amU
Aix Marseille Université

NNT : 2024AIXM0001

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 19 septembre 2025 par

Clotilde LEMETTAIS

Search for $B^0 \to K^{*0}\tau^\pm\ell^\mp$ lepton flavour violating decays and tracking performance studies at the Belle and Belle II experiments

Discipline
Physique et Sciences de la Matière

Spécialité
Physique des Particules et Astroparticules

École doctorale
Physique et Sciences de la Matière (ED352)

Laboratoire/Partenaires de recherche
Centre de Physique des Particules de Marseille (CPPM), UMR7346
Centre National de la Recherche Scientifique (CNRS)
Conseil Européen de la Recherche (ERC)

Composition du jury

Matthew CHARLES LPNHE, Sorbonne Université, Paris	Rapporteur
Christoph SCHWANDA Institute of High Energy Physics, Vienne, Autriche	Rapporteur
Viola SORDINI IP2I, CNRS, Lyon	Examinatrice
Adrian BEVAN Queen Mary University of London, UK	Examinateur
Cristinel DIACONU CPPM, CNRS, Marseille	Président du jury
Justine SERRANO CPPM, CNRS, Marseille	Directrice de thèse

Belle II
CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE CPPM
cnrs
erc European Research Council

Affidavit

Je soussignée, Clotilde Lemettais, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Justine Serrano, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte AMU relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 8 juillet 2025

Liste de publications et participation aux conférences

Liste des publications et/ou brevets réalisées dans le cadre du projet de thèse:

1. C. Lemettais, V. Bertacchi, J. Serrano, G. Mancinelli pour les Collaborations Belle et Belle II, "Search for lepton flavor-violating decay modes $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ ($\ell = e, \mu$) with hadronic B-tagging at Belle and Belle II", publication soumise au Journal of High Energy Physics (JHEP), mai 25, arXiv preprint: arXiv:2505.08418

2. C. Lemettais, V. Bertacchi, J. Serrano, G. Mancinelli, "Search for lepton flavor violating decay $B^0 \to K^{*0}\tau\ell$ using the Belle and Belle II data sample", Note interne à Belle II, août 24, BELLE2-NOTE-PH-2024-042

3. C. Lemettais, J. Serrano, L. Zani, "Fake tracks rate measurement using 2019-2022 data and simulation produced with release 6", Note interne à Belle II, janvier 24, BELLE2-NOTE-PH-2024-002

Participation aux conférences et écoles d'été au cours de la période de thèse:

Conférences :

1. **59th Rencontres de Moriond, EW Interactions, and Unified Theories**, mars 2025, La Thuile, Italie. C. Lemettais pour la collaboration Belle II. *"Search for Lepton Flavour Violating $B^0 \to K^{*0}\tau\ell$ decays with the Belle and Belle II experiments"*. Présentation orale. Proceedings à paraître.

2. **50th Belle II General Meeting**, février 2025, Tsukuba, Japon. C. Lemettais. *"Search for Lepton Flavour Violating $B^0 \to K^{*0}\tau\ell$ decays with the Belle and Belle II experiments"*. Présentation orale.

3. **Workshop annuel du Groupement de Recherche "Intensity Frontier"**, novembre 2024, Cabourg, France. C. Lemettais et I. Consigny pour la collaboration Belle II. *"Search for LFV in $B \to K^*\tau\ell$ and $B \to \rho\tau\ell$ decays at Belle II"*. Présentation orale.

4. **45th Belle II General Meeting**, juin 2023, Nagoya, Japon. C. Lemettais. *"Tracking performance in release-7 (s-proc4)"*. Présentation orale.

5. **Workshop annuel du Groupement de Recherche "Intensity Frontier"**, novembre 2023, Mont Saint-Odile, France. C. Lemettais, A. Thaller et K. Lautenbach pour la collaboration Belle II. *"Search of LFV decays at Belle II"*. Présentation orale.

Écoles :

1. **IN2P3 School of Statistics 2024**, mai 2024, Carry-le-Rouet, France.

2. **École de GIF 2023: Théories effectives pour la physique des particules**, septembre 2023, Annecy, France.

Résumé

Cette thèse présente la recherche de désintégrations $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$, ne conservant pas la saveur leptonique, où ℓ est soit un électron, soit un muon. L'analyse est réalisée à partir des données collectées par les expériences Belle II et Belle à la résonance $\Upsilon(4S)$, pour une luminosité intégrée de $365\,\mathrm{fb}^{-1}$ et $711\,\mathrm{fb}^{-1}$, respectivement. L'étude repose sur une technique d'étiquetage hadronique exclusive des B, permettant la reconstruction complète du méson B étiqueté. Le signal est ensuite recherché dans la distribution de masse de recul du τ, correspondant au système reculant contre le B étiqueté et le système $K^{*0}\ell$. Le processus de suppression du bruit de fond est effectué en deux étapes : d'abord avec une sélection basée sur des coupures visant à cibler des sources de bruit de fond spécifiques et localisées, puis en utilisant un arbre de décision boosté (Boosted Decision Tree) pour rejeter le bruit de fond restant. Les rapports d'embranchement $\mathcal{B}(B^0 \to K^{*0}\tau^{\pm}\ell^{\mp})$ sont extraits d'un ajustement de vraisemblance maximale à la distribution de masse de recul du τ simultanément dans les données de Belle et Belle II. Aucune preuve de désintégrations $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ n'est observée, et des limites supérieures sur les rapports d'embranchement sont établies sur l'intervalle $(2.9\text{–}6.4)\times10^{-5}$ à 90% de niveau de confiance en utilisant une méthode $\mathrm{CL_s}$ asymptotique.

En complément, deux études sur les performances de reconstruction des traces à Belle II sont présentées. La première étude est consacrée à l'optimisation de l'algorithme "Deterministic Annealing Filter" (DAF) utilisé pour améliorer la précision et l'exactitude de la reconstruction des traces. La seconde étude se concentre sur le taux de fausses traces à Belle II. Son évolution par rapport aux mesures précédentes est quantifiée, et l'impact de l'exploitation des informations temporelles du détecteur de vertex à silicium (SVD) sur le taux de fausses traces est étudié.

Abstract

This thesis presents the search for $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ lepton flavour violating decays, where ℓ is either an electron or a muon. The analysis is performed using 365 fb^{-1} and 711 fb^{-1} datasets collected at the $\Upsilon(4S)$ resonance by the Belle II and Belle experiments, respectively. The study relies on an exclusive hadronic B-tagging technique, allowing the full reconstruction of the tagged B meson. The signal is then searched for in the τ recoil mass system, corresponding to the system recoiling against the fully reconstructed tagged B and the $K^{*0}\ell$ system. The background rejection process is performed in two steps with first a cut-based selection aiming at targeting specific and localised background sources, and then using a Boosted Decision Tree to reject the remaining background. The branching fractions $\mathcal{B}(B^0 \to K^{*0}\tau^{\pm}\ell^{\mp})$ are extracted from a simultaneous maximum likelihood fit to the τ recoil mass distribution in Belle and Belle II data. No evidence for $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ decays is found, and upper limits on the branching fractions are set in the range of $(2.9\text{–}6.4)\times 10^{-5}$ at 90% confidence level using a $\mathrm{CL_s}$ asymptotic method.

In addition, two studies on Belle II tracking performance are presented. The first study is dedicated to the optimisation of the Deterministic Annealing Filter (DAF) algorithm used to improve the accuracy of track reconstruction. The second one focuses on the fake tracks rate at Belle II. Its evolution is measured with respect to the previous measurements, and the impact of exploiting the silicon vertex detector (SVD) timing information on the fake tracks rate is studied.

Remerciements

Parvenue au terme de ces trois années de thèse, il est temps pour moi d'exprimer ma gratitude envers toutes les personnes qui m'ont soutenue et accompagnée durant cette période.

Ma première pensée va évidemment à Justine, pour m'avoir fait confiance et m'avoir offert l'opportunité de réaliser cette thèse sous sa direction. J'ai pu grâce à toi découvrir et mettre un pied dans la collaboration Belle II, au sein de laquelle j'ai pu m'épanouir et apprendre énormément. Merci pour ta disponibilité, ta bienveillance et tes conseils avisés. Je n'aurai pas pu rêver d'un meilleur encadrement de thèse, tant d'un point de vue scientifique qu'humain.

Je tiens également à remercier chaleureusement mon jury de thèse, Matthew Charles, Christoph Schwanda, Viola Sordini, Adrian Bevan et Cristinel Diaconu, pour avoir accepté d'évaluer mon travail et de participer à ma soutenance, ainsi que pour leurs remarques et questions pertinentes.

Un immense merci s'impose aussi pour l'ensemble du groupe Belle II au CPPM, pour l'accueil que j'y ai reçu et pour l'ambiance de travail agréable et les précieux échanges que j'ai pu y avoir. En particulier, je tiens à remercier Valerio, avec qui j'ai travaillé sur la majorité du travail présenté dans ce manuscrit. Merci pour ton investissement, ta patience, tes conseils pertinents et les discussions toujours scientifiquement enrichissantes que j'ai pu avoir avec toi. Merci aussi à la team Belle II France, merci les étoiles pour votre accueil et votre bonne humeur.

Je souhaitais aussi dire un mot pour les conveners des groupes de travail de Belle II dans lesquels j'ai été impliquée, en particulier Giulia pour le tracking et Sally et Gaetano pour EWP, pour avoir toujours été non seulement de bon conseil mais également disponibles, à l'écoute et encourageants.

Merci aussi évidemment à tous les thésards, postdocs et autres membres du CPPM que j'ai pu côtoyer durant ces trois ans. Merci pour les discussions, les cafés, les rires, les apéros, les soirées qui ont rythmé ma thèse et m'ont permis de vraiment profiter de ces années. Shoutout à Marie, Alicia et Malau, merci pour votre amitié. J'en suis, il est vrai, privilégiée.

Merci aussi à mes amis de plus ou moins longue date. On a partagé ensemble la douleur des études de physique, l'ennui du lycée ou même l'insouciance de l'école maternelle, mais vous avez toujours été un soutien, une épaule, une base solide sur laquelle je sais que je pourrais toujours compter. Merci pour ça.

Finalement, mes derniers mots vont évidemment à ma famille. Merci de m'avoir toujours encouragée et soutenue dans mes choix. Mais surtout merci de m'avoir offert ces choix, de m'avoir donné le goût de la science, de la curiosité, de comprendre le monde qui nous entoure. Et merci Maman, j'ai beau avoir juré que je ne serai jamais prof de maths, grâce à toi l'idée ne m'a jamais traversé l'esprit que femme et science pouvaient être incompatibles. J'espère qu'un jour moi aussi, juste d'exister suffira à montrer que c'est possible.

Résumé étendu

Le Modèle Standard et la violations de la saveur leptonique

En physique des particules, le Modèle Standard (MS) [1-4] est la théorie qui décrit les particules élémentaires et leurs interactions fondamentales. Il est basé sur la théorie quantique des champs et unifie les interactions électromagnétiques, faibles et fortes. Le MS a été développé tout au long du XX$^{\text{e}}$ siècle et est aujourd'hui la théorie la plus aboutie en physique des particules.

Il permet de décrire deux types de particules : Les fermions, qui sont les constituants de la matière, eux-mêmes divisés en quarks et leptons, et les bosons qui sont les médiateurs des interactions fondamentales. Les quarks peuvent être classés en deux catégories. La première est le type "up" caractérisée par une charge électrique de +2/3 et comprenant les quarks up, charm et top (u, c, t), et la seconde est le type "down" caractérisée par une charge électrique de -1/3 et comprenant les quarks down, strange et bottom (d, s, b). Chaque quark possède alors sa propre saveur. Les leptons, quant à eux, sont divisés en trois familles ou saveurs leptoniques dont chacune est constitué d'un lepton de charge électrique -1 et d'un neutrino de charge électrique nulle. Il existe donc six leptons, de saveur électronique (e, ν_e), muonique (μ, ν_μ) et tauique (τ, ν_τ). Chacun de ces fermions possède également une antiparticule associée, qui possède la même masse, mais des charges opposée. Les interactions fondamentales entre les fermions sont médiées par le second type de particules : les bosons. En particulier, le photon et les bosons $W^\pm$ et Z^0 sont responsables des interactions électromagnétiques et faibles (réunies sous le terme électrofaible), tandis que les gluons sont responsables de l'interaction forte. Enfin, le boson de Higgs, observé pour la première fois en 2012 par les expériences ATLAS [5] et CMS [6] au LHC, finalise le modèle standard en permettant aux autres particules d'acquérir une masse via le mécanisme de Higgs.

Néanmoins, malgré son excellente description de la physique subatomique et ses multiples validations expérimentales, le Modèle Standard reste incomplet et ne permet pas d'expliquer certains phénomènes observés dans la nature. C'est le cas par exemple de son incapacité à inclure la force de gravité parmi les interactions fondamentales, ou de fournir une explication à la présence de matière et énergie noires dans l'univers, pourtant attestée par les observations cosmologiques. De plus, le MS n'explique pas l'asymétrie entre les quantités de matière et d'antimatière dans l'univers, ni l'apparente arbitrarité de certains paramètres du modèle, tels que les masses des particules, le nombre de générations de fermions ou la hiérarchie de leurs masses.

Ces questionnements et pièces manquantes au modèle ouvrent la voie à la recherche de "Nouvelle physique" (NP), au-delà du Modèle Standard. Cette recherche peut se faire via des études de preuves directes, par exemple en essayant d'observer de nouvelles particules, ou indirectes, en cherchant des effets de la NP dans les processus connus du SM. L'une de ces pistes de recherche indirecte, et qui fait l'objet de cette thèse, est la violation de la saveur leptonique (LFV), qui est interdite dans le Modèle Standard. La saveur est une propriété physique intrinsèque des fermions. Dans le cas des quarks, elle correspond aux six types de quarks, tandis que pour les leptons, elle correspond aux trois familles électronique, muonique et tauique et est dite saveur leptonique.

Dans le cadre de cette thèse, nous nous intéressons en particulier à la saveur leptonique et à sa conservation. Dans le Modèle Standard, l'interaction électrofaible à laquelle sont sensibles les leptons, ne permet pas de transitions entre les saveurs leptoniques, c'est-à-dire qu'un lepton d'une saveur donnée ne peut pas se désintégrer en un lepton d'une autre saveur. Bien que l'oscillation des neutrinos [7-9] puisse théoriquement permettre une transition entre deux leptons de saveur différente (via l'émission d'un neutrino qui oscillerait avant d'être réabsorbé), ce type de processus est extrêmement rare et estimé à une probabilité d'occurrence (ou rapport d'embranchement) de l'ordre de 10^{-55} [10], ce qui est bien en dessous de la sensibilité des expériences actuelles. Une observation de violation de la saveur leptonique serait donc un signe clair et indiscutable de la présence de nouvelle physique.

Dans ce manuscrit, nous étudions la désintégration d'un méson B^0 en un méson K^{*0} et deux leptons de saveurs différentes, τ et ℓ, où ℓ peut être un électron ou un muon. De manière générale, de nombreuses recherches ont lieu dans le domaine de la saveur, en particulier dans les désintégrations de mésons B, offrant une sensibilité accrue à la NP. Dans ce cadre, des divergences par rapport aux prédictions du modèle standard ont été observées par exemple dans les rapports des probabilités d'occurrence des désintégrations $B \to D^{(*)}\tau\nu_\tau$ et $B \to D^{(*)}\ell\nu_\ell$ [11-15] et dans le rapport d'embranchement des désintégrations $B \to K\nu\overline{\nu}$ [16]. Ces résultats encouragent à poursuivre des recherches dans le secteur de la saveur. Actuellement, une violation de la saveur leptonique n'a jamais été observée, mais l'étude de désintégrations ne conservant pas la saveur leptonique impliquant des transitions $B^{+/0} \to K^{+/0}_{(S)}\tau^\pm\ell^\mp$ a permis de d'obtenir des limites supérieures de l'ordre de 10^{-5} [17-19] sur les rapports d'embranchement de ses désintégrations. Dans le cas de la désintégration $B^0 \to K^{*0}\tau^\pm\ell^\mp$, l'expérience LHCb a estimé une limite supérieure à 1×10^{-5} (0.8×10^{-5}) à 90% de niveau de confiance pour le rapport d'embranchement de la désintégration $B^0 \to K^{*0}\tau^+\mu^-$ ($B^0 \to K^{*0}\tau^-\mu^+$) [19]. Les modes de désintégration contenant un électron dans l'état final ont également fait l'objet d'une étude par LHCb très récemment, en parallèle de l'analyse réalisée pour cette thèse, et les résultats préliminaires indiquent des limites à 4.9 et 5.9×10^{-6} à 90% de niveau de confiance [20].

Les expériences Belle et Belle II

L'expérience Belle II [21, 22] est une expérience de physique des particules située au Japon, au laboratoire KEK à Tsukuba, et visant à étudier entre autres les propriétés des mésons B. Elle consiste en un détecteur éponyme installé sur le collisionneur d'électrons et de positrons SuperKEKB [23, 24]. Belle II prend la suite en 2019 de l'expérience Belle [25], ayant fonctionné de 1998 à 2010, et dont le détecteur était installé sur le collisionneur KEKB [26, 27]. Belle II et SuperKEKB réutilisent ainsi en partie les infrastructures de Belle et KEKB, tout en y apportant de nombreuses et significatives améliorations.

Comme son prédécesseur KEKB, SuperKEKB est un collisionneur d'électrons et de positrons asymétrique opérant à une énergie dans le centre de masse de $\sqrt{s} = 10.58$ GeV, correspondant à la masse de la résonance $\Upsilon(4S)$. Cette dernière est un état lié de quarks bottom et anti-bottom ($b\bar{b}$) dont le rapport d'embranchement de désintégration en mésons B ($B^0\overline{B}^0$ ou B^+B^-) est proche de 96% [28], donnant ainsi le surnom d'"usine à B" aux expériences telles que Belle et Belle II. SuperKEKB se démarque des autres collisionneurs, passés et présents, par sa conception innovante permettant d'atteindre une luminosité instantanée record. Si l'objectif est d'atteindre une luminosité instantanée $\mathscr{L} = 6 \times 10^{35}$ cm^{-2}s^{-1}, correspondant à 30 fois la luminosité atteinte par KEKB, la luminosité maximale atteinte à ce jour est 5.1×10^{34} cm^{-2}s^{-1}, constituant le record mondial de luminosité. Une telle luminosité permet la production d'une quantité considérable de données, améliorant significativement la sensibilité des expériences. En effet, sur la période d'activité de l'expérience Belle II de 2019 à 2023, 365 fb^{-1} de données ont été collectées à la résonance $\Upsilon(4S)$, soit l'équivalent d'environ 390 millions de paires de $B\overline{B}$, l'objectif étant de collecter à terme 50 ab^{-1} de données. L'expérience Belle avait de son côté collecté 711 fb^{-1} de données, soit environ 770 millions de paires de $B\overline{B}$, durant ses 12 années d'activité. Cette luminosité instantanée record est atteinte grâce à une conception particulière du collisionneur, dans laquelle les courants des faisceaux d'électrons et de positrons sont augmentés (avec objectif d'être doublés par rapport à KEKB), et où les faisceaux sont fortement focalisés dans la région de collision. L'angle de croisement des faisceaux au point de collision est également optimisé conjointement à la réduction de la taille des faisceaux afin de maximiser le taux de collisions.

De la même manière que pour KEKB et SuperKEKB, le détecteur Belle II succède à Belle en réutilisant la même logique de conception, mais en intégrant de nombreuses améliorations. En particulier, Belle II est designé pour être plus résistant aux radiations et gérer un plus haut taux d'occupation, afin de pouvoir supporter le bruit de fond généré par les collisions à très haute luminosité. Dans les deux cas, le détecteur entoure presque complètement le point d'interaction, couvrant la totalité de l'angle azimutal et de 17° à 150° sur l'angle polaire, et est arrangé en différentes couches de sous-détecteurs remplissant chacun des fonctions complémentaires. Cette géométrie particulière permet de reconstruire la totalité des évènements issus des collisions, laissant comme seule énergie manquante celle due à l'émission de neutrinos. Les sous-détecteurs de Belle et Belle II sont les suivants (depuis l'intérieur vers l'extérieur) :

- Le détecteur de vertex (VXD) : Dans le cas de Belle, il s'agit seulement d'un détecteur de vertex à silicium (SVD), tandis que pour Belle II, un détecteur utilisant des capteurs à pixel (PXD) a été ajouté à l'intérieur du SVD afin de supporter les très hauts taux d'occupation et de radiations dus à l'augmentation de la luminosité instantanée. Dans les deux cas, le rôle du VXD est de reconstruire la position du point d'interaction et les points de désintégration des particules, en particulier des mésons B. Il est également utilisé pour la reconstruction des trajectoires des particules chargées.

- La chambre à dérive centrale (CDC) : Il s'agit d'une chambre à dérive gazeuse constituée de capteurs filaires sous tension. Les CDC de Belle et Belle II sont très similaires, à la différence que celle de Belle II a un plus grand volume de détection et un nombre accru de capteurs filaires. La CDC est un élément crucial de la trajectographie des particules chargées à Belle et Belle II, et joue également un rôle important dans l'identification des particules.

- Le système d'identification des particules : Pour Belle, celui-ci est constitué d'un compteur Cherenkov à aerogel (ACC) et d'un compteur de temps de vol (TOF). Le premier détecte la lumière Cherenkov émise par les particules chargées lorsqu'elles traversent l'aerogel, permettant la séparation des kaons et des pions au vu de leur différence de masse. Le second rempli la même fonction pour des particules moins énergétiques en mesurant leur temps de vol. Pour Belle II, le système d'identification des particules est amélioré en rempla*c*cant les compteurs ACC et TOF par un détecteur Cherenkov à imagerie annulaire (ARICH) et un détecteur de temps de propagation (TOP). Les deux sous-détecteur replissent la même fonction que leurs prédécesseurs, en exploitant l'effet Cherenkov.

- Le calorimètre électromagnétique (ECL) : Le calorimètre de Belle II réutilise la structure et les matériaux de détection de celui de Belle, à savoir les cristaux à scintillation servant à mesurer l'énergie des particules stoppées dans le détecteur. Cependant, afin de faire face aux très hauts taux d'occupation, le système de collecte et traitement du signal de l'ECL de Belle II a été significativement amélioré.

- Le détecteur de K_L^0 et de muons (KLM) : Pour ce sous-détecteur aussi, Belle II réutilise la structure et une partie du système de détection de Belle, mais ce dernier est partiellement remplacé afin de supporter les hauts taux d'occupation. Dans les deux cas les couches de détection actives sont alternées avec des couches de plaques de fer afin de d'arrêter ou ralentir les K_L^0 et muons, qui sont les seules particules atteignant le KLM.

Enfin, les détecteurs Belle et Belle II sont équipés d'un système de déclenchement (trigger) permettant de sélectionner les évènements à enregistrer, et d'acquisition de données (DAQ) permettant de stocker et traiter les évènements sélectionnés. Le système de déclenchement de Belle II est composé de deux éléments : un déclencheur matériel de bas niveau L1 et un déclencheur logiciel de haut niveau HLT, sur lequel une première reconstruction rapide des évènements est effectuée avant leur sélection et stockage définitif.

Pour reconstruire et analyser les données collectées par Belle et Belle II, un logiciel dédié appelé "Belle (II) Analysis Software" (basf(2))[29, 30] est utilisé. Le logiciel de Belle II basf2 permet également de traiter les données de Belle, via un module de compatibilité et est utilisé dans cette thèse. Un autre module important pour cette étude est l'algorithme de reconstruction de second méson B dit "étiqueté" de la paire produite lors de la désintégration du $\Upsilon(4S)$, i.e. le méson dans lequel on ne cherche pas le signal. Cet algorithme, appelé "Full Event Interpretation" (FEI) [31], est basé sur des techniques d'apprentissage automatique et permet de reconstruire une sélection de modes de désintégrations, soit semi-leptoniques, soit hadroniques, des mésons B. Dans le cas de l'analyse principale de cette thèse, le FEI hadronique est utilisé, permettant ainsi une connaissance totale des énergies et impulsions des produits de désintégration du B étiqueté.

Finalement, afin d'étudier les différents processus attendus dans les données collectées par Belle et Belle II, des simulations Monte-Carlo (MC) des collisions, désintégrations des particules et réponse du détecteur sont produites en reproduisant les conditions réelles de bruit de fond lors des prises de données.

Études de performances des algorithmes de reconstruction des traces dans le détecteur Belle II

La première partie de cette thèse est consacrée à l'étude des performances des algorithmes de reconstruction des traces dans le détecteur Belle II. Pour cela, une première étude est réalisée dans le but d'optimiser le temps d'exécution d'un algorithme utilisé pour améliorer la précision de la reconstruction des traces, appelé "DAF" pour "Deterministic Annealing Filter". Ensuite, une seconde étude consistant à mesurer le taux de fausses traces reconstruites dans les évènements finaux est réalisée.

Optimisation de l'algorithme d'ajustement des traces (DAF)

À Belle II, afin de ne stocker que les évènements les plus intéressants, les systèmes de déclenchement décrits précédemment sont utilisés. En particulier, pour sélectionner ces évènements au niveau du HLT, les traces des particules issues des collisions sont d'abord reconstruites une première fois dans une reconstruction dite *rapide* ou *en ligne* ne faisant intervenir que les informations des sous-détecteurs SVD et CDC. Après la sélection du HLT, la reconstruction complète des traces, dite *hors-ligne*, utilisant également les informations du PXD est effectuée.

Cependant, lors de la première période de prise de donnée de Belle II, le débit de donnée à l'entrée du HLT (soit le débit de sortie du déclencheur L1) était déjà très proche de la capacité de traitement du HLT. Avec l'augmentation de la luminosité instantanée, des évènements d'une plus grande complexité due au haut bruit de fond sont attendus, en plus de l'augmentation absolue du nombre d'évènements à traiter. Bien que des solutions matérielles pour réduire la pression sur le HLT ait été appliquées (augmentation du nombre d'unités de calcul du HLT par exemple), elles ne sont que temporaires et une optimisation au niveau logiciel des algorithmes de reconstruction est donc à considérer. Pour pallier cette saturation potentielle du HLT, le travail effectué dans cette section de la thèse consiste ainsi à optimiser un algorithme utilisé dans l'ajustement des traces.

Lors de la reconstruction, les traces sont régulièrement réajustées par rapports aux points d'impact dans les sous-détecteurs afin d'obtenir la meilleure approximation des trajectoires des particules chargées. Cependant, certains impacts qui n'appartiennent pas à la trace d'une particule (bruit de fond du faisceau, impacts d'autres traces...) peuvent être pris en compte dans l'ajustement à tort. Pour rejeter les impacts incohérents et améliorer la précision de l'ajustement, un algorithme appelé "Deterministic Annealing Filter" (DAF) est utilisé. L'idée du DAF est d'assigner des poids entre 0 et 1 aux impacts en fonction du résidu entre l'impact et la prédiction de l'ajustement de la trace, conditionnant l'importance desdits impacts dans l'ajustement final de la trajectoire. Le DAF procède ainsi de manière itérative en réajustant les traces plusieurs fois et en réévaluant les poids jusqu'à ce qu'un critère de convergence soit atteint. À chaque itération, l'assignation des poids est faite de plus en plus durement selon un paramètre de température T, incitant les poids à se diriger vers 0 ou 1. Il existe trois critères de convergence permettant d'interrompre l'algorithme dès lors que l'un des trois est rempli :

- Sur l'ensemble des impacts d'une trace, la différence maximum entre le poids d'un impact et celui de l'itération précédente est inférieure à un seuil Δw.
- L'itération $N_{\min}^{p-val}$ est atteinte et la différence entre la p-value de l'ajustement et celle de l'itération précédente est inférieure à un seuil Δp.
- Le nombre d'itérations $N_{\max}$ est atteint.

Dans la configuration initiale (aussi appelée "baseline") du DAF, les valeurs des paramètres de température, de seuils et d'itérations sont tels que le premier critère n'est jamais satisfait et que le second est quasiment toujours immédiatement satisfait dès l'itération $N_{\min}^{p-val}$. L'optimisation proposée consiste donc à relâcher le premier critère de convergence en augmentant le seuil Δw, à resserrer le second critère en diminuant le seuil Δp, à réduire $N_{\min}^{p-val}$ pour diminuer le nombre moyen d'itérations, et à modifier les paramètres de température pour accélérer la convergence de l'algorithme.

Les versions initiale et optimisée du DAF sont alors comparées grâce à des simulations Monte-Carlo, en particulier sur des évènements $B\overline{B}$ simulés avec un bruit de fond représentatif de celui observé dans les données réelles, et dans une configuration où seulement le SVD et la CDC sont utilisés, reproduisant les conditions de la *reconstruction rapide* du HLT. Un gain de temps d'exécution de 14% est obtenu sur l'ensemble des modules faisant appel au DAF dans la chaine de reconstruction des traces, et le nombre moyen d'itération est réduit de 5 à 3.3, comme l'on peut voir sur la Figure 1.

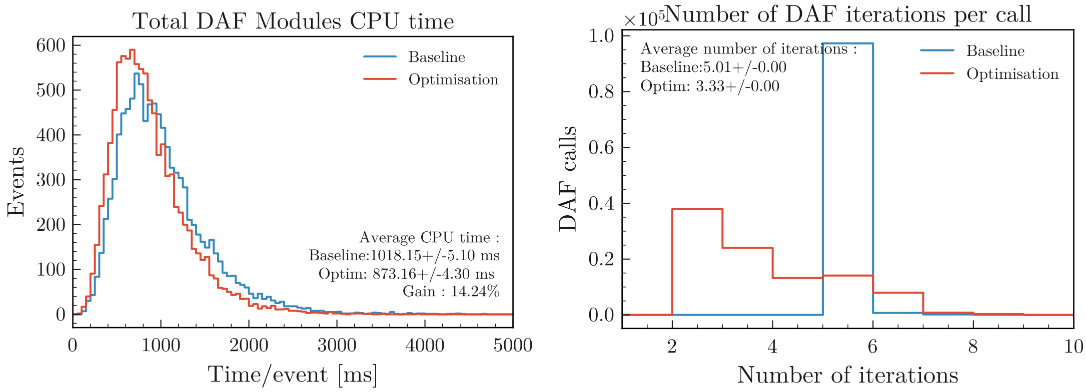


FIGURE 1 : Temps de calcul de l'ensemble des modules appelant le DAF (gauche) et nombre moyen d'itérations du DAF (droite) dans des évènements $B\overline{B}$ simulés avec le bruit de fond des données et les sous-détecteurs SVD et CDC.

En parallèle, de nombreux paramètres reflétant la qualité de la reconstruction des traces sont mesurés dans les versions initiale et optimisée du DAF afin de s'assurer que l'optimisation ne dégrade pas les performances de la reconstruction. Cette étude est toutefois réalisée entièrement à l'aide de simulations Monte-Carlo, et mériterait d'être confirmée par une étude sur les données réelles.

Mesure du taux de fausses traces

Malgré les algorithmes comme le DAF visant à maximiser la qualité des traces reconstruites, certaines traces peuvent être reconstruites à tort, on les appelle alors les fausses traces. Celles-ci peuvent être reconstruites à partir de points d'impacts dus au bruit de fond du faisceau, provenir de mauvaises combinaisons d'impacts de particules, ou être ce que l'on appelle des traces clones. Ces dernières sont des traces reconstruites à partir d'impacts appartenant à une trace déjà existante, mais sans être fusionnées avec celle-ci, et ainsi interprétées comme deux ou plusieurs traces distinctes. Mesurer le taux de fausse trace, ainsi que l'efficacité de reconstruction, est donc un paramètre important pour évaluer la qualité de la reconstruction des traces à Belle II.

Cette étude est réalisée en utilisant des processus $e^+e^- \to \tau^+\tau^-$. La section efficace de production des paires de leptons τ étant proche de celle des paires de mésons B à l'énergie de collision, et les désintégrations du τ contenant peu de particules, projetées dans des directions opposées avec une large gamme d'énergie, ce type de processus est un excellent candidat pour des études de performance de la reconstruction des traces. L'un des τ est alors contraint à se désintégrer en un lepton ($\tau^\pm \to \ell^\pm \nu_\ell \overline{\nu}_\tau$, $\ell = e, \mu$, appelé 1−prong), tandis que l'autre τ est contraint à se désintégrer en trois pions chargés ($\tau^\pm \to 3\pi^\pm \overline{\nu}_\tau$, appelé 3−prongs).

Le taux de fausses traces peut alors être mesuré en utilisant une méthode dite *tag-*

and-probe. Dans cette méthode, les 4 traces attendues dans l'évènement sont appelées "tags" et on cherche la présence d'une trace additionnelle appelée "probe". Le taux de fausses traces est ainsi défini comme :

$$r_{\text{fake}} = \frac{N_5}{N_4 + N_5},$$

où N_4 est le nombre d'évènements dans lesquels les quatre traces attendues sont reconstruites et N_5 est le nombre d'évènements dans lesquels une cinquième trace (probe) est reconstruite.

La mesure est effectuée sur des simulations Monte-Carlo, ainsi que sur les données de Belle II récoltées durant la première période de prise de données (2019-2022). Cette étude ayant déjà été réalisée par le passé, l'objectif de cette mise à jour est d'une part de contrôler l'évolution de taux de fausses traces dans le temps, mais également de vérifier que les changements apportés au logiciel de reconstruction n'ont pas dégradé les performances de reconstruction des traces. Pour le premier aspect, l'augmentation de la luminosité instantanée de SuperKEKB entraine des plus hauts taux de bruit de fond, résultant en une augmentation du taux de fausses traces. Si ce taux est aujourd'hui raisonnable et ne pose pas de problème pour les analyses physiques, il est important de surveiller son évolution. Pour le second aspect, le logiciel de Belle II basf2, et en particulier la chaine de reconstruction des traces, est régulièrement mis à jour et amélioré. Certains paramètres, dont le taux de fausses traces fait partie, doivent donc être régulièrement mesurés afin d'évaluer la variation des performances lors de ces changements, et aider à décider des modifications à implémenter dans le logiciel.

Pour l'étude réalisée dans cette thèse, le taux de fausses traces dans les données de Belle II avec la version du logiciel tel qu'il était pour le traitement des données et campagnes de simulations correspondant à la période 2019-2022 est d'environ 4%. Cette valeur est comparable à celle mesurée dans les données de la précédente campagne de traitement de données, correspondant à celles récoltées entre 2019 et 2021. Pour l'étude du taux de fausses traces dans différentes configurations du logiciel, un échantillon de données collectées à très haute luminosité (la plus élevée de la première période) est utilisé. Dans cet échantillon, le taux de fausses traces est d'environ 9% dans la configuration originale du logiciel. Différentes modification permettant d'exploiter les informations temporelles du SVD pour rejeter les traces produites en décalé de l'instant de la collision sont testées. Le taux de fausses traces est ainsi réduit à des valeurs entre 1.3% et 3.7% selon la configuration. Ces mesures sont également complémentées par des études d'autres quantités telles que l'efficacité de reconstruction, afin de d'évaluer la configuration optimale du logiciel et l'implémenter pour le traitement des données futures.

Recherche de désintégrations $B^0 \to K^{*0}\tau^\pm\ell^\mp$ ne conservant pas la saveur leptonique à Belle et Belle II

L'objectif principal de cette thèse est la recherche de désintégrations $B^0 \to K^{*0}(\to K^+\pi^-)\tau^\pm\ell^\mp$ ne conservant pas la saveur leptonique, où ℓ peut être un électron ou un muon. Pour cela, les 365 fb^{-1} de données collectées par Belle II durant la période 2019-2022 sont utilisées, ainsi que les 711 fb^{-1} de données collectées par Belle durant son activité entre 1998 et 2010. Les désintégrations $B^0 \to K^{*0}\tau^\pm\ell^\mp$ consistent en quatre états finaux distincts selon la saveur du lepton ℓ et le signe de sa charge électrique par rapport à celle du kaon produit du K^{*0} : les modes de signe opposé $OS\ell$ $B^0 \to K^{*0}(\to K^+\pi^-)\tau^+\ell^-$, et les modes de même signe $SS\ell$ ($\ell = e, \mu$) $B^0 \to K^{*0}(\to K^+\pi^-)\tau^-\ell^+$. La conjugaison de charge est implicitement appliquée à tous les modes de désintégration étudiés dans cette thèse.

Le processus d'étiquetage ("*tagging*") du second méson B décrit plus haut est utilisé pour reconstruire ce dernier (dit B_{tag}) via des désintégrations hadroniques. L'autre méson B (dit B_{sig}) est reconstruit à partir du K^{*0} et du lepton ℓ. Le signal est alors cherché dans le système reculant par rapport au reste de l'évènement, constitué du B_{tag} complètement reconstruit, et du B_{sig} comprenant le K^{*0} et le lepton ℓ. Ce système correspond à la masse de recul du lepton τ et est défini par l'équation (0.1) :

$$M_\tau = \left[m_B^2 + M(K^{*0}\ell)^2 - 2(E_{\text{beam}}E_{K^{*0}\ell} + |\vec{p}_{B_{\text{tag}}}||\vec{p}_{K^{*0}\ell}|\cos\theta)\right]^{1/2}. \quad (0.1)$$

Ici, m_B est la masse mesurée du méson B^0 [28], $M(K^{*0}\ell)$, $E_{K^{*0}\ell}$, $\vec{p}_{K^{*0}\ell}$ sont les masse invariante, énergie et quantité de mouvement du système $K^{*0}\ell$, E_{beam} est l'énergie du faisceau, correspondant à la moitié de l'énergie de centre de masse, $\vec{p}_{B_{\text{tag}}}$ est la quantité de mouvement du méson B étiqueté, et θ est l'angle entre les vecteurs $\vec{p}_{B_{\text{tag}}}$ et $\vec{p}_{K^{*0}\ell}$. Toutes les quantités sont définies dans le référentiel du centre de masse. En présence de signal, la distribution de masse du système reculant M_τ doit présenter un pic centré sur la masse du τ connue $m_\tau = 1.777\,\text{GeV}/c^2$ [28]. Les rapports d'embranchement $\mathcal{B}(B^0 \to K^{*0}\tau^\pm\ell^\mp)$ sont alors mesurés grâce à un ajustement de la distribution de masse du système reculant M_τ, simultanément sur les données de Belle et Belle II :

$$\mathcal{B}(B^0 \to K^{*0}\tau^\pm\ell^\mp) = \frac{N_{\text{sig}}}{\varepsilon_{\text{sig}} \times 2 \times N_{\Upsilon(4S)} \times f_{00} \times \mathcal{B}(K^{*0} \to K^+\pi^-)} \quad (0.2)$$

où N_{sig} est le nombre d'évènements de signal mesuré, ε_{sig} est l'efficacité de reconstruction du signal, $N_{\Upsilon(4S)}$ est le nombre de $\Upsilon(4S)$ produits, f_{00} est la fraction de paires de mésons B^0 venant de la désintégration de l'$\Upsilon(4S)$ et $\mathcal{B}(K^{*0} \to K^+\pi^-)$ est le rapport d'embranchement du K^{*0} en $K^+\pi^-$.

Le méson B étiqueté B_{tag} est reconstruit à partir de désintégrations hadroniques à l'aide de l'algorithme FEI [31] de Belle II. Les candidats sont ensuite sélectionnés selon plusieurs critères, reposant notamment sur leur masse contrainte par l'énergie du faisceau $M_{bc} = \sqrt{E^2_{\text{beam}}/c^4 - p^2_{B_{\text{tag}}}/c^2}$, leur différence d'énergie par rapport à celle du faisceau $\Delta E = E_{B_{\text{tag}}} - E_{\text{beam}}$, et le cosinus de l'angle entre la direction moyenne du B_{tag} et celle du reste des traces et foyers d'impacts dans l'ECL. De plus, les candidats B_{tag} sont classés selon leur probabilité d'être correctement reconstruits, et seulement les meilleurs candidats sont conservés.

Les candidats B_{sig} sont ensuite reconstruits en combinant trois traces correspondant au kaon, au pion et au lepton ℓ. Les candidats K^{*0} sont choisis tels que la masse invariante du système $K\pi$ les constituant soit dans une fenêtre de $\pm 50\,\text{MeV}/c^2$ autour de la masse connue du K^{*0} ($= 892\,\text{MeV}/c^2$). Comme dit précédemment, le lepton τ n'est pas explicitement reconstruit puisque le signal est extrait de sa masse de recul. Cependant, sa présence est déduite de la présence d'une trace t_τ de bonne qualité et de charge opposée à celle du lepton ℓ. Cette trace n'est pas utilisée dans l'extraction du signal mais sert dans la réduction du bruit de fond.

Finalement, les traces et foyers d'impacts dans l'ECL non utilisés pour le B_{tag}, le B_{sig} et la trace t_τ constituent le reste de l'évènement (*Rest Of Event*, ROE). Dans le cas d'une reconstruction idéale et au vu des rapports d'embranchement des désintégrations du τ, le ROE ne devrait comprendre aucune trace de particule chargée dans 85% des cas et deux traces dans 15% des cas. Néanmoins, la reconstruction n'étant pas parfaite, il se peut que des traces appartenant au B_{tag} se retrouvent dans le ROE, créant ainsi des évènements avec une seule trace dans le ROE ou deux traces ne correspondant pas à des produits du τ. Afin de garder ces évènements mal reconstruits, mais potentiellement intéressants et piquant à la masse du τ dans la distribution de masse du système reculant, seuls les évènements avec trois traces ou plus dans le ROE sont rejetés.

Un schéma de la reconstruction de l'évènement est présenté sur la Figure 2.

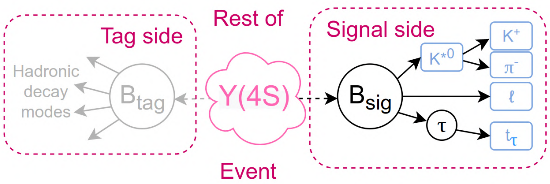


FIGURE 2 : Schéma de la reconstruction de l'évènement.

Pour rejeter le bruit de fond, une première sélection est faite via l'application de coupures sur certaines variables, puis un arbre de décision boosté (BDT) est entrainé et une deuxième sélection est effectuée. Pour commencer, tous les évènements se trouvant en dehors de l'intervalle $[1.0, 2.5]\,\text{GeV}/c^2$ de la masse M_τ sont rejetés. Par ailleurs, la région du signal $[1.65, 1.90]\,\text{GeV}/c^2$ est masquée jusqu'à la fin de l'analyse afin de ne pas introduire de biais. Certains bruits de fond correspondant à des contributions physiques sont très localisés dans certaines variables, et donc facile à rejeter via des coupures. C'est le cas par exemple des évènements $B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-)$, qui apparaissent lorsque les produits du J/ψ sont reconstruits comme le lepton ℓ et la trace t_τ. Ces évènements sont supprimés en appliquant un veto à la masse du

J/ψ dans la distribution de la masse invariante du système ℓt_τ. Deux types de bruits de fond physique dominent alors les modes $OS\ell$ et $SS\ell$. Dans le cas des modes $OS\ell$, les bruits de fond dominants sont du type $B^0 \to D^{(*)-}(\to K^{*0}\ell^-\nu_\ell)X^+$, où X peut être n'importe quel système de particules. Des désintégrations hadroniques de type $B^0 \to D^-(\to K^+\pi^-\pi^-)X$ peuvent également être reconstruites comme le signal des modes $OS\mu$ lorsque qu'un pion est identifié de manière erronée à un muon. Ces évènements sont alors rejetés par un veto sur la masse invariante $M(K^{*0}\ell)$. Dans le cas des modes $SS\ell$, les bruits de fond dominants sont des désintégrations de type $B^0 \to D^{(*)-}(\to K^{*0}t_\tau^-)\ell^+\overline{\nu}_\ell$. En particulier, les évènements dont le méson D se désintègre en $K\pi\pi$ sont rejetés par un veto sur la masse invariante $M(K^{*0}t_\tau)$.

