, \mathbf{1}, 2/3)$ This weak singlet vector LQ has gained a large attention as it provides the needed operators to explain the anomalies in FCCC and FCNC [4, 20, 142]. The interaction of U_1 with the SM fermions is encoded in the currents

$$J_\mu^U = \beta_L^{i\alpha}(\bar{q}_L^i\gamma_\mu\ell_L^\alpha) + \beta_R^{i\alpha}(\bar{d}_R^i\gamma_\mu e_R^\alpha)$$

with $i = \{d, s, b\}$ and $\alpha = \{e, \mu, \tau\}$ and

$$q_L^i = \begin{pmatrix} V_{ji}^* u_L^j \\ d_L^i \end{pmatrix}\,, \ell_L^i \begin{pmatrix} \nu_L^i \\ e_L^i \end{pmatrix}\,.$$

[1] $v = (\sqrt{2}G_F)^{-2} = 246\,\text{GeV}$.

This means that the relevant couplings are $(\beta^{b\tau}, \beta^{s\ell})$ for the OS_ℓ case and $(\beta^{b\ell}, \beta^{s\tau})$ for the SS_ℓ one. Moreover, as opposed to the S_3 case, we expect $\mathcal{B}(B^\pm \to K^\pm \tau^\pm \mu^\mp) \ll \mathcal{B}(B^\pm \to K^\pm \tau^\mp \mu^\pm)$.
There are two UV realizations that are equally viable and make different predictions (the fact of having a scalar WC or not is model-dependent). In fact the U_1 mediator can couple left- and right-handed fermions, inducing both vector and scalar WCs [143]:

- $C_9^{\ell_1\ell_2} = -C_{10}^{\ell_1\ell_2} = -\frac{\pi v^2}{\lambda_t \alpha_{\text{em}} m_U^2} x_L^{b\ell_1 *} x_L^{s\ell_2}$
- $C_9^{\prime\ell_1\ell_2} = C_{10}^{\prime\ell_1\ell_2} = -\frac{\pi v^2}{\lambda_t \alpha_{\text{em}} m_U^2} x_R^{b\ell_1 *} x_R^{s\ell_2}$
- $C_S^{\ell_1\ell_2} = -C_P^{\ell_1\ell_2} = \frac{2\pi v^2}{\lambda_t \alpha_{\text{em}} m_U^2} x_R^{b\ell_1 *} x_L^{s\ell_2}$
- $C_S^{\prime\ell_1\ell_2} = C_P^{\prime\ell_1\ell_2} = \frac{2\pi v^2}{\lambda_t \alpha_{\text{em}} m_U^2} x_L^{b\ell_1 *} x_R^{s\ell_2}$

Where $\lambda_t = V_{tb} V_{ts}^*$. As just shown the U_1 mediator can also generate the C_i' coefficients but they can be neglected in the scenarios explaining the anomalies.

We now consider the U_1 best fit obtained by Ref. [20] with the hypothesis of maximal right-handed current (RH+LH): $\beta_L^{b\tau} = 1 = -\beta_R^{b\tau}$ (i.e. max RH currents). The needed fit values are

$$\{C_U = 0.005, \beta_L^{b\mu} = -0.21, \beta_L^{s\tau} = 0.21, \beta_L^{s\mu} = 0.02\},$$

where $C_U = \frac{g_U^2 v^2}{4 M_U^2}$.
In the $\mathbf{OS}_\mu$ case ($b \to s\tau^- \mu^+$) the WCs, rewritten according Ref. [20], assume the values:

- $C_9^{\tau\mu} = -C_{10}^{\tau\mu} = -\frac{\pi v^2 g_U^2}{\lambda_t \alpha_{\text{em}} m_U^2} \beta_L^{b\tau *} \beta_L^{s\mu} = -\frac{4\pi C_U}{\lambda_t \alpha_{\text{em}}} \beta_L^{b\tau *} \beta_L^{s\mu} = -4.21$
- $C_S^{\tau\mu} = -C_P^{\tau\mu} = +\frac{\pi v^2 g_U^2}{2\lambda_t \alpha_{\text{em}} m_U^2} \beta_R^{b\tau *} \beta_L^{s\mu} = +\frac{2\pi C_U}{\lambda_t \alpha_{\text{em}}} \beta_R^{b\tau *} \beta_L^{s\mu} = -2.11 = C_9^{\tau\mu}/2.$

In the $\mathbf{SS}_\mu$ case ($b \to s\tau^+ \mu^-$):

- $C_9^{\mu\tau} = -C_{10}^{\mu\tau} = -\frac{\pi v^2 g_U^2}{\lambda_t \alpha_{\text{em}} m_U^2} \beta_L^{b\mu *} \beta_L^{s\tau} = -\frac{4\pi C_U}{\lambda_t \alpha_{\text{em}}} \beta_L^{b\mu *} \beta_L^{s\tau} = 9.28$
- $C_S^{\mu\tau} = -C_P^{\mu\tau} = +\frac{\pi v^2 g_U^2}{2\lambda_t \alpha_{\text{em}} m_U^2} \beta_R^{b\mu *} \beta_L^{s\tau} = +\frac{2\pi C_U}{\lambda_t \alpha_{\text{em}}} \beta_R^{b\mu *} \beta_L^{s\tau} \approx 0,$

where we follow the assumption of Ref. [20] $\beta_R^{b\mu *} << 1$.

- Extended scalar sectors: We do not treat them in this study.

The scenarios explored above lead us to consider the benchmark case where:

$$\begin{cases} C_S = -C_P \equiv C_{S_L} \in \mathbb{R} \\ C_9 = -C_{10} \equiv C_{V_L} \in \mathbb{R} \end{cases} \tag{A.11}$$

Therefore the coefficients of Eq. A.5 can be written as:

$$\begin{aligned}
\mathcal{K} &\to 2|C_{V_L}|^2 + 2|C_{S_L}|^2 = 2\frac{|C_{V_L}|^2}{r_V} \\
r_V &\to \frac{|C_{V_L}|^2}{|C_{V_L}|^2 + |C_{S_L}|^2} = \frac{1}{1 + \left|\frac{C_{S_L}}{C_{V_L}}\right|^2} \\
r_S &\to \frac{|C_{S_L}|^2}{|C_{V_L}|^2 + |C_{S_L}|^2} = \frac{1}{1 + \left|\frac{C_{V_L}}{C_{S_L}}\right|^2} = 1 - r_V \\
r_I &\to f(m_1, m_2)
\end{aligned} \tag{A.12}$$

$$f(m_1, m_2) = \begin{cases} 0 & m_1 < m_2 \quad (SS) \\ \frac{C_{V_L} C_{S_L}}{|C_{V_L}|^2 + |C_{S_L}|^2} = \frac{C_{S_L}/C_{V_L}}{1 + \left|\frac{C_{S_L}}{C_{V_L}}\right|^2} = \pm 2\sqrt{r_V - r_V^2} & m_1 > m_2 \quad (OS) \end{cases} \tag{A.13}$$

In the scenario of Eq. A.11, the expression in Eq. A.7 has only two unknown parameters: the normalization $|C_{V_L}|^2 + |C_{S_L}|^2$ and the ratio $\frac{C_{V_L}}{C_{S_L}}$, or, equivalently, C_{V_L} and r_V. The values of the $r_{V,S,I}$ coefficients as functions of $\frac{C_{V_L}}{C_{S_L}}$ are shown in Fig. A.4.

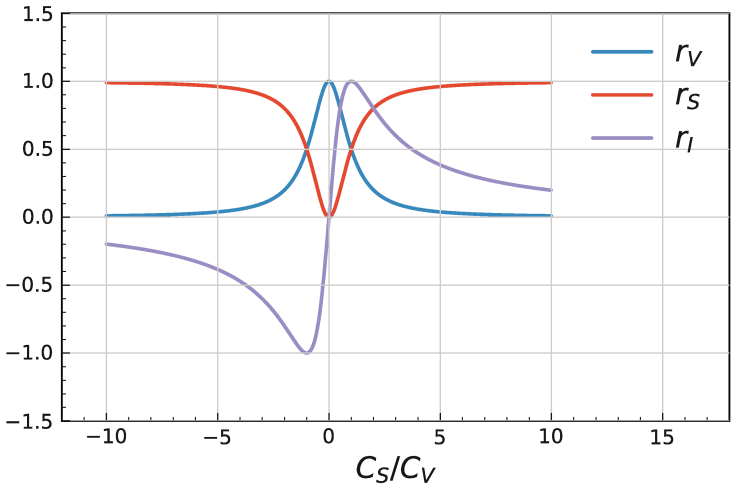


FIGURE A.4: $r_{V,S,I}$ coefficients as functions of the C_S/C_V ratio, in the simplified scenario described in Eq. A.11.

We consider now three notable values for r_V:

- $\mathbf{r_V = 0}$
  It corresponds to having a maximum scalar component. Such situation, not compatible with current anomalies, can appear in models with additional (pseudo) scalars - axion-like particles or two-Higgs-doublet models where lepton flavor is violated (see [5]).

- $\mathbf{r_V = 0.8}$
  Obtained for the U_1 model with max RH currents ($\beta_R = -1$). This is numerically true only for the OS_μ case.

- $\mathbf{r_V = 1}$
  It is compatible with scalar mediators (S_3 or Z' with LH couplings) and can explain the current deviation in R_K. $r_V \sim 1$ is also obtained in case of U_1 with max RH couplings (but only for the SS_μ mode), or a U_1 with only left-handed couplings to fermions, as the scalar coefficients are zero by definition. We recall that U_1 is compatible with the $R_{K^{(*)}}$ and $R_{D^{(*)}}$ deviations.