Les bruits de fond restants sont rejetés grâce à un BDT. En particulier, huit BDTs sont entrainés séparément pour les modes $OS\ell$ et $SS\ell$, et pour Belle et Belle II, à l'aide d'échantillons de données simulées correspondant soit à du signal, soit à des évènements $B\overline{B}$ et $q\overline{q} = u\overline{u}, d\overline{d}, s\overline{s}, c\overline{c}$. Pour chaque BDT, un sous-ensemble de douze à quatorze variables est utilisé parmi un ensemble plus large comprenant des informations cinématiques sur le signal (masses invariantes $M(K^{*0}\ell)$ et $M(K^{*0}t_\tau)$, énergies du lepton ℓ et de la trace t_τ, qualité du vertex $K^{*0}\ell$), des informations sur le ROE (nombre de traces et de foyers d'impacts dans l'ECL, énergie de ces foyers, quantité de mouvement totale des traces), et des variables caractérisant la forme de l'évènement, ces dernières visant à rejeter les évènements du type $q\overline{q}$. La distribution du score de BDT pour les simulations, les données et le signal est présentée sur la Figure 3 pour le mode OSe comme exemple.

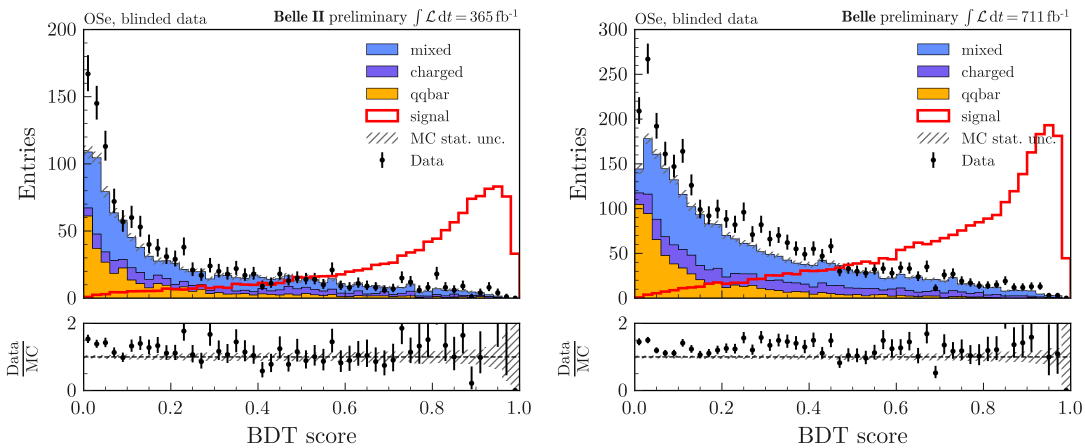


FIGURE 3 : Distribution du score de BDT pour le mode OSe pour Belle II (gauche) et Belle (droite). Les labels "mixed", "charged" et "qqbar" désignent les simulations pour les évènements de type $B^0\overline{B}^0$, B^+B^- et $q\overline{q}$ respectivement.

Après la coupure sur le score du BDT, optimisée de manière à préserver l'efficacité du signal, la dernière étape de la sélection consiste à ne garder qu'un seul candidat par évènement dans les cas où plusieurs auraient passé toutes les sélections.

Le signal peut alors être extrait des évènements restants grâce à un ajustement par maximum de vraisemblance simultané aux distributions de M_τ dans les données de Belle et Belle II. La fonction de densité de probabilité (PDF) du signal est modélisée par une fonction de Johnson [32], et le bruit de fond est décrit par un polynôme du deuxième ordre. La PDF totale est alors donnée par l'équation (0.3) :

$$\mathscr{P}_{\text{total}}(M_\tau) = \mathscr{B}(B^0 \to K^{*0}\tau^\pm\ell^\mp) \cdot \beta^{\text{exp}} \cdot \mathscr{P}^{\text{exp}}_{\text{Johnson}}(M_\tau) + N^{\text{exp}}_{\text{bkg}} \cdot \mathscr{P}^{\text{exp}}_{\text{Poly2}}(M_\tau) \tag{0.3}$$

où exp se réfère à Belle et Belle II et $\beta^{\text{exp}} = \varepsilon^{\text{exp}}_{\text{sig}} \cdot 2 \cdot f_{00} \cdot N^{\text{exp}}_{\Upsilon(4S)} \cdot \mathscr{B}(K^{*0} \to K^+\pi^-)$. Les paramètres de la PDF du signal sont ajustés à l'aide des simulations de signal et fixées dans l'ajustement aux données, alors que les coefficients de la PDF du bruit de fond sont laissé libres lors de l'ajustement aux données.

Lors de la mesure du rapport d'embranchement, certaines quantités sont sujettes à des incertitudes systématiques. C'est notamment le cas de l'efficacité de reconstruction du signal ε_{sig}, qui est mesurée à l'aide de simulations Monte-Carlo. Des incertitudes systématiques sont donc considérées sur les efficacités d'identification des particules, de reconstruction du B_{tag} par l'algorithme FEI, de reconstruction des traces, ainsi que sur l'efficacité de sélection du BDT. Des incertitudes sont également considérées sur la modélisation du signal et du bruit de fond, ainsi que sur les différents paramètres externes intervenant dans le calcul du rapport d'embranchement. Pour évaluer les incertitudes systématiques sur la sélection du BDT et la modélisation du signal, le canal de contrôle $B^0 \to D^- D_s^+ (\to K^{*0}K^+/\phi\pi^+)$ avec $K^{*0} \to K^+\pi^-$ et $\phi \to K^+K^-$ est utilisé. Le méson D_s^+ est utilisé comme proxy pour le système $K^{*0}\ell$ et la trace t_τ est obtenue en sélectionnant au hasard une trace du D^- ayant la charge adéquate. Les sources principales d'erreur systématique sont l'efficacité de la sélection du BDT, la résolution du signal et la modélisation du bruit de fond, bien que la mesure reste dominée par les incertitudes statistiques.

Finalement, les rapports d'embranchement sont mesurés dans les données et les limites supérieures à 90% de confiance sont dérivées. Dans un premier temps, les limites attendues sont mesurées grâce à un ajustement de la distribution M_τ dans une pseudo-expérience générée grâce aux données des régions extérieures à la région masqué et en faisant l'hypothèse d'absence de signal. Ensuite, après dévoilement de la région du signal, les rapports d'embranchement sont mesurés en ajustant la distribution M_τ dans les données simultanément pour Belle et Belle II. Les résultats sont visibles sur la Figure 4 pour les modes *OSe* et *SSe*. Puisqu'aucun signal significatif n'est observé, les limites supérieures à 90% de confiance sont mesurées à l'aide de la méthode CLs asymptotique [33, 34]. Les efficacités de reconstruction du signal finales, les rapports d'embranchement mesurés, ainsi que les limites attendues et observées sont présentés dans le Tableau 1.

Les limites obtenues sont les premiers résultats de recherche de violation de la saveur leptonique dans les désintégrations $B^0 \to K^{*0}\tau^\pm\ell^\mp$ effectuée dans une usine à B. Les limites ne sont pas compétitives avec les résultats préexistants de LHCb pour les modes muoniques [19], ni avec les récents résultats préliminaires de LHCb pour les modes électroniques [20]. Néanmoins, cette recherche permet de corroborer les

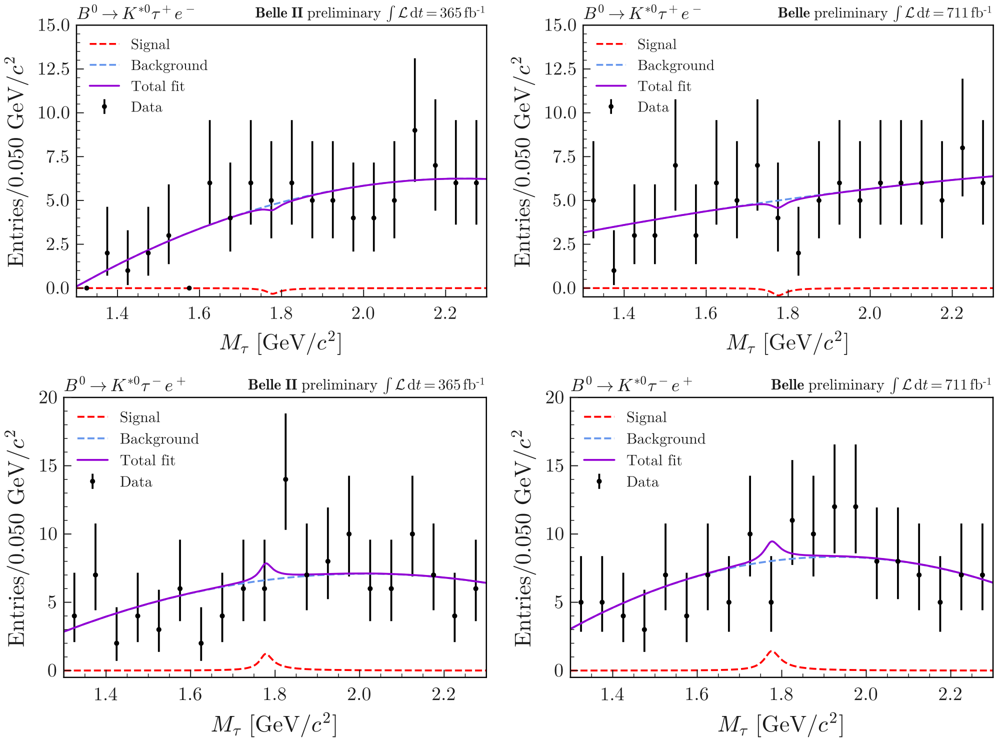


FIGURE 4 : Ajustements simultanés entre Belle II (gauche) et Belle (droite) des distributions de masse du système reculant M_τ dans les données pour les modes *OSe* (haut) et *SSe* (bas).

	$\varepsilon_{\text{sig}}^{\text{Belle II}}$ [%]	$\varepsilon_{\text{sig}}^{\text{Belle}}$ [%]	$\mathscr{B}^{\text{fit}}$ $(\times 10^{-5})$	$\mathscr{B}^{UL}_{\text{exp}}$ $(\times 10^{-5})$	$\mathscr{B}^{UL}_{\text{obs}}$ $(\times 10^{-5})$
OSe, $B^0 \to K^{*0}\tau^+ e^-$	0.075	0.046	-0.24 ± 1.46	2.8 (2.7)	2.9 (2.6)
SSe, $B^0 \to K^{*0}\tau^- e^+$	0.056	0.038	1.17 ± 2.77	4.4 (4.3)	6.4 (5.4)
$OS\mu$, $B^0 \to K^{*0}\tau^+ \mu^-$	0.060	0.053	1.07 ± 1.80	3.0 (2.9)	4.2 (3.8)
$SS\mu$, $B^0 \to K^{*0}\tau^- \mu^+$	0.051	0.024	0.48 ± 2.61	5.5 (4.9)	5.6 (5.1)

TABLE 1 : Efficacités de reconstruction du signal, rapports d'embranchement mesurés et limites supérieures attendures (exp) et observées (obs). La valeur entre parenthèses est la valeur sans incertitudes systématiques.

résultats de LHCb en ayant utilisé une méthode d'analyse différente et des données indépendantes. Elle pourra être mise à jour avec les données enregistrées dans le futur, et bénéficier des améliorations apportées aux outils d'analyse, tels que le FEI.

Par ailleurs, les limites présentées dans le Tableau 1 sont évaluées en faisant l'hypothèse que le signal suit une distribution dans l'espace des phases. Elles ont également été recalculées pour deux modèles de nouvelle physique.

Contents

Introduction

In particle physics, the Standard Model (SM) is the theory describing the fundamental particles and their interactions. This model was developed along the 20$^{\text{th}}$ century, culminating in the discovery of the Higgs boson in 2012 at CERN's Large Hadron Collider (LHC). The description of the three fundamental forces, namely the electromagnetic, weak and strong interactions, governing the interactions between fermions, and mediated by gauge bosons has proven itself many times and by many experiments. However, the SM is today considered incomplete. Indeed, many observations cannot find an explanation through it. For instance, its inability to include gravitation, to describe dark matter and dark energy or to explain the matter-antimatter asymmetry are among the most striking limitations.

An active field of particle physics is thus to test the SM by searching for deviations from its predictions. A very promising path is the search for Lepton Flavour Violation (LFV). In the SM, leptons carry an intrinsic property called flavour, which is conserved in all interactions. If charged lepton flavour violation is theoretically possible through neutrino oscillations, these processes are extremely suppressed and expected to occur at rates around 10^{-55}, far below the current experimental sensitivity. Any observation of LFV would therefore be a clear and indisputable sign of new physics (NP) beyond the SM. In addition, many NP scenarios allow for LFV processes, and so at rates relatively close to the current experiments' sensitivity. In the flavour sector, many searches are also ongoing in B meson decays, in particular in $b \to s\ell\ell^{(\prime)}$ ($\ell^{(\prime)} = e, \mu, \tau$) decays, probing for NP in various observables. Search for LFV in $B \to s\tau\ell$ ($\ell = e, \mu$) decays, specifically $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ decays for this thesis, is thus particularly interesting regarding the current flavour anomalies observed in flavour-changing neutral currents and its ability to probe third generation couplings.

In this context, the study for $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ decays is performed within the Belle II collaboration, using data collected with the Belle II detector at the SuperKEKB accelerator complex, as well as data collected with its predecessor, the Belle detector at KEKB.

This manuscript describes the work performed for this thesis and is structured in four chapters. In Chapter 1, we present the theoretical framework for the analysis, including the SM description and its limits, as well as NP scenarios for LFV and the current status of searches in this field. In Chapter 2, the Belle and Belle II experimental setups are presented, with descriptions of the KEKB and SuperKEKB colliders, and Belle and Belle II detectors. Chapter 3 is dedicated to two tracking performance studies that were performed as part of the qualification task to obtain the Belle II authorship. The first one consists in the optimisation of an algorithm used in the Belle II tracking chain, while the second one is a study of the fake tracks rate at Belle II. Finally, in

Chapter 4, we present the main part of this thesis, the search for $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ lepton flavour violating decays using hadronic $B-$tagging at Belle and Belle II.

1. The Standard Model, New Physics and Lepton Flavour Violation

Summary

The Standard Model (SM) is the theory describing the elementary particles and their interactions. Despite being the most successful theory in particle physics, it is not complete and does not explain some of the phenomena observed in nature. Therefore, current searches in particle physics aim at finding signs of New Physics (NP) beyond the standard model (BSM). Such evidences could arise from Lepton Flavour Violation (LFV), which is forbidden in the SM but predicted by many NP models and is the subject of this thesis. In this chapter, we will introduce the SM and its limitations. After a focus on flavour in the SM, we will discuss the lepton flavour violation, and its implications for new physics.

1.1. The Standard Model of particle physics and its limitations

The Standard Model of particle physics (SM) is a model based on a quantum field theory describing the elementary particles and their interactions. In this theory, particles are represented as excitations of their corresponding fields, and are classified into two distinct categories characterised by their spin: fermions with spin 1/2, and bosons with spin 0 or 1. Fermions are the matter (and anti-matter) particles, and bosons are the force carriers of the three interactions described in the SM: the electromagnetic, weak and strong interactions. The interactions are thus represented through the exchange of virtual particles. The SM features a set of symmetries and conservation laws governing the interactions between particles. It is a gauge theory[1], and it is invariant under the discrete CPT symmetry[2]. The success of the SM lies in its ability to describe a wide range of phenomena at the subatomic scale, and its validation through numerous experimental results. However, despite being largely supported by experiments, the SM shows some weaknesses when it comes to explaining certain observed phenomena, such as dark matter or matter/anti-matter asymmetry in the Universe.

In this section, we will first introduce the theoretical aspects of the elementary particles and their interactions in the SM. Then its limitations will be discussed, and we will briefly summarise how these limitations open the path for searches for new physics. For the theoretical overview of the SM, various references have been used, including textbooks, theses and lecture notes [1–4, 35, 36].

1.1.1. Elementary particles

The Standard Model is composed of 37 fundamental particles and anti-particles: 12 fermions and their anti-particles, distinguished into 6 quarks (+ 6 anti-quarks) and 6 leptons (+ 6 anti-leptons [3]); 12 gauge bosons mediating the electromagnetic, weak and strong interactions; and the Higgs boson, known to give their masses to the other particles of the SM. Although particles having a mass is predicted theoretically through the interactions with the Higgs field, their mass values, as well as other SM properties described in Section 1.2, are not predicted by the theory and must be measured experimentally. A summary of the SM particles and their properties is given in Figure 1.1.

Fermions

Fermions are the matter particles of the SM. They are characterised by a spin of 1/2 and are distinguished into quarks and leptons, each with three generations of different mass.

There are 6 quark flavours, which can be separated into the “up-type” category with a +2/3 electric charge (up u, charm c, top t) and the “down-type” category with a -1/3 charge (down d, strange s, bottom b). In addition to their flavour and electric charge, quarks also carry a “colour” charge that may take three values, commonly denoted

[1] This means that the Lagrangian density of the theory is invariant under local gauge transformations of the fields. It is detailed in Section 1.1.2.

[2] CPT stands for Charge, Parity and Time reversal. This symmetry is defined in Section 1.1.3.

[3] In the SM, neutrinos and anti-neutrinos are different particles but some theories consider neutrinos as Majorana particles, i.e. being their own anti-particles [37].

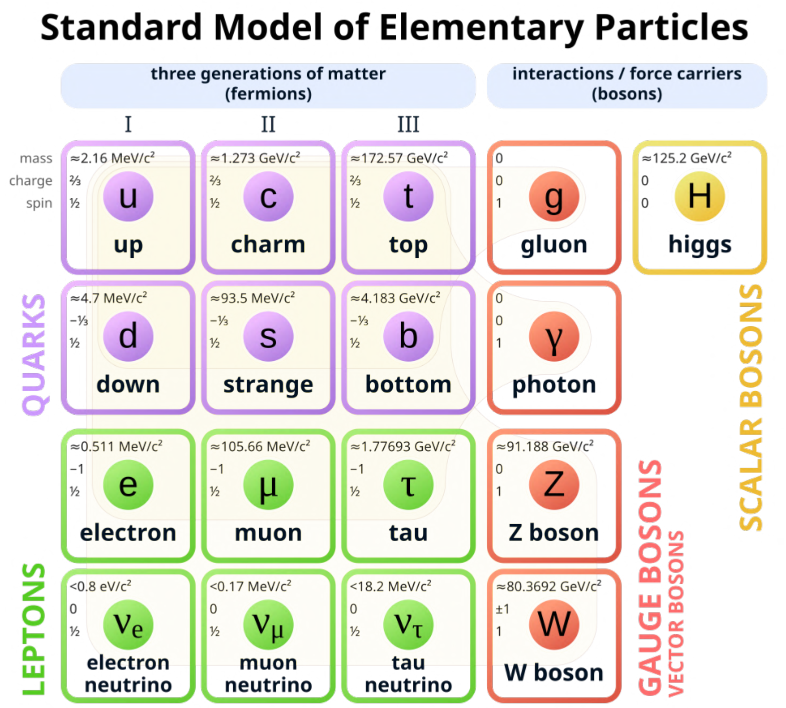


Figure 1.1: The Standard Model of particle physics and elementary particles properties (From Ref. [38])

as "red", "green" and "blue", or anti-colour for their anti-matter counterparts, which is responsible for the strong interaction described by quantum chromodynamics (QCD). Quarks are thus sensitive to the three fundamental interactions of the SM: the electromagnetic, weak and strong forces. In nature, quarks are always found in bound states, called hadrons, which are colour-neutral combinations of quarks bound by the strong interaction. An exception is observed for the top quark, which is so massive that it decays before hadronising[4]. The hadrons composed of an odd [5] number of quarks and anti-quarks (typically three) are called baryons (e.g. protons and neutrons), while those composed of even [6] numbers (quark/anti-quark pairs, typically one pair) are called mesons (e.g. pions and kaons). However, only the light quarks u and d are stable in hadrons, forming the only two stable baryons: the proton (uud) and the neutron [7] (udd) at the basis of the atomic structure. The other hadrons are unstable and decay into lighter particles. This is the case of the B mesons B^0 ($d\bar{b}$) and B^+ ($u\bar{b}$), (and their respective anti-mesons) for which the decay of the former is the subject of this thesis. In hadron decays, the quark flavour is not conserved, as it is not conserved

[4]Some recent studies by the CMS experiment suggest the observation of a $t\bar{t}$ pseudo-bound state [39].
[5]Most of the baryons contain 3 quarks but exotic baryons with 5 quarks and anti-quarks, called pentaquarks, are theoretically possible and have been observed at the LHC [40, 41].
[6]Similarly, mesons are usually composed of a quark and an anti-quark but exotic mesons with 4 quarks and anti-quarks (tetraquarks) have been observed [42–44].
[7]Neutrons are only stable in atoms nuclei.

by the weak interaction (while it is by strong and electromagnetic interactions), but the baryon number (1 for baryons, -1 for anti-baryons and 0 for mesons) is.

The other half of the fermions are the leptons, also divided into two charge categories with each three lepton flavours: the charged leptons (electron e, muon μ, tau τ) with a -1 charge (+1 for the anti-leptons) and the neutral leptons (neutrinos ν_e, ν_μ, ν_τ) with no charge. Contrary to quarks, the two leptons of the same family (e.g. e and ν_e) carry the same flavour number, which is conserved in interactions involving leptons. Leptons are not sensitive to the strong interaction and only interact through the weak force and the electromagnetic one for the charged leptons. An additional difference with quarks is that, although the charged leptons are also massive and organised in increasing mass generations, their corresponding neutrinos are massless in the SM [8].

In field theory, fermions (as any spin 1/2 fields) are represented by Dirac spinors ψ and their Lagrangian density for a free particle is given by:

$$\mathscr{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{1.1}$$

where $\bar{\psi} = \psi^\dagger\gamma^0$ is the Dirac conjugate of the spinor, m is the mass of the particle, ∂_μ is the derivative operator and γ^μ are the Dirac matrices. The Einstein summation convention and the natural units $\hbar = c = 1$ are implied throughout this chapter.
Fermions having a left-handed and a right-handed component (left-handed only for neutrinos), it is convenient to decompose the Dirac spinor ψ into two Weyl spinors ψ_L and ψ_R, for the left-handed and right-handed components, respectively:

$$\psi = \begin{pmatrix}\psi_L\\ \psi_R\end{pmatrix} = \begin{pmatrix}\psi_L\\ 0\end{pmatrix} + \begin{pmatrix}0\\ \psi_R\end{pmatrix}, \quad \text{with} \quad \begin{pmatrix}\psi_L\\ 0\end{pmatrix} = P_L\psi = \frac{\mathbb{1}-\gamma_5}{2}\psi, \quad \begin{pmatrix}0\\ \psi_R\end{pmatrix} = P_R\psi = \frac{\mathbb{1}+\gamma_5}{2}\psi, \tag{1.2}$$

where P_L, P_R are the chirality projection operators, $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is the chirality operator and $\mathbb{1}$ is the identity matrix.
In interactions, particles are called left-handed or right-handed depending on which component contributes to the interaction.

Bosons

Bosons are particles characterised by an integer spin with value 0 (scalar bosons) or 1 (vector bosons) for elementary bosons. In the SM, the elementary vector bosons (also called gauge bosons) carry the three fundamental interactions and the Higgs boson is the only elementary scalar field and is responsible for the mass of the other elementary particles.

The photon (γ) is the mediator of the electromagnetic interaction, acting on charged particles. It is a massless vector field described by the Maxwell Lagrangian:

$$\mathscr{L}_{\text{Maxwell}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.3}$$

[8] The observation of neutrino oscillations implies that neutrinos have a mass. The neutrino oscillation and its implications are discussed in Section 1.2.

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength tensor and A_μ is the electromagnetic potential.

The weak interaction is mediated by the $W^\pm$ and Z^0 bosons, which are massive vector bosons. On one hand, interactions with the W bosons are called charged currents and are responsible for the charged to neutral lepton transitions and vice versa in the lepton sector and flavour changing transitions in the quark sector. On the other hand, the Z boson mediates neutral currents between same flavour and same charge leptons and quarks. The expression of a free massive vector field A_μ is obtained by adding a mass term to the massless Lagrangian, giving the Proca Lagrangian:

$$\mathscr{L}_{\text{Proca}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} m^2 A_\mu A^\mu, \tag{1.4}$$

where m is the mass of the vector boson.

In the strong interaction, the gauge bosons are the 8 gluons (g), which are massless vector fields. Since the strong interaction is described by quantum chromodynamics (QCD) with three different conserved colour charges (and three corresponding anti-colours), gluons must carry a colour charge corresponding to one colour and one anti-colour. This leads to 9 possible colour/anti-colour combinations. However, only 8 are linearly independent as the ninth combination can be written such that it corresponds to a colour-neutral state, which does not participate in strong interaction.

Finally, the Higgs boson (H) is the massive scalar boson that finalises the SM by providing the mass to the other elementary particles (quarks, charged leptons, W and Z bosons) through the Higgs mechanism. The masses of the particles are proportional to their coupling to the Higgs field. The Lagrangian density of a free real scalar field is given by the Klein-Gordon Lagrangian:

$$\mathscr{L}_{\text{KG}} = \frac{1}{2} (\partial_\mu \phi \partial^\mu \phi - m^2 \phi^2), \tag{1.5}$$

where ϕ is the scalar field and m is its mass.

1.1.2. Interactions in the SM

In the previous section, we defined and described the fields populating the space-time. Here we will study how these fields interact with each other. The interactions between the fields correspond to the fundamental forces of nature. However, among the four fundamental forces (electromagnetic, weak, strong and gravitation), only the first three are described by the SM. The inability of the SM to describe gravitation, resulting from the incompatibility of quantum field theory with general relativity, is one of the limitations of the SM and will be briefly discussed in the next section.

The SM is a gauge theory. This means that the SM Lagrangian density is invariant under local gauge transformations of the fields. By introducing these gauge symmetries, interaction terms between the fields are generated, thus providing the description of the SM interactions. In particular, the SM Lagrangian density is invariant under the gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$, where $SU(3)_c$ is the gauge group of the strong interaction, characterised by the colour c, and $SU(2)_L \times U(1)_Y$ is the gauge group resulting from the combination of the electromagnetic and weak interactions, referred to as electroweak (EW). The subscript L refers to the left-handed chirality of the weak interaction and Y is the weak hypercharge.

Electromagnetic interaction

We stated in the previous section that the Lagrangian density of a free fermion is given by the Dirac Lagrangian (1.1). When applying a gauge transformation from the electromagnetic group $U(1)_{EM}$ to the Dirac Lagrangian, additional terms are generated making the Lagrangian non-invariant. In order to restore the gauge invariance, one must introduce new terms to the original Lagrangian to cancel out the non-invariant terms after the gauge transformation. This is done by changing the derivative operator ∂_μ into a covariant derivative operator D_μ, defined as:

$$D_\mu = \partial_\mu + iqA_\mu. \tag{1.6}$$

This new term involving a vector field A_μ, the free Lagrangian of a vector field must be added too. Since only the massless vector field is invariant under the gauge transformation, the Maxwell Lagrangian is considered and the resulting gauge boson of the electromagnetic interaction has to be massless. The Lagrangian density of the electromagnetic interaction (also called quantum electrodynamics QED) is therefore given by:

$$\mathscr{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.7}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength tensor, $D_\mu = \partial_\mu + ieA_\mu$ is the covariant derivative operator, e is the electric charge and ψ is the Dirac spinor of the fermion field. When distributing the covariant derivative D_μ, the decomposition of the Lagrangian into free spinor term, an interaction term and a free vector field term is clearly visible. We recover here the well known massless photon mediating the electromagnetic interaction characterised by the electron charge e.

Electroweak unification

Similarly to the electromagnetic interaction, one can require the Dirac Lagrangian to be invariant under the gauge transformation of the weak interaction group $SU(2)_L$, whose generators T^a (weak isospin operators) are expressed in terms of the Pauli matrices σ^a such that $T^a = \sigma^a/2$. The gauge invariance is then satisfied by introducing three gauge bosons W^1, W^2 and W^3. Given the 2×2 matrix nature of the group generators, the spinor fields ψ must be written with two components. In addition, the charged current weak interaction is observed to couple only to left-handed particles and right-handed anti-particles. For this reason, left-handed (LH) particles are placed in doublets with weak isospin $I_W = 1/2$ and right-handed (RH) particles in singlets with weak isospin $I_W = 0$:

$$\ell_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \quad q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \quad \text{and} \quad \ell_R = e_R, \quad u_R, \quad d_R, \tag{1.8}$$

where e represents the charged leptons e, μ, τ, ν is the corresponding neutrino, and u and d are the up-type and down-type quarks, respectively.

By requiring local gauge invariance under the $SU(2)_L$ group, one can identify the gauge bosons W^1 and W^2 as linear combinations of the known charged $W^\pm$ bosons. The gauge invariance also implies the existence of a weak neutral current, carried by a neutral gauge boson W^3. On the other hand, since the RH particles and LH anti-particles are not affected by the $SU(2)_L$ gauge transformation, they do not couple to the $W^{1,2,3}$ gauge bosons. The identification of the W^3 boson with the known Z^0 boson would thus seem natural. However, this is in contradiction with experiments showing that, although it does so with different strengths, the Z^0 boson couples to both LH and RH particles.

This issue is solved by the unification of the electromagnetic and weak interactions, developed by Weinberg, Salam and Glashow. This is done by replacing the $U(1)_{EM}$ gauge symmetry, with a $U(1)_Y$ gauge symmetry, where Y is the weak hypercharge. From the invariance under the $U(1)_Y$ arises a new gauge boson B_μ coupling to the weak hypercharge analogously to the photon coupling to the electric charge. The photon and Z^0 boson (expressed as A_μ and Z_μ) are then obtained as linear combinations of the W^3_μ and B_μ gauge bosons with the introduction of a new parameter θ_W, called the Weinberg angle or weak mixing angle. In addition, the weak hypercharge Y, the third component of the weak isospin T^3 and the electric charge Q are related by the Gell-Mann-Nishijima relation $Q = T^3 + Y/2$. The EW bosons $W^\pm_\mu$, Z_μ and A_μ are then expressed as:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \mp i W^2_\mu), \\ Z_\mu &= \cos\theta_W W^3_\mu - \sin\theta_W B_\mu, \\ A_\mu &= \sin\theta_W W^3_\mu + \cos\theta_W B_\mu. \end{aligned} \tag{1.9}$$

The resulting Lagrangian density of the electroweak interaction is given by:

$$\begin{aligned} \mathscr{L}_{\text{EW}} =& \bar{\ell}_L i\gamma_\mu D^\mu \ell_L + \bar{e}_R i\gamma_\mu D^\mu e_R \\ &+ \bar{q}_L i\gamma_\mu D^\mu q_L + \bar{u}_R i\gamma_\mu D^\mu u_R + \bar{d}_R i\gamma_\mu D^\mu d_R \\ &- \frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \end{aligned} \tag{1.10}$$

where D_μ is the covariant derivative operator defined as:

$$D_\mu = \partial_\mu + i g_W T^a W^a_\mu + i g' Y B_\mu. \tag{1.11}$$

$W^a_{\mu\nu}$ and $B_{\mu\nu}$ are the field tensors of the $W^a (a = 1,2,3)$ and B gauge bosons, respectively. T and Y are the weak isospin operators and weak hypercharge, with respective couplings g_W and g'.

Strong interaction

The strong interaction is described by QCD, which is a gauge theory based on the $SU(3)_c$ gauge group, where c refers to the colour charge. This group has eight generators, corresponding to the eight massless gluons. Similarly to the QED Lagrangian, the QCD Lagrangian can be expressed as the sum of a modified Dirac Lagrangian and a free massless vector Lagrangian, with the subtlety that one has to sum over the quark colours and flavours:

$$\mathscr{L}_{\text{QCD}} = \sum_{f=1}^{6} \bar{q}_{f\alpha}(i\gamma_\mu D^\mu_{\alpha\beta} - m_f \delta_{\alpha\beta}) q_{f\beta} - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{1.12}$$

where q designates the quark fields, f corresponds to the flavour, α, β to the colour, a to the gluon indices, and $\delta_{\alpha\beta}$ is the Kronecker delta. The covariant derivative operator $D^\mu_{\alpha\beta}$ is defined as:

$$D^\mu_{\alpha\beta} = \partial^\mu \delta_{\alpha\beta} + i g_s \frac{1}{2} \lambda^a_{\alpha\beta} G^a_\mu, \tag{1.13}$$

where g_s is the strong coupling constant and $\lambda^a_{\alpha\beta}$ are the Gell-Mann matrices, which are the generators of the $SU(3)_c$ group. We note here that the quark mass term appears by analogy with the QED Lagrangian, but we didn't define yet any mechanism to provide the quark masses, neither generate the charged leptons and EW boson masses.

Interaction with the Higgs field

The masses of the particles, fermions or bosons, arise from the interaction of their fields with the Higgs field.

Let's first consider the interaction of a Higgs field ϕ with the EW gauge bosons. We can write the scalar field as a $SU(2)_L$ doublet $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$. The Lagrangian of this Higgs field can be written as:

$$\mathscr{L}_{\text{Higgs}} = |D_\mu \phi|^2 - V(|\phi|), \quad V(|\phi|) = \mu^2 |\phi| + \lambda |\phi|^4, \tag{1.14}$$

where D_μ is the covariant derivative operator defined in Eq. (1.11), $V(|\phi|)$ is the Higgs potential, $|\phi|^2 = \phi^\dagger \phi$, $\mu \in \mathbb{C}$ and $\lambda \in \mathbb{R}*_+$.

If $\mu^2 \geq 0$, the potential $V(|\phi|)$ has a minimum at $|\phi| = 0$, i.e. the Higgs field is not excited. However, if $\mu^2 < 0$, the potential $V(|\phi|)$ has an infinite set of minima at $|\phi| = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}$ where v is called the vacuum expectation value (vev). Choosing such minimum corresponds to a spontaneous symmetry breaking (SSB) of the global $U(1)$ symmetry. Since the photon should remain massless after symmetry breaking, it should not couple to the Higgs field and therefore the potential minimum should correspond to a non-zero vev only for the neutral component ϕ^0 of the scalar field ϕ. The Higgs fields can then be written in the so-called "unitary gauge" as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.15}$$

By expanding the Higgs Lagrangian with this new expression, we can identify the mass terms of the gauge bosons as well as their interaction terms with the Higgs boson h. First, from the $(W^1_\mu W^{1\mu} + W^2_\mu W^{2\mu})$ term, we obtain the mass term of the $W^\pm$ bosons $m_W = \frac{1}{2} g_W v$. Then, by moving from the (W^3_μ, B_μ) basis to the (Z_μ, A_μ) basis and using the relations (1.9), we can identify the Z^0 boson mass $m_Z = \frac{1}{2}\sqrt{g_W^2 + g'^2}\, v = \frac{1}{2}\frac{g_W}{\cos\theta_W} v$ and the photon mass $m_A = 0$. Finally, the mass term of the Higgs boson is obtained from the $-\mu^2|\phi|^2$ term, giving $m_H = \sqrt{2\lambda}\, v$.

This mass generation Higgs mechanism can also be used to generate the mass of the fermions. Given the chirality properties of the fermions, their mass term in the Dirac Lagrangian is $-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$, which is not invariant under the $SU(2)_L \times U(1)_Y$ gauge transformation. Since the Higgs field is placed in a $SU(2)$ doublet, one can apply a $SU(2)$ local gauge transformation such that the combination of the Higgs field ϕ with a left-handed doublet L and a right-handed singlet R and its hermitian conjugate $\bar{L}\phi R$ and $\bar{R}\phi^\dagger L$ are invariant under $SU(2)_L$ and $U(1)_Y$ gauge transformations. After spontaneous symmetry breaking, the fermions mass terms and interactions with the Higgs field can be expressed as:

$$\mathscr{L}_f = -\frac{g_f}{\sqrt{2}} v(\bar{f}_L f_R + \bar{f}_R f_L) - \frac{g_f}{\sqrt{2}} h(\bar{f}_L f_R + \bar{f}_R f_L), \tag{1.16}$$

where g_f is called the Yukawa coupling of the fermion f with the Higgs field. The mass of the fermion f is then given by $m_f = \frac{g_f}{\sqrt{2}} v$. This can be summarised in the Yukawa Lagrangian:

$$\mathscr{L}_{\text{Yukawa}} = -\sum_{i,j} (g^\ell_{ij} \bar{\ell}_{iL} \phi \ell_{jR} - g^u_{ij} \bar{q}_{iL} \phi^\dagger u_{jR} - g^d_{ij} \bar{q}_{iL} \phi d_{jR}) + h.c., \tag{1.17}$$

where i, j correspond to the lepton or quark generations, $g^{\ell,u,d}_{ij}$ are the Yukawa couplings of the leptons and quarks with the Higgs field, $\phi^\dagger = i\sigma^2\phi$ is the $SU(2)$ conjugate of ϕ and $h.c.$ is the hermitian conjugate of the Lagrangian. Since the non-zero vev of the Higgs field is obtained in the lower component of the doublet, the fields' combination with ϕ can only generate masses for the lower components of the doublets, i.e. the charged lepton and down-type quarks. The masses of the up-type quarks are thus obtained using the $SU(2)$ conjugate of the Higgs field, hence the $\phi^\dagger$ term in the Lagrangian for up-type quarks.

Summary

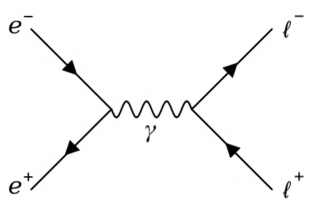
The Standard Model theory, defined as a gauge theory invariant under the $SU(3)_c \times$

$SU(2)_L \times U(1)_Y$ group, can be summarised by the Lagrangian density:

$$\mathscr{L}_{\text{SM}} = \mathscr{L}_{\text{EW}} + \mathscr{L}_{\text{QCD}} + \mathscr{L}_{\text{Higgs}} + \mathscr{L}_{\text{Yukawa}}, \tag{1.18}$$

where we remove from the Lagrangian densities described above the redundant terms, i.e. the EW boson fields between the EW and Higgs Lagrangians, and the fermion kinetic and mass terms between the Yukawa, EW and QCD Lagrangians.

To express the interactions of the particles in time, we can use the Feynman rules, which are a set of rules allowing us to calculate the probability amplitudes of particle interactions. A convenient way to represent these interactions is through Feynman diagrams, where time is usually represented from left to right, the outer legs represent real particles, the inner legs represent virtual particles that mediate the interaction, and the points where the legs meet are called vertices' couplings. The vertices are associated with the interaction strength, and are thus proportional to the coupling constant of the interaction involved. The probability of a process is thus proportional to the product of the vertices. Feynman diagrams can be called "tree-level" or "loop-level" depending on the presence or not of virtual particles loops as can be seen in Figure 1.2. For electromagnetic and weak interactions, the coupling constants being smaller than 1, the tree-level diagrams are significantly more probable than loop-level diagrams in the absence of other sources of suppression.



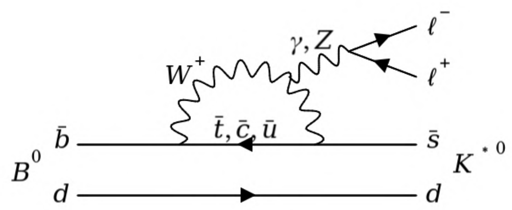


Figure 1.2: Examples of Feynman diagrams for the tree-level (left) representing the $e^+e^- \to \ell^+\ell^-$ process and loop-level (right) representing the $B^0 \to K^{*0}\ell\ell$ process with a W loop in the $b \to s$ transition. The right diagram is also called "penguin" diagram.

1.1.3. Discrete symmetries

In addition to the gauge symmetries, the SM is also invariant under discrete symmetries, which are transformations that change the sign of certain quantum numbers. In particular the SM is invariant under the CPT symmetry, which is the combination of three discrete symmetries: charge conjugation C, parity transformation P and time reversal T.

The charge conjugation C is a transformation changing the sign of all charges of the particles (electric, flavour, colour...), thus transforming the particles into anti-particles. The parity transformation P is a spatial inversion changing the sign of the spatial coordinates, i.e. it transforms a particle into its mirror image. Finally, the time reversal

T is a transformation changing the direction of time. As a consequence of the CPT symmetry, the anti-particles can be seen as particles travelling backward in time, hence their representation with an inverted arrow in Feynman diagrams. Although the SM is invariant under CPT, it is not necessarily the case for each of the three symmetries separately. For instance, the weak interaction is known to violate the combination CP of parity and charge conjugation symmetries.

1.1.4. Limits of the SM

The Standard Model is a well-established theory that has been tested and validated by many experiments. The discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations [5, 6] provided the missing experimental piece to confirm the SM picture. However, the SM is not yet a complete theory and several experimental observations cannot be explained by it.

Probably the most well known issue with the SM is its inability to include gravitation. Indeed, the quantum field theory framework used to describe the SM is incompatible with general relativity, which describes gravitation as a curvature of space-time. If some theories postulate the existence of a graviton, which would be the mediating particle of the gravitational interaction, the SM does not provide any reliable description of this particle and interaction. In addition, the SM does not explain the existence of dark matter and dark energy, constituting respectively 27% and 68% of the energy content of the universe according to cosmological observations [45]. Searches for dark matter candidates are a very active field of research [46], but no conclusive evidence has been found yet.

Another limit is that the SM doesn't provide an explanation for the matter-antimatter asymmetry observed in the universe. The SM predicts that matter and antimatter should have been produced in equal amounts during the Big Bang, but the universe is predominantly made of matter, and the CP violation observed in the electroweak sector is not sufficient to explain this asymmetry. Also related to CP violation and known as the "strong CP problem", the strong interaction is observed to conserve CP symmetry, but the QCD Lagrangian can be parametrised such that a complex phase encoding the CP violation is introduced[9]. This phase being very close to 0 is thus only an experimental result and is not constrained by the theory.

Finally, many parameters of the SM seem arbitrary and are not predicted or constrained by the theory. It is the case for instance of the masses of the different particles, their hierarchy between the generations, or the reason why there are three fermion generations and the connections between them.

All these questions and missing pieces open the way to searches for new physics beyond the SM (BSM). This is done through the experimental search for direct or indirect evidences of new particles or new interactions. For instance, any significant

[9] Such complex phase, also called CP phase, is introduced in the EW sector and its non-zero value is responsible for the CP violation observed in weak interactions. More details are given in Section 1.2.1.

deviation from the SM predictions in experimental measurement would be a clear sign of NP. In parallel, many theoretical models are proposed to explain the limits of the SM and are evolving with the increasingly precise experimental measurements.

1.2. Flavour in the Standard Model

Flavour is a category of quantum numbers associated to quarks and leptons in the SM. In this section, we will define and characterise the flavour of a fermion beginning with quarks and then addressing leptons. Finally, we will discuss the flavour anomalies observed in some experimental measurements, their implications, and how flavour properties can be used to probe new physics.

1.2.1. Flavour in the quark sector

In the quark sector, each quark carries its own flavour quantum number, corresponding to the 6 quark types u, d, s, c, b and t. Change of flavour is only permitted via the weak interaction, through the exchange of a W boson. Since the W boson is electrically charged, only charged current transitions, i.e. transitions between up-type quarks and down-type quarks or vice versa, are allowed at tree level. The existence of flavour changing transitions, also called quark mixing, is a consequence of the quark mass eigenstates being different from their flavour eigenstates. The rotation matrix from the mass eigenstates to the flavour eigenstates is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{pmatrix}, \tag{1.19}$$

where the V_{ij} complex matrix element corresponds to the transition amplitude of the quark i to the quark j and the numerical values correspond to the absolute values of the CKM matrix elements. The matrix is not diagonal but flavour changing transitions within the same generation are strongly favoured. The CKM matrix elements are not predicted by the theory and have to be determined experimentally.