The q^2 shapes are shown in Figs. A.5, A.6 under the simplifications expressed in the Equations A.11- A.13. For the SS modes, the interference function is approximated to zero, in agreement with Eq.!A.13.

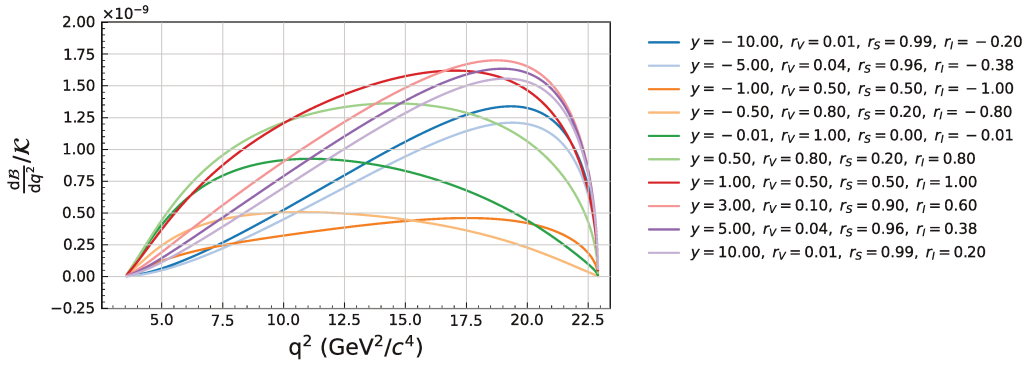


FIGURE A.5: Differential $\bar{B} \to \bar{K}\tau\ell$ decay rate as a function of the parameter $y = C_S/C_V$, in the simplified scenario described in Eq. A.11. The curves refer to the $\mathbf{OS}_\mu$ mode and are defined up to the normalization factor $\mathcal{K}$.

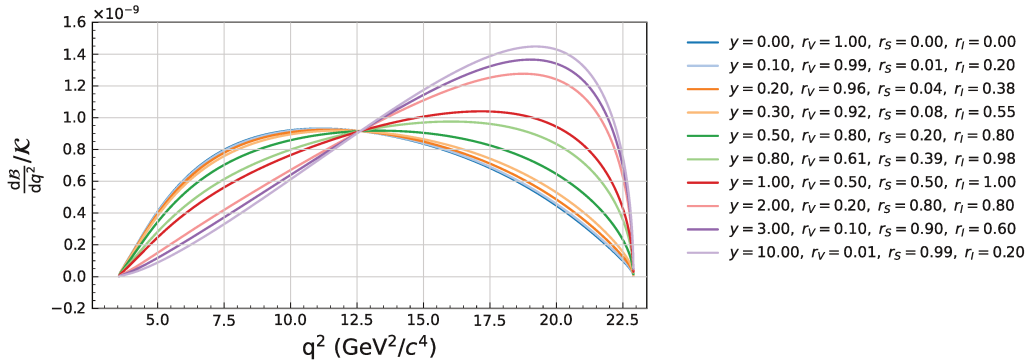


FIGURE A.6: Differential $\bar{B} \to \bar{K}\tau\ell$ decay rate as a function of the parameter $y = C_S/C_V$, in the simplified scenario described in Eq. A.11. The curves refer to the $\mathbf{SS}_\mu$ mode and are defined up to the normalization factor $\mathcal{K}$.

The intent of this study is to propose an alternative model to the phase-space for $B \to K\tau\ell$ MC event generation. This is motivated by the fact that the reconstruction efficiency, and therefore the quoted BF UL, they depend on the kinematic of the signal particles, i.e. on the q^2 distribution. Here we provide a q^2 shape that depends on the operators $\mathcal{O}^{(\prime)}_{9,10,S,P}$ and the respective WCs. Under a certain amount of assumptions, suggested by the current B-anomalies, the number of unknown parameters is such that a new, more realistic efficiency $\varepsilon_{\mathrm{BSM}}$ can be evaluated for several scenarios. The scaling of $\varepsilon_{\mathrm{BSM}}$ with respect to the efficiency evaluated with the phase-space model ($\varepsilon_{\mathrm{PHSP}}$) is shown in Fig. A.7.

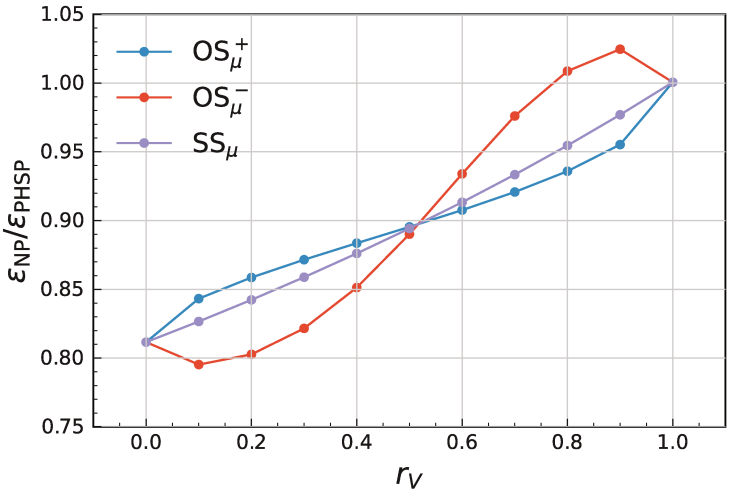


FIGURE A.7: Ratio of the reconstruction efficiencies between the BSM model and the PHSP initial assumption is displayed as a function of the r_V parameter ($\ell = \mu$ modes).

Bibliography

[1] A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski, and J. Rosiek, *Lepton-flavor violating B decays in generic Z' models*, Phys. Rev. D **92** (Sep, 2015) 054013. https://link.aps.org/doi/10.1103/PhysRevD.92.054013.

[2] G. Hiller, D. Loose, and K. Schönwald, *Leptoquark flavor patterns and B decay anomalies*, Journal of High Energy Physics **2016** (2016) no. 12, 27. https://doi.org/10.1007/JHEP12(2016)027.

[3] L. Calibbi, A. Crivellin, and T. Li, *Model of vector leptoquarks in view of the B-physics anomalies*, Phys. Rev. D **98** (Dec, 2018) 115002. https://link.aps.org/doi/10.1103/PhysRevD.98.115002.

[4] A. Angelescu, D. Bečirević, D. A. Faroughy, F. Jaffredo, and O. Sumensari, *Single leptoquark solutions to the B-physics anomalies*, Phys. Rev. D **104** (Sep, 2021) 055017. https://link.aps.org/doi/10.1103/PhysRevD.104.055017.

[5] C. Cornella, P. Paradisi, and O. Sumensari, *Hunting for ALPs with lepton flavor violation*, Journal of High Energy Physics **2020** (jan, 2020) . https://doi.org/10.1007/JHEP01(2020)158.

[6] M. Bordone, C. Cornella, J. Fuentes-Martín, and G. Isidori, *A three-site gauge model for flavor hierarchies and flavor anomalies*, Physics Letters B **779** (apr, 2018) 317–323. https://doi.org/10.1016/j.physletb.2018.02.011.

[7] D. Bečirević, I. Doršner, S. Fajfer, D. A. Faroughy, F. Jaffredo, N. Košnik, and O. Sumensari, *Model with two scalar leptoquarks: R2 and S3*, Phys. Rev. D **106** (2022) no. 7, 075023, arXiv:2206.09717 [hep-ph].

[8] 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.

[9] 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].

[10] G. Punzi, *Sensitivity of searches for new signals and its optimization*, 2003. https://arxiv.org/abs/physics/0308063.

[11] J. Schwab, *Calibration of the Full Event Interpretation for the Belle and the Belle II Experiment*, Master's thesis, Karlsruhe Institute of Technology (KIT), 2017.

[12] D. Bečirević, O. Sumensari, and R. Z. Funchal, *Lepton flavor violation in exclusive $b \to s$ decays*, The European Physical Journal C **76** (2016) no. 3, 134. https://doi.org/10.1140/epjc/s10052-016-3985-0.

[13] P. D. Group, P. A. Zyla, et al., *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (08, 2020) . https://doi.org/10.1093/ptep/ptaa104. 083C01.

[14] S. Mantry, *Heavy quark symmetry in nonleptonic B-decays to excited charmed mesons*, Phys. Rev. D **70** (2004) 114006, arXiv:hep-ph/0405290.

[15] A. Lenz, *Some comments on the missing charm puzzle*, in *UK Phenomenology Workshop on Heavy Flavor and CP Violation*. 9, 2000. arXiv:hep-ph/0011258.

[16] M. Neubert, *Theoretical analysis of $\overline{B} \to D^{**}\pi$ decays*, Phys. Lett. B **418** (1998) 173–180, arXiv:hep-ph/9709327.

[17] D. Guadagnoli, C. B. Park, and F. Tenchini, *$\tau \to \ell^+$ invisible through invisible-savvy collider variables*, Physics Letters B **822** (nov, 2021) 136701. https://doi.org/10.1016/j.physletb.2021.136701.

[18] G. de Marino, D. Guadagnoli, C. B. Park, and K. Trabelsi, *A new approach to semi-leptonic tags in B-meson semi-invisible decays*, arXiv:2209.03387 [hep-ph].