The CKM matrix can be parametrised in terms of three mixing angles θ_{12}, θ_{23} and θ_{13} and one complex phase δ, encoding the CP violation in the quark sector:

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.20}$$

where $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$. This parametrisation is called the "standard parametrisation".

Flavour transitions between quarks of the same charge such as $b \to s$ transitions, called flavour changing neutral current (FCNC), are only possible through loop processes involving a W boson and a virtual quark of the other charge, as illustrated in Figure 1.2 (right). Because of the < 1 value of the weak interaction coupling constant and the small amplitude of generation changing CKM elements, these processes are suppressed and therefore relatively rare.

1.2.2. Flavour in the lepton sector

In case of leptons, the flavour quantum number is associated with the three lepton families e, μ and τ. This means that the charged lepton ℓ and neutrino ν_ℓ of the same generation have the same lepton flavour number $L_\ell = +1$ for particles and $L_\ell = -1$ for anti-particles. For such leptons ℓ or ν_ℓ, the lepton numbers $L_{\ell'}$ with $\ell' \neq \ell$ are equal to 0. In the SM, the lepton flavour number is conserved through all interactions and lepton transitions through the weak interaction are only allowed between leptons of the same flavour, i.e. $\ell \leftrightarrow \nu_\ell$. We call Lepton Flavour Violation (LFV) any process that violates this conservation law.

Neutrino oscillations

We stated in the previous section that neutrinos are massless in the SM. Indeed, since there is no right-handed neutrinos, the Yukawa coupling does not apply to neutrinos, and therefore they do not acquire mass through the Higgs mechanism.

However, hints of neutrino oscillations, i.e. neutrinos changing flavour as they propagate in space, arose in the 1970s from the observation of a deficit in the flux of solar neutrinos [47]. In 1998, the Super-Kamiokande experiment provided the first experimental evidence of neutrino oscillations [7], which was confirmed by several experiments using various neutrino sources [8, 9] since then.

The oscillation phenomenon can only occur if neutrinos have non-zero masses that differ from one another. The neutrinos flavour and mass eigenstates can therefore be related through a mixing matrix similar to the CKM matrix and called the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix:

$$U_{\text{PMNS}} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}, \quad \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U_{\text{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \tag{1.21}$$

where ν_i, $i = 1,2,3$ are the mass eigenstates and ν_ℓ, $\ell = e,\mu,\tau$ the flavour eigenstates, and $U_{\ell i}$ are the transition amplitudes between the flavour and mass eigenstates.

The probability of transition from a neutrino of flavour ℓ to a neutrino of flavour ℓ' thus depends on the PMNS matrix elements involving ℓ and ℓ' as well as the squared mass difference $\Delta m^2_{ij} = m^2_i - m^2_j$ between the mass eigenstates i and j, for $i, j = 1,2,3$.

The mass differences can be measured through oscillation observations, but the absolute neutrino masses are so small that they cannot be measured with the current experimental sensitivity, and the mass ordering hierarchy is still an open question. From the theoretical point of view, the question of the mass origin is also unanswered. One of the most common explanations is the so-called “seesaw mechanism”, which postulates the existence of heavy right-handed neutrinos that couple to the Higgs field, acquiring in that way a very large mass, inversely proportional to the known left-handed neutrino masses. No evidence of such right-handed neutrinos has been observed, however, and other complex mechanisms could be involved in the origin of neutrino masses.

Taking into account neutrino oscillations, lepton flavour violation is theoretically permitted in the SM. Neutral LFV exists by definition as neutrino oscillations involve changes in flavour. Charged LFV (LFV between charged leptons) is theoretically possible in a process where a neutrino is emitted, undergoes oscillation and then is reabsorbed by a W boson, resulting in the emission of a charged lepton of a different flavour, as can be seen in Figure 1.3. However, the probability for such process to appear is extremely low and much beyond the reach of current experimental sensitivity, estimated to be around 10^{-55} [10]. Any observation of LFV in the charged lepton sector would thus be a clear evidence for physics beyond the standard model.

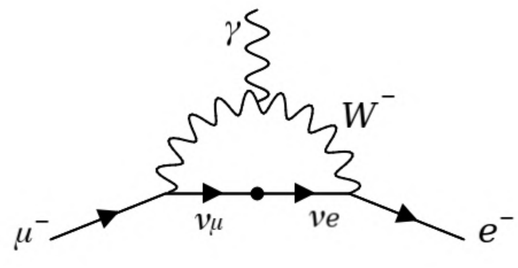


Figure 1.3: Feynman diagram for charged lepton flavour violation in $\mu \to e\gamma$ decay through neutrino oscillation (denoted by a dot).

1.2.3. LFV and tensions in LFU – Flavour anomalies

Another SM property related to lepton flavour is lepton flavour universality (LFU). In the SM Lagrangian, the coupling of the gauge bosons to the leptons is independent of their flavour. However, this universal behaviour is an accidental symmetry of the SM and is not constrained by any wider symmetry principle. Also, in case of lepton flavour non-universality, lepton flavour conservation is not guaranteed.

As a consequence of LFU, the decay rates of leptonic and semi-leptonic processes with different lepton flavours in the final state are expected to be equal up to phase spaces factors. This provides a powerful test of the SM, as any significant deviation from the expected decay rates would indicate the presence of unknown processes. The most common LFU tests are the ratios of decay rates of leptonic and semi-leptonic B meson decays, such as the so-called "R ratios" $R_{D^{(*)}}$ and $R_{K^{(*)}}$. Studying ratios of decay rates is particularly interesting because it allows to cancel a large part of experimental and theoretical uncertainties, leading to very precise measurements. In particular, one can distinguish the LFU tests with semi-leptonic flavour-changing charged currents $b \to c\ell\nu_\ell$ and the ones with leptonic flavour-changing neutral currents $b \to s\ell\ell$. In the first case, the observables are the $R_{D^{(*)}}$ ratios given by Eq. (1.22), and usually probe the $\tau - \ell$ universality.

$$R_{D^{(*)}} = \frac{\mathscr{B}(B \to D^{(*)}\tau\nu_\tau)}{\mathscr{B}(B \to D^{(*)}\ell\nu_\ell)}, \quad \ell = e, \mu. \tag{1.22}$$

In the second case, the $R_{K^{(*)}}$ ratios given by Eq. (1.23) are considered and probe the $\mu - e$ universality.

$$R_{K^{(*)}} = \frac{\mathscr{B}(B \to K^{(*)}\mu^+\mu^-)}{\mathscr{B}(B \to K^{(*)}e^+e^-)}. \tag{1.23}$$

If some tensions were present in both tests for a long time [11–14, 48, 49], recent results in $R_{K^{(*)}}$ measurements showed good agreement with the SM expectations [50, 51]. A summary of the tests and SM deviations is presented in Figure 1.4.

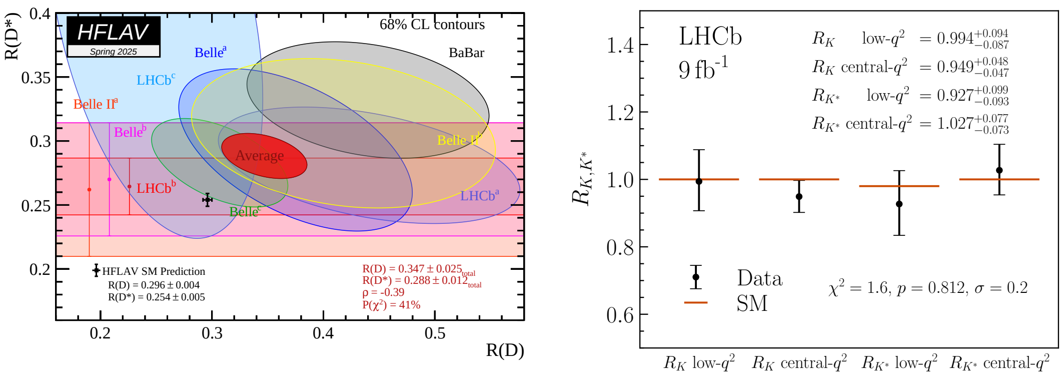


Figure 1.4: Summary of the $R_{D^{(*)}}$ (left [52]) and $R_{K^{(*)}}$ (right [51]) measurements and SM predictions.

Some tensions are nonetheless still present in the $R_{D^{(*)}}$ measurements, with a deviation of 3.8σ between the experimental average and the SM prediction [15, 52], or in other observables such as angular distributions in $B \to K^*\ell^+\ell^-$ decays [53–55], which are sensitive to the same kind of NP contributions as the $R_{K^{(*)}}$ ratios.

Another flavour anomaly probing observable is the $B^+ \to K^+\nu\bar{\nu}$ decays, for which an enhanced branching fraction at 2.7σ above the SM predictions was recently measured by the Belle II experiment [16]. Here, the lepton flavour universality or violation is not directly tested, but the BSM contributions in FCNC that could explain this observation could also be responsible for LFV processes [56].

Many observables are thus used to probe physics beyond the SM through flavour measurements. Among them, the study of LFV decays benefits from a very clean experimental signature and direct evidence for NP in case of signal observation.

1.3. Lepton Flavour beyond the Standard Model and status of the searches of LFV

Many new physics models include new interaction mediators that do not restrain LFV processes to appear. In some of them, the LFV processes are even predicted to be enhanced to branching fractions close to the current experimental sensitivity. In addition to probing directly possible NP, measuring LFV processes is thus also crucial to help to constrain the theoretical models. In this section, we will first present a brief overview of some theoretical models that can allow LFV processes. Then, the current status of LFV searches, in particular in B decays will be presented, and we will discuss the motivations to study the $B^0 \to K^{*0}\tau\ell$ decay modes.

Effective Field Theory

Before focusing on specific BSM models, let's first discuss the theoretical interpretation of LFV processes. To do so, a common approach is to use an effective field theory (EFT) framework [57, 58], which allows one to describe the low energy limit of a more fundamental theory. The idea behind effective theories is that a complex interaction, for instance a complex loop process in a Feynman diagram, can be represented at lower energy scales by a simpler theory, resulting in a point-like interaction. In this framework, the effective Lagrangian is expressed as a sum of the SM Lagrangian and a series of higher-order operators describing the new physics contributions:

$$\mathscr{L}_{\text{EFT}} = \mathscr{L}_{\text{SM}} + \frac{1}{\Lambda^2}\sum_i C_i O_i + \mathcal{O}\left(\frac{1}{\Lambda^3}\right), \tag{1.24}$$

where Λ is the energy scale of the new physics, O_i are the effective operators encoding the new physics interactions and C_i are the Wilson coefficients encoding the strength of the NP contributions.

When considering the lepton flavour violation, only a reduced set of operators is relevant, leading to the effective LFV Hamiltonian given in Eq. (1.25):

$$\mathscr{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{tq}^* \sum_{i=9,10,S,P} \left(C_i^{q\ell\ell'} O_i^{q\ell\ell'} + C_i'^{q\ell\ell'} O_i'^{q\ell\ell'}\right) + h.c., \tag{1.25}$$

where G_F is the Fermi constant, V_{tb} and V_{tq} are the CKM matrix elements, $q \in \{d, s\}$ stands for the quark flavours and $\ell, \ell' \in \{e, \mu, \tau\}$ for the lepton flavours, and the primed operators are chirality-flipped compared to the unprimed ones. The operators O_i, $i = 9, 10, S, P$ are defined in Eq. (1.26) and correspond to the vector, axial vector, scalar and pseudo-scalar operators, respectively:

$$\begin{aligned} O_9^{q\ell\ell'} &= \frac{e^2}{16\pi^2}(\bar{q}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu \ell'), & O_S^{q\ell\ell'} &= \frac{e^2}{16\pi^2}(\bar{q}P_R b)(\bar{\ell}\ell'), \\ O_{10}^{q\ell\ell'} &= \frac{e^2}{16\pi^2}(\bar{q}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu\gamma^5 \ell'), & O_P^{q\ell\ell'} &= \frac{e^2}{16\pi^2}(\bar{q}P_R b)(\bar{\ell}\gamma^5 \ell'), \end{aligned} \tag{1.26}$$

where P_L and P_R are the left-handed and right-handed projection operators defined in Eq. (1.2). The O' operators are obtained by replacing P_L with P_R and vice versa. In the SM, all Wilson coefficients $C_i^{(\prime)}$ are equal to 0 for $\ell \neq \ell'$.

To study NP processes such as LFV from a theoretical point of view, one can use model-independent or model-dependent approaches. In the first case, the idea is to fit the experimental data with the Wilson coefficients as free parameters. Only one or multiple coefficients can be considered simultaneously, and their variation range can be limited according to BSM hypothesis. For instance, the branching fraction of $B^0 \to K^{*0}\tau\ell$ decays can be written as (1.27):

$$\begin{aligned}\mathscr{B}(B^0 \to K^{*0}\tau\ell) \times 10^9 = {} & a_9|C_9 + C_9'|^2 + a_{10}|C_{10} + C_{10}'|^2 + b_9|C_9 - C_9'|^2 + b_{10}|C_{10} - C_{10}'|^2 \\ & + b_S|C_S - C_S'|^2 + b_P|C_P - C_P'|^2 \\ & + c_{9S}\,\mathrm{Re}[(C_9 - C_9')(C_S - C_S')^*] + c_{10P}\,\mathrm{Re}[(C_{10} - C_{10}')(C_P - C_P')^*],\end{aligned} \tag{1.27}$$

where the $a/b/c_{9,10,S,P,9S,10P}$ coefficients are calculated using QCD sum rules [59, 60] and are given in [58]. The q, ℓ, ℓ', corresponding here to respectively $s, \tau, \ell' = e, \mu$ are omitted for clarity. Assumptions on the type of NP interactions can then be made. If we consider NP interactions to be left-handed only, we can set $C_S^{(\prime)}, C_P^{(\prime)}, C_9', C_{10}' = 0$, while keeping $C_9 = -C_{10} \neq 0$. Similarly, by assuming NP to be mediated by scalar operators only, the Wilson coefficients $C_9^{(\prime)}, C_{10}^{(\prime)}, C_P', C_S'$ are set to 0, and $C_S = -\Delta C_P \neq 0$. The branching fraction of $B^0 \to K^{*0}\tau e$ and $B^0 \to K^{*0}\tau\mu$ decays can then be estimated as a function of the Wilson coefficients NP contributions $\Delta C_i = C_i - C_i^{SM}$ using the `flavio` package [61], as shown in Figure 1.5 for the left-handed and scalar NP scenarios. Note that the Wilson coefficients C_i^{SM} are equal to 0 for $\ell \neq \ell'$, so in the particular case of this analysis $\Delta C_i = C_i$.

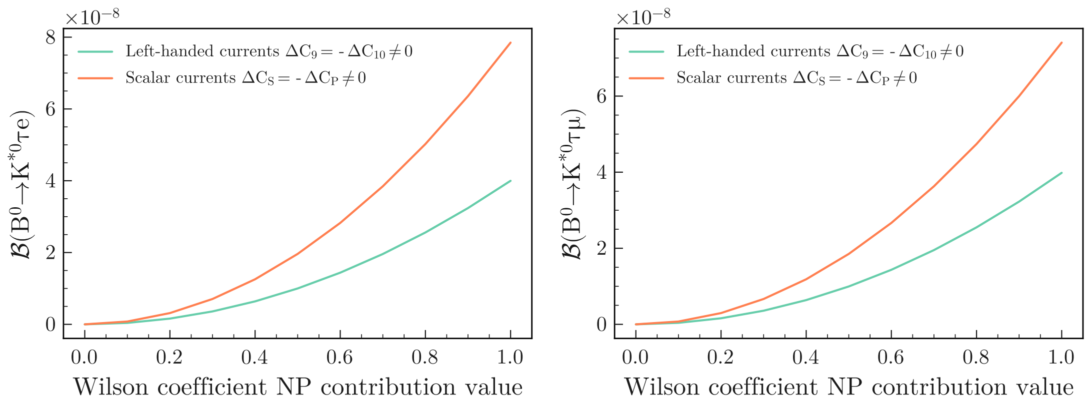


Figure 1.5: Prediction of branching fractions of $B^0 \to K^{*0}\tau^\pm e^\mp$ (left) and $B^0 \to K^{*0}\tau^\pm\mu^\mp$ (right) as a function of the Wilson coefficients values.

In the second case, specific models are built using the already existing experimental constraints. The two approaches are complementary and are usually used together to provide the best interpretation of the experimental data.

1.3.1. Promising BSM models

Several BSM models allow lepton flavour violating processes. Here we focus on two of the most widely studied scenarios: leptoquarks and Z' bosons. Figure 1.6 illustrates the Feynman diagrams depicting their contributions to $B^0 \to K^{*0}\tau\ell$ decays.

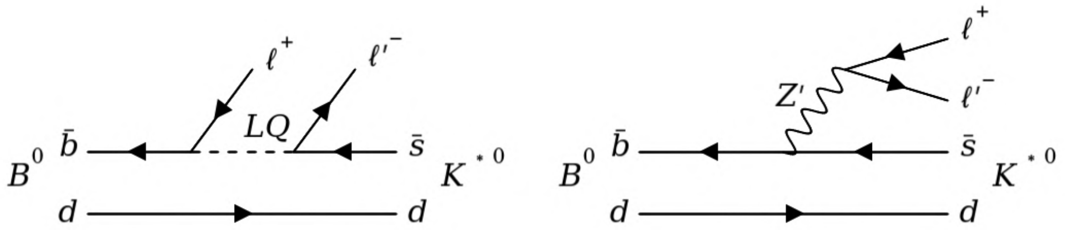


Figure 1.6: Feynman diagrams for the leptoquark (left) and Z' (right) contributions to $B^0 \to K^{*0}\tau\ell$ decays.

To begin with, leptoquarks (LQs) are hypothetical particles that would couple to both leptons and quarks, thus carrying both a baryon and lepton number. This particular property makes them interesting candidates to accommodate the deviations observed in the $R_{D^{(*)}}$ ratio, as well as the recent deviations in $B \to K\nu\bar{\nu}$ decays [56, 62–68]. Leptoquarks could be scalar or vector fields, and their mass is expected to be in the TeV range, which explains why they have not been observed at the LHC. Different LQ models exist, involving a single or multiple leptoquark fields, where the most popular ones are either a single vector LQ model [69] or a two scalar LQ model [70]. An important feature of LQs models, is that in most of them, lepton flavour conservation is not guaranteed any more, leading to possibly enhanced LFV processes.

When building the LQ models, one difficulty resides in handling all anomalies observed experimentally while keeping the model as simple as possible. It has been shown in [71] that the only viable vector LQ model with a single leptoquark field is renormalizable only under certain conditions, leading to the introduction of additional parameters and assumptions. An alternative to this model is proposed in [64, 70], where the combination of two scalar LQs with mass $\mathcal{O}(1\,\text{TeV})$ is considered. Indeed, a combination of the $R_2(\mathbf{3},\mathbf{1},7/6)$ and $S_3(\mathbf{3},\mathbf{1},1/3)$ LQs, where $(\mathbf{a},\mathbf{b},c)$ is the LQ representation under the SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$, would allow to address the current flavour anomalies while keeping the model renormalizable. In this model, the branching fraction of $B^0 \to K^{*0}\tau\mu$ would lie between 1.2 and 6.2×10^{-7}, and be linearly correlated with an enhancement of the $B \to K^{(*)}\nu\bar{\nu}$ branching fraction [70]. Other studies performed in [72], provide predictions for the branching fraction of $B^0 \to K^{*0}\tau\mu$ for the single scalar LQ model $\Delta^{(1/6)}(\mathbf{3},\mathbf{2},1/6)$ and the triplet vector LQ model $U_3(\mathbf{3},\mathbf{3},2/3)$, other considered model being discarded due to their inability to explain some experimental observations or not being renormalizable as discussed previously. The $\Delta^{(1/6)}$ model predicts the branching fraction of $B^0 \to K^{*0}\tau\mu$ to be $< 2\times10^{-5}$ in case of 1 TeV LQ mass and $< 3.7\times10^{-7}$ in case of 10 TeV LQ mass; while in the U_3 vector LQ model, such branching fraction is predicted to be $\lesssim 5.4\times10^{-6}$ [72].

On the other hand, the Z' boson is a hypothetical additional neutral gauge boson, heavier than the known Z and W bosons, that would couple to the SM fermions. It would not allow transitions between leptons and quarks, but could couple to leptons of different flavours. Such boson would thus contribute to LFV decays at tree level, leading to a possible significant enhancement of the branching fractions [58, 73, 74]. In particular, if LFV arises from couplings to the Z' boson, the Wilson coefficients C_S and C_P are 0. Among the remaining $C_{9,10}^{(\prime)}$ coefficients, different scenarios are explored, such as vectorial couplings, corresponding to $C_{10}^{(\prime)} = 0$ and leading to $\mathcal{B}(B^0 \to K^{*0}\tau\mu) < 8 \times 10^{-5}$ [73], left-handed couplings only, corresponding to $C_9 = -C_{10}$ and leading to $\mathcal{B}(B^0 \to K^{*0}\tau\mu) < 1.5 \times 10^{-7}$ [58], or right-handed couplings only, corresponding to $C_9' = C_{10}'$ and leading to $\mathcal{B}(B^0 \to K^{*0}\tau\mu) < 3 \times 10^{-6}$ [58].

1.3.2. Experimental status of LFV searches

To this day, the most stringent experimental limits on LFV processes are set in muon decays, with upper limits at 90% confidence level (CL) of the order of 10^{-13} for $\mu \to e\gamma$ by the MEG collaboration [75], and $\mu N \to eN^{(\prime)}$ by the SINDRUM II collaboration [76]. Other experiments such as COMET [77], DeeMe [78] and Mu2e [79] aim at studying $\mu \to e$ transitions in the near future. LFV is also actively searched for in τ decays, with upper limits of the order of 10^{-8} at 90% CL set on approximately 50 different decay channels mostly by the Belle experiment, but also by BaBar and recently by the Belle II experiment. Searches of LFV in Higgs, Z boson, top quark and meson other than B (kaon, D mesons, J/ψ, $\Upsilon(1,2,3S)$) decays were also performed by various experiments.

LFV searches in B-meson decays

In the B sector, stringent upper limits are set by the LHCb, BaBar and Belle experiments for modes with LFV involving e and μ in the final state, with upper limits on branching fractions of the order of 10^{-9} at 90% CL, benefiting from the full reconstruction of the final state. However, in B decays with τ in the final state, the searches are more challenging due to the missing energy of neutrino(s) appearing in the τ decay. At B factories, such as Belle II and previously Belle and BaBar, one can take advantage of the well-known initial state and the ability to fully reconstruct the event to extract the signal from the recoil mass of the τ. However, in experiments like LHCb, the τ lepton must be reconstructed from its decay products, despite the presence of missing energy. Studying LFV B decays with τ in the final state is particularly interesting as it allows to probe the third generation coupling, which is expected to be enhanced in many BSM models. In addition, $b \to s\tau\ell$ decays are particularly interesting as they allow to probe at the same time FCNC and LFV processes.

Upper limits on the branching fraction of $B \to \tau\ell$ and $B \to \pi\tau\ell$ decays are set at the level of 10^{-5} at 90% CL. Here and in the rest of the section, ℓ stands for e or μ, and when no charges are specified for the leptons, we consider both $\tau^+\ell^-$ and $\tau^-\ell^+$ decays.

Concerning $b \to s\tau\ell$ decays, the $B^+ \to K^+\tau\ell$ decays were searched for by BaBar, LHCb and lastly by the Belle experiment with upper limits results between 0.6 and 2.5×10^{-5} at 90% CL [17]. Upper limits on the $B^0 \to K_S^0\tau\ell$ decays were also set using a combined dataset of the Belle and Belle II experiments in the range of 0.8 to 3.6×10^{-5} at 90% CL [18]. Finally, the $B^0 \to K^{*0}\tau^\pm\mu^\mp$ decays were searched for by LHCb which set upper limits of 0.8 and 1.0×10^{-5} at 90% CL [19] and a very recent search by LHCb for $B^0 \to K^{*0}\tau^\pm e^\mp$ decays set preliminary upper limits of 4.9 and 5.9×10^{-6} at 90% CL [20]. These limits are the most stringent to date for $b \to s\tau\ell$ decays. However, at the time of the development of the analysis presented in this thesis, no search had been performed for the $B^0 \to K^{*0}\tau e$ decays, and $B^0 \to K^{*0}\tau\ell$ decays in general had never been searched for at B factories. Searching for such decays is thus particularly interesting as it would complement existing LHCb LFV limits, using a totally different approach for the signal extraction.

As already mentioned, the particular geometry of the B factory detectors allows them to measure the branching fraction of $B^0 \to K^{*0}\tau\ell$ decays using the recoil mass of the τ lepton. In this way, a peak at the τ mass in the recoil mass distribution is a direct and clear signature of the signal, and thus NP. This approach nonetheless relies on the full reconstruction of the event (called B-tagging, see Section 2.4.2), which has a very low efficiency, typically of the order of 0.2% when considering only hadronic B decays for the partner B meson[10]. In the case of LHCb, the τ lepton must be reconstructed from its decay products, and then combined with the other tracks to reconstruct the B meson. The signal is then searched for in the B meson mass distribution, which makes the signal extraction more challenging. Moreover, the τ lepton decays into a single charged track in approximately 85% of the cases, thus not providing any vertex information for the τ reconstruction. The common approach is therefore to use events in which the τ lepton decays into three charged pions (and potentially additional neutral pions), representing approximately 15% of τ decays, in order to reconstruct the τ decay vertex. In addition to this large efficiency cost, the missing energy due to the neutrino from the τ decay (or neutral pions, which are not reconstructed) introduces some uncertainty in the kinematics of the event and a degradation of the signal resolution. Nevertheless, thanks to the large b-hadron production cross-section at the LHC and advanced analysis techniques, the LHCb experiment was able to set stringent upper limits on $B^0 \to K^{*0}\tau^\pm\mu^\mp$ decays, in the same range as limits set by B factories on similar modes [19]. The recent analysis [20] even achieved upper limits of the order of 10^{-6} on the branching fraction of $B^0 \to K^{*0}\tau^\pm e^\mp$ decays, representing the best limits in $B \to s\tau\ell$ decays to date.

A summary of the current experimental limits in $b \to s\tau\ell$ decays is presented in Figure 1.7.

[10] When considering semi-leptonic decays for the partner B meson, the efficiency is of the order of 2%, but a large uncertainty on the kinematics of the event is introduced due to the missing energy of the neutrinos from semi-leptonic decays, leading to a poor signal resolution.

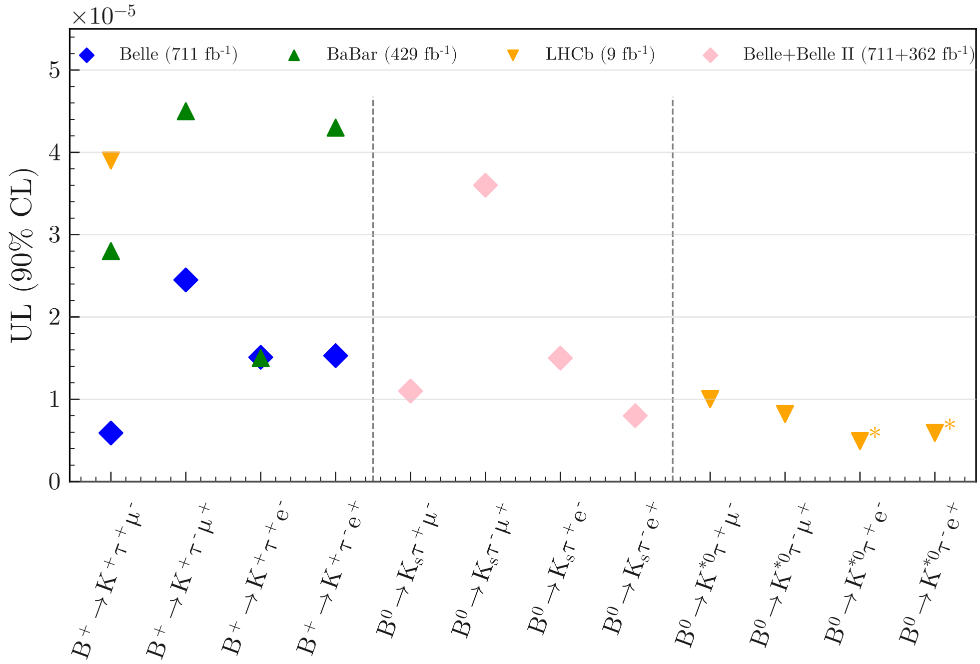


Figure 1.7: Summary of the current experimental limits on LFV $b \to s\tau\ell$ decays. The star denotes preliminary results.

As we saw in the previous section, the theoretical BSM expectations on the branching fractions can reach the $10^{-6} - 10^{-7}$ range, which is close to the current experimental sensitivity. Even without any signal observation, the derived experimental limits on the different modes will have a significant impact on constraining theoretical models.

2. The Belle and Belle II experiments

Summary

The Belle II experiment [21, 22] is a particle physics experiment designed to study, among others, B meson properties. It consists of a particle detector of the same name, surrounding the interaction point of the SuperKEKB electron-positron collider [23, 24]. It is the successor of the Belle experiment [25] which operated on the KEKB collider [26, 27], and it reuses most of the infrastructures built for the Belle experiment at the KEK laboratory in Tsukuba, Japan. The Belle experiment operated from 1998 to 2010, and Belle II took over in 2019.

In the analysis presented in Chapter 4, data from both experiments are used in order to benefit from all available statistics for the study. However, the analysis is performed using the Belle II analysis tools, and the performance studies described in Chapter 3 are also related to Belle II development. For this reason, this chapter will present both experiments, but will emphasize Belle II, in particular concerning data collection and analysis.

2.1. The KEKB and SuperKEKB electron-positron colliders

After operating from December 1998 to June 2010, the KEKB collider was replaced by its upgraded version, the SuperKEKB collider, which began its first tests in February 2016, recorded its first collisions in April 2018 and started operation for Belle II in 2019.

The operation periods of SuperKEKB are separated into so-called Phase 1, 2 and 3.

- Phase 1 (2016) corresponds to a commissioning phase with the BEAST II detector [80] and without any collision in order to estimate the beam background.
- Phase 2 (2018) is also a commissioning phase, but with the Belle II detector partially installed and with first collisions.
- Phase 3 (2019-present) is the phase corresponding to the physics data taking period with the full detector installed[1].

In June 2022, SuperKEKB and Belle II entered a shut-down period called Long Shutdown 1 (LS1) to perform technical upgrades on both the collider and the detector. Data-taking resumed in February 2024 and is planned to continue until 2032, before a second Long Shutdown (LS2). Data collected before LS1 is referred to as Run 1 data and data collected after as Run 2 data.

2.1.1. The KEKB and SuperKEKB accelerators

KEKB and SuperKEKB are asymmetric electron-positron colliders each consisting of two rings with a circumference of 3 km and an injection linear accelerator (LINAC). In SuperKEKB, the 7 GeV electrons circulate in the high-energy ring (HER) and the 4 GeV positrons in the low-energy ring (LER). In contrast in KEKB, the electrons (positrons)

[1] Except for a part of the pixel detector, see Section 2.2.1.1.

had an energy of 8 GeV (3.5 GeV). The SuperKEKB accelerator was build using the tunnel from KEKB and reuses several already existing infrastructures. A schematic view of the SuperKEKB complex can be seen in Figure 2.1.

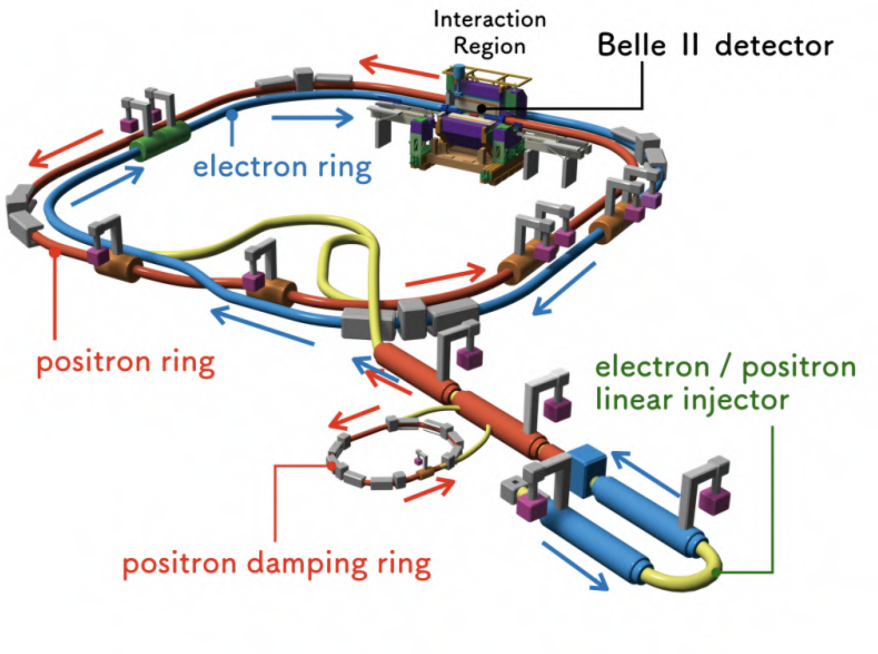


Figure 2.1: Schematic view of the SuperKEKB electron-positron collider.

The collision energy is set to be $\sqrt{s} = 10.58$ GeV, which corresponds to the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ is a bound $b\bar{b}$ state whose branching fraction to $B\bar{B}$ pairs ($B^0\bar{B}^0$ or B^+B^-) is close to 96% [28]; This makes the chosen collision energy highly suitable for B meson studies, hence the term "B factories" used for experiments such as Belle and Belle II.

Due to the different beam energies of the electron and positron beams, the $B\bar{B}$ pairs are produced with a Lorentz boost $\beta\gamma \approx 0.28$ (0.42 for KEKB) (2.1) along the beam axis. This boost allows for the separation of the two produced B mesons, making the setup particularly suitable for CP-violation studies:

$$\beta\gamma = \frac{|\vec{p_{e^-}} + \vec{p_{e^+}}|c}{\sqrt{s}} \approx \frac{E_{e^-} - E_{e^+}}{2\sqrt{E_{e^-}E_{e^+}}}. \tag{2.1}$$

Electrons are produced in the pre-injector from the interaction of a pulsed laser onto a cathode target and are accelerated to 7 GeV in the linear accelerator (LINAC) before being injected in the high-energy ring (HER). The electron beam is also used to create positrons by interacting with a tungsten target located halfway through the LINAC. The resulting positron beam circulates through a damping ring in order to reduce the beam emittance and is accelerated to 4 GeV before being injected in the low-energy ring (LER). The two beams are focused along the beam axis by quadrupole magnets and collide at the so-called interaction point (IP).

2.1.2. The electron-positron beams and particle production

The SuperKEKB collider is designed to reach an instantaneous luminosity of $\mathscr{L} = 6 \times 10^{35}\ \text{cm}^{-2}\text{s}^{-1}$, which corresponds to 30 times the luminosity record achieved by KEKB. In terms of integrated luminosity, the goal is to collect 50 ab^{-1} of data by 2043, corresponding to about 55 billions $B\overline{B}$ pairs, while the total integrated luminosity collected at Belle with KEKB is 711 fb^{-1} for approximately 770 millions $B\overline{B}$ pairs. During the first Belle II data taking period (Run 1), SuperKEKB has delivered an integrated luminosity of 365 fb^{-1} at the $\Upsilon(4S)$ resonance, corresponding to approximately 390 millions $B\overline{B}$ pairs. To achieve this, it reached an instantaneous luminosity of $4.71 \times 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$ in June 2022, setting a new world record (with the detector off). In December 2024, this record was surpassed with an instantaneous luminosity of $5.1 \times 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$.

The instantaneous luminosity $\mathscr{L}$ and the integrated luminosity L are related to the number of produced particles N and their cross-section σ by the following expressions:

$$\mathscr{L} = \frac{1}{\sigma}\frac{dN}{dt}, \qquad L = \frac{N}{\sigma}, \tag{2.2}$$

where the $B\overline{B}$ production cross-section σ in e^+e^- collision at $\sqrt{s} = 10.58$ GeV is 1.1 nb.

To achieve such high luminosity, the SuperKEKB collider, and KEKB before it, relies on the fine tuning of several parameters. In particular, the nano-beam scheme, originally developed in [81] for the Italian SuperB project and later adopted by SuperKEKB has enabled remarkable performance.

The collider luminosity can be written as:

$$\mathscr{L} = \frac{\gamma_\pm}{2er_e}\left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)\left(\frac{I_\pm \xi_{y\pm}}{\beta_y^*}\right)\left(\frac{R_{\mathscr{L}}}{R_{\xi_{y\pm}}}\right) \propto \frac{I_\pm \xi_{y\pm}}{\beta_y^*}, \tag{2.3}$$

where "+" stands for positrons and "−" for electrons, e and r_e are the electron charge and radius, γ the Lorentz factor, $\sigma_{x,y}^*$ the beam size at the IP in the horizontal (x) and vertical (y) plan, I the beam current, $\xi_{y\pm}$ the vertical beam-beam parameter, β_y^* the vertical beta function at the IP, and $R_{\mathscr{L}}$ and $R_{\xi_{y\pm}}$ the reduction factors for the luminosity and the beam-beam parameter.

To increase the luminosity, the beam currents in SuperKEKB plan to be increased to twice the KEKB current and some effort was put in the minimisation of the vertical beta function following the modifications proposed by SuperB [81]. Minimising such function can be seen as minimising the length d of the beam overlap area, as can be seen in Figure 2.2. Since this distance can be expressed as:

$$d = \frac{\sigma_x^*}{\sin\phi} \approx \frac{\sigma_x^*}{\phi}, \tag{2.4}$$

the optimal luminosity design for SuperKEKB/Belle II is achieved by increasing the half-crossing angle ϕ to four times the KEKB/Belle one. However, the value of ϕ is constrained by the magnet design and the detector background level and cannot be increased above a certain limit without degrading the beam-beam parameters $\xi_{y\pm}$.

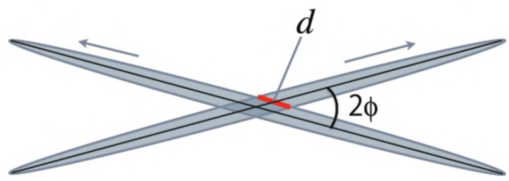


Figure 2.2: Schematic view of the two beams colliding at SuperKEKB in the nano-beam scheme.

Concerning the vertical beam-beam parameter $\xi_{y\pm}$, representing the focusing force induced by one beam electromagnetic field on the opposite beam, its value is limited in a circular collider by the nonlinear force. It can be increased by increasing the number of particles in the bunches, but the vertical emittance grows with the beam-beam parameter, thus stopping the luminosity increase.

The different settings of the KEKB and SuperKEKB colliders are presented in Table 2.1. In case of SuperKEKB, both the designed parameter values and the ones used during the first data-taking period are reported.

Parameter	KEKB (HER/LER)	SuperKEKB (HER/LER)	
	Achieved	LS1	Design
Beam energy [GeV]	8/3.5	7/4	7/4
Beam current I [A]	1.19/1.64	1.10/1.32	2.62/3.60
Vertical beam-beam parameter ξ_y	0.090/0.129	0.028/0.041	0.081/0.088
Vertical beta function at IP β_y^* [mm]	5.9	1.0	0.30/0.27
Lorentz factor $\beta\gamma$	0.43	0.28	0.28
Half crossing-angle [mrad]	11	41.5	41.5
Max luminosity $\mathscr{L}$ [10^{34}cm^{-2}s^{-1}]	2.1	4.71	60

Table 2.1: KEKB and SuperKEKB machine parameter settings. The values are given in the final configuration for KEKB, and both at LS1 and for the final design for SuperKEKB.

The beam collision at $\sqrt{s} = 10.58$ GeV is designed to produce $B\overline{B}$ meson pairs, as $2m_B \approx m_{\Upsilon(4S)}$, but several other particles are also produced, including light quark pairs ($q = u, d, s, c$, often referred to as "continuum") and lepton pairs as can be seen in Table 2.2. The dominant process is Bhabha scattering but a dedicated trigger and selection recipe described later on is applied to reject this background as well as the so-called "low multiplicity" processes such as $\mu^+\mu^-$, $e^+e^-e^+e^-$ or $e^+e^-\mu^+\mu^-$.

In addition, KEKB and SuperKEKB also allow collecting data off-resonance and at other resonances than $\Upsilon(4S)$. In particular KEKB and Belle collected large samples of data at the $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ resonances visible in Figure 2.3, as well as at the $\Upsilon(5S)$

$e^+e^- \rightarrow$	Cross-section [nb]
$\Upsilon(4S)$	1.1
$u\overline{u}$	1.6
$d\overline{d}$	0.4
$s\overline{s}$	0.4
$c\overline{c}$	1.3
$\tau^+\tau^-$	0.9
$\mu^+\mu^-$	1.1
e^+e^-	300
$\gamma\gamma$	5

Table 2.2: Cross-section of different physics processes resulting from an e^+e^- collision at $\sqrt{s} = 10.58$ GeV.

resonance. In case of Belle II, SuperKEKB is not designed to scan all resonances as KEKB, but it collected ~20 fb^{-1} of data around 10.7 GeV for quarkonium studies.

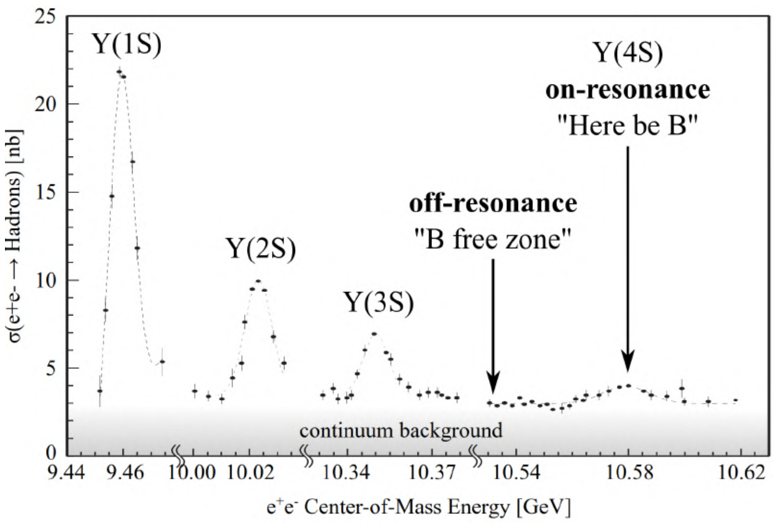


Figure 2.3: Cross-section around the Υ resonances as a function of the e^+e^- centre of mass energy.

Including all on- and off- resonance scans, SuperKEKB has delivered 575 fb^{-1} data to this day of which 424 fb^{-1} correspond to the Run 1 dataset. The details of instantaneous and integrated luminosities since 2019 are visible in Figure 2.4. In case of Belle and KEKB, the total delivered luminosity including all on- and off-resonance scans is 980 fb^{-1}.

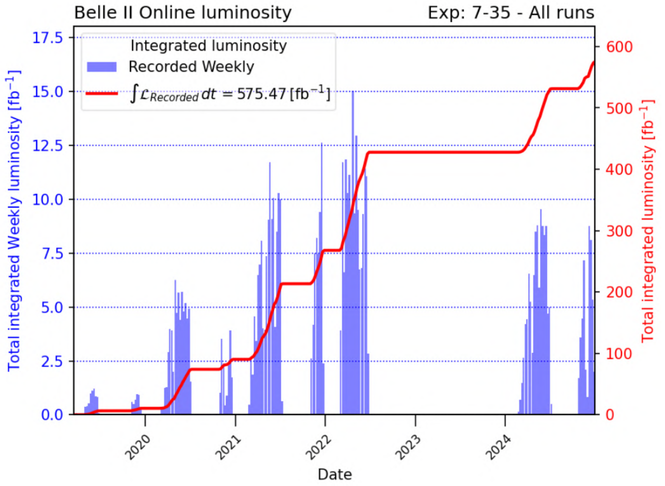


Figure 2.4: Weekly collected data luminosity since Belle II operation beginning in 2019 (blue) and total integrated luminosity (red).

2.2. The Belle and Belle II detectors

In the same way as KEKB and SuperKEKB, Belle II is the successor of Belle where Belle (II) designates both the experiment and the particle detector. Therefore, the Belle II detector is built with the same logic and mainly with upgraded versions of Belle sub-detectors. Indeed, both are hermetic and almost 4π detectors surrounding the interaction point, allowing the reconstruction of the full events with only missing energy due to neutrinos.

The sub-detectors are arranged in layers around the interaction point, starting from tracking detectors, close to the collision, then different sub-detectors participating in particle identification and finally surrounded by a muon and K_L detector for the long-lived particles. In addition to the sub-detectors, a super-conducting solenoid is placed before the muon detector and generates a uniform magnetic field of 1.5 T. This allows charged particles to follow helical trajectories, from which their charge and momentum can be deduced. A view of the Belle II and Belle detectors side by side is presented in Figure 2.5. 3D views of both detectors are also shown in Figure 2.6.

In total, the Belle and Belle II detectors are about 7 m high and 7.5 m long. They cover the full azimuthal angle and from 17° to 150° in the zenith angle. To describe the geometry of the detectors, it is convenient to use both a Cartesian (x, y, z) coordinate system and a polar (r, θ, ϕ) one:

- $(0, 0, 0)$: coordinates of the interaction point (IP). As the beams crossing angle is very small we call "beam axis" the bisectrix of the crossing angle
- x: in the horizontal plane, perpendicular to the beam axis, points towards the outside of the rings
- y: in the vertical plane, points up

- z: in the horizontal plane, along the beam axis, points to the electron beam direction
- r: radial distance $\sqrt{x^2 + y^2}$ from the IP
- θ: zenith angle from 0° to 180° with respect to the z axis, 0° corresponds to positive z coordinate
- ϕ: azimuthal angle from 0° to 360° with respect to the x axis

The "longitudinal" direction corresponds the direction along z, equally referred to as "beam axis" or "beam direction" while the "transverse" one is the direction in the (x, y) or (r, ϕ) plane. The "forward" and "backward" labels respectively stand for positive and negative z directions.

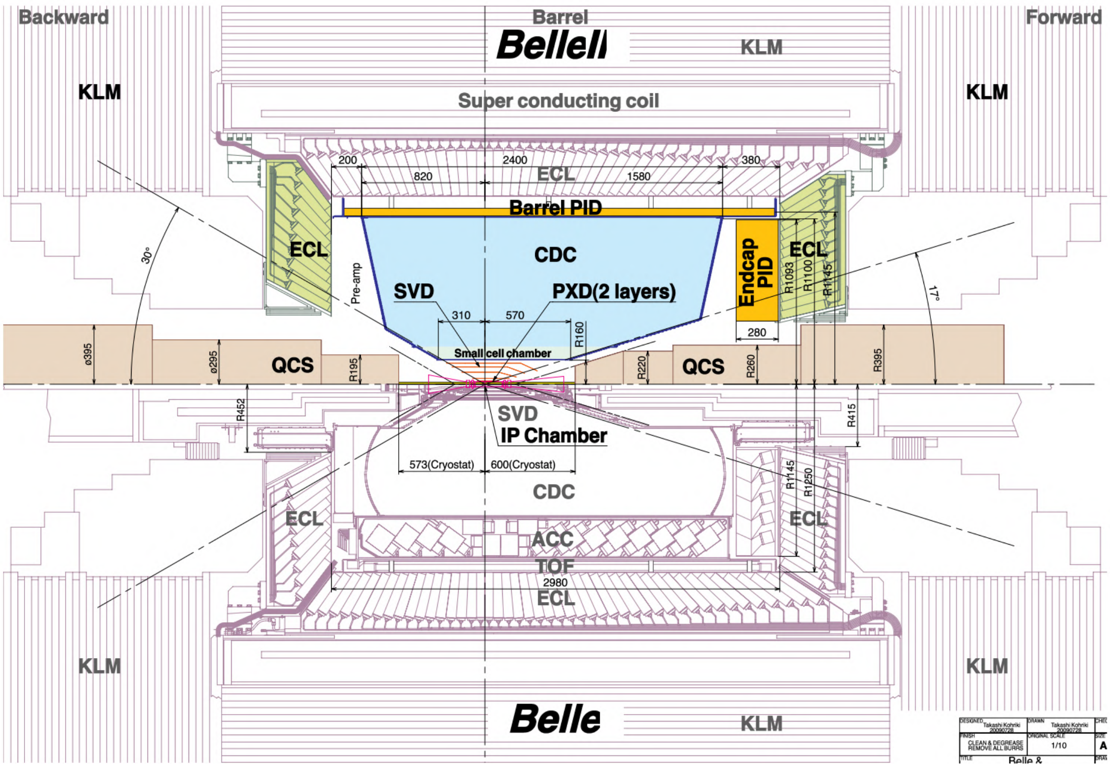


Figure 2.5: Top view of the Belle (bottom half) and Belle II (top half) detectors (From Ref. [82]).

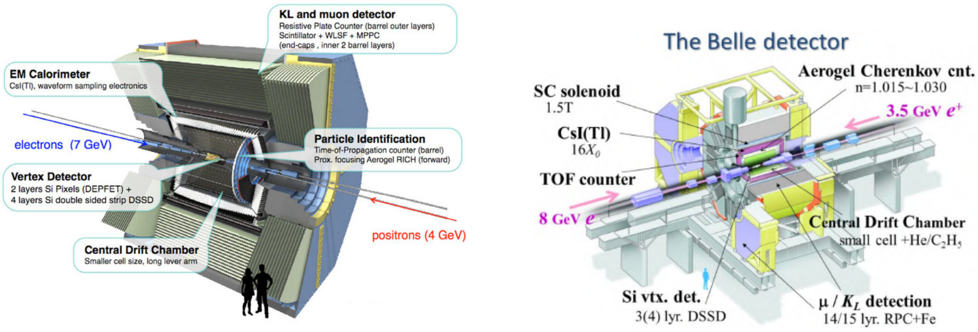


Figure 2.6: 3D view of the Belle II (left) and Belle (right) detectors (From Ref. [83]).

2.2.1. The Tracking system

2.2.1.1. The vertex detectors

The most inner part of the Belle and Belle II detectors is the vertex detector. In Belle, it consisted of a Silicon Vertex Detector (SVD) only, while for Belle II a PiXel Detector (PXD) is added as inner layers of the SVD, even closer to the interaction point, forming with the SVD the VerteX Detector (VXD). The VXD can be seen in Figure 2.7 with at its centre the fully installed PXD and at the outer layers, the SVD. Both detectors aim at locating the particle vertices with a high precision.

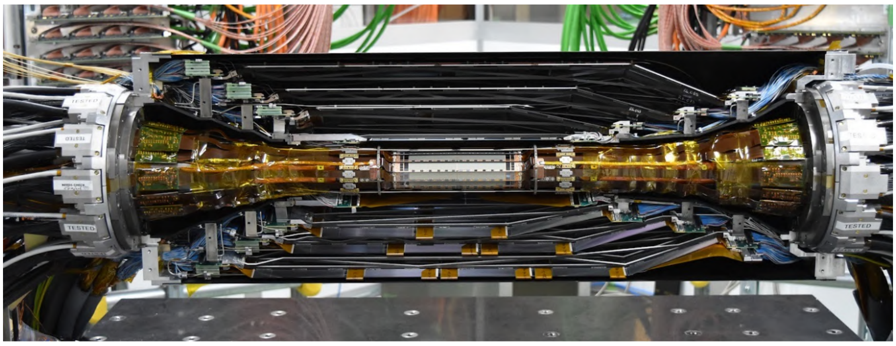


Figure 2.7: Picture of the Belle II VerteX Detector (VXD). The centre part is the PiXel Detector (PXD) and the outer part is the Silicon Vertex Detector (SVD) (From Ref. [84]).

The nano-beam scheme used in SuperKEKB and Belle II results in a much smaller beam pipe (around 10 mm radius) compared to Belle (15 mm after 2003 and 20 mm before). This geometry allows the tracking system to be very close to the interaction point but is balanced by the extremely high hit rate that the detector will have to handle. For this reason, a new tracking sub-detector based on pixel sensors technology is introduced.

The so-called PiXel Detector (PXD) [85] consists of two layers radii of at 14 mm and 22 mm, made up of 8 and 12 ladders respectively covering a polar angle from 17° to 150°, as shown in Figure 2.8. The PXD uses the DEPFET (DEPleted Field Effect

Transistor) technology, which significantly reduces the thickness of the sensors (50-75 μm), thereby minimising the possibility of multiple scattering. The sensors consist of 800 pixel rows on each side, for a total of 1600, with each pixel measuring $50 \times 55 - 85$ μm^2, allowing a spatial hit resolution of 10 μm. Each pixel has a readout time of 100 ns. Despite four rows of pixels being treated simultaneously and resulting in a 20 μs readout time per module, the PXD readout time is about 5 times larger than the other sub-detectors. For this reason the PXD information is treated separately in the event reconstruction, as detailed in section 2.3.

Due to some issues in the assembly of the PXD layers, only the first layer and two ladders of the second layer were installed during Run 1 (2019-2022). The installation was completed during the LS1.

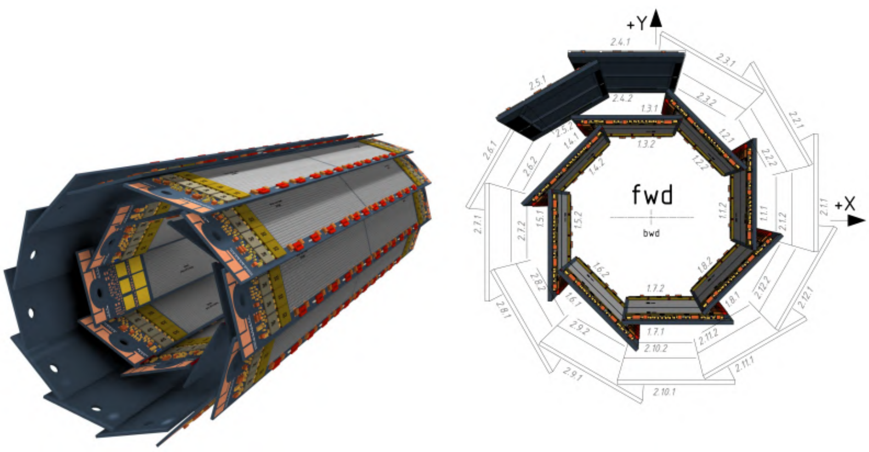


Figure 2.8: 3D (left) and front (right) view of Belle II PiXel Detector (PXD) (From Ref. [85]).

The SVD is made of 4 layers of double-sided silicon strips detectors (DSSDs). In case of Belle SVD [86], the layers are arranged in four cylinders at radii 20, 43.5, 70 and 88 mm around the 15 mm radius beryllium beam pipe. The length of the layers increases with the distance to IP in order to cover the full region from 17° to 150°, as can be seen on Figure 2.9. The layers are composed of 6, 12, 18 and 18 ladders, each having 2, 3, 5 and 6 DSSDs respectively. This configuration is valid from the installation of the so-called SVD2 in 2003. Before that, the SVD1 was composed of only 3 layers located at larger radii due to the larger beam pipe, and covered a polar angle from 23° to 139° [86].

In the Belle II SVD [84, 87], the four layers are located at radii 39, 80, 105 and 135 mm and are labelled from layer 3 to layer 6. Indeed, since the Pixel Detector PXD is installed at the innermost part of the detector, its two layers are layers 1 and 2 of the total vertex detector, also called VXD. The SVD layers are arranged in a cylindrical way with increasing length towards the outer layers similarly to Belle, but in addition, the layers 4, 5 and 6 present an inclined edge with respect to the beam axis in the forward region. As shown in Figure 2.10, this allows to keep a 17° to 150° coverage

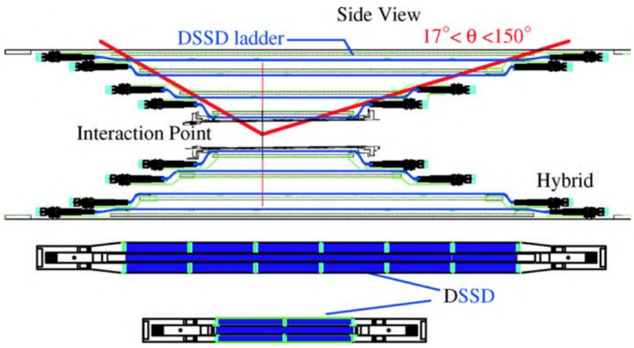


Figure 2.9: Side view of Belle Silicon Vertex Detector (SVD), and ladder configuration of the 1st and 4th layers (From Ref. [86]).

while minimising the traversed material. Similarly to Belle the layers are composed of multiple ladders, here respectively 7, 10, 12 and 16 containing 2, 3, 4 and 5 DSSDs sensors.

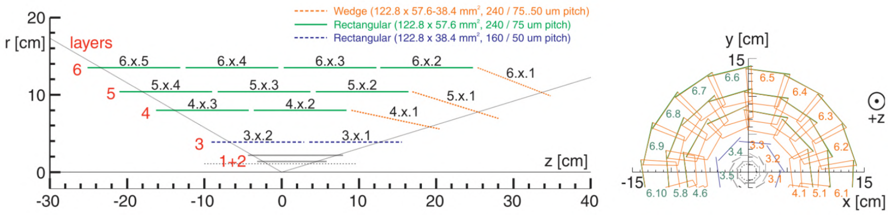


Figure 2.10: Side (left) and front (right) view of Belle II Silicon Vertex Detector (SVD) (From Ref. [87]).

In terms of performance, Belle SVD achieved a hit resolution of 12 µm (19 µm) and an impact parameter resolution of 20-80 µm (30-80 µm) in the transverse (longitudinal) direction, where the poor resolutions correspond to the low momentum region. In addition to handling a much higher hit rate, Belle II SVD reached a cluster resolution of 15-25 µm (7-12 µm) in the transverse (longitudinal) direction. To cope with the increasing detector occupancy that grows with luminosity, the hit time information from the SVD is also exploited in Belle II. Indeed, the hit-time resolution for signal hits is less than 3 ns and signal hits are "on-time" (at $t = 0$ ns) while beam-induced background hits span on a range of about 100 ns and are distributed uniformly over time.

2.2.1.2. The Central Drift Chamber

The Central Drift Chamber (CDC) plays a key role in charged particles tracking as well as in particle identification (PID) for which the energy loss dE/dx is evaluated. The Belle II CDC [88] reuses the same technology as the Belle CDC [89] (visible in Figure 2.11) but aims at better performance, in particular in momentum and dE/dx

resolution by increasing the number of sensors and the size of the detection volume by 22%.

Figure 2.11: Belle Central Drift Chamber (CDC) exhibited at KEK.

Indeed, Belle II CDC extends from radii of 16 cm to 113 cm (vs 8 to 88 cm for Belle) for a length of 232.5 cm (240 cm for Belle) allowing to cover the polar angle from 17° to 150°. The Belle II (Belle) CDC, is composed of 14336 (8400) gold-plated tungsten sense wires arranged in 56 (50) layers constituting 9 (11) super-layers alternating between axial and stereo wires. In axial super-layers, the wires are parallel to the z axis while in stereo super-layers, wires are twisted by an angle of at most 80 mrad. The stereo super-layers alternate between clockwise and counter-clockwise twist. The layer configuration of the Belle and Belle II CDCs and the difference between axial and stereo super-layers are shown in Figure 2.12.

The CDC volume is filled with a gas mixture of helium and ethane at a 50% proportion each and aluminium field wires distributed homogeneously in the different layers which provide an electric field. This allows the electrons from the gas ionised by the passage of a charged particle to drift towards the sense wires and thus retrieve the trajectory of the original particle. The charge is determined thanks to the curvature of the trajectory and the spatial position of the particle is obtained in the transverse plane from both axial and stereo wires, while the z position is determined thanks to the stereo wires angle.

In Belle II (Belle) the achieved spatial resolution is about 2 mm in z and 100 µm (130 µm) in r direction. In terms of transverse momentum p_T resolution, it is 0.73% in Belle II for $p_T = 5$ GeV/c vs 1.85% for p_T between 4 and 5.2 GeV/c in Belle.

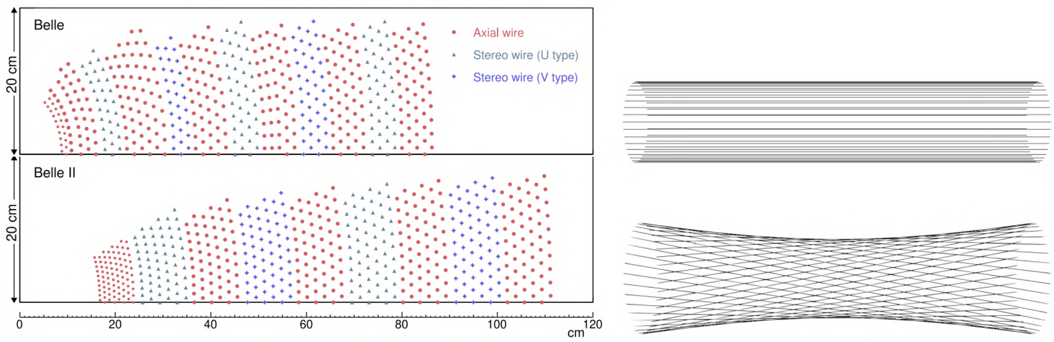


Figure 2.12: Central Drift Chamber (CDC) layer configuration for Belle II and Belle (left) and schematic view of the axial (top right) and stereo (bottom right) super-layers (exaggerated skew) (From Ref. [88, 90]).

2.2.2. The Particle Identification system

The main part of the particle identification (PID) system in both Belle and Belle II relies on two detectors : the Aerogel Cherenkov Counter (ACC) and the Time Of Flight counter (TOF) for Belle, and the Aerogel Ring-Imaging CHerenkov (ARICH) and the Time Of Propagation (TOP) detectors for Belle II.

Despite targeting the same goal of particle identification in the two experiments, in particular discriminate kaons from pions, the two couples of detectors show notable differences both about the global geometry and the used detecting technologies. ACC, ARICH and TOP all make use of the Cherenkov principle. However, the three sub-detectors use a different technology to exploit the emitted Cherenkov light.

Cherenkov light is emitted when a particle travels faster than the speed of light in a given medium. For a particle of mass m and momentum p, this happens if $m < p\sqrt{n^2-1}$, where n is the refractive index of the medium. The opening angle of the cone created by the emitted Cherenkov photons is called Cherenkov angle θ and is defined as $\cos\theta = \frac{1}{\beta n}$, with $\beta = v/c$ where v is the velocity of the particle and n is the refractive index of the medium.

2.2.2.1. The Aerogel Cherenkov Counter and Time of Flight Counter at Belle

The Aerogel Cherenkov Counter (ACC) [91] system is composed of a barrel ACC and a forward endcap ACC. There is no detector in the backward endcap as the particles are boosted in the forward direction and thus the absence of PID system in this region should not affect the detector performance. Moreover, only about 10% of the CDC acceptance is outside the PID system reach.

The barrel part of the ACC is made of 960 counter modules arranged in 60 and 15 cells in respectively the ϕ and z directions, while the endcap part consists of 228 modules arranged in 5 layers of concentric rings. They can be seen in Figure 2.13. The modules contain aerogel with refractive index from 1.01 to 1.03 and the Cherenkov light is detected with one or two photomultiplier tubes (PMTs). The choice of the refractive index allows to create a threshold to separate particle species according to their mass. In particular, the pions and kaons are separated at more than 3σ in the 1.2-3.6 GeV/c momentum region in the barrel ACC and in the 0.8-2.2 GeV/c momentum region in the endcap ACC.

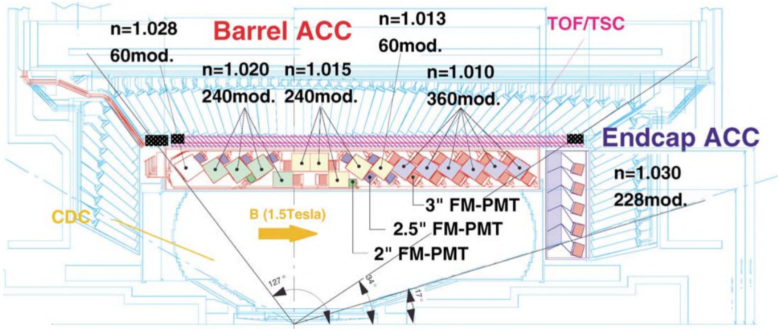


Figure 2.13: Schematic view of the Belle Aerogel Cherenkov Counter (ACC) (From Ref. [25]).

To provide pion and kaon separation for particles with a momentum below 1.2 GeV/c, the Time Of Flight counter (TOF) [92] is used. It is located in the barrel region of the detector at approximately 1.2 m from the interaction point and consists of 64 modules made of two Time Of Flight (TOF) counters and one Trigger Scintillation Counters (TSC), covering the polar angle from 34° to 120°. The TOF measures the time it takes to charged particles to reach the detector from the interaction point. The mass of the particle is then inferred from its momentum, its time of flight and the distance of the TOF to the IP. The TOF modules provide a time resolution of 100 ps, which allow an effective separation of pions and kaons with momentum from 0.28 GeV/c (minimum momentum to reach the TOF detector) to 1.2 GeV/c, below which 90% of particles produced in $\Upsilon(4S)$ decays are. The TSC counters are introduced in order to provide fast timing signals for the data acquisition trigger system. The scintillating light is collected by photomultiplier tubes attached to the ends of the TOF and TSC counters.

2.2.2.2. The Aerogel Ring-Imaging CHerenkov and the Time Of Propagation detectors at Belle II

Contrary to the Belle ACC which has a component in the barrel, the Belle II Aerogel Ring-Imaging CHerenkov detector (ARICH) [93] is located only in the forward endcap region. It is composed of a 4 cm large aerogel radiator for the particles to emit Cherenkov light, a 20 cm large expansion volume for the Cherenkov rings to be properly formed, and a single photon detector based on the Hybrid Avalanche PhotoDetector (HADP) technology to detect Cherenkov photons. In order to increase the resolution of the Cherenkov rings, the aerogel volume is divided into two layers of 2 cm width each with different refractive indices $n_1 = 1.045$ and $n_2 = 1.055$. In this way, the photon spread in the detector plane is reduced and the Cherenkov rings from the two layers overlap, thus increasing the number of photons to be detected and reducing the uncertainty on the emission point. The Ring-Imaging Cherenkov principle and the ARICH design are presented in Figure 2.14 while the dual-layer focusing scheme is illustrated in Figure 2.15.

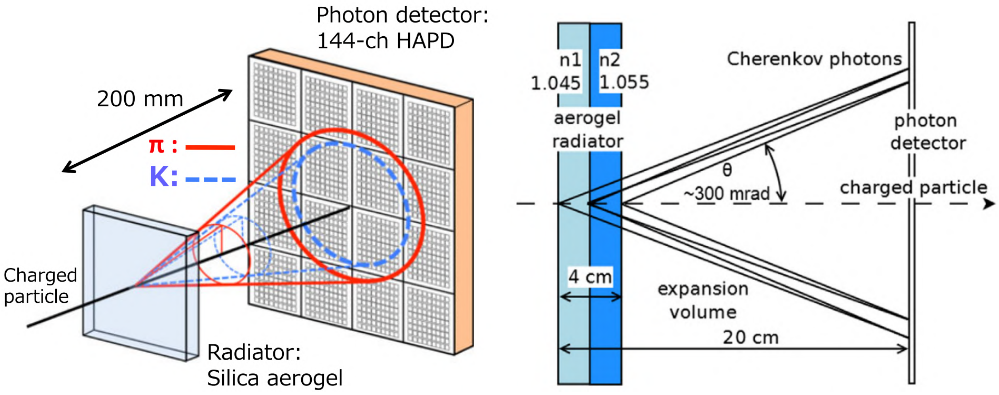


Figure 2.14: General view (left) and detailed scheme (left) of the Ring-Imaging Cherenkov principle used in Belle II ARICH detector (From Ref. [93, 94]).

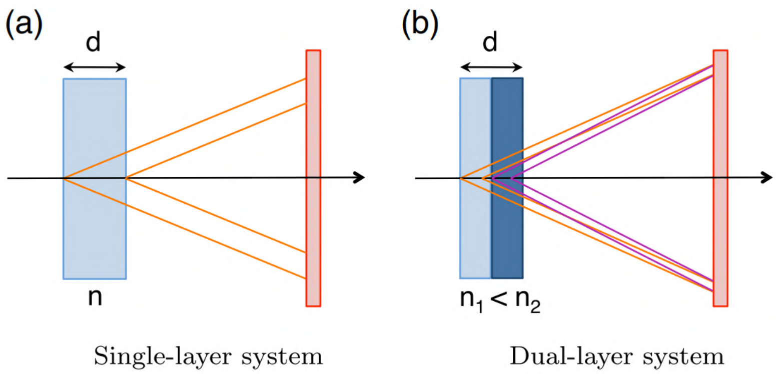


Figure 2.15: The dual-layer focusing scheme: (a) Normal Cherenkov counter with a single layer; (b) Focus with dual layer in which different refractive indices $n_1 < n_2$ are used. (From Ref. [94]).

Thanks to this design, the Cherenkov angle resolution is measured to be $\sigma_\theta = 14.03$ mrad for single photons and allows a separation $S = \frac{\Delta\theta_C}{\sigma_\theta}\sqrt{N_\gamma}$ of the pions and kaons at more than 5σ, where $\Delta\theta_C$ is the measured difference between the pions and kaons Cherenkov angle and N_γ is the number of detected photons.

The particle identification in the barrel of the detector is performed using the Time Of Propagation (TOP) detector [95]. It consists of 16 modules of 2.5 m × 44 cm × 2 cm made of two quartz radiators joined together and arranged in cylinder around the CDC barrel. On the forward extremity, a mirror is placed, and micro-channel plate photomultiplier tubes (MCP-PMTs) are installed on the backward end. When charged particles travel through the quartz radiator, they emit Cherenkov photons that are collected by the MCP-PMTs (after being returned by the mirror for the ones being emitted in the forward direction) which provide timing and spatial information in the (x, y) plane. As is illustrated in Figure 2.16, different particle species lead to different Cherenkov angles, resulting in different time of propagation. The time resolution of the PMTs is thus a crucial element for particle identification, and the use of MCP-PMTs allows to achieve a 50 ps time resolution for single photon detection.

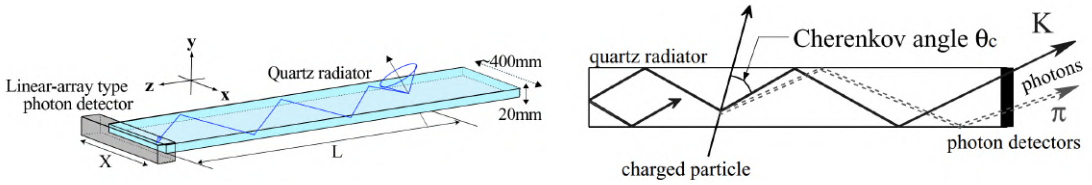


Figure 2.16: View of a single module of the Time Of Propagation (TOP) detector (left) and schematic view of the internal reflecting Cherenkov photons (right) (From Ref. [22]).

2.2.3. The Electromagnetic Calorimeter

Electromagnetic calorimeters are crucial elements in detectors aiming at studying B physics. Indeed, about one third of the B mesons decay products are π^0 or other neutral particles producing photons with an energy range from 20 MeV to 4 GeV. In Belle and Belle II, the role of the Electromagnetic CaLorimeter (ECL) is thus to detect such photons and measure their energy as well as participate in electron and K_L^0 identification (with the KLM for the latter). In addition, it contributes to the trigger system and the luminosity measurement.

The Belle II electromagnetic calorimeter [82] reuses the structure and detecting material of the Belle ECL [96] while upgrading the readout electronics in order to handle the much higher background rates.

The ECL is composed of a 3 m long and 1.25 m inner radius barrel part covering a polar angle between 32° and 129°, and two end-caps for a total angular coverage between 12° and 155°. The detector is made of 8736 CsI(T) crystals with a truncated pyramid shape of average size $6 \times 6 \times 30$ cm^3. On the rear surface of each crystal, two 10×20 mm^2 photodiodes are attached followed by two preamplifiers in order to collect the scintillation light.

The Belle ECL achieved energy resolutions from 4% to 1.6% for energies of 100 MeV to 8 GeV. In the absence of background, the Belle II ECL performance is very similar. However, the high background rates from SuperKEKB cause so-called pile-up noise consisting of random photon clusters in the calorimeter. To cope with this increase in combinatorial background, the readout electronics are upgraded in order to exploit the time information to suppress the fake clusters by rejecting wrong time clusters. The pile-up noise can thus be reduced by a factor 1.5.

2.2.4. The K_L^0 and Muon detector

Long-lived and high penetration power particles like muons and K_L^0 are not stopped in the inner layers of the Belle (II) detector and reach the K_L^0 and muon detector (KLM), whose goal is to detect them. Muons need a momentum of 600 MeV/c to reach the KLM and can travel through it depending on their momentum and polar angle (they escape the KLM if their momentum is > 1.5 GeV/c). K_L^0 however are likely to interact with the detector material and create hadronic showers in the ECL and/or the KLM.

The Belle II [22] and Belle [25] KLM use the same structure, but the detection system from Belle is partially replaced for Belle II in order to cope with the high background rates.

The KLM is located outside the superconducting solenoid and consists of an octagonal barrel region divided in 16 sectors covering a polar angle between 45° and 125°. Two endcaps extend the angular coverage to 20°–155°. Each section (barrel, forward endcap, backward endcap) comprises alternating layers of 4.7 cm thick iron plates and active detector elements. There are 15 (14) detector layers and 14 iron layers in the barrel (endcaps) region. The iron plates contribute as 3.9 interaction lengths in addition to the 0.8 from the ECL, enabling the K_L^0 to produce hadronic showers. In the Belle KLM, the detection system consisted of glass-electrode resistive plate chambers (RPCs). However, since the RPCs have a long dead-time after discharge, this technology is unsuitable for the high background rate in Belle II. Therefore, in the Belle II KLM, the active detecting system is replaced with scintillator strips read out by silicon photomultipliers (SiPMs) in the two innermost layers of the barrel and all layers of the endcaps. The SiPMs provide time resolution of 0.7 ns, and their use is planned to be extended to the whole KLM.

2.3. Collecting data at Belle II

2.3.1. Trigger and data acquisition (DAQ) systems

In addition to all sub-detectors, some crucial elements of the Belle II detector are the trigger [22] and data acquisition (DAQ) [97] systems. Because of the increasing luminosity, the experiment has to cope with very large event rates, estimated at 20 kHz for the designed instantaneous luminosity for a bunch-crossing rate of 250MHz. The dominant beam-induced and luminosity-dependent backgrounds are (radiative) Bhabha ($e^+e^- \to e^+e^-(\gamma)$) and two-photon ($e^+e^- \to e^+e^-e^+e^-$) processes, which are easily discriminated from $B\overline{B}$ events thanks to the track multiplicity but can be difficult to distinguish from interesting low multiplicity events such as $\tau-$pair decays or events with dark photons. The trigger system is designed to filter out the background events and retain only events of interest by reducing the event rate to a maximum of 10 kHz. This value corresponds to the maximum incoming rate capacity of the DAQ system, whose role is to read all information from the sub-detectors and store the events for later analysis.

The trigger system is composed of two elements: the hardware Level 1 (L1) trigger and the software High Level Trigger (HLT), which are shown in Figure 2.17 together with the general trigger logic.

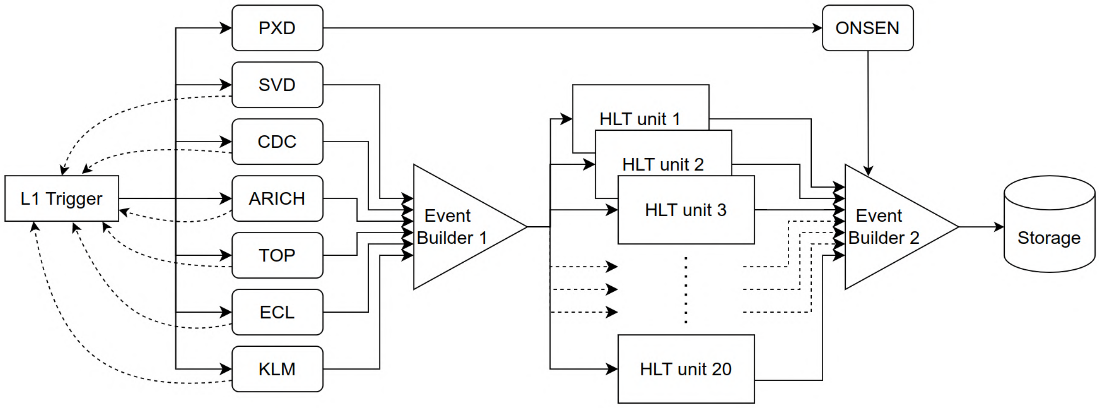


Figure 2.17: Schematic view of the Belle II trigger system with the Level 1 (L1) trigger and the High Level Trigger (HLT).

2.3.1.1. Hardware Level 1 trigger

The Level 1 trigger (L1) [98], also called low-level or online trigger, is a hardware trigger collecting information from all sub-detectors (except the PXD) and aiming at suppressing beam-induced, radiative Bhabha and two-photon background events while retaining more than 99.9% of $B\overline{B}$ events. It is designed to achieve a maximum output rate of 30 kHz. The L1 trigger uses in particular the information from the CDC and the ECL detectors. The CDC trigger line is mainly based on 3D tracking information while the ECL trigger line uses the total energy and the number of clusters in the calorimeter. The CDC and ECL information is then complemented by the TOP and KLM systems and all sub-detectors triggers are combined in a global L1 trigger decision within a 5 μs latency time. If the decision is positive, the information is transferred to the front-end readout boards of all sub-detectors.

2.3.1.2. Software High Level Trigger

The High Level Trigger (HLT) [99], turned on in 2021, is a software trigger receiving as input the data passing the L1 trigger decision and aiming at reducing the event rate to 10 kHz.

The information from all sub-detectors (except the PXD) is gathered at software level and is used to perform a so-called "fast" or "online" event reconstruction to decide whether or not events should be stored. During fast reconstruction, the PXD information is not taken into account. Indeed, since the PXD is very close to the interaction point, it is subject to extremely high background rates and the readout of the PXD would result in a huge amount of data to be stored. From the events reconstructed during fast reconstruction, PXD Regions Of Interest (ROIs) are defined by extrapolating the SVD and CDC tracking information to the PXD. These ROIs, consisting in rectangular regions on each layer of the PXD, are then given to the PXD readout system and only the hits in those regions are stored for the full (or "offline") reconstruction. PXD ROIs are also estimated from SVD-only information and the found ROIs from the two methods are merged together with the ONSEN system.

To perform the fast reconstruction, the HLT aims at using 20 computing units, each composed of 16 workers with 20 CPUs, which gives a total of 6400 parallel processes. Given the expected input rate at nominal luminosity of 20 kHz and the available computing power, the HLT has about 300 ms per event to calculate the trigger decision. However, it was measured during the first data taking period that the average HLT processing time per event was already around 300 ms. During Run 1, only 9 HLT units were used, and 3 more units were added for Run 2, which reduces the computing burden. Nevertheless, software optimisation of the HLT is still needed to cope with the increasing event rate. A study focused on the optimisation of an algorithm having a high impact on the HLT computing time is presented in Chapter 3.1.

2.3.2. Track reconstruction

The track reconstruction, also called "tracking" is the process of combining the spatial information ("hits") from the different sub-detectors to build "tracks" representing the particle trajectories in the detector and measure their kinematics and properties.

The track reconstruction [90] is composed of two main tasks: track finding and track fitting. Track finding consists of combining the spatial information from the different sub-detectors into tracks, while track fitting is the adjustment of the measured trajectory and the measurement of the tracking parameters.

In the tracking chain, CDC and SVD informations are first treated in an independent way. On one hand, SVD clusters are reconstructed and combined to create space-points on each sensor. The space-points are three-dimensional hits in SVD that are created from every possible combination of clusters from both side of the SVD sensors. On the other hand, using independents Legendre and cellular automaton algorithms, CDC hits are combined together to form the CDC tracks. They are then extrapolated to the SVD and attached to SVD space-points using a combinatorial Kalman Filter (CKF) algorithm. "SVD-only" tracks are built from the remaining (not attached) SVD space-points. Then, the SVD stand-alone tracks and the CDC tracks attached to SVD space-points are merged together. Finally, hits are searched again in the CDC to be attached to the merged tracks, resulting in final CDC-SVD tracks to which the PXD information is added using a second CKF to form the final tracks. The PXD information is attached to the tracks only in the offline reconstruction. During fast reconstruction, the rest of the workflow is identical as the one described above. Tracks are fitted all along the tracking procedure in order to keep the best approximation of the particle trajectory. More details on track fitting along the tracking chain and the optimisation of one of the fitting algorithm are presented in Chapter 3.1. A schematic view of the tracking chain is presented in Figure 2.18.

In the helix parametrisation of the track trajectories, one can define the Point Of Closest Approach (POCA) that corresponds to the point where the (extrapolated) track

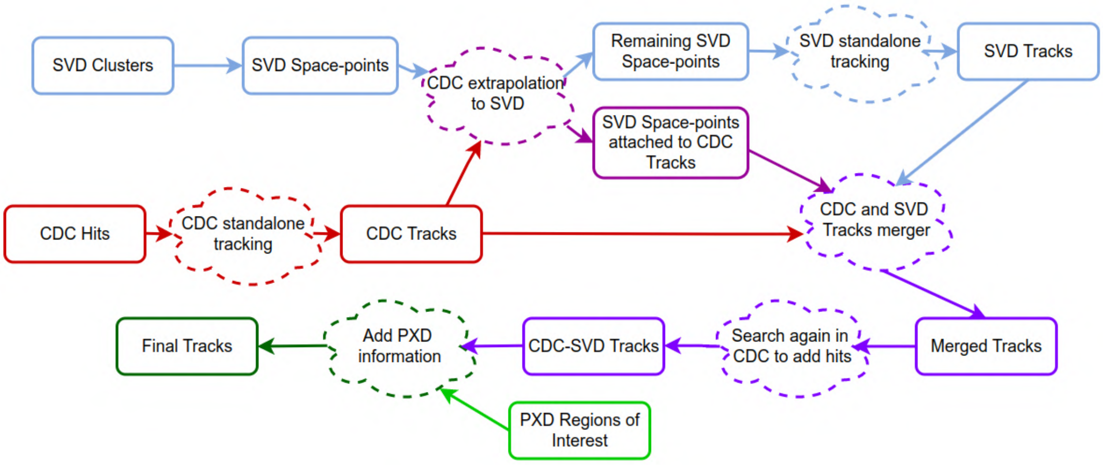


Figure 2.18: Schematic view of the simplified tracking chain. The solid line rectangles represent tracking objects, while the dotted line clouds represent tracking algorithms.

is the closest to the origin of the coordinate system, i.e. the interaction point. Note that the exact position of the interaction point is remeasured for each event and may vary slightly from the (0,0,0) position. The final tracks are characterised by several parameters, among which the position of the POCA and the curvature of the track (see Figure 2.19):

- d_0 (dr) is the signed (radial) distance of the POCA to the origin (measured IP) in the (x, y) plane.
- z_0 (dz) is the z coordinate of the POCA with respect to the origin (measured IP).
- ϕ_0 is the angle defined by the x axis and the track transverse momentum at the POCA.
- ω is the track curvature signed with the particle charge.

The formed final tracks are unfortunately not always perfect. Indeed, some tracks can result from a wrong combination of hits, with hits from beam-induced background or mixing hits from multiple tracks. In some cases, two tracks can also be reconstructed from the same particle, leading to a "clone" track. These misreconstructed tracks are called "fake tracks" and a dedicated study is performed for their rate measurement is presented in Chapter 3.2.

Concerning the tracking efficiency, it was measured with a tag-and-probe method in τ-pair events $e^+e^- \to \tau^\pm[\to \pi^+\pi^-\pi^\pm\nu_\tau]\tau^\mp[\to \ell^\mp\bar{\nu}_\ell\nu_\tau]$ ($\ell = e, \mu$) evaluating the ratio of events properly reconstructed to the total number of events [100, 101]. The tracking efficiency varies from 75% at low momentum to 95% at high momentum (4.5 GeV/c) and is constant at 90% along the azimuthal angle, with a data/simulation discrepancy of $\sim 0.05\%$.

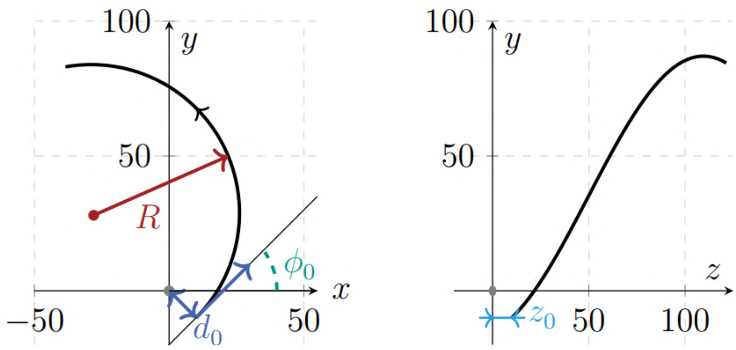


Figure 2.19: Schematic view of the tracks' trajectory in the (x, y) plane (left) and in the (y, z) plane (right) together with the d_0, z_0, ϕ_0 and $R = 1/|\omega|$ parameters (From Ref. [90]).

2.3.3. Particle identification

The particle identification (PID) at Belle II uses the information provided by the sub-detectors SVD, CDC, ARICH, TOP, ECL and KLM to define the likelihood $\mathscr{L}_i^D$ of a particle to be of the specie $i = e, \mu, \pi, K, p, d$ according to the sub-detector D. These likelihoods are combined into a global likelihood ratio called PID indicating the probability of a track to originate from a particle i:

$$PID(i) = \frac{\mathscr{L}_i}{\mathscr{L}_e + \mathscr{L}_\mu + \mathscr{L}_\pi + \mathscr{L}_K + \mathscr{L}_p + \mathscr{L}_d}, \tag{2.5}$$

where $\mathscr{L}_i = \prod_D \mathscr{L}_i^D$, with D the sub-detectors.

In this case, the likelihood $\mathscr{L}_i$ is tested against the likelihoods of all other particle types likelihoods and is thus referred to as "global" PID. However, PID can also be defined as "binary" for a separation between two specific particle types. For example $\mathscr{L}_{K/\pi} = \frac{\mathscr{L}_K}{\mathscr{L}_K + \mathscr{L}_\pi}$ is used to distinguish between kaons and pions.

For lepton identification, an alternative PID method based on machine learning is available. It relies on various different ECL measurements, combined with likelihoods from the other sub-detectors, which are then processed with a boosted decision tree (BDT). The BDT-based PID is available for both global and binary (lepton against pion) PID.

Studies performed with $J/\psi \to \ell^+\ell^-$ events [102] showed that the lepton identification efficiency is 86% for electrons and 89% for muons in a momentum range from 0.4 GeV/c to 7 GeV/c when using a $PID > 0.9$ selection. The corresponding fake rates against pions (i.e. the rate of mis-identification of a pion into the lepton of interest) are 0.4% and 7% for electrons and muons respectively. For the hadron identification, measurements performed with $D^{*+} \to D^0[K^-\pi^+\pi^0]\pi^+$ control samples [103] showed that the kaon identification efficiency is approximately 85% with a pion fake rate of $\sim 5\%$ for a $PID > 0.6$ selection for particles with momentum between 0 and 4 GeV/c.