[19] B. Abi et al., Muon $g-2$ Collaboration, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Phys. Rev. Lett. **126** (Apr, 2021) 141801. https://link.aps.org/doi/10.1103/PhysRevLett.126.141801.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] S. Pohl, *Track reconstruction at the first level trigger of the Belle II experiment*. PhD thesis.

[26] 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.

[27] M. Gell-Mann and M. Levy, *The axial vector current in beta decay*, Nuovo Cim. **16** (1960) 705.

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

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

[30] 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.

[31] Belle Collaboration, G. Caria, et al., *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a Semileptonic Tagging Method*, Phys. Rev. Lett. **124** (Apr, 2020) 161803. https://link.aps.org/doi/10.1103/PhysRevLett.124.161803.

[32] 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.

[33] 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.

[34] Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, arXiv:2206.07501 [hep-ex].

[35] R. Watanabe, *New Physics effect on $B_c \to J/\psi\tau\bar{\nu}$ in relation to the $R_{D^{(*)}}$ anomaly*, Phys. Lett. B **776** (2018) 5–9, arXiv:1709.08644 [hep-ph].

[36] B. Capdevila, A. Crivellin, S. Descotes-Genon, L. Hofer, and J. Matias, *Searching for New Physics with $b \to s\tau^+\tau^-$ processes*, Phys. Rev. Lett. **120** (2018) no. 18, 181802, arXiv:1712.01919 [hep-ph].

[37] M. Algueró, B. Capdevila, S. Descotes-Genon, J. Matias, and M. Novoa-Brunet, *$b \to s\ell^+\ell^-$ global fits after R_{K_S} and $R_{K^{*+}}$*, Eur. Phys. J. C **82** (2022) no. 4, 326, arXiv:2104.08921 [hep-ph].

[38] R. Aaij et al., LHCb, *Angular Analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ Decay*, Phys. Rev. Lett. **126** (2021) no. 16, 161802, arXiv:2012.13241 [hep-ex].

[39] S. Wehle et al., Belle, *Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$*, Phys. Rev. Lett. **118** (2017) no. 11, 111801, arXiv:1612.05014 [hep-ex].

[40] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y. M. Wang, *Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$*, JHEP **09** (2010) 089, arXiv:1006.4945 [hep-ph].

[41] M. Beneke, C. Bobeth, and R. Szafron, *Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$*, Journal of High Energy Physics **2019** (oct, 2019) . https://link.springer.com/article/10.1007/JHEP10(2019)232.

[42] F. U. Bernlochner, M. F. Sevilla, D. J. Robinson, and G. Wormser, *Semitauonic b-hadron decays: A lepton flavor universality laboratory*, Reviews of Modern Physics **94** (feb, 2022) . https://doi.org/10.1103/RevModPhys.94.015003.

[43] *Heavy Flavor Averaging Group (HFLAV)*, https://hflav.web.cern.ch/content/semileptonic-b-decays. Accessed: 2022-06-10.

[44] 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.

[45] W. Altmannshofer and P. Stangl, *New physics in rare B decays after Moriond 2021*, The European Physical Journal C **81** (oct, 2021) . https://doi.org/10.1140/epjc/s10052-021-09725-1.

[46] T. Hurth, F. Mahmoudi, D. M. Santos, and S. Neshatpour, *More Indications for Lepton Nonuniversality in $b \to s\ell^+\ell^-$*, Phys. Lett. B **824** (2022) 136838, arXiv:2104.10058 [hep-ph].

[47] I. Doršner, S. Fajfer, A. Greljo, J. Kamenik, and N. Košnik, *Physics of leptoquarks in precision experiments and at particle colliders*, Physics Reports **641** (jun, 2016) 1–68. https://doi.org/10.1016/j.physrep.2016.06.001.

[48] M. Aaboud et al., ATLAS, *Search for lepton-flavor violation in different-flavor, high-mass final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **98** (2018) no. 9, 092008, arXiv:1807.06573 [hep-ex].

[49] A. M. Sirunyan et al., CMS, *Search for singly and pair-produced leptoquarks coupling to third-generation fermions in proton-proton collisions at s=13 TeV*, Phys. Lett. B **819** (2021) 136446, arXiv:2012.04178 [hep-ex].

[50] R. Aaij et al., LHCb, *Search for Lepton-Flavor Violating Decays $B^+ \to K^+\mu^\pm e^\mp$*, Phys. Rev. Lett. **123** (2019) no. 24, 241802, arXiv:1909.01010 [hep-ex].

[51] LHCb Collaboration, LHCb, *Search for the lepton-flavour violating decays $B^0 \to K^{*0}\mu^\pm e^\mp$ and $B_s^0 \to \phi\mu^\pm e^\mp$*, arXiv:2207.04005 [hep-ex].