Particle identification in Belle is also based on likelihoods from sub-detectors. However, hadron PID is a binary separation of pions and kaons, electron PID is defined against pions, and muon PID is defined against pions and kaons. For the electron identification, the E/p ratio is also used, as electrons are expected to deposit all their energy in the ECL, resulting in $E/p \approx 1$.

In terms of performance, the electron (muon) identification efficiency is 92% (89%) with a pion fake rate of 0.3% (1.5%) for a $PID > 0.5$ ($PID > 0.9$) selection and for electrons (muons) with momentum range 1-3 GeV/c [104, 105]. The main difference with Belle II is that the muon ID from Belle is only evaluated with the information from the KLM. The PID efficiency for muons with a momentum below 0.6 GeV/c is thus 0 (and drops rapidly going down from 1 GeV/c to 0.6 GeV/c, where it is about 40%). In Belle II, machine learning techniques are used to exploit information from the other detectors, leading to a non-zero efficiency for muons with a momentum below 0.6 GeV/c (but resulting in a higher fake rate).

2.4. Data production at Belle II

2.4.1. The Belle II analysis software basf2

The Belle II analysis software [30, 106], called *basf2*, is used for data processing and data analysis. It is based on its predecessor *basf* [29] used for the Belle experiment.

The software is written in C++ and is based on a modular structure where each module has a specific function or task in the reconstruction chain and is processed one after the other in a *path*, as can be seen in Figure 2.20. The data to be processed is read from the storage instance called *dataStore* and the modules can access information and settings conditions stored in a central database *DBStore* usually consisting of *payloads*. The payloads are ROOT files containing settings defined by their name and revision number and valid for a given data-taking period specified in an *interval of valididty (iov)*. The payloads can be grouped into sets called *globaltags*. Loading a specific global tag allows to include all payloads stored in it and needed for a specific data processing.

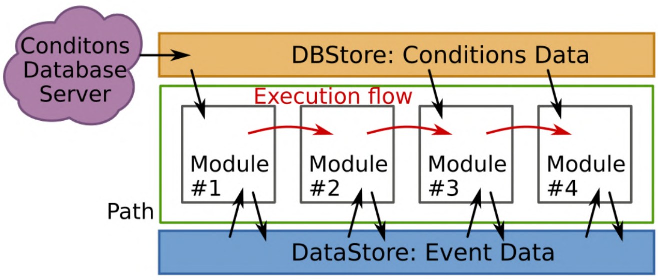


Figure 2.20: Schematic view of the basf2 data processing workflow (From Ref. [106]).

At the end of the "raw" data processing, i.e. the full reconstruction performed after the HLT decision, the events are stored in *mdst* (mini data-summary table) files.

From the analysis point of view, events from the mdst files are selected using a so-called "steering file". The steering file is a user-friendly python script that allows to create final particle lists from reconstructed tracks with given conditions and selections. The particle lists are then combined into intermediate particles and so on until the desired level of the decay chain. The selections can be applied at the track level, the combined track level (intermediate particles) or the event level to filter events and remove background. The variables of interest (kinematics, PID...) are attached to the relevant particles and a loop is performed over the events, which are processed one by one and stored into a ROOT file [107], usually called "ntuple".

The software management and upgrades are organised into full *releases* and *light releases*. Full releases contain all the software needed for the reconstruction and analysis, while light releases contain only improvements, updates or new features for the analysis level. The light releases are thus released much more often (about once a month) where full releases are released about once a year. At the time of the LS1, the last full release was release 6. Releases 7 and 8 were developed during LS1 mainly to test and calibrate new features for the post-LS1 data taking, and the release 9 is currently in preparation.

2.4.1.1. Using Belle data in basf2 : B2BII

In order to be able to use Belle data in the basf2 software, a specific framework called *B2BII* [108] has been developed to convert the Belle mdst (basf data format) format [29] into Belle II mdst (basf2 data format) format based on ROOT. B2BII allows to read and convert Belle mdst files within basf2 without any intermediate conversion to output ROOT-based file. All the Belle II tools are thus available for Belle data, including the B-tagging tool presented in the next section, and the same scripts can be used for the Belle and Belle II analyses.

2.4.2. Hadronic B-tagging

When working with $B\overline{B}$ events at B factories, it is often interesting to use the information coming from the partner B meson to exploit the well-known $\Upsilon(4S)$ initial state and the almost 4π hermetic detector. Indeed, the goal of the B-tagging is to reconstruct the full event, separating the B "signal" (channel of interest, B_{sig}) from the B "tag" (partner B, B_{tag}). This helps to improve the signal reconstruction by providing kinematic constraints from the tagged B meson for modes with missing energy in the signal side.

The tagged B meson can be reconstructed into different decay modes, which can be hadronic or semi-leptonic. On one hand, B semi-leptonic decays, which represent about 1/4 of all B decays, have a large branching fraction ($\mathcal{B}(B \to X\ell\nu_\ell) \approx 10\%$), but suffer from missing energy due to the neutrino in the final state.

On the other hand, B hadronic decays, despite representing the remaining 3/4 of B decays, have much smaller branching fractions (typical hadronic B decays have branching fractions of the order of 10^{-3}) and consist of events with a high track multiplicity, leading to a more complex reconstruction. However, thanks to the absence of missing energy, the four-momentum of the tagged B meson is well-known and allows good constraints on the B signal. In analysis with missing energy on the signal side like $B^0 \to K^{*0}\tau\ell$, the hadronic tagging is thus preferred.

2.4.2.1. The Full Event Interpretation

In order to perform B-tagging at Belle II, the *Full Event Interpretation* (FEI) [31] tool was developped for the basf2 software, as an upgraded version of its Belle predecessor, the *Full Reconstruction* (FR) [109]. The FEI is an algorithm based on machine learning techniques used to reconstruct a selection of exclusive B decay modes. In case of semi-leptonic tagging, it consists of 8 $B \to D^{(*)}(\pi)\ell\nu_\ell$ modes, while for hadronic tagging, it consists of 32 (36) fully hadronic modes for B^0 (B^+) tagging. The reconstructed modes for the B^0 hadronic tagging are displayed in Table 2.3.

The FEI is based on a hierarchical approach, illustrated in Figure 2.21, where the final state particles are reconstructed using low-level information such as tracks and clusters, and combined into intermediate particles and so on until the B meson candidates. A multivariate classifier is used to assign to each candidate its probability to be correct. In order to use all available information at each stage of the decay chain, a network of multivariate classifiers is built, and a classifier is trained for each final state particle and for each decay channel of an intermediate particle. The input features of the classifiers are based on PID, track kinematics and quality of the final state particles, neutral clusters timing, energy and direction as well as kinematics and vertex fit information of the intermediate particle candidates. Finally, the B_{tag} candidates are classified according to a quantity called signal probability, $P_{sig}(FEI)$, which is the probability of the B_{tag} candidate to be correctly reconstructed.

Hadronic B^0 decay modes			
0. $B^0 \to D^-\pi^+$	8. $B^0 \to D^{*-}D^0K^+$	16. $B^0 \to \overline{D}^{*-}\pi^+\pi^0$	24. $B^0 \to J/\psi K^+\pi^-$
1. $B^0 \to D^-\pi^+\pi^0$	9. $B^0 \to \overline{D}^{*-}D^{*0}K^+$	17. $B^0 \to \overline{D}^{*-}\pi^+\pi^0\pi^0$	25. $B^0 \to J/\psi K_s^0\pi^+\pi^-$
2. $B^0 \to D^-\pi^+\pi^0\pi^0$	10. $B^0 \to D^-D^+K_s^0$	18. $B^0 \to \overline{D}^{*-}\pi^+\pi^+\pi^-$	26. $B^0 \to \overline{\Lambda}_c^- p^+\pi^+\pi^-$
3. $B^0 \to D^-\pi^+\pi^+\pi^-$	11. $B^0 \to \overline{D}^{*-}D^+K_s^0$	19. $B^0 \to \overline{D}^{*-}\pi^+\pi^+\pi^-\pi^0$	27. $B^0 \to \overline{D}^0 p^+p^-$
4. $B^0 \to D^-\pi^+\pi^+\pi^-\pi^0$	12. $B^0 \to D^-D^{*+}K_s^0$	20. $B^0 \to D_s^{*+}D^-$	28. $B^0 \to D^-p^+p^-\pi^+$
5. $B^0 \to \overline{D}^0\pi^+\pi^-$	13. $B^0 \to \overline{D}^{*-}D^{*+}K_s^0$	21. $B^0 \to D_s^+\overline{D}^{*-}$	29. $B^0 \to D^{*-}p^+p^-\pi^+$
6. $B^0 \to D^-D^0K^+$	14. $B^0 \to D_s^+D^-$	22. $B^0 \to D_s^{*+}D^{*-}$	30. $B^0 \to \overline{D}^0 p^+p^-\pi^+\pi^-$
7. $B^0 \to D^-\overline{D}^{*0}K^+$	15. $B^0 \to \overline{D}^{*-}\pi^+$	23. $B^0 \to J/\psi K_s^0$	31. $B^0 \to \overline{D}^{*0}p^+p^-\pi^+\pi^-$

Table 2.3: List of reconstructed hadronic B^0 decay modes in the Full Event Interpretation (FEI).

The FEI is trained on Monte-Carlo (MC) simulations and thus need to be calibrated with data to account for possible data/simulation discrepancies. For that, calibration factors are derived from the study of $B \to D^{(*)}\pi$ and $B \to X\ell\nu_\ell$ decays [110, 111] for

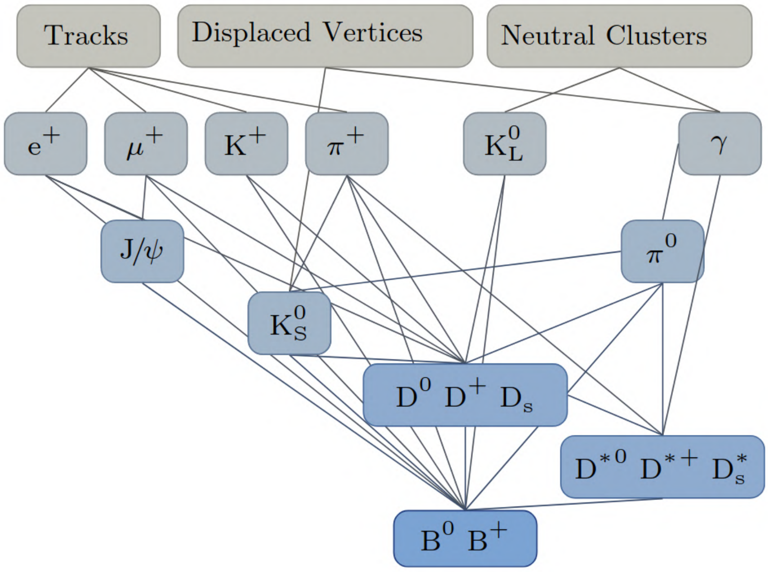


Figure 2.21: Schematic view of the Full Event Interpretation (FEI) algorithm (From Ref. [31]).

hadronic tagging. Moreover, the FEI is trained with Belle simulations and calibrated with the corresponding data. In addition to provide an accurate comparison of the FEI performance with respect to the previous Belle FR algorithm, it allows to use the FEI on Belle analyses, which is a non-negligible improvement, especially at low purity (see Figure 2.22). Studies performed with Belle data showed that the maximum tag-side efficiency of the FEI for B^0 hadronic tagging was around 0.3% for a purity of 15% and 0.2% for a purity of 50%.

2.4.3. Dataset production campaigns

2.4.3.1. Data

Before the Long Shutdown 1, the Belle II experiment collected a total of 424 fb^{-1} of data, including 365 fb^{-1} at the $\Upsilon(4S)$ resonance. The data is divided into different "experiments", each experiment being a set of runs, which are continuous data collection periods ranging from a few minutes to eight hours. The data collected before LS1 is divided into experiments up to 26, with the interesting physics events being in experiments 7, 8, 10, 12, 14, 16, 17, 18, 20, 22, 24, 26, with increasing instantaneous luminosity along the experiments.

At each new basf2 full release, the full Belle II dataset is reprocessed with the new release to benefit from all the software improvements and new features. The already existing dataset is compiled in a new "proc" dataset while data still being collected constitutes the "prompt" dataset and is processed in parallel of data taking with the most recent software release. The prompt data is divided into "buckets" for which all data in the same bucket is processed together while data for the next bucket is still

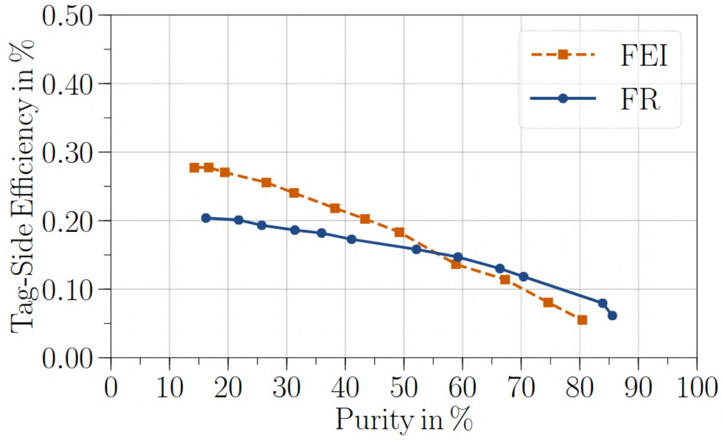


Figure 2.22: Comparison of the performance of the Full Event Interpretation (FEI) and the Full Reconstruction (FR) algorithms in terms of efficiency and purity for B^0 hadronic tagging. The FEI outperforms the FR algorithm performance at low and intermediate purity. At high purity, the tag-side efficiency cannot be extracted reliably due to the low statistics (From Ref. [31]).

being collected. This prompt data can be used for physics studies to prepare the next software releases or develop calibrations.

At the time of LS1, the 365 fb^{-1} collected during Run 1 are thus available in the proc13 (exp7-18) and prompt2022 (exp20-26) datasets, processed with the basf2 release 6.

In case of Belle, the 711 fb^{-1} of data collected at the $\Upsilon(4S)$ resonance are divided into 26 experiments between exp7 and exp65, following the same logic as Belle II.

2.4.3.2. Monte-Carlo simulations

In parallel to the data taking, Monte-Carlo (MC) simulations campaigns are run to produce data-like MC events. The MC production is necessary to get information not available in data, like the true identity of particles in events decay chains. For that, different event types, such as $B\bar{B}$ events but also $q\bar{q}$, τ-pairs or low multiplicity events, are generated using various MC packages. In particular EVTGEN [112] and PYTHIA [113] are used for $B\bar{B}$ and $q\bar{q}$ events while other packages are used for τ-pairs, Bhabha and low multiplicity events and the detector interactions are simulated with GEANT4 [114]. The set of MC samples containing all event types is referred to as "generic MC".

In case of Belle, the $B\bar{B}$ and $q\bar{q}$ events were also generated with EVTGEN and PYTHIA and the detector interactions were simulated with GEANT3 [115].

The MC simulations are divided into two categories: MCri (run-independent) and MCrd (run-dependent). The difference between the two categories lies in the simulation of the background: in MCri, events are simulated with a simulated background overlay, while in MCrd, events and detector interactions are simulated following the same experimental conditions as data. Thus, for each data run, an equivalent MC run can be created using the corresponding data background. The MCrd is therefore expected to be more accurate than MCri, but has the disadvantage of being much more complicated to produce due the splitting of the MC generation into numerous runs. For Belle II, the equivalent of 4 times the data luminosity (also referred to as 4 "streams" of data luminosity) was simulated for $B\overline{B}$ and $q\bar{q}$ events in MCrd, for a total of 1444 fb^{-1}. For Belle, only MCrd was produced, and we use four "streams" of data luminosity as well.

In addition to the generic MC, signal MC is produced on demand for analysts. It can be run-independent (MCri) or run-dependent (MCrd) (only MCrd for Belle) and is generated with EVTGEN for $B\overline{B}$ events.

Similarly to data, the generic and signal MC are reprocessed at each new release. At LS1, the MC campaign processed with release 6 and corresponding to the proc13 and prompt2022 data is MC15.

2.4.3.3. Skimming

At each data and MC production campaign, an additional step is performed called "skimming". The idea is to pre-filter data (and MC) according to some specific analyses such that this step is done centrally and not repeated by each analyst needing this selection, aiming at a more efficient management of the computing time and resources. Skimming is performed from mdst data (MC) and stored in *user-defined summary table* (udst) format, which allow the storage of analysis-like objects such as particle lists, and results in samples usually 90% smaller than the original datasets.

Skimming can be performed for very specific analysis types or in more general cases as for the FEI skims. Indeed, the FEI algorithm is long and heavy to run with a very low efficiency, and many analyses require B-tagging, which makes the FEI application and pre-selection a typical candidate for a dedicated skim. Analysts can therefore directly use the FEI (hadronic or semi-leptonic) skims, which are significantly lighter than the original datasets and in which one B candidate is already selected to be the B_{tag}, and only the signal part has to be selected in the steering file.

3. Study of the Belle II tracking performance

Summary

Tracking is a crucial part of the event reconstruction at Belle II. Studying its performance is therefore necessary to ensure the quality of the data and the physics results obtained from it, and continue to improve the event reconstruction. In this chapter, we present the study of Belle II tracking performance through two different analyses: the optimisation of the tracking algorithm and the measurement of the fake tracks rate. In the first part of this chapter, we will detail the track reconstruction chain and present an optimisation of an algorithm called DAF (Deterministic Annealing Filter) used to improve the tracks building accuracy. In the second part, tracking performance at a higher level is presented, as we evaluate at analysis level the fake tracks rate in the Belle II experiment, i.e. the rate of reconstructed tracks that do not correspond to a particle created in the e^+e^- collision.

3.1. Optimisation of the tracking reconstruction algorithm

In the previous chapter, we presented the main ideas of track reconstruction at Belle II. In this section, the fitting part of the tracking chain will be detailed, and we present a study to optimise the computing time of the tracking reconstruction chain by improving the so-called Deterministic Annealing Filter (DAF) algorithm used in the track fitting procedure. Indeed, the tracking reconstruction chain can be improved in many aspects which lead to consequences on the global tracking performance, and computing time improvements are particularly interesting as they affect the whole chain.

3.1.1. Tracking reconstruction workflow and the Deterministic Annealing Filter (DAF)

For this study, we need to distinguish the two tracking chains corresponding to the *fast reconstruction* (performed on HLT) and the *offline reconstruction* (performed after the HLT decision) defined in Section 2.3.

In case of *fast reconstruction*, only the information from CDC and SVD are used. As mentioned in Section 2.3, CDC tracks are created and extrapolated to the SVD to be attached to SVD space-points. SVD-only tracks are also created with the unattached SVD space-points. The CDC-only tracks, SVD-only tracks and CDC tracks attached to SVD space-points are merged together to form the hereafter called CDC-SVD tracks. At that point, the timing information is added to the tracks. All along the track finding chain, the tracks are fitted with a default pion mass hypothesis, but at this stage the tracks are refitted with the kaon and with the proton mass hypothesis. Finally, the HLT filter is applied, and the remaining tracks are refitted one last time when searching for K_s and Λ vertices. From this *fast reconstruction*, Regions of Interest (ROI) in the PXD are defined by extrapolating the tracks to the PXD and only PXD hits from these regions are saved and used for the offline reconstruction.

In the *offline reconstruction*, the CDC and SVD tracking is the same, but before adding the timing information, the PXD contribution is added. To do so, the CDC-SVD tracks are refitted and extrapolated to the PXD ROIs to create PXD tracks, which are merged with the CDC-SVD tracks. In addition, a final step is performed in which tracks with wrong charge assignment are identified. For these tracks, the charge is flipped, and the tracks are refitted with the three π, K, p mass hypotheses.

Along the tracking reconstruction chain, the tracks are fitted several times to keep the best approximation of the particles' trajectory, using a combinatorial Kalman filter (CKF) [116]. This iterative algorithm starts with a first approximation of the track parameters, which is extrapolated in the detector volume and corrected for the relevant effects (magnetic field non-uniformity, energy loss, multiple scattering). From the new extrapolated track candidate's position and uncertainty, hit candidates are selected, the track parameters are updated and the procedure is repeated. However, hits that do not belong to the particle track can be wrongly added to the reconstructed track. They can be for instance beam background hits, hits from other tracks in the event, or δ-rays. To reject these background hits and improve the fit accuracy, within the fitting procedure, an algorithm called Deterministic Annealing Filter (DAF) is used to improve the fitting process by removing outlier points used in the track fitting. The DAF is also used to remove the left-right ambiguity on CDC hits, corresponding to if the track is going through the left or right of the CDC wire that recorded the hit.

A schematic view of the full tracking chain for the *fast* and *offline reconstruction* is shown in Figure 3.1. The steps where the DAF algorithm is called are highlighted with the "DAF" label. A more detailed view of the DAF calls is presented in Table 3.1 with, in addition to the tracking steps, the *basf2* modules in which the DAF is called and the number of times the algorithm is run depending on the detector information used.

Tracking step	Module name	CDC	CDC + SVD	CDC + SVD + PXD
Extrapolate CDC tracks to SVD	`CDCToSVDSpacePointCKF_backward`	–	✓	✓
Merge CDC and SVD tracks	`RelatedTracksCombiner`	–	✓*	✓*
Refit CDC+SVD tracks before extrapolating to PXD	`DAFRecoFitter`	–	–	✓
Merge CDC+SVD and PXD tracks	`RelatedTracksCombiner`	–	–	✓*
Refit tracks with timing information	`Combined_DAFRecoFitter`	✓	✓	✓
Refit tracks with K and p mass hypotheses	`TrackCreator`	✓ ×2	✓ ×2	✓ ×2
Search for K_s and Λ vertices	`V0Finder`	✓*	✓*	✓*
Flip track charge and refit it	`Combined_DAFRecoFitter_flipped`	–	–	✓
Refit the flipped tracks with K and p mass hypotheses	`TrackCreator_flipped`	–	–	✓ ×2
Number of DAF calls per track (up to)		4	6	11

Table 3.1: Summary of the DAF calls along the tracking chain with basf2 modules in which are the calls. The asterisk indicates that the DAF is not called for every track but only if some condition is filled. The *fast reconstruction* (on HLT) corresponds to the CDC+SVD columns while the *offline reconstruction* corresponds to the CDC+SVD+PXD column. The CDC-only column is shown for information at it is the simplest case, used to develop the optimisation.

The idea of the Deterministic Annealing Filter is to assign weights between 0 and 1 to the hits used to build a track accordingly to the residuals between the measurement and the fit prediction. The DAF then works in an iterative way, by refitting the tracks several times (between N_{min} and N_{max}) adjusting the weights until a convergence criterion is reached. At each iteration, a parameter called *annealing temperature* T is lowered. A high temperature corresponds to a soft assignment, making the weights tend to 0.5, while a low temperature is a harder assignment where the weights tend to be 0 or 1. The annealing temperature is therefore reduced from T_{max} to T_{min} in N_{iter} steps. For the steps between N_{iter} and N_{max}, the temperature is kept constant and

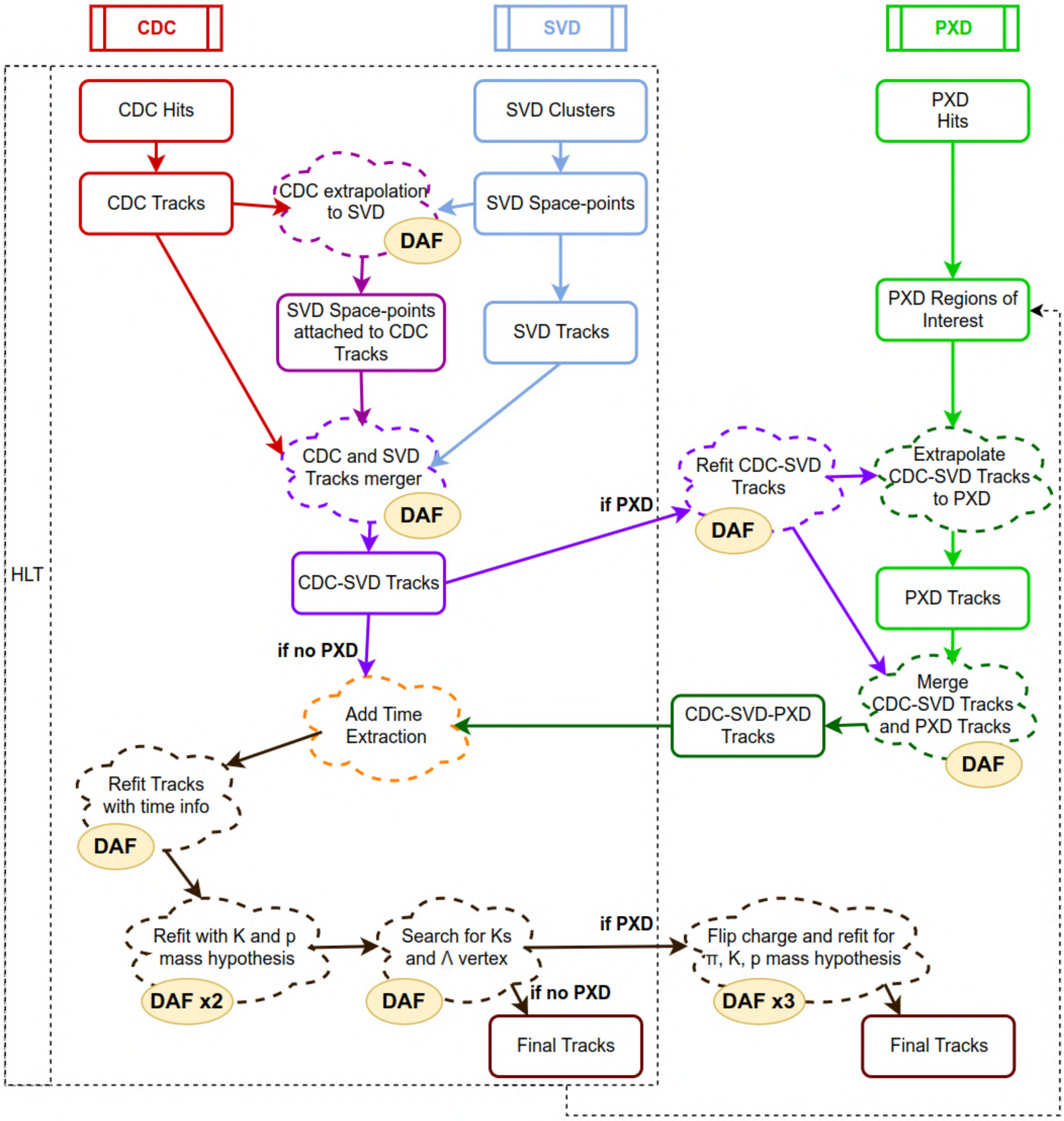


Figure 3.1: Schematic view of the tracking chain. The solid line rectangles represent tracking objects while the dotted line clouds represent tracking algorithms. The "DAF" mentions indicate each module in which the DAF is run. The part of the chain in the "HLT" box is the part that is processed during *fast reconstruction* while the total chain is processed during *offline reconstruction.*

when the weights reach a given threshold, they are forced to 0 using a damping factor regulated by a probability cut P. To stop iterations, three convergence criteria are defined based on the variation Δw of the weights w_j, the variation Δp of the p-value of the fit p_j and the minimum and maximum number of iterations N_{min} and N_{max}.

- Criterion 1: $\max_{i \in \text{track}} (|w^i_{j-1} - w^i_j|) < \Delta w$

- Criterion 2: $j > N_{\text{min}}^{\text{p-val}}$ and $|p_{j-1} - p_j| < \Delta p$
- Criterion 3: $j > N_{\text{max}}$

where i corresponds to the hits associated to a given track, j corresponds to the iterations, and Δw, Δp, N_{min} and N_{max} are defined beforehand. $N_{\text{min}}^{\text{p-val}}$ corresponds to a minimal number of iterations before checking the p-value convergence criterion. At least one of the criteria should be satisfied to stop the iterations.

In the initial configuration of the DAF (detailed in Table 3.2), the values of the T_{max}, T_{min}, N_{iter}, Δw, Δp, N_{min}, N_{max} and $N_{\text{min}}^{\text{p-val}}$ parameters are such that criterion 1 is never satisfied and criterion 2 is almost always immediately satisfied as soon as checked, so at the $N_{\text{min}}^{\text{p-val}}$ iteration. The DAF therefore almost always runs $N_{\text{min}}^{\text{p-val}}$ iterations, which reflects a clearly suboptimal behaviour.

3.1.2. The timing challenge and optimisation idea

During the first period of data taking, the L1 trigger output rate, corresponding to the HLT input, was very close to the HLT throughput limit, i.e. the number of processes performed by the HLT in a given time, as can be seen in Figure 3.2. For the future data taking at higher instantaneous luminosity, we expect events with an increasing complexity due to higher background rates, leading to a longer processing time, in addition to the larger absolute number of events. During LS1, pressure on HLT was released by adding three more units to HLT allowing a higher throughput. New collimators installed on SuperKEKB will also help to reduce the total input rate, thus reducing the L1 output [117].

However, this is only a temporary solution and the input rate and complexity of events will continue to grow with the increase of luminosity. Any improvement in the computing time is thus worth to consider. In addition, several modules used in track fitting and in which the DAF is called are among the most time-consuming modules on HLT, as can be seen in Figure 3.3.

As it happens, the DAF has never been optimised since its implementation thus leading to the suboptimal behaviour described earlier. The goal of the optimisation is therefore to reduce the average number of iterations of the DAF by tuning the convergence criteria for a more efficient behaviour, while keeping the tracking performance unchanged. The optimisation idea is therefore the following:

- Loosen criterion 1 by increasing the Δw threshold and allow convergence by weight.
- Tighten criterion 2 by decreasing the Δp threshold avoiding an automatic convergence by p-value.
- Decrease the N_{min} and $N_{\text{min}}^{\text{p-val}}$ number of iterations to reduce the average number of iterations.

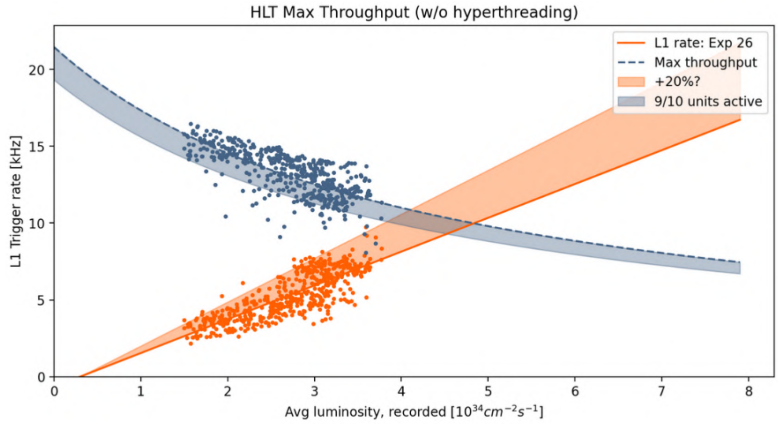


Figure 3.2: HLT maximum throughput (blue points and trend) and L1 output (orange points and trend) according to recorded luminosity for the experiment 26 (From Ref. [118]).

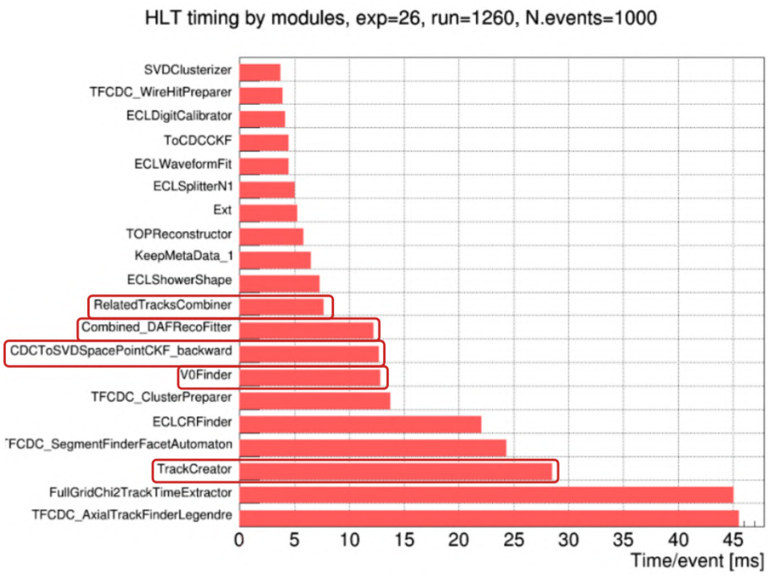


Figure 3.3: Average CPU time of the most time-consuming modules in the HLT chain for simulated muon events. The modules calling the DAF are circled in red (From Ref. [117]).

- Lower the temperature range and in particular the initial temperature T_{max} to accelerate the convergence.

The values of the different parameters in the initial (baseline) and in the optimised configurations are presented in Table 3.2. In addition, in the optimisation, the iterations parameters N_{iter}, N_{min}, N_{max} and $N_{\text{min}}^{\text{p-val}}$ are decorrelated such that N_{min}, N_{max} and $N_{\text{min}}^{\text{p-val}}$ are independent of N_{iter}. To achieve this optimisation, the DAF parameters are modified and the impact on the computing time and the tracking performance is evaluated in MC simulations reproducing a very simplified version of data. Once a suitable configuration is found, the performance are evaluated in more complex simulations. These simulations and the method for the performance evaluation is detailed in the next section. The initial configuration is referred to as *baseline* and the configuration with the parameters in the *optimised* column from Table 3.2 is referred to as *optimised* in the next sections. Note that we call "optimised" the new configuration with respect to the baseline configuration, but it is not necessarily the best possible configuration and further improvements could be made.

As the DAF is run multiple times along the tracking chain, and this for each track in the event, any small improvement should still have a non-negligible impact on the total time of the tracking reconstruction.

DAF parameter	Description	Baseline	Optimised
Δw	weight threshold	0.001	**0.06**
Δp	p-value threshold	1	**0.001**
P	Probability cut	0.001	0.001
$(T_{\max}, T_{\min}, N_{\text{iter}})$	Annealing scheme	(100, 0.1, 5)	(**2**, **0.01**, 5)
$N_{\min}$	Minimum number of iterations	$N_{\text{iter}} (= 5)$	**1**
$N_{\min}^{\text{p-val}}$	Minimum iterations before p-value criterion checking	$N_{\min} (= 5)$	**1**
$N_{\max}$	Maximum number of iterations	$N_{\text{iter}} + 4 (= 9)$	9

Table 3.2: DAF parameters and their value for the initial (baseline) and the optimised configurations. The modified parameters are highlighted in bold.

3.1.3. Setup and measurement strategy

To evaluate the timing improvement in the tracking chain, we measure in MC simulations the execution time of each *basf2* module calling the DAF as well as the total time improvement adding all the module's computing time together. To estimate the timing improvement and evaluate the tracking performance, we start with a very simple simulation, and we increase progressively the complexity until reaching a simulation close to what is expected in the upcoming data taking. We distinguish three different levers to create the simulated data:

- Event type: Muon beam or $B\overline{B}$ events (all simulated).
- Background conditions: Absence of background, background corresponding to the recorded background in experiment 26 which is the highest background recorded in data (peak luminosity of $4.7\times10^{34}\text{cm}^{-2}\text{s}^{-1}$), and "nominal" background which corresponds to the expected background at nominal luminosity ($6\times10^{35}\text{cm}^{-2}\text{s}^{-1}$).

- Sub-detector used for reconstruction: CDC only, CDC+SVD which corresponds to what is used during *fast reconstruction*, CDC+SVD+PXD which is used during *offline reconstruction*.

Reconstructions with all CDC, SVD and PXD information is independent of the computing time in HLT, but the timing improvement of the offline reconstruction also has indirect impact for the future MC production and data reprocessing. Simulations using a muon gun, CDC only and no background are used to define the first optimisation, before testing it on more complex simulations and refine it. The optimised values shown in Table 3.2 were chosen arbitrarily in the first place, and the ones leading to a suboptimal behaviour of the DAF or to a lower tracking performance with respect to the baseline were modified to reach a more efficient behaviour.

For every simulated sample (μ-gun, $B\overline{B}$) × (no background, exp26 background, nominal background) × (CDC, CDC+SVD, CDC+SVD+PXD), the CPU time of each basf2 module calling the DAF is measured in the initial and the optimised version of the DAF parameters. The CPU times of all modules are also added together and the total timing improvement of these modules is evaluated. Additionally, we estimate the average number of iterations of the DAF.

In parallel several quantities accounting for the tracking performance are verified, as the main challenge of the DAF optimisation is to not degrade the tracking performance. For that, we estimate the residuals and pull distributions of parameters such as the track momentum or the distance of the point-of-closest-approach of the tracks to the origin, the efficiency and fake rate, and the p-value of the track fit. Note that in this case, the fake rate represents the rate at which a reconstructed track doesn't match any true track or is paired with a wrong true track, so is different to the fake tracks rate evaluated in Chapter 3.2. Similarly, the efficiency here corresponds to the rate at which true tracks are matched with the correct reconstructed track.

3.1.4. Results

The simplest case of 1000 simulated muon events, with the CDC tracking only and with an absence of background scenario was used to define a first optimisation, before testing it on more complex simulations and refining it. Figure 3.4 shows the timing difference between the original DAF configuration and the optimised one in the `Combined_DAFRecoFitter` module and for the sum of all basf2 module calling the DAF.

Configurations closer to reality can then be studied. To mimic the real data, simulated $B\overline{B}$ events are used and the timing improvement and tracking performance of the fast reconstruction chain, corresponding to CDC and SVD tracking are evaluated in the experiment 26 and nominal background scenarios. For each simulation, 10000 $B\overline{B}$ events are used. The total CPU time of the baseline and optimised DAF parameters configurations for the exp26 and nominal background scenarios are shown in Figure 3.5, and the average numbers of DAF iterations in each case are shown in Figure 3.6. For the experiment 26 background scenario, the detail of CPU time of each module calling the DAF is shown in Figure 3.7.

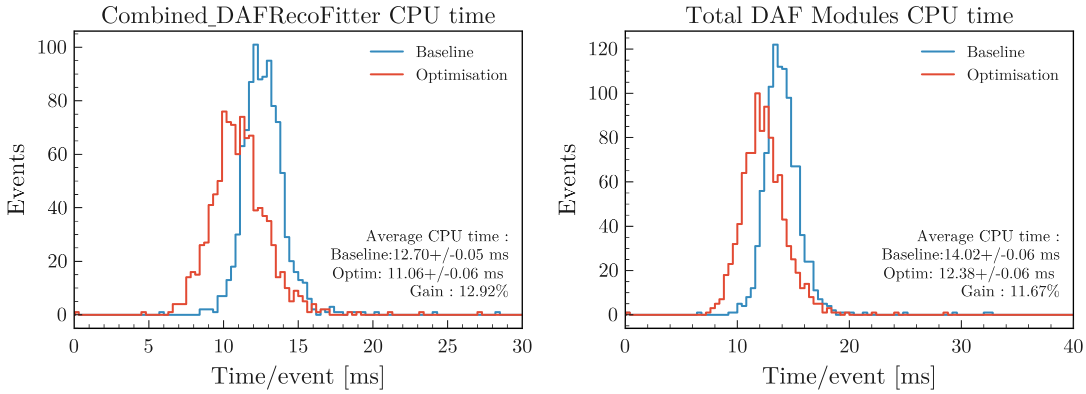


Figure 3.4: CPU time difference between the original DAF configuration (baseline) and the optimised one in the `Combined_DAFRecoFitter` basf2 module (left) and the sum of all the modules calling the DAF (right) for the μ-gun, CDC only, no background configuration.

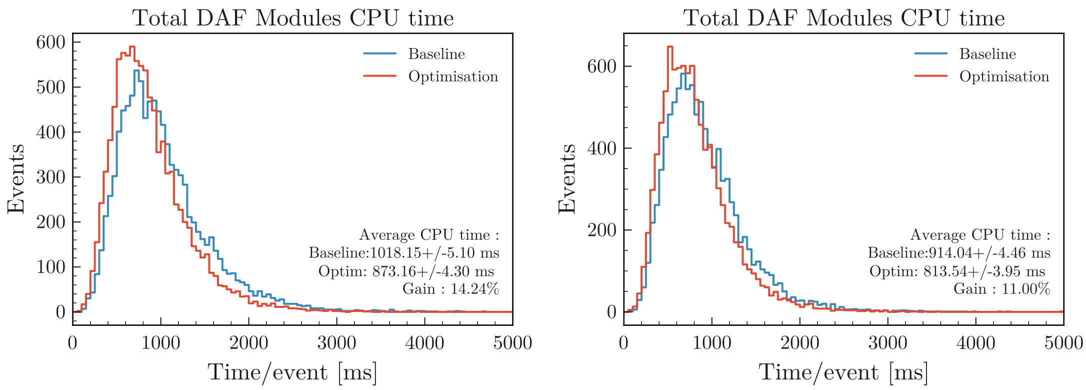


Figure 3.5: CPU time of all modules calling the DAF in the $B\overline{B}$ events, CDC+SVD tracking and experiment 26 (left) and nominal (right) backgrounds configurations.

The timing improvement varies from 0% to ~20% in the different modules and results into a 14% total improvement in case of experiment 26 background, and 11% total improvement in case of nominal background. Only the V0Finder module shows no improvement with the optimised version of the DAF. However, this module calls the DAF only under certain conditions and its total CPU time is thus dominated by other processes than the action of the DAF. Also, its absolute CPU time being in average larger than for the other modules calling the DAF (up to 1s for V0Finder vs < 500ms for other modules), its contribution to the total CPU time is large. The average total improvement visible in Figure 3.5 is thus closer to 10-15% than 20%. In addition, the average number of iterations run by the DAF is reduced from 5 to 3.3 (3.2 with nominal background), as seen in Figure 3.6.

In parallel of the timing optimisation, one must make sure the tracking performance

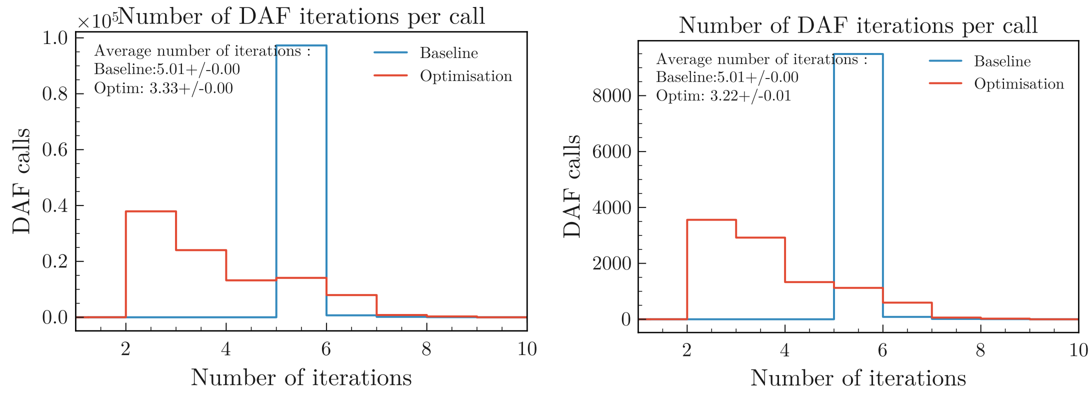


Figure 3.6: Number of DAF iterations per call in the $B\bar{B}$ events, CDC+SVD tracking and experiment 26 (left) and nominal (right) backgrounds configurations.

is not affected by the changes in the algorithm. For that, different parameters are studied. In particular, the following parameters are measured and shown in Figure 3.8:

- d_0 residuals: Residuals (estimate - truth) of the signed distance to the track point-of-closest-approach (POCA) in the $r-\phi$ plane (where the distance is measured with respect to the origin (0,0,0)).
- p_T pulls: Pull distribution ((estimate - truth)/error) of the tracks transverse momentum
- Finding efficiency (with respect to p_T): Rate at which the true MC tracks are matched with the correct reconstructed tracks
- Fake rate (with respect to p_T): Rate at which a reconstructed track doesn't match any MC track or is paired with the wrong MC true track.
- p-value: p-value of the track fit

Apart from the p-value, no degradation of the tracking parameters is observed with the optimised configuration. In case of the p-value parameter, we observe some excess in the ~ 0 p-value bin for the optimised configuration. This excess represents 3% more events at ~ 0 p-value in the optimised configuration compared to the baseline. This difference is observed in every tracking chain × background scenario configuration for $B\bar{B}$ events with the same amplitude. A part of the 0 p-value events were due to an immediate convergence with the criterion 2 when both p-values were 0, thus stopping the algorithm at first iteration and resulting into a track with a 0 p-value. To mitigate this behaviour, we asked in criterion 2 that in addition, both p_{j-1} and p_j should be larger than 10^{-10} to allow the algorithm to stop. With this change the discrepancy in the ~ 0 p-value bin was reduced to approximately 1.5%[1]. As the 0 p-value effect is not

[1] This change was reversed after tracking validation tests with Run 2 data, the reason and consequences are described in Section 3.1.5.

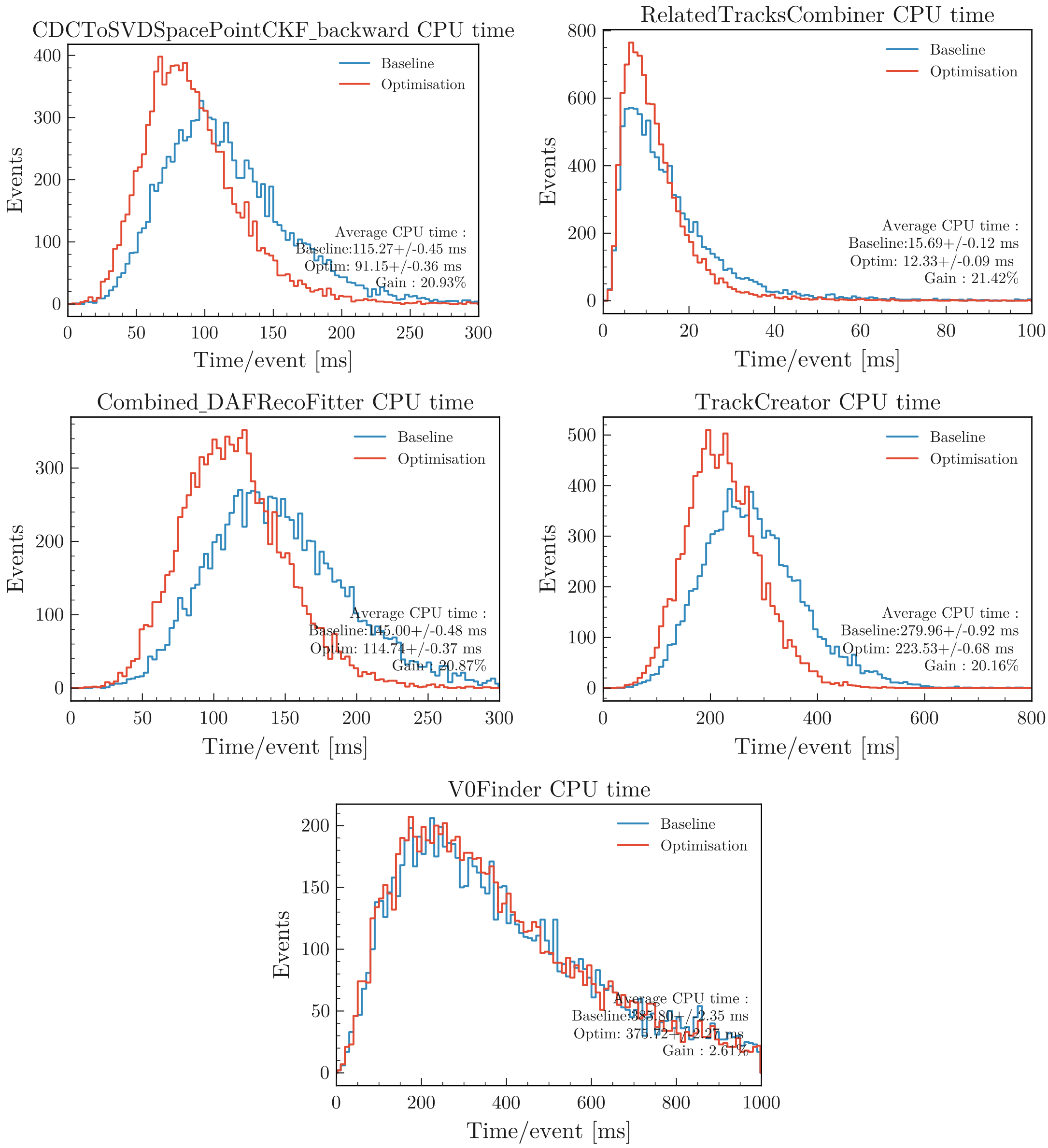


Figure 3.7: CPU time difference between the original DAF configuration (baseline) and the optimised one in the different basf2 modules calling the DAF for the $B\overline{B}$, CDC+SVD, exp26 background configuration.

observed to impact any other tracking parameter and as no selection is made in the tracking chain based on the track fit p-value, this effect is considered as not degrading the global performance.

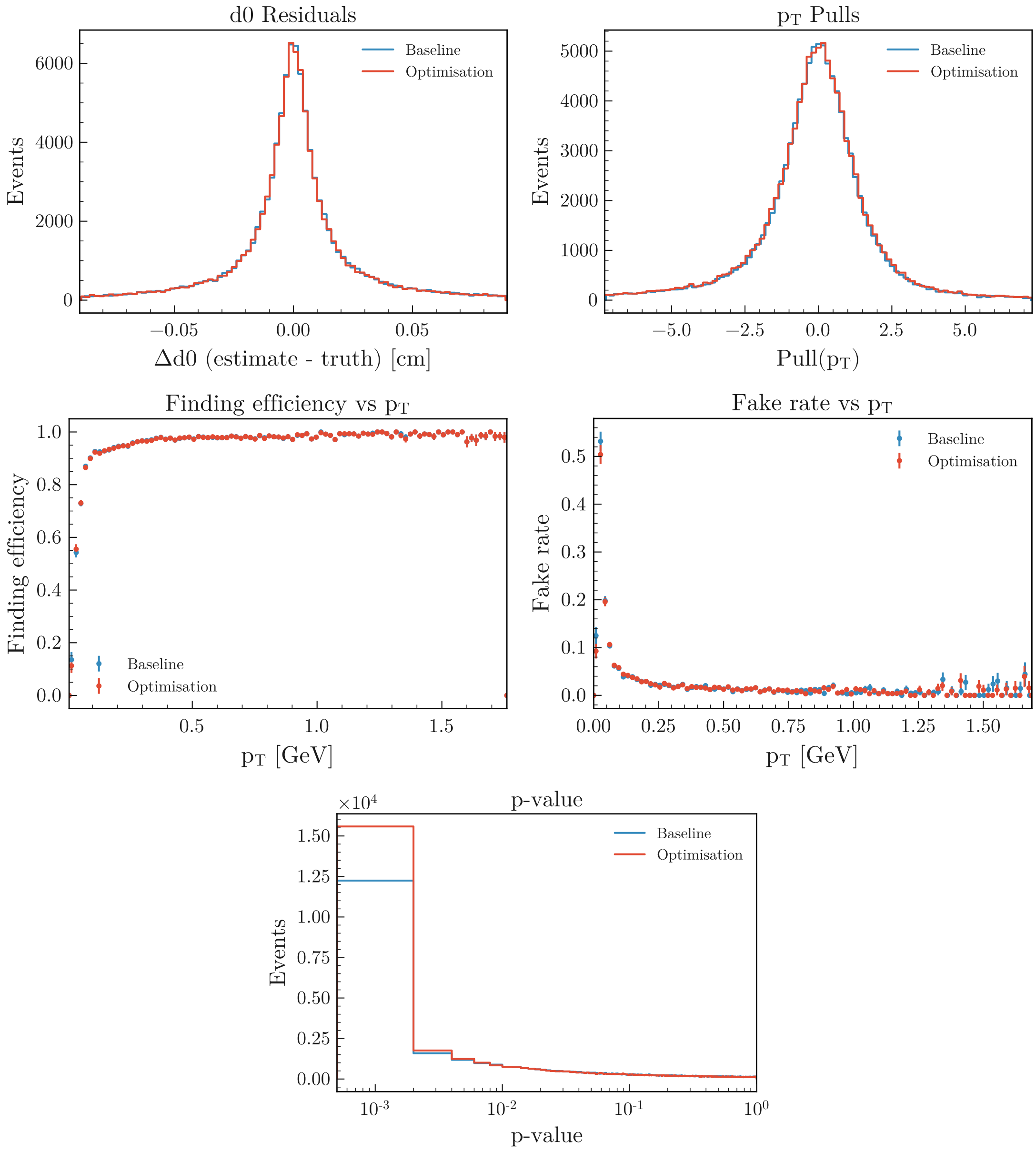


Figure 3.8: Tracking parameters for the original and optimised DAF configurations, in the $B\overline{B}$ events, CDC+SVD tracking, experiment 26 background scenario.

Despite the workload tensions are mainly affecting the HLT reconstruction, one can evaluate the time gain also in the offline reconstruction. Note that at the time of the study the "flip and refit" algorithm, which consists into flipping the track charge and refit it with the π, K and p mass hypothesis, was not implemented for the CDC+SVD+PXD full tracking chain. The global timing improvement is thus expected to be larger than the measured ones, which are shown in Figure 3.9. For reference, the timing improvement in the modules calling the DAF during flip and refit are shown in Figure 3.10 for the $B\overline{B}$ events, CDC+SVD tracking scenarios. Note again that now, the flip and refit algorithm is not applied on HLT, so the CPU timing of these modules is not taken into account in the results presented previously for the fast reconstruction tracking chain.

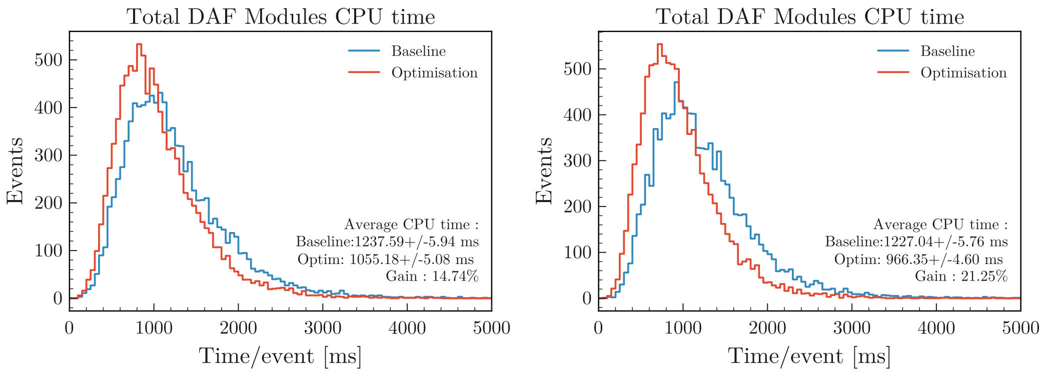


Figure 3.9: CPU time of all modules calling the DAF in the $B\overline{B}$ events, CDC+SVD+PXD tracking and experiment 26 (left) and nominal (right) backgrounds configurations.

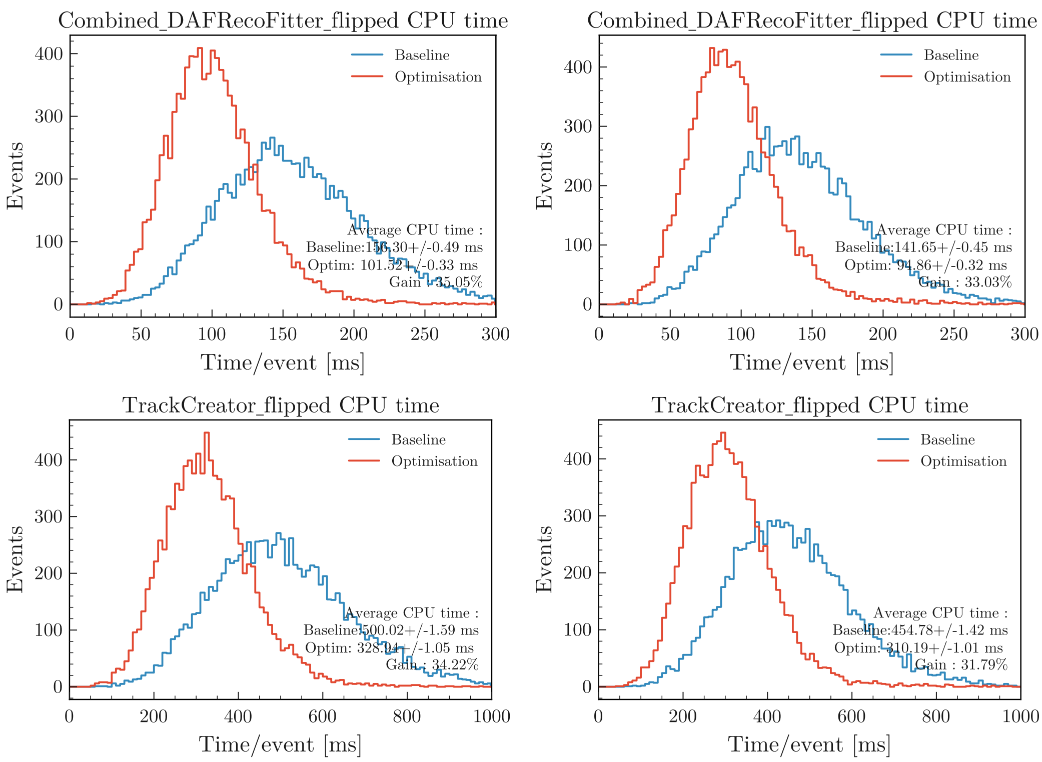


Figure 3.10: CPU time of the modules calling the DAF during the flip and refit algorithm in the $B\overline{B}$ events, CDC+SVD tracking and experiment 26 (left) and nominal (right) backgrounds configurations.

To conclude, this study lead to changes in the basf2 software to include the DAF parameters as a payload instead of having them directly coded in the software. Both the original and the new (optimised) version of the parameters are now available and easily accessible if one wants to use one version or the other. In addition, custom values of the parameters can be used easily, which will help in case of further improvement studies in the future.

Nevertheless, this study was performed only in MC simulations and the optimised parameters have not been validated in data in a dedicated analysis.

3.1.5. The DAF today

The optimised DAF parameters presented in this chapter were used for the first tests of Run 2 data processing. During validations of tracking performance, a loss in CDC tracking efficiency was observed. Indeed, for high momentum tracks ($p > 1.3$ GeV/c) in a very specific angular region ($\lambda = \pi/2 - \theta \approx 55°$ where θ is the radial angle), the CDC-only tracking fails to reconstruct tracks. These tracks are recovered by SVD-only tracking leading to a preserved tracking efficiency when using CDC and SVD information, but a degraded efficiency for CDC-only tracking. The reason of this loss was found to be the change made in DAF convergence criterion 2 so that p_{j-1} and p_j should be larger than 10^{-10} instead of 0. Several studies showed that the CDC track loss effect was much reduced in Run 1 compared to Run 2 data, and in MC compared to data, as shown in Figure 3.11, hence the non-observation of this effect when validating the changes made for DAF optimisation in Run 1-like MC simulations.

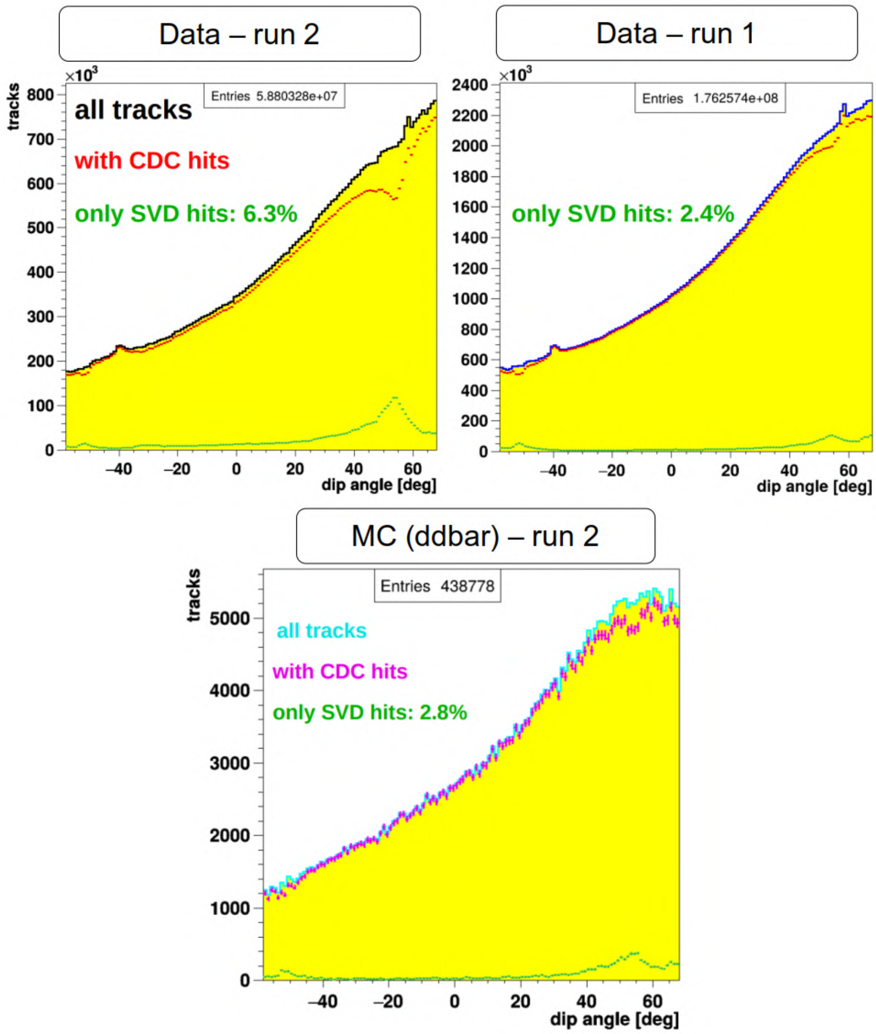


Figure 3.11: Track distribution in Run 2 data (top left), Run 1 data (top right) and Run 2 MC (bottom). The DAF optimised parameters and p-value $> 10^{-10}$ requirement are used in all cases. The loss of CDC tracks (see red/pink histogram) is much reduced in Run 1 data and in Run 2 MC (From Ref. [119]).

To solve this issue and recover the CDC tracking efficiency, the 10^{-10} p-value cut was removed from the software and added as DAF parameter in the payload developed in the optimisation study. In addition, custom sets of DAF parameters were defined to allow the use of a given set of parameters for different purposes in the tracking chain and further studies will be performed in the future to reintegrate the p-value selection with an adapted optimisation of the parameters.

3.2. Measurement of the fake tracks rate

In the previous section, we studied tracking performance from the software point of view. Nonetheless, direct impacts of the tracking quality are also visible at analysis level, and having a follow-up of tracking performance through the study of various parameters after reconstruction of events and along the data taking periods is necessary. In this context, one can study the fake tracks rate. This quantity has been measured in the past for the previous data and MC production campaigns, in particular the procedure was optimised with the MC14 and proc12 campaign. With the new MC15 and proc13+prompt2022 MC and data processing campaign, a reevaluation of the fake tracks rate is needed to follow the evolution of the tracking performance with new data and new software updates.

It is also important to monitor the fake tracks rate according to the beam conditions. Indeed, the fake tracks rate is currently low enough to not be a concern for the tracking performance, but can become a problem in the future with higher luminosities and thus higher beam-induced background rates.

In addition, the fake tracks rate can be used, among others, as an estimator of the tracking performance when deciding which update to implement in the Belle II software to improve the tracking reconstruction algorithms.

In this section, we present the procedure to estimate the fake tracks rate in MC and data collected during the first data-taking period from 2019 to 2022. A study of the impact on fake tracks rate of various tracking reconstruction configurations is also presented, with final aim to decide which configuration will be implemented in the new software version for future data processing.

3.2.1. Fake tracks rate definition and analysis strategy

Fake tracks refer to reconstructed tracks that do not correspond to tracks from the actual physics event. These tracks can be reconstructed with hits from beam-induced background, tracks originating from incorrect hit combinations, or what is known as "clone tracks". Clone tracks are reconstructed from hits belonging to another track but remaining unmerged, thus interpreted as two (or more) distinct particles instead of a single one. This is typically observed for low-momentum particles curling inside the detector. Hence, the fake tracks rate, together with the track finding efficiency, are good indicators of the tracking performance at Belle II.

This measurement is carried out in $e^+e^- \to \tau^+\tau^-$ processes as the cross-section for τ-pair production (0.9 nb) is large and close to that of B meson pairs at the $\Upsilon(4S)$ resonance energy (1.1 nb). In addition, the jet-like topology of τ decay products and their wide track momentum range coverage (200 MeV to around 3.5 GeV) are suitable to study tracking performance in a high-density track environment while still taking advantage of the event low multiplicity.

In these τ-pair events, one tau lepton is required to decay leptonically ($\tau^\pm \to \ell^\pm \nu_\ell \bar{\nu}_\tau$, $\ell = e, \mu$) while the other should decay hadronically into three charged pions ($\tau^\pm \to$

$3\pi^\pm\nu_\tau$). These decays are referred to as the *1−prong* and *3−prong* τ-decays, respectively. The case where the leptonic decay is $\tau \to e^\pm \nu_e \bar{\nu_\tau}$ is labelled as *electron mode* and the case where it is $\tau \to \mu^\pm \nu_\mu \bar{\nu_\tau}$ is labelled as *muon mode*. A schematic view of such event, called 3 × 1−prong event, is presented in Figure 3.12.

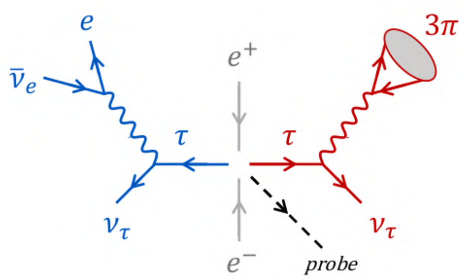


Figure 3.12: Schematic view of a 3 × 1−prong τ-pair event. The τ leptons are produced back to back in the centre of mass frame.

The fake tracks rate is then measured using a tag-and-probe method. For that, a first sample labelled as N_4 is composed of events in which exactly 4 tracks corresponding to the 1−prong and 3−prong decays are reconstructed. A second sample labelled as N_5 is composed of events for which an additional 5th track is found. In this sample, the 1−prong and 3−prong τ-decays form together the tag particles while the additional track is the probe particle. The fake tracks rate r_{fake} is therefore defined by:

$$r_{\text{fake}} = \frac{N_5}{N_4 + N_5}, \tag{3.1}$$

where N_4 and N_5 are the numbers of events where four and five tracks are found, respectively.

To mitigate the effect of background contamination, the fake tracks rate in signal data is estimated rescaling the N_4 and N_5 yields by the signal purity $\frac{S}{S+B}$ measured in MC simulations, where S and B are the numbers of signal and background events respectively.

This measurement has been carried out in both electron and muon modes. However, the selection used was optimised only for the electron channel, following the previous work done in [120]. The muon channel giving satisfying results in terms of purity despite the not optimised selection, the fake tracks rate results for this channel are presented here as well. Nevertheless, a more in-depth study would be required to define a more efficient background suppression selection for the muon mode.

3.2.2. Datasets

The data samples correspond to the Moriond2023 dataset, i.e. the proc13 reprocessing of the 2019-2021 data as well as 2022 prompt processing. The total integrated luminosity is 362 fb^{-1}.

MC samples are taken from the run-independent (MCri) and run-dependent (MCrd) MC15 production campaign. The list of samples used and the corresponding luminosity is presented in Table 3.3.

	MC15ri [fb^{-1}]	MC15rd [fb^{-1}]
taupair	1000	1455.052
uubar, ddbar, ssbar, ccbar, mixed, charged	1000	1455.052
ee	100	36.376
mumu	1000	1455.052
eeee, eemumu	200	–
eepipi	1000	–
eeKK, eepp, mumumumu, eetautau	2000	–
tautautautau	10000	–
llXX	–	363.763

Table 3.3: Luminosity available for the MC15 samples used.

3.2.3. Reconstruction and selection

3.2.3.1. Tracks and event selection

For the tag part of the event, 4 tracks corresponding to 3 pions (3−prongs tracks) and 1 lepton (electron or muon) (1−prong track) are reconstructed with the following requirements:

- All tracks should come from the interaction point. This translates into having radial distance to the IP $dr < 1$cm and longitudinal distance $|dz| < 3$cm.
- All tracks should have a transverse momentum larger than 200 MeV/c.
- The 1−prong track should have its particleID larger than 0.9, i.e. electronID> 0.9 for the electron mode, muonID> 0.9 for the muon mode.
- The 3−prongs tracks do not have pionID requirement but their muonID should be lower than 0.9. In addition, the track with an opposite charge sign compared to the two other tracks should have a kaonID lower than 0.6. This requirement was added as it was observed in MC simulations that in background, this track often corresponded to a kaon.
- The 3−prongs tracks and the muon track in muon modes should have E/p smaller than 0.6 to separate them from electrons and cluster energy larger than 0.

For the reconstruction of N_5 events, a fifth track is reconstructed in addition to the 4 tag tracks. This track is only asked to fulfill the interaction point criterion, i.e. $dr < 1$cm and $|dz| < 3$cm.

Finally, all events are required to fire at least one of the low multiplicity triggers. The track selection criteria are summarised in Table 3.4

	Probe track	Tag pion track	Tag electron track	Tag muon track
$\lvert dz \rvert$ [cm]	< 3	< 3	< 3	< 3
dr [cm]	< 1	< 1	< 1	< 1
p_T [MeV]	–	> 200	> 200	> 200
electronID	–	–	> 0.9	–
muonID)	–	< 0.9	–	> 0.9
$\frac{E_{\text{cluster}}}{p}$	–	< 0.6	–	< 0.6
E_{cluster}	–	> 0	–	> 0

Table 3.4: Track selection criteria.

3.2.3.2. Background suppression

The dominant backgrounds are mainly continuum ($q\overline{q}$ pairs) and low multiplicity backgrounds (ee, $\mu\mu$, $eeee$, $ee\mu\mu$...). The background suppression strategy developed is thus focused on exploiting the jet-like tau pair decay topology and the kinematics of its products.

Event shape:

- The opening angle between the 1−prong track and each of the 3−prongs tracks in the centre-of-mass frame (CMS) should satisfy $cos(\theta^*) < -0.5$.
- The transverse momentum of the combined 3−prongs tracks should be larger than 3 GeV/c and that of the 1−prong track should be larger than 1 GeV/c.

τ daughters kinematics:

- The 1−prong track CMS momentum should be lower than 80% and larger than 20% of the beam energy. The upper threshold significantly reduces the radiative dilepton background, while the lower threshold helps to suppresses the continuum and $e^+e^- \to e^+e^-\mu^+\mu^-$.
- The difference between the invariant mass of the two pion tracks from the 3−prongs tracks that have opposite sign charges (OS) should be within ±100 MeV/c^2 around the ρ meson mass $m_\rho = 775.26$ MeV/c^2. Since there are 3 pion tracks, there are two different $\pi^+\pi^-$ invariant mass combinations, to which the selection is applied. This targets hadronic τ-decays with an intermediate ρ meson ($\rho \to \pi^+\pi^-$), and significantly reduces background from continuum.
- The invariant mass of the two pion tracks from the 3−prongs tracks that have same sign charge (SS) should be larger than 300 MeV/c^2 and lower than the τ mass $m_\tau = 1776.86$ MeV/c^2. This helps to reduce background from continuum.

- The invariant mass of the 3−prongs tracks $M_{\pi\pi\pi}$ is required to be smaller than 1.3 GeV/c^2.

Event clean-up:

- There shouldn't be any π^0 in the event. This limits the number of photons in the event that could convert and create additional tracks, and improves the signal purity of events with five tracks.
- There shouldn't be any photon in the event for the electron mode. This avoids to count as fake tracks electrons due to photon conversion.

The selection was optimised with the previous production campaign proc12/MC14 for the electron mode and was mainly kept as it was for this update of the fake rate study. The differences between last study [120] and this one are the requirement on the absence of photon in the event for the electron mode, and the belonging of the probe track to a specific region of phase space. For this last point, in the previous studies, the probe track was required to belong to the same hemisphere as the 3−prongs tracks, while it has now no more spatial requirement. The fake tracks rate is thus expected to increase proportionally compared to the previous measurements, as the full phase space is now considered for the probe track.

Although the selection was optimised targeting the electron mode, it is used as is for the muon mode. A dedicated study and optimisation would be required to perform a more accurate fake tracks rate analysis on the muon mode.

3.2.3.3. Definition of signal events in MC

Signal events in the simulated τ-pair samples have been defined by exploiting the available MC truth matching information.

Events must satisfy the following requirements to be considered as signal:

1. All tag tracks must originate from truth-matched τ^+ or τ^-.
2. The 1−prong track should be matched to an electron (muon) and the 3−prongs tracks should be matched with pions
3. The decay mode of one τ should be $\tau \to \ell \nu_\ell \nu_\tau$ ($\ell = e, \mu$) and the decay mode of the other τ should be $\tau \to 3\pi^\pm \nu_\tau$ according to [121]

An event that doesn't meet these requirements is tagged as background. The last two points were added compared to the previous measurement with proc12+prompt2021 and MC14 for which only the 3×1 topology of the τ decay was required.

The addition of new selection cuts and the tightening of the signal criterion are the consequence of the study of the MC truth of the reconstructed probe tracks. With the selection developed in the previous analysis, it was found that the amount of probe tracks matched to a MC particle was at the level of 30%. Among those tracks, about 40% consisted in electrons/positrons originating from the conversion of a photon from a neutral pion. Another 40-50% were pions coming from a 1−prong decay $\tau \to \pi\pi^0\nu_\tau$ where the charged pion was identified as the probe track, and the tag electron in the event was coming from a photon conversion from the neutral pion.

In order to veto these events, the requirement of the absence of photon in the event was implemented. The matching of the exact τ decay modes was also added as a matter of safety to prevent from having any other reconstructed 3×1 events not corresponding to the signal falling in the signal category. With this new selection, the amount of probe particles matched to a MC track fell to approximately 12%. This value includes the clone tracks, which are expected to be matched.

3.2.4. Data-MonteCarlo comparison

3.2.4.1. Electron mode

The data/MC agreement is studied in different variables such as the 3−prong tracks invariant mass ($M_{\pi\pi\pi}$), and the 1−prong and probe tracks azimuthal and radial angles and transverse momentum in the laboratory frame (ϕ^{lab}, θ^{lab} and p_T respectively). The distributions and the Data-MC ratios for the 3−prong tracks invariant mass ($M_{\pi\pi\pi}$) and the ϕ^{lab}, θ^{lab} and p_T variables of the 1−prong and probe tracks after the application of the selections described in Section 3.2.3.2 are displayed in Figures 3.13, 3.14, 3.15 and 3.16, respectively. The overall (signal and background combined) yields in MC and data are shown in Table 3.5.

Table 3.6 shows the MC signal purity $S/(S+B)$ where S ans B are the number of signal and backgrounds events respectively. The estimated signal yields in data, computed as the product of the total data yields and the MC signal purity, are also presented.

Figure 3.17 shows the distributions of the minimum angle between the probe track and the closest pion from the 3−prong tracks (top) and the angle between the probe track and the 1−prong track (bottom). The events located at very low angle can be interpreted as clones of the corresponding track.

The run-independent and run-dependent MC simulations show similar data/MC agreement for the 4- and 5-tracks samples combined (largely dominated by the 4-tracks sample). However, in the plots displaying the 5-tracks sample only, a small discrepancy is visible between the two categories of MC simulations, especially on Figure 3.14. Indeed, the data/MC agreement is observed to be better with the MCri for values of $\phi^{\text{lab}}_{\text{probe}}$ close to zero, while better with MCrd otherwise. The peaking behavior around zero in data is due to the machine background which is higher on the horizontal plane. However, the reason why the simulations are not correctly reproducing this effect, and why they show behaviors going in opposite directions (MCri seems to accentuate the effect while MCrd seems to flatten it) is not fully understood yet and would require more investigation.

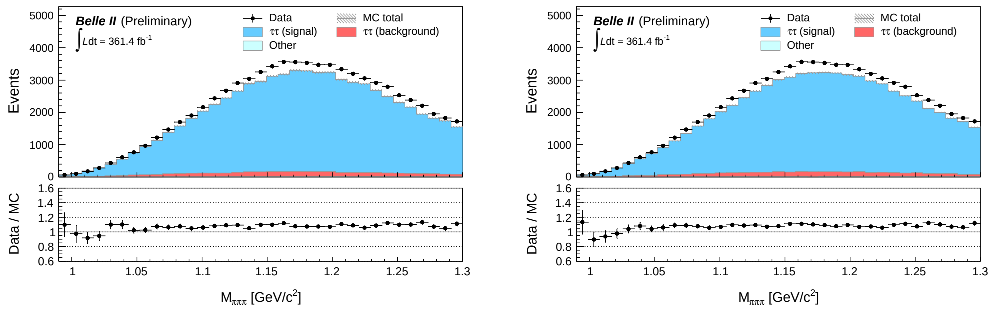


Figure 3.13: Distributions of $M_{\pi\pi\pi}$ for the combined 4- and 5-track samples in the electron mode, using MCri (left) or MCrd (right) simulation.

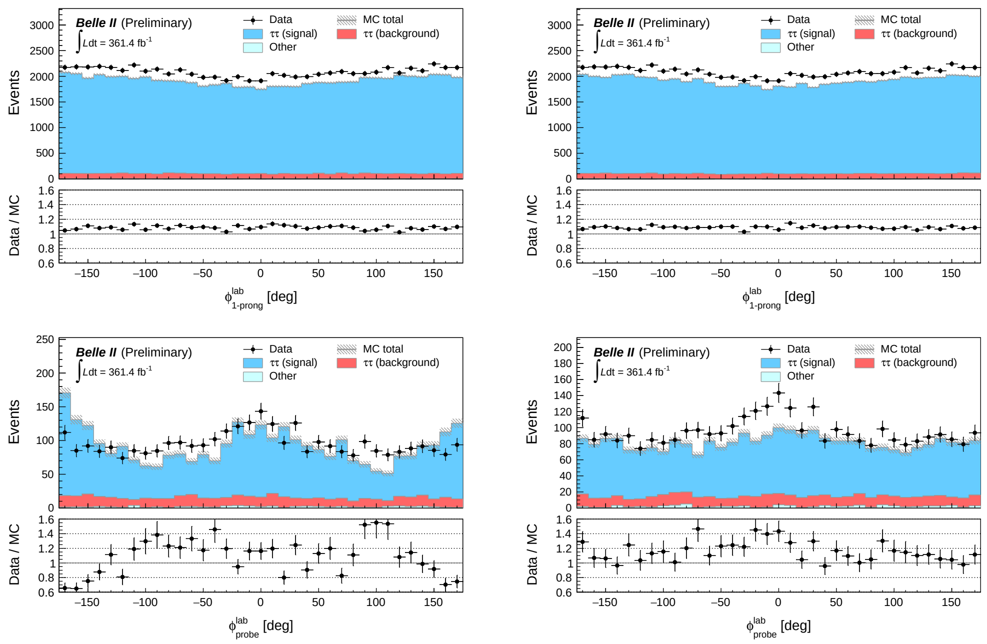


Figure 3.14: Distributions of the 1−prong track (top) and probe track (bottom) ϕ^{lab}, for the combined 4- and 5-track samples (top) or 5-track sample only (bottom) in the electron mode, using MCri (left) or MCrd (right) simulation.

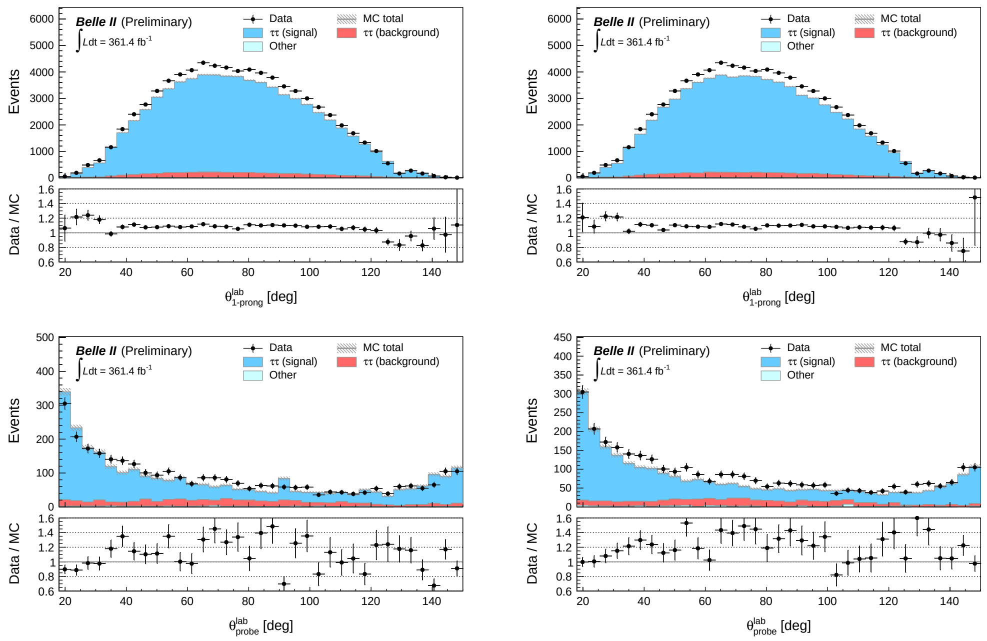


Figure 3.15: Distributions of the 1−prong track (top) and probe track (bottom) θ^{lab}, for the combined 4- and 5-track samples (top) or 5-track sample only (bottom) in the electron mode, using MCri (left) or MCrd (right) simulation.

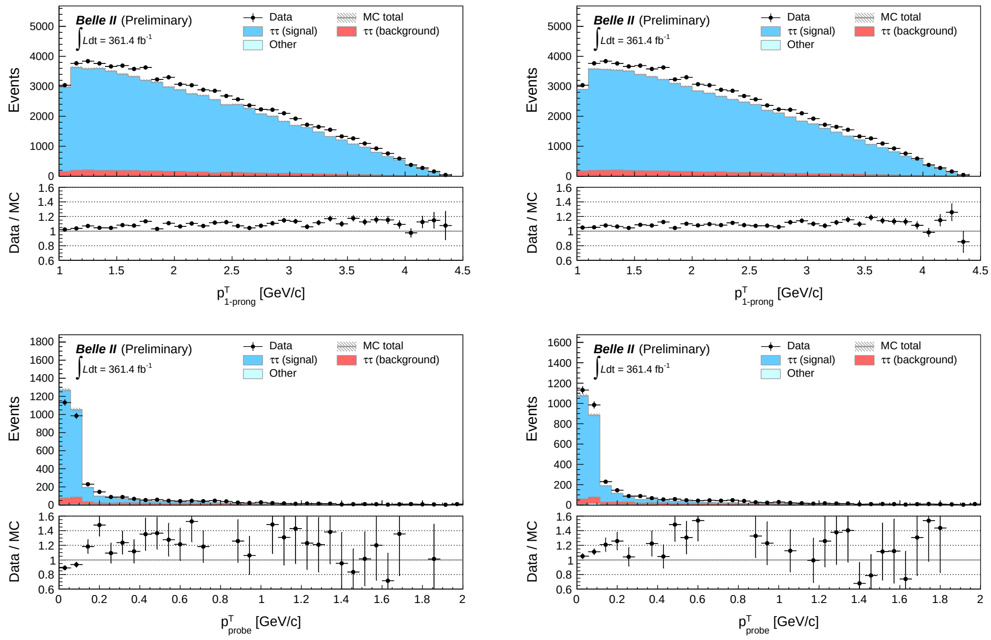


Figure 3.16: Distributions of the 1−prong track (top) and probe track (bottom) p_T, for the combined 4- and 5-track samples (top) or 5-track sample only (bottom) in the electron mode, using MCri (left) or MCrd (right) simulation.

MC run dependence	MC Signal yields		MC Background yields		Data yields	
	N_5	N_4	N_5	N_4	N_5	N_4
MCri	2786 ± 32	62768 ± 151	565 ± 14	3203 ± 34	3441 ± 59	71647 ± 268
MCrd	2436 ± 25	62970 ± 125	528 ± 13	3211 ± 29		

Table 3.5: Overall yields in the 5- and 4-track samples for the electron mode, with statistical uncertainties. The MC yields are rescaled to the data luminosity of 361.4 fb^{-1}.

MC run dependence	MC Signal purity		Data Signal yields (estimation)	
	N_5	N_4	N_5	N_4
MCri	0.831 ± 0.004	0.951 ± 0.001	2861 ± 51	68168 ± 257
MCrd	0.822 ± 0.004	0.951 ± 0.001	2828 ± 50	68171 ± 257

Table 3.6: Estimated signal yields in the 5- and 4-track data samples for the electron mode. The estimated data yields are obtained from the overall data yields from Table 3.5 scaled to the MC signal purities. The data yields correspond to an integrated luminosity of 361.4 fb^{-1}.

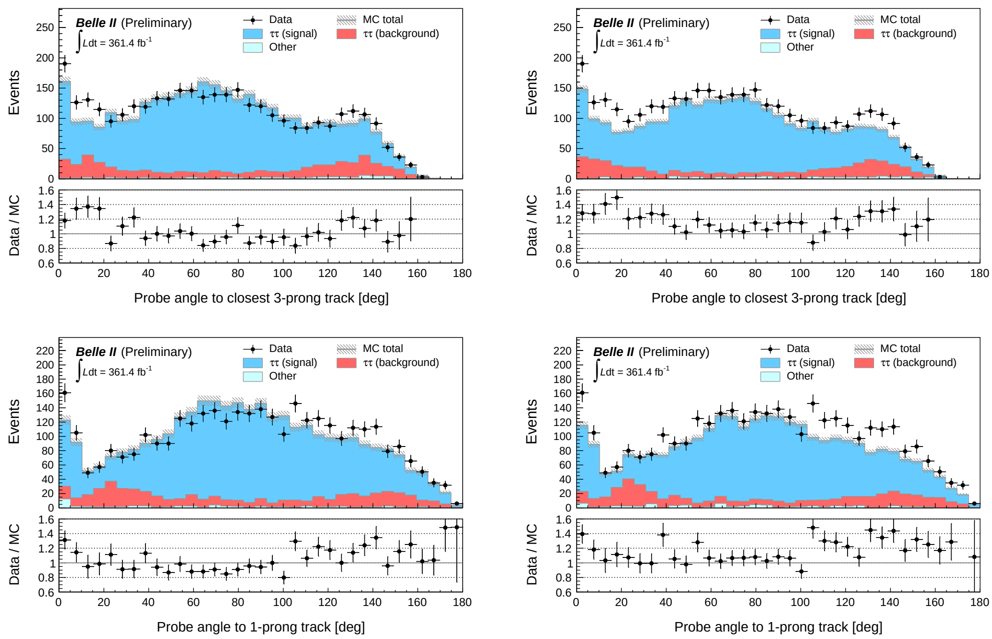


Figure 3.17: Distributions of the (minimum) angle between the probe track and the closest 3−prongs track (top) and between the probe track and the 1−prong track (bottom), for 5-track sample only in the electron mode, using MCri (left) or MCrd (right) simulation.