[52] R. Aaij et al., LHCb Collaboration, *Measurement of the relative $B^- \to D^0/D^{*0}/D^{**0}\mu^-\overline{\nu}_\mu$ branching fractions using B^- mesons from $\overline{B}_{s2}^{*0}$ decays*, Phys. Rev. D **99** (May, 2019) 092009. https://link.aps.org/doi/10.1103/PhysRevD.99.092009.

[53] B. Aubert et al., BaBar, *Study of inclusive B^- and $\overline{B}^0$ decays to flavor-tagged D, $D_{(s)}$ and $\Lambda_{(c)}^+$*, Phys. Rev. D **75** (2007) 072002, arXiv:hep-ex/0606026.

[54] E. Bagan, P. Ball, V. M. Braun, and P. Gosdzinsky, *Charm quark mass dependence of QCD corrections to nonleptonic inclusive B decays*, Nucl. Phys. B **432** (1994) 3–38, arXiv:hep-ph/9408306.

[55] M. Neubert, *Heavy quark effective theory and weak matrix elements*, in *1997 Europhysics Conference on High Energy Physics*, pp. 243–268. 1, 1998. arXiv:hep-ph/9801269.

[56] 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].

[57] E. Bagan, P. Ball, B. Fiol, and P. Gosdzinsky, *Next-to-leading order radiative corrections to the decay $b \to ccs$*, Phys. Lett. B **351** (1995) 546–554, arXiv:hep-ph/9502338.

[58] G. Brandenburg et al., CLEO Collaboration, *Charged track multiplicity in B meson decay*, Phys. Rev. D **61** (Mar, 2000) 072002. https://link.aps.org/doi/10.1103/PhysRevD.61.072002.

[59] Z. Ligeti, M. Luke, and M. B. Wise, *Comment on studying the corrections to factorization in $B \to D^{(*)}X$*, Physics Letters B **507** (May, 2001) 142–146. http://dx.doi.org/10.1016/S0370-2693(01)00424-5.

[60] 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.

[61] 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.

[62] P. del Amo Sanchez et al., BaBar, *Studies of $\tau^- \to \eta K^- nu_\tau$ $\tau^- \to \eta\pi^- nu_\tau$ at BaBar and a search for a second-class current*, Phys. Rev. D **83** (2011) 032002, arXiv:1011.3917 [hep-ex].

[63] M. Davier, A. Hocker, and Z. Zhang, *The Physics of Hadronic Tau Decays*, Rev. Mod. Phys. **78** (2006) 1043–1109, arXiv:hep-ph/0507078.

[64] 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].

[65] T. Abe, K. Akai, N. Akasaka, M. Akemoto, A. Akiyama, M. Arinaga, Y. Cai, K. Ebihara, K. Egawa, A. Enomoto, E. Ezura, J. Flanagan, S. Fukuda, H. Fukuma, Y. Funakoshi, K. Furukawa, T. Furuya, J. Haba, K. Hara, and Z. Zong, *Achievements of KEKB*, PTEP **2013** (03, 2013) .

[66] Ohnishi et al., *Accelerator design at SuperKEKB*, Progress of Theoretical and Experimental Physics **2013** (03, 2013) .

[67] K. Wille, *The physics of particle accelerators: An introduction*. 2000.

[68] 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].

[69] J. Brodzicka et al., Belle, *Physics Achievements from the Belle Experiment*, PTEP **2012** (2012) 04D001, arXiv:1212.5342 [hep-ex].

[70] A. Bevan et al., *The Physics of the B Factories*, The European Physical Journal C **74** (2014) no. 11, 3026.

[71] R. Mizuk, M. Danilov, K. Abe, I. Adachi, H. Aihara, Y. Asano, T. Aushev, S. Bahinipati, A. Bakich, V. Balagura, S. Banerjee, E. Barberio, M. Barbero, A. Bay, I. Bedny, U. Bitenc, I. Bizjak, S. Blyth, A. Bondar, and D. Zürcher, *Search for the $\Theta(1540)^+$ pentaquark using kaon secondary interactions at Belle*, Physics Letters B **632** (01, 2006) .

[72] A. Abashian et al., *The Belle detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **479** (2002) no. 1, 117–232. https://www.sciencedirect.com/science/article/pii/S0168900201020137. Detectors for Asymmetric B-factories.

[73] 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.

[74] 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.

[75] T. Abe et al., Belle II, *Belle II Technical Design Report*, arXiv:1011.0352 [physics.ins-det].

[76] Belle II, N. Braun, *Softwaretrigger Framework and Tooling*, 2019. Internal note: BELLE2-NOTE-TE-2019-019.

[77] Belle II, G. de Marino and K. Trabelsi, *HLT skims*, 2020. Internal note: BELLE2-NOTE-TE-2020-018.

[78] 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.

[79] E. Barberio and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections. Version 2.0*, Comput. Phys. Commun. **79** (1994) 291–308.