3.2.4.2. Muon mode

As for the electron mode, the distributions and Data-MC ratios of the same quantities are displayed in Figures 3.18, 3.19, 3.20 and 3.21, respectively. The overall (signal and background combined) yields in MC and data are shown in Table 3.7 and the signal purities in MC, are presented in Table 3.8, together with the estimated signal yields in data.

Figure 3.22 shows the distributions of the minimum angle between the probe track and the closest pion from the 3−prong tracks (top) and the angle between the probe track and the 1−prongs track (bottom).

As expected, the proportion of background is much larger compared to the electron mode since the selection was not optimised for muons. However, the data/MC agreement is very good, especially when using MCrd. In particular, when using MCrd, the mismodelling effect observed on $\phi^{\text{lab}}_{\text{probe}}$ in the electron mode disappears and this quantity is very well modelled.

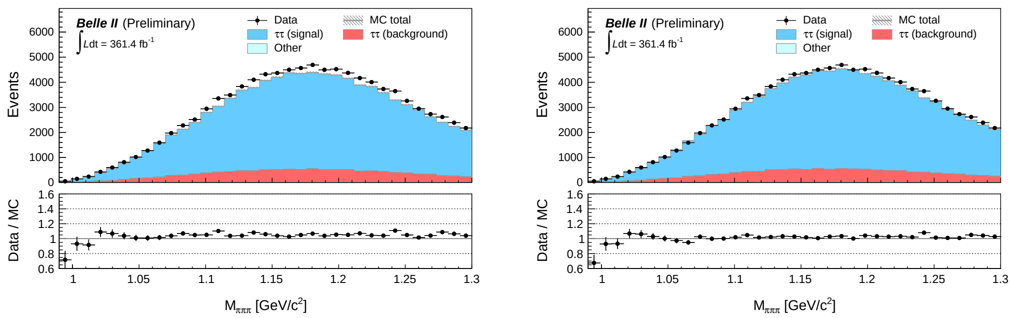


Figure 3.18: Distributions of $M_{\pi\pi\pi}$ for the combined 4- and 5-track samples in the muon mode, using MCri (left) or MCrd (right) simulation. The lower panel shows the Data/MC ratio and the error bars take into account contributions from both data and MC.

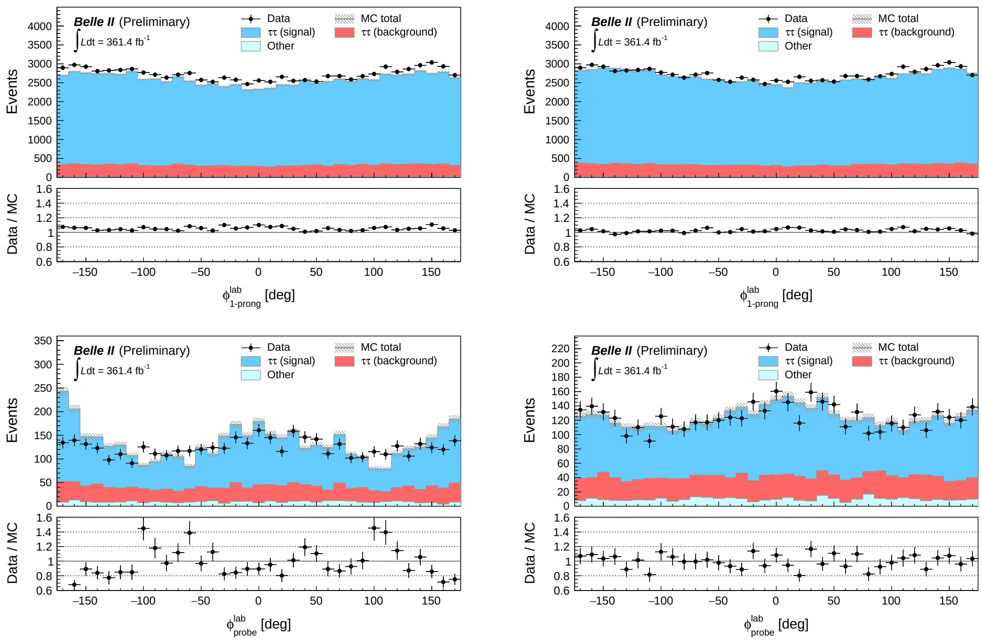


Figure 3.19: Distributions of the 1−prong track (top) and probe track (bottom) ϕ^{lab}, for the combined 4- and 5-track samples (top) or 5-track sample only (bottom) in the muon mode, using MCri (left) or MCrd (right) simulation.

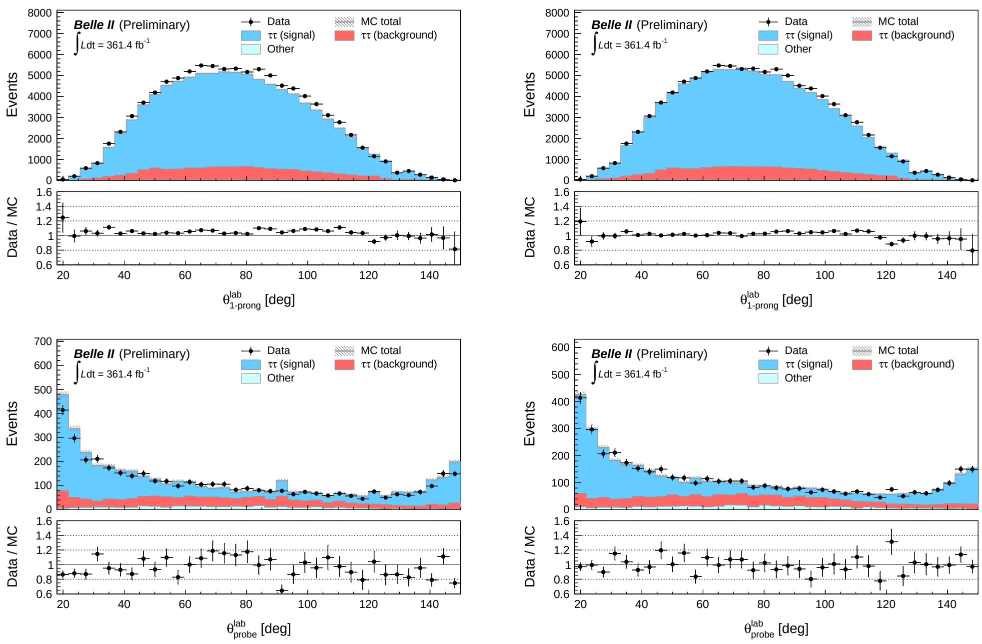


Figure 3.20: Distributions of the 1−prong track (top) and probe track (bottom) θ^{lab}, for the combined 4- and 5-track samples (top) or 5-track sample only (bottom) in the muon mode, using MCri (left) or MCrd (right) simulation.

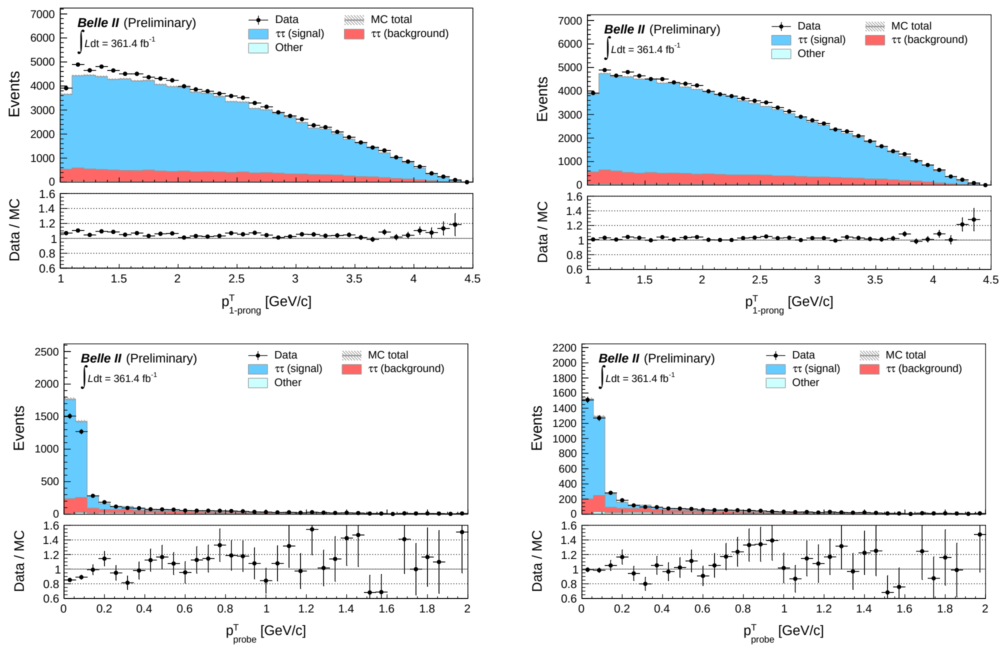


Figure 3.21: Distributions of the 1−prong track (top) and probe track (bottom) p_T, for the combined 4- and 5-track samples (top) or 5-track sample only (bottom) in the muon mode, using MCri (left) or MCrd (right) simulation.

MC run dependence	MC Signal yields		MC Background yields		Data yields	
	N_5	N_4	N_5	N_4	N_5	N_4
MCri	3368 ± 35	78135 ± 168	1487 ± 23	10338 ± 61	4484 ± 67	93679 ± 306
MCrd	2985 ± 27	80707 ± 142	1481 ± 20	10760 ± 53		

Table 3.7: Overall yields in the 5- and 4-track samples for the muon mode, with statistical uncertainties. The MC yields are rescaled to the data luminosity of 361.4 fb^{-1}.

MC run dependence	MC Signal purity		Data Signal yields (estimation)	
	N_5	N_4	N_5	N_4
MCri	0.694 ± 0.004	0.883 ± 0.001	3111 ± 50	82733 ± 277
MCrd	0.668 ± 0.004	0.882 ± 0.001	2997 ± 48	82659 ± 275

Table 3.8: Estimated signal yields in the 5- and 4-track data samples for the muon mode. The estimated data yields are obtained from the overall data yields from Table 3.7 scaled to the MC signal purities.

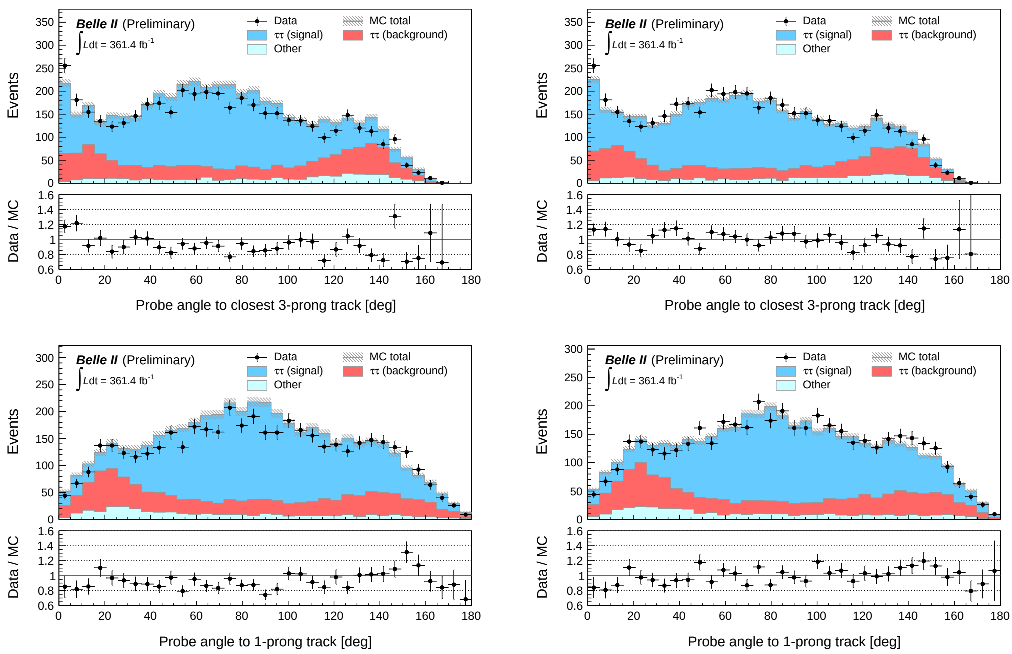


Figure 3.22: Distributions of the (minimum) angle between the probe track and the closest 3−prong track (top) and between the probe track and the 1−prong track (bottom), for 5-track sample only in the muon mode, using MCri (left) or MCrd (right) simulation.

3.2.5. Results

3.2.5.1. Electron mode

The results of the fake tracks rate measurement in MC and data are presented in Table 3.9, including only statistical uncertainties.

	$r_{\text{fake}}^{\text{MC}}$		$r_{\text{fake}}^{\text{Data}}$	
	MCri	MCrd	MCri	MCrd
Overall	4.83 ± 0.05 (stat) %	4.29 ± 0.04 (stat) %	4.58 ± 0.08 (stat) %	
Signal	4.25 ± 0.05 (stat) %	3.72 ± 0.04 (stat) %	4.03 ± 0.07 (stat) %	3.98 ± 0.07 (stat) %

Table 3.9: Fake tracks rates in MCri/MCrd ($r_{\text{fake}}^{\text{MC}}$) and data ($r_{\text{fake}}^{\text{Data}}$) for the electron mode, computed according to Eq. (3.1). The "overall" and "signal" labels designate respectively the total dataset and signal dataset (background substracted dataset for data). For data, the signal estimation is done by considering either MCri or MCrd signal purities, therefore two results are given.

Following the previous study, we can estimate the systematic uncertainty on $r_{\text{fake}}^{\text{Data}}$ as the maximum absolute difference between values evaluated using yields with or without background substraction in data, which is in this case 0.6%. Nonetheless, this uncertainty assignment could be improved by developing a proper strategy as this estimation is very dependent on the quality of the selection and the optimisation of the purity.

The fake tracks rate in signal data is thus estimated to be around 4%, which is compatible with the measurements in MC simuation given the relatively large systematic uncertainty. This matches the expectations one could have according to the values found in the previous study given the fact that the phase space considered for the probe track in the present study is larger. Indeed, the fake tracks rate was estimated to $\sim 1.3\%$ with proc12+prompt2021 data and MC14 simulations, and the probe track was required to be in the 3−prongs side, covering a region of space approximately equivalent to 1/3 of the full phase space.

As a matter of comparison, for the present study, the fake tracks rate in signal data requiring a probe in the 3−prongs space region is estimated to be $1.22{\pm}0.04(\text{stat}){\pm}0.3(\text{syst})\%$ using run-independent MC and $1.19{\pm}0.04(\text{stat}){\pm}0.3(\text{syst})\%$ using run-dependent MC.

Finally, Figure 3.23 displays the fake rate measured on signal run-dependent MC and in signal data according to the data-taking period. The higher instantaneous luminosity in the last experiments leads to a larger beam-induced background and thus, a higher fake tracks rate.

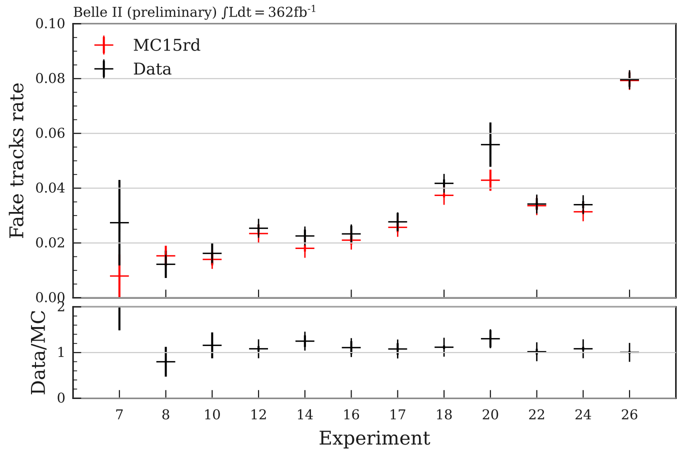


Figure 3.23: Evolution of the signal fake tracks rate in MCrd and data as a function of the data taking period for the electron mode.

3.2.5.2. Muon mode

For the muon mode, the results of the fake tracks rate measurement in MC and data are presented in Table 3.10, including only statistical uncertainties.

	$r_{\text{fake}}^{\text{MC}}$		$r_{\text{fake}}^{\text{Data}}$	
	MCri	MCrd	MCri	MCrd
Overall	5.20 ± 0.04 (stat) %	4.66 ± 0.04 (stat) %	4.57 ± 0.07 (stat) %	
Signal	4.13 ± 0.04 (stat) %	3.57 ± 0.03 (stat) %	3.62 ± 0.06 (stat) %	3.50 ± 0.06 (stat) %

Table 3.10: Fake tracks rates in MCri/MCrd ($r_{\text{fake}}^{\text{MC}}$) and data ($r_{\text{fake}}^{\text{Data}}$) for the muon mode, computed according to Eq. (3.1). The "overall" and "signal" labels designate respectively the total dataset and signal dataset (background substracted dataset for data). For data, the signal estimation is done by considering either MCri or MCrd signal purities, therefore two results are given.

Following the previous study, the systematic uncertainty on $r_{\text{fake}}^{\text{Data}}$ is estimated as the maximum absolute difference between values evaluated using yields with or without background substraction in data, that is to say 1.07%. As discussed for the electron mode, the estimation method is fully dependent on the purity, which is not optimised in this case. The resulting difference between data yields with or without background substraction is thus large, leading to a very large systematic uncertainty. The order of magnitude of the estimated fake tracks rate in the muon mode is nevertheless compatible to the fake tracks rate measured in the electron mode.

Figure 3.24 displays the fake rate measured on signal run-dependent MC and in signal data according to the experiment number for the muon mode. The higher instantaneous luminosity in the last experiments leads to a larger beam-induced background and thus, a higher fake tracks rate.

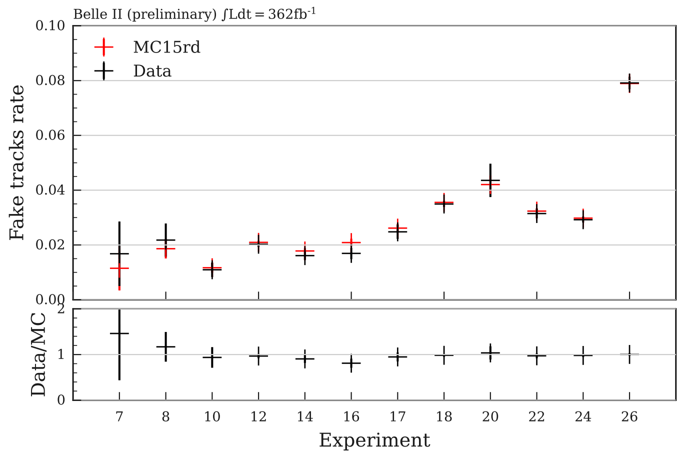


Figure 3.24: Evolution of the signal fake tracks rate in MCrd and data as a function of the data taking period for the muon mode.

3.2.6. The fake tracks rate as key parameter for performance evaluation

In addition to a study over time of the fake tracks rate, which accounts for the tracking performance depending on the background level in data, as long as there is no software change in the tracking chain, one can use this parameter as a performance indicator when trying to improve the tracking reconstruction.

In this context, a study of the fake tracks rate has been performed given different configurations of the SVD timing. As described in Section 2.3 and in the previous chapter 3.1, the tracking reconstruction chain relies highly on the SVD information. The SVD space-points and tracks are thus key ingredients for the tracking procedure and the quality of their reconstruction can have a large impact on the final tracking performance.

With the increase of the instantaneous luminosity, a larger background is expected, challenging the current tracking reconstruction. However, improvements in this area are possible by exploiting the timing information from the SVD clusters. In particular, we can focus on the following options:

- The SVDTimeSelection: Timing condition applied on the SVD clusters to remove background characterised by "off-time" clusters.
- The VXDTF2TimeFilters: Time filters that are applied on the SVD stand-alone part of the tracking chain.
- The SVD Grouping: By looking at the cluster time distribution for each event, groups of clusters belonging to the tracks coming from the same collisions are built and space-points are created with only clusters belonging to the same group.

- The SVD trackTime: Quantity built averaging the cluster times of all SVD clusters belonging to a track and saved to be used at the analysis level. It corresponds to the time at which the track is produced relative to the time of the collision given by the SVD.

During Run 1, none of the SVDTimeSelection, VXDTF2TimeFilters and SVD grouping options were used and the SVD trackTime variable was not available.

To decide on the configuration to use for future data processing, several studies are performed, including the measurement of the fake tracks rate. To do so, a subset of the available data corresponding to the bucket 36 of experiment 26 is reprocessed in different SVD timing configurations. This specific subset is used as it is a sample presenting a high background level. A MCrd sample corresponding to this data subset is also reprocessed in the reference configuration, ie. without any SVD timing improvement option activated. As the MC information is used only to estimate the purity in data, which is not expected to change in the different configurations, the same MCrd sample is used for all configurations.

Following the fake tracks measurement recipe that was described earlier, the fake tracks rate is first measured in the following configurations for the electron mode:

- "OFFOFF": SVDTimeSelection OFF, VXDTF2TimeFilter OFF → standard configuration used as reference
- "ONOFF": SVDTimeSelection ON, VXDTF2TimeFilter OFF
- "OFFON": SVDTimeSelection OFF, VXDTF2TimeFilter ON
- "ONON": SVDTimeSelection ON, VXDTF2TimeFilter ON

For all the configurations, the SVD grouping is not activated and we make no use yet of the trackTime variable at reconstruction level.

In parallel to the fake tracks rate measurement, the tracking performance is evaluated in terms of track finding efficiency. Similarly to the fake tracks rate, the track finding efficiency is measured in τ-pair events with a 3×1 topology and using a tag-and-probe method. The efficiency is measured as the ratio of events in which all four tracks are found to the events in which three or four tracks are found. Combining the fake tracks rate results and the track finding efficiency measurements with these configurations, the OFFON configuration is found to be the best balance between reducing the fake tracks rate without degrading the efficiency. From this conclusion, the bucket 36 of experiment 26 data sample and corresponding MCrd are reprocessed in the OFFON configuration, adding the SVD grouping to the reconstruction and the fake tracks rate is evaluated again in the electron mode.

In addition to this new configuration, the effect of a selection on the trackTime quantity at the analysis level can be studied. The tracks belonging to the event are expected to be produced at the same time as the collision while beam background fake tracks can be produced at any time. The distribution of the trackTime for a tag pion track and the probe track in a N5 event are shown in Figure 3.25 for the standard configuration and in 3.26 for the OFFON+Grouping configuration. In the latter configuration, the SVD trackTime distribution of the probe track doesn't show a large amount of off-time tracks, but this is expected as the VXDTF2TimeFilters and the SVD Grouping are already very powerful in rejecting off-time fake tracks, but one can still make use of the trackTime to reject even more fakes.

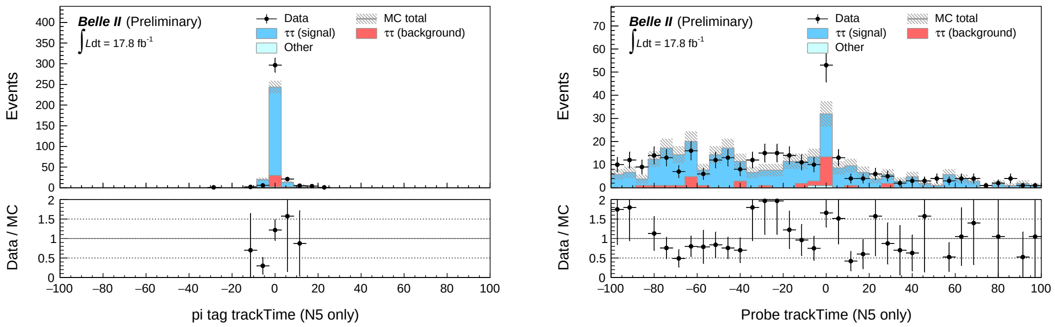


Figure 3.25: Distribution if the SVD trackTime for a tag pion (left) and the probe track (right) in a N5 event for the standard SVD timing configuration.

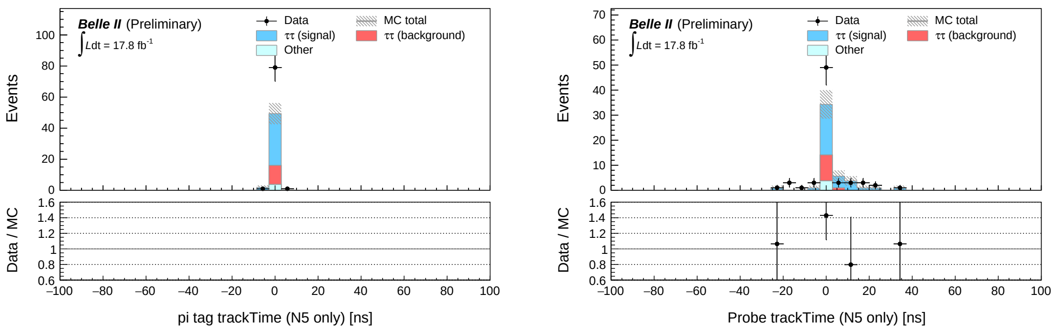


Figure 3.26: Distribution if the SVD trackTime for a tag pion (left) and the probe track (right) in a N5 event for the OFFON+Grouping SVD timing configuration.

The results for the fake tracks rate in data in all configurations are presented in Table 3.11. A measurement of the fake tracks rate with a cut on the trackTime of every track at ±20 ns is also presented.

$r_{\text{fake}}^{\text{Data}}$		Default (no trackTime)	with SVD trackTime cut
Standard	Overall	9.4±0.5 %	3.3±0.3 %
	Signal	8.8±0.5 %	2.8±0.3 %
ONOFF	Overall	2.9±0.3 %	2.1±0.2 %
	Signal	2.7±0.3 %	1.8±0.2 %
OFFON	Overall	4.0±0.3 %	1.9±0.2 %
	Signal	3.7±0.3 %	1.6±0.2 %
ONON	Overall	2.6±0.3 %	1.9±0.2 %
	Signal	2.4±0.3 %	1.6±0.2 %
OFFON+Grouping	Overall	2.2±0.2 %	1.8±0.2 %
	Signal	1.6±0.2 %	1.3±0.2 %

Table 3.11: Fake tracks rate in data for the electron mode in the different SVD timing configurations with and without a selection on the trackTime at analysis level. When applied, the cut on the trackTime is ±20 ns.

This study shows that using the SVD timing information is extremely efficient in reducing the fake tracks rate in data as one can reach a reduction by factor up to 6, without degrading the track finding efficiency or any other performance. Coupled to other key parameters measurements, in particular the efficiency measurement, the OFFON configuration without SVD grouping was chosen to be implemented in the new version of the tracking chain that will be used for future data processing. In addition, although no selection on the SVD trackTime is recommended at the moment, this information is available for analysts. Further studies are needed in this direction to optimise a selection, but this study showed that it is a powerful tool in reducing the fake tracks rate.

Nonetheless, despite SVD timing being very efficient in reducing the fake rate, clone tracks cannot be targeted with such techniques as, being clones of real tracks, they are by definition on time. The presence of these tracks among the fake tracks also explains why the trackTime distribution of the probe does not have a flat distribution and is still peaking at 0. Additional studies could be performed on clones to develop techniques to reject them.

4. Search for $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ lepton flavour violating decays

Summary

As discussed in Chapter 1, charged lepton flavour violation is forbidden in the Standard Model, and any observation of such process would be a clear evidence for new physics. In this chapter, we will present the analysis performed to search for the lepton flavour violating decays $B^0 \to K^{*0}\tau\ell$, where $\ell = e, \mu$. The analysis uses data collected from Belle and Belle II experiments in order to fully exploit the available statistics and is based on a hadronic B-tagging technique. After presenting the global strategy of the analysis and the datasets used, we will detail the event reconstruction and selection, the background suppression and the signal extraction procedures. Finally, we will present the results of the analysis and the upper limits on the branching fractions of $B^0 \to K^{*0}\tau\ell$ decays.

4.1. Analysis strategy

The goal of this analysis is to study the four decay modes $B^0 \to K^{*0}\tau^\pm\ell^\mp$ with $\ell = e, \mu$ and $K^{*0} \to K^+\pi^-$. For the rest of this chapter, we will refer to these decays as $OS\ell$ and $SS\ell$ modes, where OS and SS stand for opposite and same sign of the electric charge between the kaon from the K^{*0} and the prompt lepton $\ell = e, \mu$. A summary of the modes naming is presented in Table 4.1. The charge conjugation is implied in the notation, i.e. $B^0 \to K^{*0}\tau^\pm\ell^\mp$ is equivalent to $\bar{B}^0 \to \bar{K}^{*0}\tau^\mp\ell^\pm$, and so through this whole chapter.

Label	Decay mode
OSe	$B^0 \to K^{*0}\tau^+e^-$
SSe	$B^0 \to K^{*0}\tau^-e^+$
$OS\mu$	$B^0 \to K^{*0}\tau^+\mu^-$
$SS\mu$	$B^0 \to K^{*0}\tau^-\mu^+$

Table 4.1: Summary of naming convention for different decay channels.

In this analysis, we take advantage of the Belle and Belle II detectors' geometry to reconstruct the entirety of the event, i.e. both the B meson in which the signal is searched, and its partner also produced in the $\Upsilon(4S)$ decay, called the tagged B. Using the kinematic information from the tagged B meson, and the $K^{*0}\ell$ system of the signal side, we can compute the recoil mass of the τ system. If some signal exists, it will appear as a peak in the recoil mass distribution, centred at the known τ lepton mass $m_\tau = 1.777$ GeV/c^2 [28].

The tagged B meson is reconstructed from hadronic decays with the Full Event Interpretation (FEI) [31] Belle II tool, described in Section 2.4.2. Despite its low efficiency, using hadronic tagging allows a very good resolution of the recoil mass as no energy is missing in the tag side. From the signal side, the $K^{*0}\ell$ ($K^{*0} \to K^+\pi^-$) system is reconstructed from three tracks identified as a kaon, a pion and a lepton with the correct charges. Since the τ lepton is electrically charged and unstable, it must decay into at least one charged particle. Its presence is therefore inferred from the requirement of an additional track in the event, with a charge opposite to that of the prompt lepton. By reconstructing this track from the τ, we are selecting the decay of τ into a final state with a single charged track (called 1 – prong as in Chapter 3.2), representing approximately 85% of τ decays [28]. Furthermore, we can use the information of this track and of the tracks and neutral clusters that remain in the event, constituting the Rest Of Event (ROE), to reject background. By applying a loose selection on the number of tracks in the ROE, events corresponding to τ decays with final states with three charged particles ($\tau \to 3$ – prongs decays) can also be kept in the selection. Nevertheless, the kinematics of this track are not used in the calculation of the τ recoil mass. A schematic view of the event is shown in Figure 4.1.

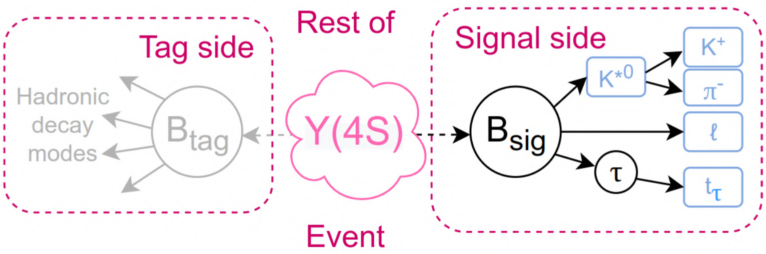


Figure 4.1: Schematic view of the event reconstruction.

The signal is then extracted from a fit to the τ recoil mass given by Eq. (4.1):

$$M_\tau = \left[m_B^2 + M(K^{*0}\ell)^2 - 2(E_{\text{beam}} E_{K^{*0}\ell} + |\vec{p}_{B_{\text{tag}}}||\vec{p}_{K^{*0}\ell}| \cos\theta) \right]^{1/2}, \tag{4.1}$$

where $m_B = 5.27926$ GeV/c^2 [28] and $M(K^{*0}\ell)$ are the known B meson mass and the invariant mass of the $K^{*0}\ell$ system, $E_{\text{beam}} = E_{\text{CMS}}/2$ is half the centre of mass energy, $E_{K^{*0}\ell}$, $\vec{p}_{B_{\text{tag}}}$ and $\vec{p}_{K^{*0}\ell}$ are respectively the energy and momentum vectors of the tag B and the $K^{*0}\ell$ system, and θ is the angle between $\vec{p}_{B_{\text{tag}}}$ and $\vec{p}_{K^{*0}\ell}$. All quantities are defined in the centre-of-mass frame.

The branching fraction of the $B^0 \to K^{*0}\tau\ell$ decays can therefore be measured as:

$$\mathcal{B}(B^0 \to K^{*0}\tau^\pm\ell^\mp) = \frac{N_{\text{sig}}}{\varepsilon_{\text{sig}} \times 2 \times N_{\Upsilon(4S)} \times f_{00} \times \mathcal{B}(K^{*0} \to K^+\pi^-)}, \tag{4.2}$$

where N_{sig} is the signal yield, ε_{sig} is the signal efficiency, $N_{\Upsilon(4S)} = (771.6 \pm 10.6) \times 10^6$ for Belle [122] and $(387.1 \pm 5.6) \times 10^6$ for Belle II [123], the number of produced $\Upsilon(4S)$, $f_{00} = 0.4861 \pm 0.008$ the fraction of $\Upsilon(4S)$ decaying into $B^0\bar{B}^0$ pairs [15], the factor 2 accounts for the fact that the signal is searched in both the B^0 and $\bar{B}^0$ decays, and $\mathcal{B}(K^{*0} \to K^+\pi^-) = 2/3$ the branching fraction of $K^{*0} \to K^+\pi^-$.

Finally, if no signal is observed, the upper limits on the branching fractions $\mathcal{B}(B^0 \to K^{*0}\tau\ell)$ are estimated with a CLs asymptotic approach [33, 34]. During the development of the analysis and until the optimisation and strategy is finalised, the signal region is hidden in order to not introduce any bias.

4.2. Datasets

The analysis is performed using data collected by the Belle and Belle II experiments in order to exploit all the available statistics. To study the signal and background components, Monte Carlo (MC) simulations are used. The event generation is performed with EvtGen [112] and Pythia [113] and the detector is simulated with GEANT3 [115] for Belle and GEANT4 [114] for Belle II.

Signal Monte Carlo samples

For the signal study, a MC sample of 20 million $B^0\overline{B}^0$ events with one B meson decaying generically and one to $K^{*0}(\to K^+\pi^-)\tau\ell$ is generated for each OSe, $OS\mu$, SSe, $SS\mu$ decay and for both Belle and Belle II. The signal is generated following a phase-space distribution, and the τ lepton decay can decay to any SM mode. Charge conjugation is implied and $B^0\overline{B}^0$ mixing is taken into account.

When comparing signal to background for the development of the selection process, we often *truthmatch* the signal for the $K^{*0}\ell$ system, meaning that we exploit the knowledge of the generated MC sample to ask the reconstructed identity of the particles to match their generated identity. In particular, we require the K^{*0} candidate to be a real K^{*0} reconstructed from true kaon and pion tracks, and the lepton to be a true electron or muon, depending on the mode. The K^{*0} and ℓ candidates are also required to originate from the same vertex, corresponding to the decay of the B meson identified as signal B.

Background Monte Carlo samples

For the background study, generic (in particular, samples for $u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$ pairs as well as B^+B^- and $B^0\overline{B}^0$ pairs) MC simulations are used. In the following, the $u\overline{u}$, $d\overline{d}$, $s\overline{s}$, and $c\overline{c}$ backgrounds are often referred to as $q\overline{q}$, and the B^+B^- and $B^0\overline{B}^0$ backgrounds are referred to as $B\overline{B}$ background when considered together or sometimes as *charged* and *mixed* respectively when treated separately.

For both Belle and Belle II, the total amount of available run-dependent MC (see Section 2.4.3.2) is used, corresponding to four times the data luminosity for each experiment. For Belle II, MC15 run-dependent samples produced with the basf2 release 06 are used. These samples are skimmed with the hadronic FEI selection described in Section 2.4.3.3. For Belle, the most up-to-date MC samples are considered. Since no FEI skims are produced for Belle, the FEI is applied at reconstruction level.

Data samples

The full Belle dataset collected at the $\Upsilon(4S)$ resonance is used together with the full data sample of Belle II collected before the LS1 at the $\Upsilon(4S)$ resonance too. The total integrated luminosity of the Belle data set is $L = 711\ \text{fb}^{-1}$, which corresponds to $(771.58 \pm 10.57) \times 10^6$ produced $\Upsilon(4S)$ [122]. The Belle II sample is $L = 365.29 \pm 1.70$

fb^{-1} corresponding to $(387.1 \pm 5.6) \times 10^6$ produced $\Upsilon(4S)$ [123]. The Belle II data corresponding to the proc13+prompt2022 dataset produced with release 06 is used. Similarly to generic MC, hadronic FEI skimmed samples are used for Belle II.

Although MC simulations are generated such that they reproduce data in the most accurate way, some effects can be mismodelled. For this reason, corrections to simulation are developed for various aspects of analysis tools by the corresponding performance groups and are provided as weights to the MC events. This is, in particular, the case for the particle identification 2.3.3 and the FEI algorithm 2.4.2. In this chapter, every efficiency and MC yield values quoted or shown in histograms are corrected with the appropriate weights and should therefore be comparable to data yields.

4.3. Reconstruction and event selection

4.3.1. Tag side reconstruction

To reconstruct the tagged B meson, labelled B_{tag}, we use the FEI B^0 hadronic tagging algorithm, described in Section 2.4.2. Once the FEI is applied to the events, a set of selections is made on the resulting sample.

In particular, one can take advantage of the distributions of the B_{tag} beam constrained mass M_{bc} ($M_{bc} = \sqrt{E^2_{\text{beam}}/c^4 - p^2_{B_{\text{tag}}}/c^2}$), the difference between the B_{tag} energy and the half of the centre-of-mass energy ΔE ($\Delta E = E_{B_{\text{tag}}} - E_{\text{beam}}$), and the cosine angle between the B_{tag} and the rest of the tracks from the event (i.e. the Rest Of Event against B_{tag}) $\cos\text{TBTO}$ that are very different for B decays and continuum events. A set of specific selections was developed by Belle II FEI experts to maximise the quality of the B_{tag} reconstruction and the purity of the sample. Following these recommendations, the above quantities are required to fulfil $M_{bc} > 5.27$ GeV/c^2, $-0.15 < \Delta E < 0.1$ GeV and $\cos\text{TBTO} < 0.9$. In addition, the signal probability $P_{\text{sig}}(FEI)$ encoding the quality of the B tagging reconstruction should fulfil $P_{\text{sig}}(FEI) > 10^{-3}$ in order to remove the bad B_{tag} candidates. Then, only the best candidate for each event according to this variable is kept.

The distributions of the variables and the applied selection requirements are shown in Figure 4.2 for Belle II and in Figure 4.3 for Belle. For the signal probability $P_{\text{sig}}(FEI)$, the selection criterion is applied directly in the FEI reconstruction, hence the absence of events with $P_{\text{sig}}(FEI) < 10^{-3}$ in the plots. The OSe mode is shown as example here and throughout the analysis, but the same selection is applied to all four modes unless specified otherwise, and the corresponding variables are similar in all four modes.

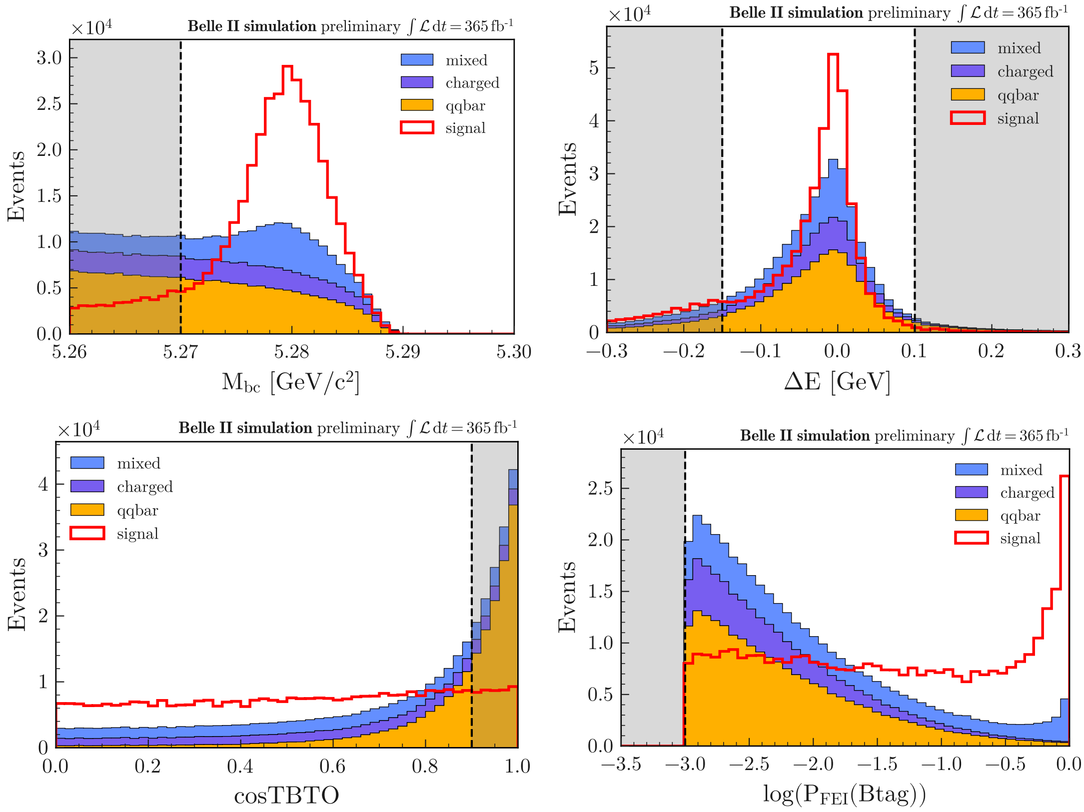


Figure 4.2: Selection cuts applied on the tag side for the *OSe* mode in Belle II MC. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison.

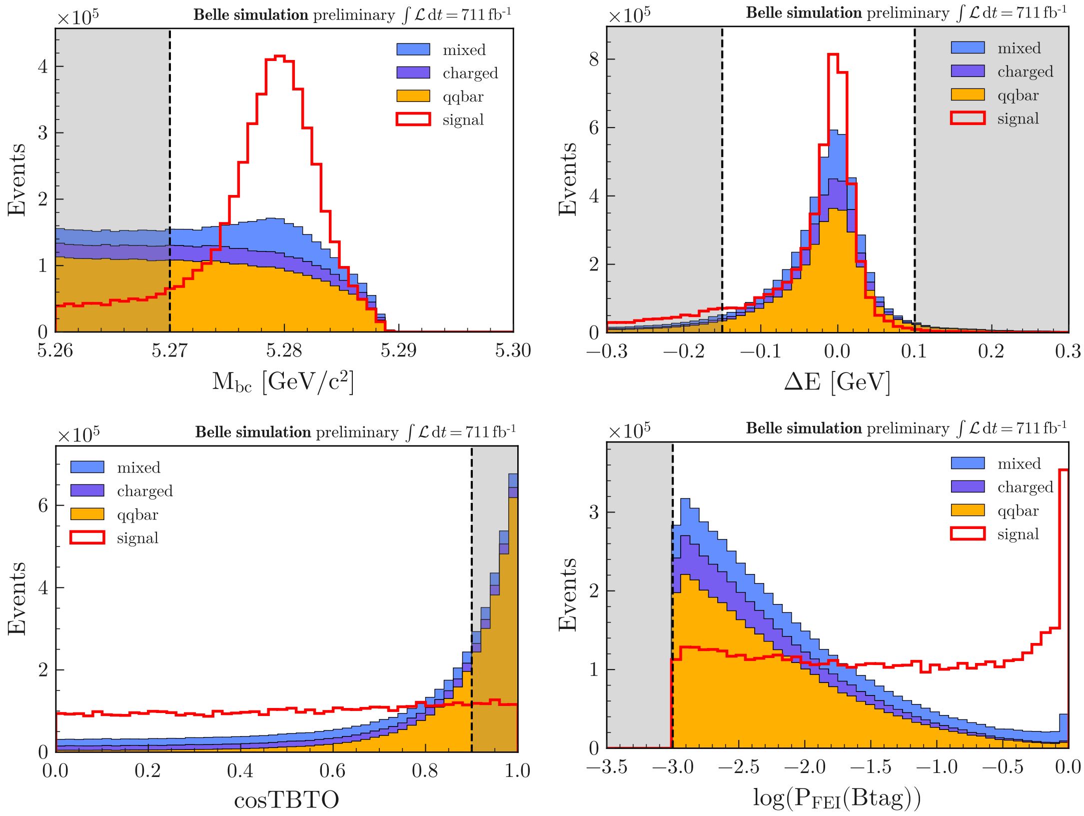


Figure 4.3: Selection cuts applied on the tag side for the *OSe* mode in Belle MC. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison.