[80] A. Sibidanov, T. E. Browder, S. Dubey, S. Kohani, S. S. R. Mandal, R. Sinha, and S. E. Vahsen, *A New Tool for Detecting BSM Physics in $B \to K^* \ell\ell$ Decays*, arXiv:2203.06827 [hep-ex].

[81] P. Ball and R. Zwicky, *New results on $B \to \pi, K, \eta$ decay form factors from light-cone sum rules*, Phys. Rev. D **71** (Jan, 2005) 014015. https://link.aps.org/doi/10.1103/PhysRevD.71.014015.

[82] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, *GEANT3*, .

[83] R. Itoh, *BASF - BELLE AnalysiS Framework*, in *9th International Conference on Computing in High-Energy and Nuclear Physics*. 4, 1997.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81–86.

[89] T. Abe et al., *Achievements of KEKB*, Progress of Theoretical and Experimental Physics **2013** (03, 2013) . https://doi.org/10.1093/ptep/pts102. 03A001.

[90] Belle, B. Casey, *HadronB*, 2001. Internal note: BELLE-390.

[91] T. Keck, *Machine learning algorithms for the Belle II experiment and their validation on Belle data*. PhD thesis, KIT, Karlsruhe, 2017.

[92] *PDG Live*, https://pdglive.lbl.gov/Viewer.action. Accessed: 2022-03-22.

[93] 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.

[94] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. ACM, New York, NY, USA, 2016. http://doi.acm.org/10.1145/2939672.2939785.

[95] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (Dec, 1978) 1581–1585. https://link.aps.org/doi/10.1103/PhysRevLett.41.1581.

[96] D. M. Asner et al., CLEO, *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53** (1996) 1039–1050, arXiv:hep-ex/9508004.

[97] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, 2003. https://arxiv.org/abs/physics/0306116.

[98] *Number of B events in Hadron B(J)*, https://belle.kek.jp/secured/nbb/nbb.html. Accessed:2022-08-24.

[99] 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.

[100] 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].

[101] S. Choudhury et al., *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, Journal of High Energy Physics **2021** (2021) no. 3, 105. https://doi.org/10.1007/JHEP03(2021)105.

[102] Belle, S. Nishida, *Study of kaon and pion identification using inclusive D^* sample*, 2005. Internal note: BELLE-779.

[103] L. Aggarwal et al., Belle-II, *Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond*, arXiv:2207.06307 [hep-ex].

[104] Belle II, V. Bertacchi, G. de Marino, K. Trabelsi, and V. S. Vobbilisetti, *Studies of charged hadronic B decays in Belle and Belle II MC*, 2022. Internal note: BELLE2-NOTE-PH-2022-002.

[105] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[106] W. Altmannshofer et al., Belle II, *The Belle II Physics Book*, arXiv:1808.10567 [hep-ex].

[107] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[108] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, arXiv:2203.11601 [hep-ph].

[109] 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].

[110] R. Fleischer, N. Serra, and N. Tuning, *Tests of factorization and $SU(3)$ relations in B decays into heavy-light final states*, Phys. Rev. D **83** (Jan, 2011) 014017. https://link.aps.org/doi/10.1103/PhysRevD.83.014017.

[111] 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].

[112] 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.

[113] 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.

[114] 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.

[115] 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)].

[116] H. Albrecht et al., ARGUS, *Exclusive Hadronic Decays of B Mesons*, Z. Phys. C **48** (1990) 543–552.

[117] B. Aubert et al., BABAR Collaboration, *Measurement of the absolute branching fractions $B \to D\pi$, $D^*\pi$, $D^{**}\pi$ with a missing mass method*, Phys. Rev. D **74** (Dec, 2006) 111102. https://link.aps.org/doi/10.1103/PhysRevD.74.111102.

[118] H. Atmacan et al., The Belle Collaboration, *Search for $B^0 \to \tau^\pm \ell^\mp (\ell = e, \mu)$ with a hadronic tagging method at Belle*, Phys. Rev. D **104** (Nov, 2021) L091105. https://link.aps.org/doi/10.1103/PhysRevD.104.L091105.

[119] M. Gelb et al., Belle, *Search for the rare decay of $B^+ \to \ell^+ \nu_\ell \gamma$ with improved hadronic tagging*, Phys. Rev. D **98** (2018) no. 11, 112016, arXiv:1810.12976 [hep-ex].

[120] G. Majumder, Belle, *Observations of $B \to D^{*-}5\pi$ and $B^+ \to D^{*-}\pi^+\pi^+\pi^+\pi^-$*, Belle internal note 620 .

[121] A. Drutskoy et al., Belle, *Observation of $B \to D^{(*)}K^-K^{0(*)}$ decays*, Phys. Lett. B **542** (2002) 171–182, arXiv:hep-ex/0207041.

[122] 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].

[123] 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].

[124] M. T. Prim et al., Belle Collaboration, *Search for $B^+ \to \mu^+ \nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging*, Phys. Rev. D **101** (Feb, 2020) 032007. https://link.aps.org/doi/10.1103/PhysRevD.101.032007.

[125] F. Abudinén et al., Belle II Collaboration, *Search for $B^+ \to K^+ \nu\bar{\nu}$ Decays Using an Inclusive Tagging Method at Belle II*, Phys. Rev. Lett. **127** (Oct, 2021) 181802. https://link.aps.org/doi/10.1103/PhysRevLett.127.181802.

[126] M. Kaleta, *Search for lepton flavour violating B meson decays at the Belle experiment*, Master's thesis, AGH University of Science and Technology, 2020.

[127] W. Dungel, C. Schwanda, et al., The Belle Collaboration, *Measurement of the form factors of the decay $B^0 \to D^{*-}\ell^+\nu_\ell$ and determination of the CKM matrix element $|V_{cb}|$*, Phys. Rev. D **82** (Dec, 2010) 112007. https://link.aps.org/doi/10.1103/PhysRevD.82.112007.

[128] C. Lester and D. Summers, *Measuring masses of semi-invisibly decaying particle pairs produced at hadron colliders*, Physics Letters B **463** (sep, 1999) 99–103. https://doi.org/10.1016/S0370-2693(99)00945-4.

[129] A. Barr, C. Lester, and P. Stephens, *m_{T2}: the truth behind the glamour*, Journal of Physics G: Nuclear and Particle Physics **29** (sep, 2003) 2343–2363. https://iopscience.iop.org/article/10.1088/0954-3899/29/10/304.

[130] A. J. Barr, T. J. Khoo, P. Konar, K. Kong, C. G. Lester, K. T. Matchev, and M. Park, *Guide to transverse projections and mass-constraining variables*, Physical Review D **84** (nov, 2011) . https://doi.org/10.1103/PhysRevD.84.095031.

[131] G. G. Ross and M. Serna, *Mass determination of new states at hadron colliders*, Physics Letters B **665** (jul, 2008) 212–218. https://doi.org/10.1016/j.physletb.2008.06.003.

[132] W. S. Cho, J. S. Gainer, D. Kim, K. T. Matchev, F. Moortgat, L. Pape, and M. Park, *On-shell constrained M 2 variables with applications to mass measurements and topology disambiguation*, Journal of High Energy Physics **2014** (aug, 2014) . https://doi.org/10.1007/JHEP08(2014)070.

[133] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park, $M_{(T2)}$*-assisted on-shell reconstruction of missing momenta and its application to spin measurement at the LHC*, Phys. Rev. D **79** (2009) 031701, arXiv:0810.4853 [hep-ph].

[134] C. B. Park, *Reconstructing the heavy resonance at hadron colliders*, Physical Review D **84** (nov, 2011) . https://doi.org/10.1103/PhysRevD.84.096001.

[135] C. B. Park, *YAM2: Yet another library for the M_2 variables using sequential quadratic programming*, Comput. Phys. Commun. **264** (2021) 107967, arXiv:2007.15537 [hep-ph].

[136] S. G. Johnson, *The NLopt nonlinear-optimization package*, http://github.com/stevengj/nlopt.

[137] B. Aubert et al., BABAR Collaboration, *Measurement of the $\overline{B}^0$ lifetime and the $B^0\overline{B}^0$ oscillation frequency using partially reconstructed $\overline{B}^0 \to D^{*+}\ell^-\overline{\nu}_\ell$ decays*, Phys. Rev. D **73** (Jan, 2006) 012004. https://link.aps.org/doi/10.1103/PhysRevD.73.012004.

[138] K. Belous et al., Belle Collaboration, *Measurement of the τ-lepton Lifetime at Belle*, Phys. Rev. Lett. **112** (Jan, 2014) 031801. https://link.aps.org/doi/10.1103/PhysRevLett.112.031801.

[139] LHCb Collaboration, LHCb, *Search for the lepton-flavour violating decays $B^0 \to K^{*0}\tau^\pm\mu^\mp$*, arXiv:2209.09846 [hep-ex].

[140] D. Bečirević, I. Doršner, S. Fajfer, D. A. Faroughy, N. Košnik, and O. Sumensari, *Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies*, Physical Review D **98** (sep, 2018) . https://doi.org/10.1103/PhysRevD.98.055003.

[141] A. Crivellin, D. Müller, and F. Saturnino, *Flavor Phenomenology of the Leptoquark Singlet-Triplet Model*, 2019. https://arxiv.org/abs/1912.04224.

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

[143] D. Bečirević, N. Košnik, O. Sumensari, and R. Z. Funchal, *Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive $b \to s\ell_1\ell_2$ modes*, 2016. https://arxiv.org/abs/1608.07583.