4.3.2. Signal side reconstruction

The B_{sig} candidate is reconstructed in two parts.

On one hand, the $K^{*0}\ell$ system, also referred to as B_{sig}, is reconstructed from K, π and $\ell = e, \mu$ candidates with the following requirements:

- Each of the K, π, ℓ tracks should come from the interaction point (radial distance from the IP $dr < 0.5$cm and longitudinal distance from the IP $|dz| < 5$cm), be in the CDC angular acceptance and in case of Belle II, hadron tracks should have $nCDCHits > 20$, which is the default requirement for good hadron tracks.
- The hadron identification likelihoods kaonID and pionID should be larger than 0.6 for the kaon and pion respectively.
- The lepton identification likelihoods electronID and muonID should be larger than 0.9 for the electron or muon respectively according to the mode.
- The invariant mass of the $K\pi$ system should be within 50 MeV/c^2 around the K^{*0} nominal mass $m_{K^{*0}} = 892$ MeV/c^2.

For Belle II, the PID variables for the kaon, pion and muon use a global likelihood-based particle identification, and the electron PID is global and BDT-based. For Belle, the kaon and pion variables use a binary separation between the two particle types based on likelihoods while for the electron and muons, global separation likelihood-based variables are used. In case of electron mode, the prompt electron is corrected for bremsstrahlung radiations, where the recovered bremsstrahlung photons must satisfy $E > 50$MeV. Regarding the K^{*0} mass selection, the selection criterion was optimised with the "Punzi" figure of merit $\frac{\varepsilon_{\text{sig}}}{\alpha/2+\sqrt{N_{\text{bkg}}}}$ [124], where ε_{sig} is the signal efficiency and N_{bkg} is the number of background events, and $\alpha = 3$ is the desired significance in terms of σ where σ is the Gaussian distribution's standard deviation. The optimisation is performed on the four modes for Belle II and Belle and the best cut is found to be in a window between 40 and 50 MeV/c^2 from the known K^{*0} mass depending on the mode and the experiment. In order to apply the same selection on all samples, we choose the loosest cut from all the optimised values, corresponding to 50 MeV/c^2.

In addition to this selection, a vertex fit of the $K^{*0}\ell$ system is performed, and the tracks' parameters are updated with the information resulting from the vertex fit. Events with a failed $K^{*0}\ell$ vertex fit are removed. They represent less than 0.1% of all reconstructed events in signal MC and around 0.1% in generic MC. Their removal has no impact on the signal efficiency. The distribution of the K^{*0} mass with the applied selection cut is visible in Figure 4.4 for Belle II *OSe* mode as example, and the corresponding efficiency is presented in Table 4.2.

On the other hand, the τ lepton is reconstructed from a single track, targeting the $\tau \to 1-$prong decays. The $1-$prong track, hereafter referred to as t_τ, is only required to have $dr < 0.5$cm, $|dz| < 5$cm and be of the correct charge.

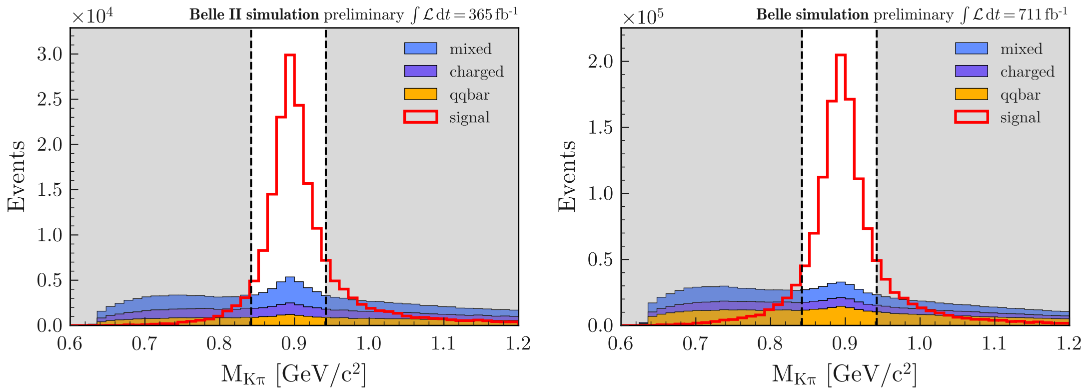


Figure 4.4: Selection cut applied on the $K\pi$ invariant mass for the OSe mode in Belle II (left) and Belle (right) MC. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison. The FEI selection criteria described previously are applied.

4.3.3. Event level selection

After reconstructing the tag and signal sides, we build the Rest Of Event, which consists of every track that is reconstructed but associated neither to the tag side nor the signal side, as well as the neutral clusters found in the ECL. In case of perfectly reconstructed events, the ROE should contain 0 tracks for events with a $\tau \to 1-$prong and 2 tracks for events with a $\tau \to 3-$prongs.

To be considered as part of the ROE tracks and ECL clusters should satisfy certain conditions. To be kept, tracks should have $dr < 0.5$ cm and $|dz| < 5$ cm, and ECL clusters must originate from photon candidates with $E > 50$MeV ($E > 50$MeV) in the barrel, $E > 75$MeV ($E > 100$MeV) in the forward endcap and $E > 100$MeV ($E > 150$MeV) in the backward endcap for Belle II (Belle). The ECL clusters should also have a score > 0.3 on the output of two multivariate classifiers (MVAs) aiming at rejecting beam background and fake photons, respectively.

We choose to keep events with one or two tracks in the ROE as it allows including $\tau \to 3-$prongs events as well, significantly increasing the signal efficiency. Events that are accepted with this looser criterion compared to a requirement of 0 tracks in the ROE are still peaking at the τ mass with a very small degradation of the resolution as can be seen in Figure 4.5, and are therefore safe to keep. However, in order to reduce the background yields as much as possible in the Belle samples, while keeping a rather high efficiency, the total charge of the ROE is required to be 0. This selection is not applied to the Belle II samples as the background yield is already reasonably low and such a cut would affect the signal efficiency too much. The full analysis was derived with this cut also applied to Belle II and the upper limit results found were worse, indicating that the efficiency loss vs background rejection gain balance was not optimal with such cut. We also tested to not apply the cut neither to Belle nor Belle II data and found that the best combination was to apply it on Belle data only.

Finally, a selection based on the global shape of the event can be applied. To separate the $B\overline{B}$ events from the continuum background made of $q\overline{q}$ events, a set of observables usually called "event shape variables" [125–128] was developed to characterise the geometry and structure of the final states of e^+e^- annihilations. One of these variables is the sphericity of the event (defined as a linear combination of the sphericity tensor eigenvalues), which is required to be > 0.2 in this analysis. Other event shape variables will be used in later stages of the analysis, as input variables to multivariate classifiers.

The distributions of the sphericity and number of charged tracks in the Rest Of Event are displayed in Figures 4.6 for Belle II and 4.7 for Belle, together with the total charge of the ROE for the latter.

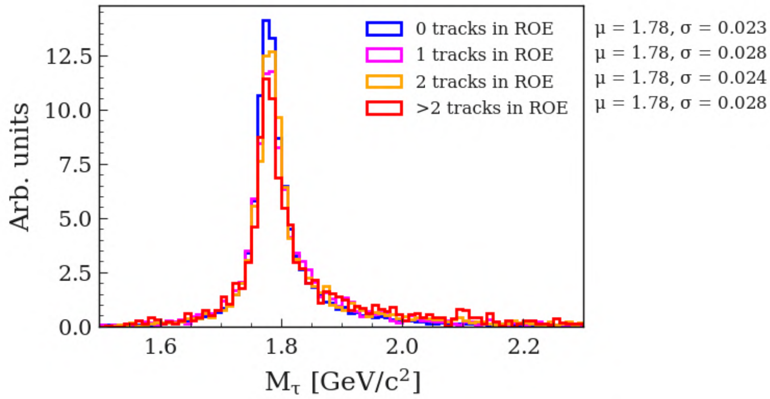


Figure 4.5: Impact of the number of accepted tracks in the ROE on the τ recoil mass for the *OSe* mode in Belle II signal MC. The signal MC is truthmatched for the $K^{*0}\ell$ system and the histograms are normalised to each others for shape comparison.

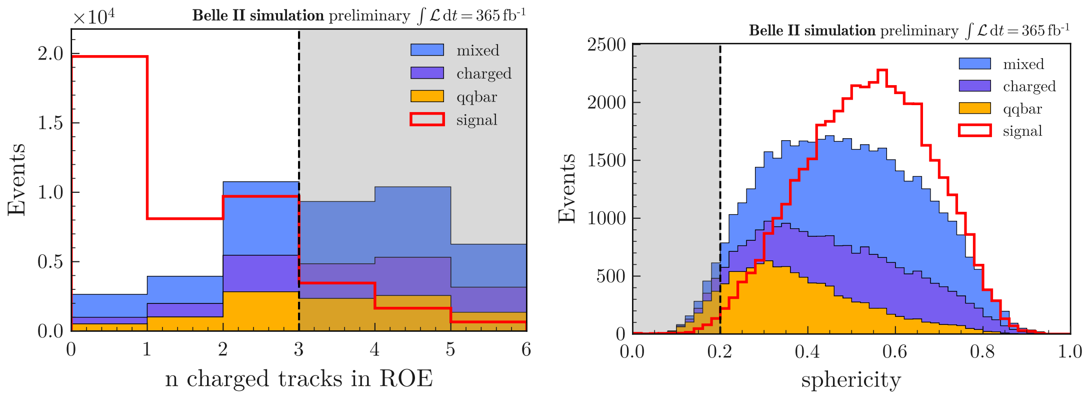


Figure 4.6: Selection cuts applied at the event level on the number of charged tracks in the ROE (left) and on the sphericity (right) for the *OSe* mode in Belle II MC. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison. The FEI and K^{*0} mass selection criteria described previously are applied.

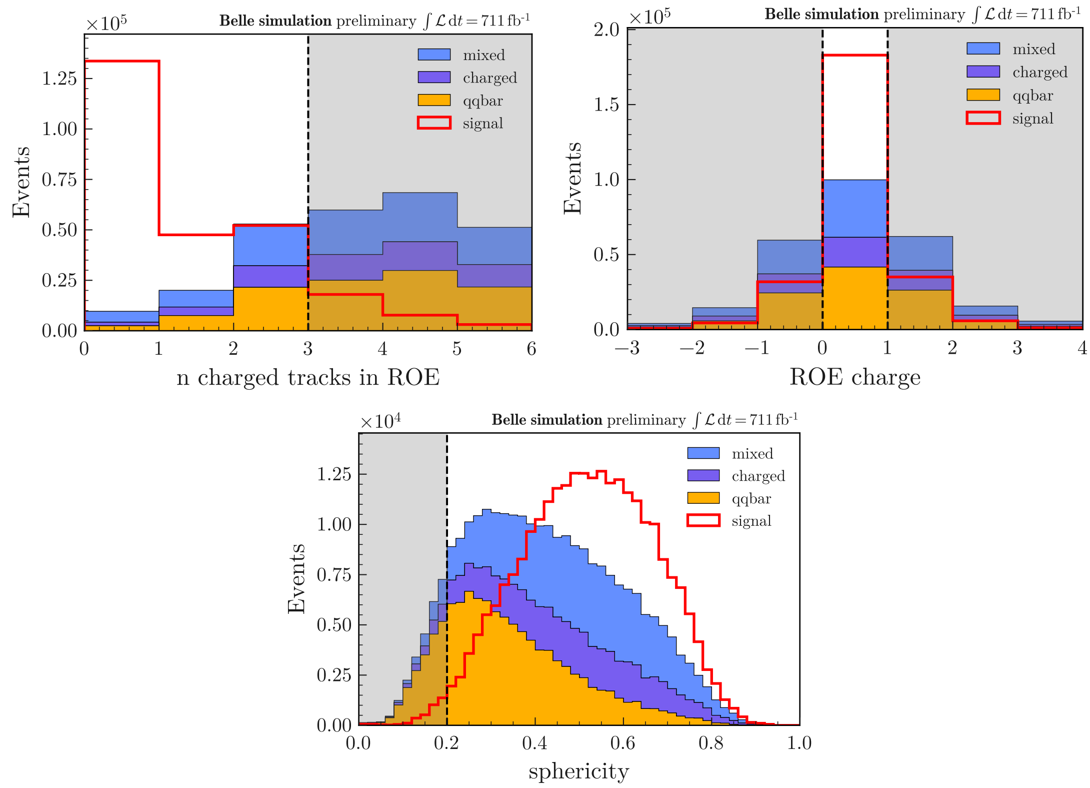


Figure 4.7: Selection cuts applied at the event level on the number of charged tracks in the ROE, total charge of the ROE (top) and on the sphericity (bottom) for the *OSe* mode in Belle MC. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison. The FEI and K^{*0} mass selection criteria described previously are applied.

4.3.4. Signal study and efficiency

The signal efficiency ε_{sig} is defined as the ratio of the number of reconstructed signal events passing the selection described above compared to the number of generated signal events. The largest component of the efficiency drop is the reconstruction of the tag side with the hadronic FEI. Indeed, as discussed in Section 2.4.2, after the selection on ΔE and M_{bc}, the neutral hadronic FEI is known to have a maximum efficiency of ~0.3% for a purity of 15% and only ~0.2% for a purity of 50% where the purity is defined as the fraction of events with a correct tag [31]. In this analysis, the efficiency of the reconstruction, which includes the tag side reconstruction and selection and the reconstruction of signal particle tracks, is presented in Table 4.2 for Belle II and 4.3 for Belle. The additional preselection cuts efficiencies are also presented in these tables. The difference between the electron and muon modes for Belle can be explained by the Belle muon PID whose performance is, on average, worse since only muons reaching the KLM are considered.

	OSe	SSe	$OS\mu$	$SS\mu$
Reconstruction (FEI + track reco.)	0.30% (0.17)	0.42% (0.18)	0.31% (0.17)	0.42% (0.17)
K^{*0} mass window	0.15% (0.14)	0.16% (0.13)	0.15% (0.13)	0.16% (0.13)
Sphericity	0.15% (0.14)	0.16% (0.13)	0.15% (0.13)	0.15% (0.12)
Tracks in ROE	0.12% (0.11)	0.13% (0.12)	0.11% (0.10)	0.12% (0.11)

Table 4.2: Efficiency of the signal MC selection cuts for Belle II. The value in parentheses is for truthmatched $K^{*0}\ell$ system. For each step, the efficiency takes into account the previous steps.

	OSe	SSe	$OS\mu$	$SS\mu$
Reconstruction (FEI + track reco.)	0.33% (0.18)	0.51% (0.18)	0.26% (0.15)	0.40% (0.15)
K^{*0} mass window	0.16% (0.14)	0.17% (0.14)	0.13% (0.11)	0.14% (0.11)
Sphericity	0.16% (0.13)	0.17% (0.14)	0.12% (0.11)	0.14% (0.11)
Tracks in ROE	0.14% (0.12)	0.15% (0.12)	0.11% (0.10)	0.12% (0.10)
ROE charge	0.11% (0.10)	0.11% (0.10)	0.08% (0.08)	0.09% (0.08)

Table 4.3: Efficiency of the signal MC selection cuts for Belle. The value in parentheses is for truthmatched $K^{*0}\ell$ system.

After the preselection is applied, the signal MC samples still contain some self cross-feed events. They consist of events which are true signal events (as they are generated as such) and reconstructed as signal (so present in the samples), but with a wrong B_{sig} or B_{tag} candidate, according to the MC truth. The proportions of correct and wrong B_{sig} ($K^{*0}\ell$ system) and B_{tag} candidates are shown in Table 4.4 (resp. 4.5) for the Belle II (resp. Belle) OSe mode. For both Belle II and Belle, the proportion of misreconstructed B_{sig} is reasonably small. However, the proportion of wrong B_{tag} is not negligible. Figure 4.8 shows that despite a poorer resolution, the events with a wrong B_{tag} are still peaking at the τ mass. They are thus treated as part of the signal template.

In all four modes, the ratio of good B_{sig} vs. wrong B_{sig} is approximately 95% / 5% for Belle II and 90%/10% for Belle, and the ratio of good B_{tag} vs. wrong B_{tag} is approximately 60% / 40% for Belle II and 50%/50% for Belle, with a slightly worse ratio for the SS modes compared to OS.

TAG \ SIG	Correct B_{sig}	Wrong B_{sig}	Total
Correct B_{tag}	53.4%	2.1%	55.5%
Wrong B_{tag}	39.1%	5.4%	44.5%
Total	92.5%	7.5%	100.0%

Table 4.4: Self cross-feed events for the OSe mode in Belle II signal MC.

TAG \ SIG	Correct B_{sig}	Wrong B_{sig}	Total
Correct B_{tag}	47.3%	2.6%	49.9%
Wrong B_{tag}	43.1%	7.0%	50.1%
Total	90.4%	9.6%	100.0%

Table 4.5: Self cross-feed events for the OSe mode in Belle signal MC.

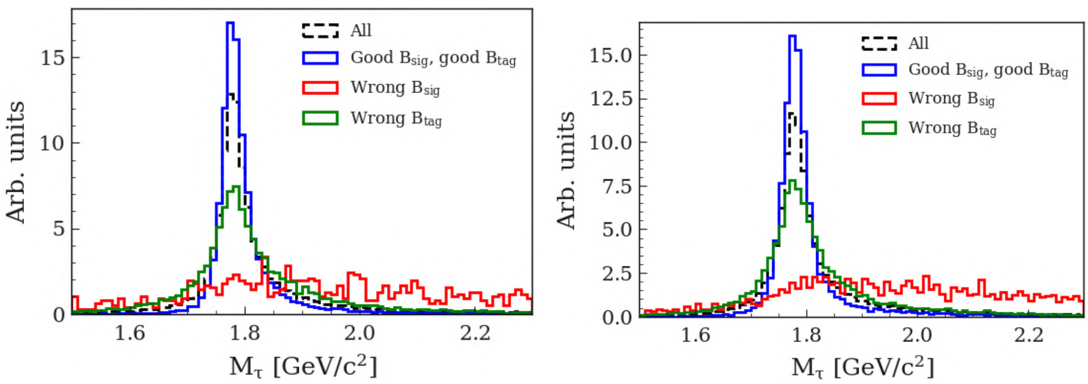


Figure 4.8: Effect of the misreconstructed events on the τ recoil mass for the OSe mode in Belle II (left) and Belle (right) signal MC. The histograms are normalised to the same area for shape comparison.

4.3.5. Summary

After this preselection summarised in Table 4.6, the signal efficiency and number of background events remaining are presented in Tables 4.7 for Belle II and 4.8 for Belle.

The τ recoil mass distribution after preselection is also shown in Figure 4.9 for Belle II and on Figure 4.10 for Belle. As can be seen in these figures, the background distribution is different for the *OS* and *SS* modes, the latter having a higher background yield in the signal region.

Variable	Description	Selection	Applied to
M_{bc} [GeV/c^2]	Beam-constrained mass	> 5.27	B_{tag}
ΔE [GeV]	$E_{B_{\text{tag}}} - E_{\text{CMS}}/2$	$-0.15 < \Delta E < 0.1$	B_{tag}
cos TBTO	Cosine between B_{tag} and (ROE+B_{sig})	< 0.9	B_{tag}
$P_{\text{sig}}(FEI)$	FEI signal probability	$> 10^{-3}$	B_{tag}
dr [cm]	Transverse distance between IP and POCA	< 0.5	K, π, ℓ, t_τ
dz [cm]	Distance between IP and POCA along z coordinate	$-5 < dz < 5$	K, π, ℓ, t_τ
$nCDCHits$	Number of hits in the CDC	> 20 (Belle II only)	K, π
kaon ID	Kaon particle identification	≥ 0.6	K
pion ID	Pion particle identification	≥ 0.6	π
electron ID	Electron particle identification	≥ 0.9	e
muon ID	Muon particle identification	≥ 0.9	μ
$dM(K^{*0})$ [GeV/c^2]	Difference between reconstructed mass and nominal mass	$-0.05 < dM < 0.05$	K^{*0}
$sphericity$	Event sphericity	> 0.2	event
$N_{\text{ROE}}^{\text{charged}}$	Number of charged tracks in ROE	< 3	event
ROE charge	Total charge of the ROE	$== 0$ (Belle only)	event

Table 4.6: Selection summary and variable descriptions.

Belle II	OSe	SSe	$OS\mu$	$SS\mu$
$N_{q\bar{q}}$	2730 (750)	3120 (860)	11270 (3090)	12190 (3340)
$N_{B\bar{B}}$	8800 (2410)	13570 (3720)	16630 (4560)	21920 (6010)
$N_{\text{bkg}}^{\text{tot}}$	11530 (3160)	16690 (4580)	27900 (7650)	34110 (9360)
ε_{sig}	0.123%	0.128%	0.112%	0.117%

Table 4.7: Signal MC efficiency and number of background events left in generic MC scaled to the data luminosity after the preselection for Belle II. The value in parentheses is scaled to 100 fb^{-1}.

Belle	OSe	SSe	$OS\mu$	$SS\mu$
$N_{q\bar{q}}$	10330	12750	6260	6990
	(1450)	(1790)	(880)	(980)
$N_{B\bar{B}}$	21450	34230	8960	23180
	(3020)	(4820)	(1260)	(3260)
$N_{\text{bkg}}^{\text{tot}}$	31780	46980	15220	30170
	(4470)	(6610)	(2140)	(4240)
ε_{sig}	0.105%	0.114%	0.084%	0.093%

Table 4.8: Signal MC efficiency and number of background events left in generic MC scaled to the data luminosity after the preselection for Belle. The value in parentheses is scaled to 100 fb^{-1}.

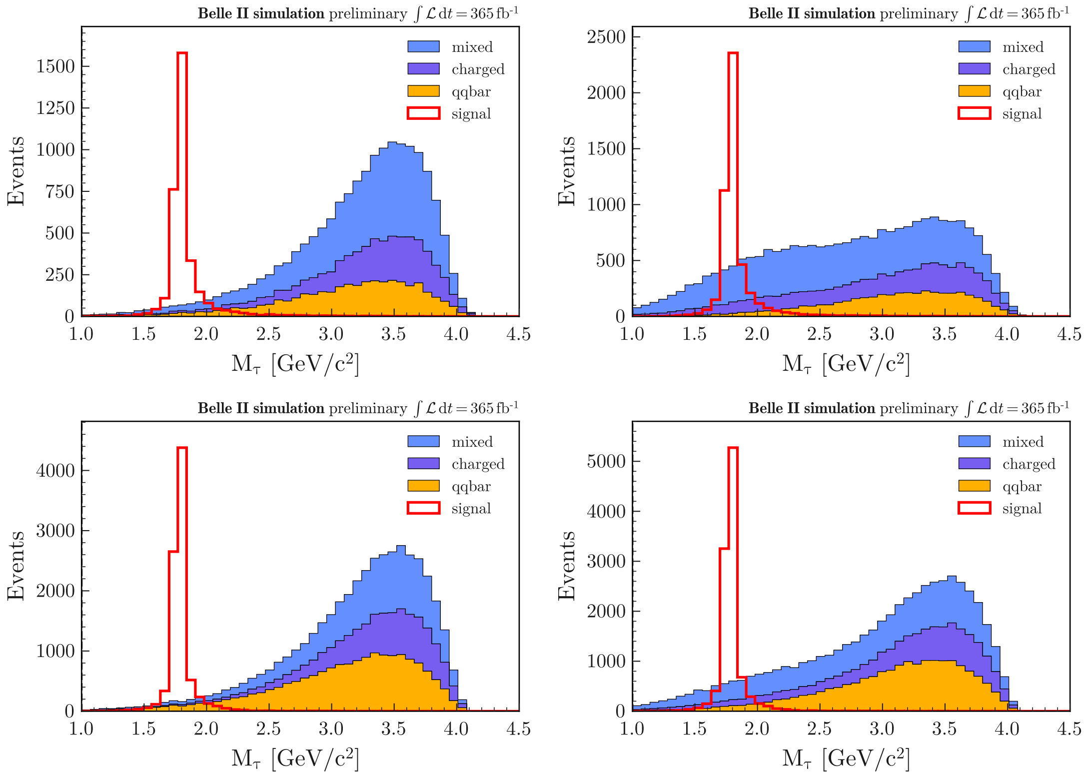


Figure 4.9: τ recoil mass distribution after the preselection for the four modes (from top left to bottom right : OSe, SSe, $OS\mu$, $SS\mu$) in Belle II MC. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and arbitrary scaled for better visualisation and shape comparison.

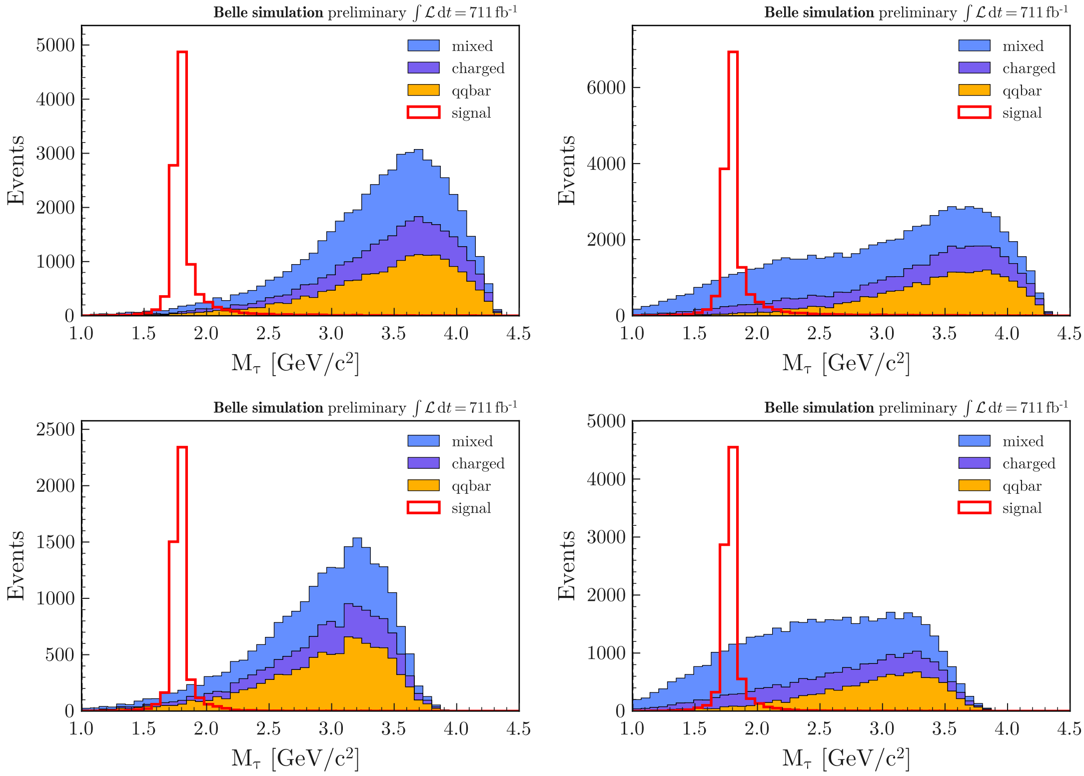


Figure 4.10: τ recoil mass distribution after the preselection for the four modes (from top left to bottom right : OSe, SSe, $OS\mu$, $SS\mu$) in Belle MC. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and arbitrary scaled for better visualisation and shape comparison.

4.4. Background suppression

The background suppression is performed in two steps. First, a cut-based selection is applied to remove certain specific physics backgrounds, that are easy to target. Then, a Boosted Decision Tree (BDT) is trained to separate the signal from the remaining backgrounds. Unless specified otherwise, the same selection is applied to Belle and Belle II datasets.

4.4.1. Signal region

Before studying the different physics backgrounds, we can exclude all events that are very far from the expected signal position in the τ recoil mass. In this way, all events that do not fall in the $M_\tau \in [1.0, 2.5]\mathrm{GeV}/c^2$ region are rejected. The corresponding selection is shown in Figure 4.11 for Belle and Belle II *OSe* mode.

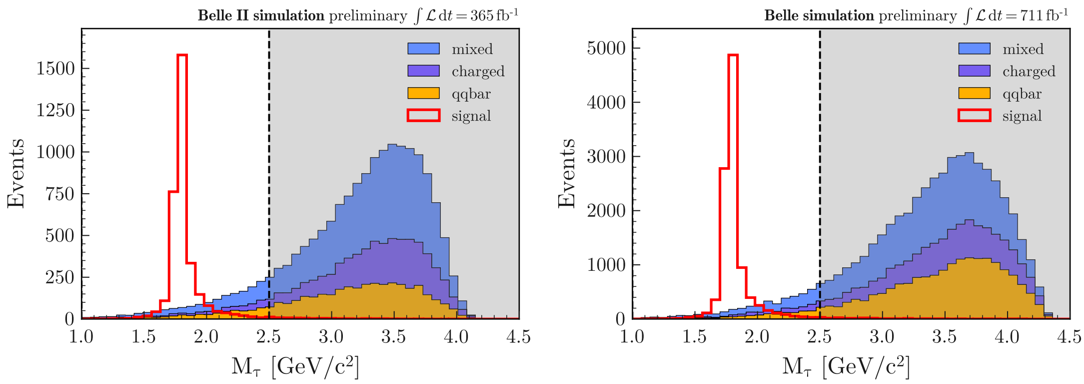


Figure 4.11: Selection cut applied on τ recoil mass distribution for the *OSe* mode in Belle II (left) and Belle (right) MC. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and arbitrary scaled for better visualisation and shape comparison.

4.4.2. J/ψ–meson veto

A contribution of $B \to K^{*0}J/\psi(\to \ell\ell)$ can arise when the 1 – prong track from the τ is an electron in the *OSe* and *SSe* modes, and a muon in the $OS\mu$ and $SS\mu$ modes. This results into a peak at the J/ψ mass in the invariant mass of the ℓt_τ system, where t_τ is the 1 – prong track. This peaking background is removed by vetoing the region $M(\ell t_\tau) \in [3.05, 3.15]\mathrm{GeV}/c^2$ for all modes, as illustrated on Figure 4.12.

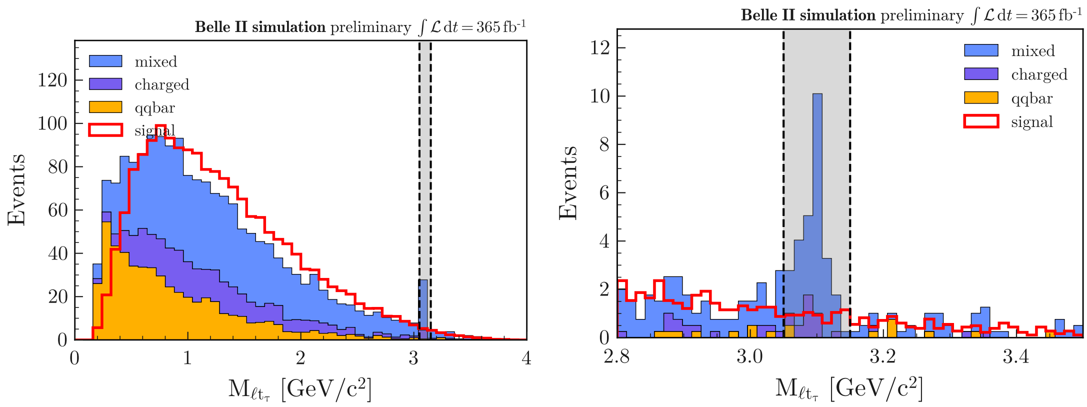


Figure 4.12: Selection cut applied on the invariant mass of the ℓt_τ system for the OSe mode in Belle II MC. The right plot is a zoomed view of the left plot. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison.

4.4.3. $D-$meson semi-leptonic backgrounds

The dominant backgrounds left at this stage of the selection are modes involving a $b \to c$ quark transition.

OS modes

In OS modes, the $K^{*0}\ell$ system comes from a $D^{(*)}$ meson decaying semi-leptonically and results in a peaking behaviour in the $K^{*0}\ell$ invariant mass. A schematic view of the signal and dominant backgrounds can be seen on Figures 4.13. In the $OS\mu$ mode, the misidentification of a pion as muon can result into a peak at the known $D-$meson mass corresponding to the $D \to K\pi\pi$ decay. The distribution of the $K^{*0}\ell$ system invariant mass is shown in Figure 4.14 for the electron and muon modes for Belle II as an example.

A veto on the $K\pi\pi$ peak is only applied in the muon mode such that events with $M(K^{*0}\ell) \in [1.83, 1.9]\text{GeV}/c^2$ are rejected. The lighter peaking structure between 1 and 2 GeV/c^2 is composed of various semileptonic D decays, but since it is made of a very large number of different final states, no selection can be applied to target any specific mode. Only a strict selection criterion around 1.9 GeV/c^2 could be applied to remove all the D semileptonic backgrounds at once, but at a cost of a very large drop in efficiency. Thus, no additional selection is applied to keep the efficiency as high as possible and preserve enough events in background to train a multivariate selection described below, in which the $K^{*0}\ell$ invariant mass will be used.

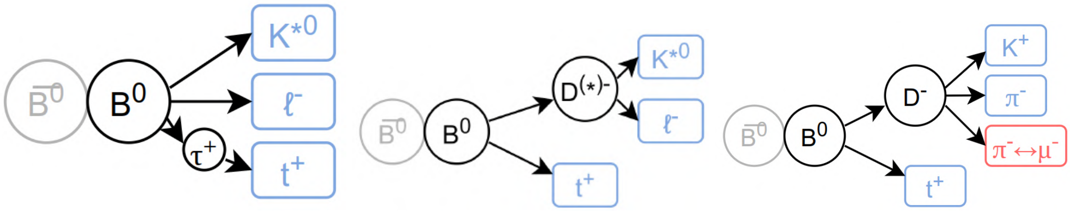


Figure 4.13: Schematic view of the signal (left) and dominant backgrounds (middle and right) decays for the *OS* modes.

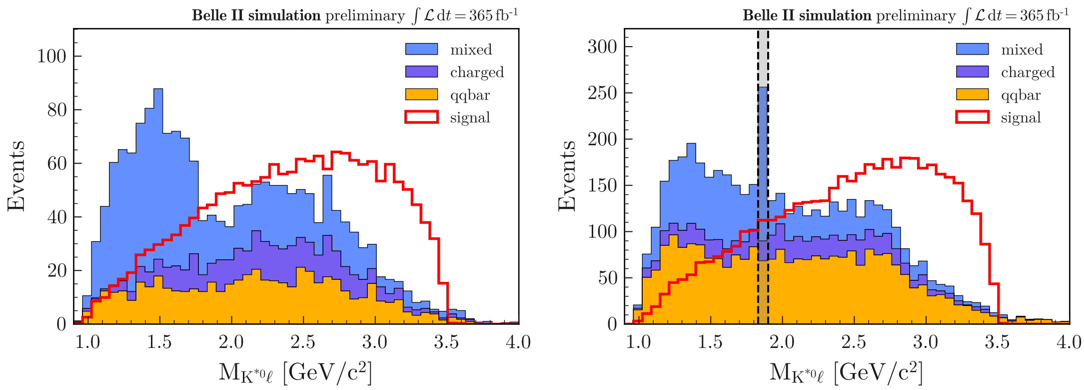


Figure 4.14: $K^{*0}\ell$ invariant mass distributions for OSe (left) and $OS\mu$ (right) in Belle II MC and veto applied. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison.

SS modes

In *SS* modes, the *B* meson is decaying semi-leptonically into the prompt lepton and a $D^{(*)}$ meson, which can decay into the $K^{*0}t_\tau$ system or into $K\pi\pi$ when the $1-$prong track is a pion or a misidentified muon. A schematic view of the signal and dominant backgrounds can be seen in Figures 4.15. The distribution of the invariant mass of the $K^{*0}t_\tau$ system is shown in Figure 4.16 for the electron and muon modes for Belle II as an example.

Similarly to the *OS* modes, only a veto is applied on the $K\pi\pi$ peak such that events with $M(K^{*0}t_\tau) \in [1.83, 1.9]\text{GeV}/c^2$ are rejected. As the peaking $K\pi\pi$ is independent of the prompt lepton flavour, this veto is applied to both $SS\mu$ and SSe. For the same reason as for the *OS* modes, no additional selection is applied and the $K^{*0}t_\tau$ invariant mass will be used in a BDT.

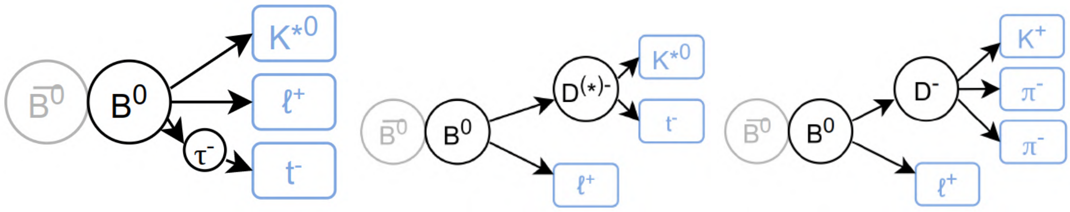


Figure 4.15: Schematic view of the signal (left) and dominant backgrounds (middle and right) decays for the SS modes.

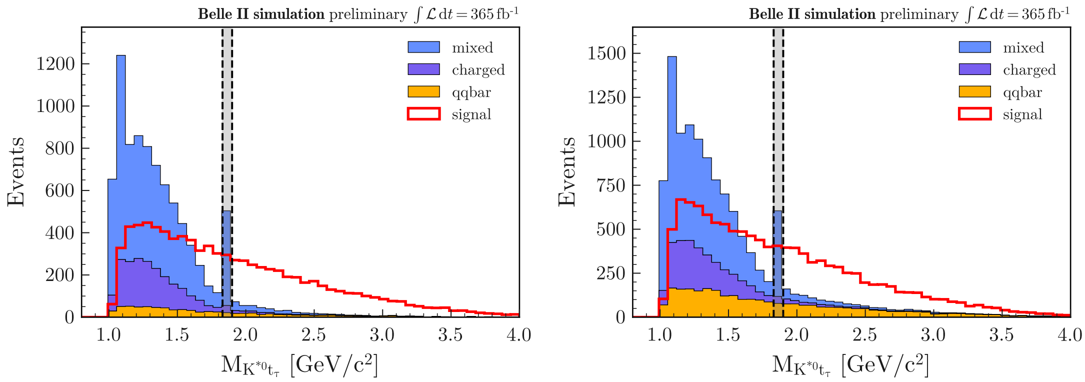


Figure 4.16: $K^{*0}t_\tau$ invariant mass distributions for SSe (left) and $SS\mu$ (right) in Belle II MC and veto applied. The shaded area is rejected. Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and scaled to the same area as background for the shape comparison.

The distribution of the τ recoil mass after the cut-based selection is presented in Figures 4.17 (Belle II) and 4.18 (Belle). In addition, the signal efficiency and number of background events left at this stage of the selection are displayed in Tables 4.9 and 4.10 for Belle II and Belle respectively. The efficiencies and background yields are now coherent between Belle II and Belle, with more background for the SS modes compared to the OS, and higher efficiencies in the electron modes compared to the muon ones.

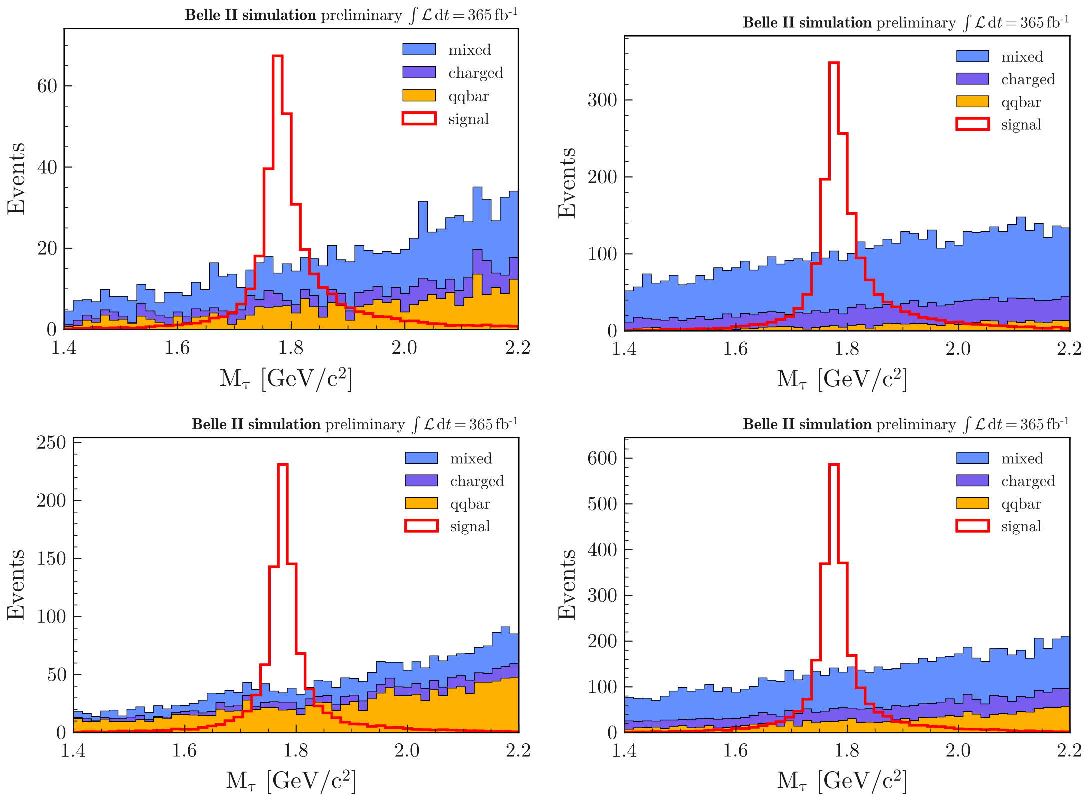


Figure 4.17: τ recoil mass after the cut-based selection for Belle II (from top left to bottom right to OSe, SSe, $OS\mu$, $SS\mu$). Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and arbitrarily scaled for better visualisation and shape comparison.

Belle II	OSe	SSe	$OS\mu$	$SS\mu$
$N_{q\bar{q}}$	357 (98)	499 (140)	1610 (441)	1760 (480)
$N_{B\bar{B}}$	835 (229)	4960 (1360)	1213 (332)	5570 (1530)
$N_{\text{bkg}}^{\text{tot}}$	1191 (327)	5460 (1500)	2820 (773)	7330 (2010)
ε_{sig}	0.117%	0.115%	0.103%	0.105%

Table 4.9: Signal MC efficiency and number of background events left in generic MC scaled to the data luminosity after the cut-based selection for Belle II. The values in parentheses are scaled to 100 fb^{-1}.

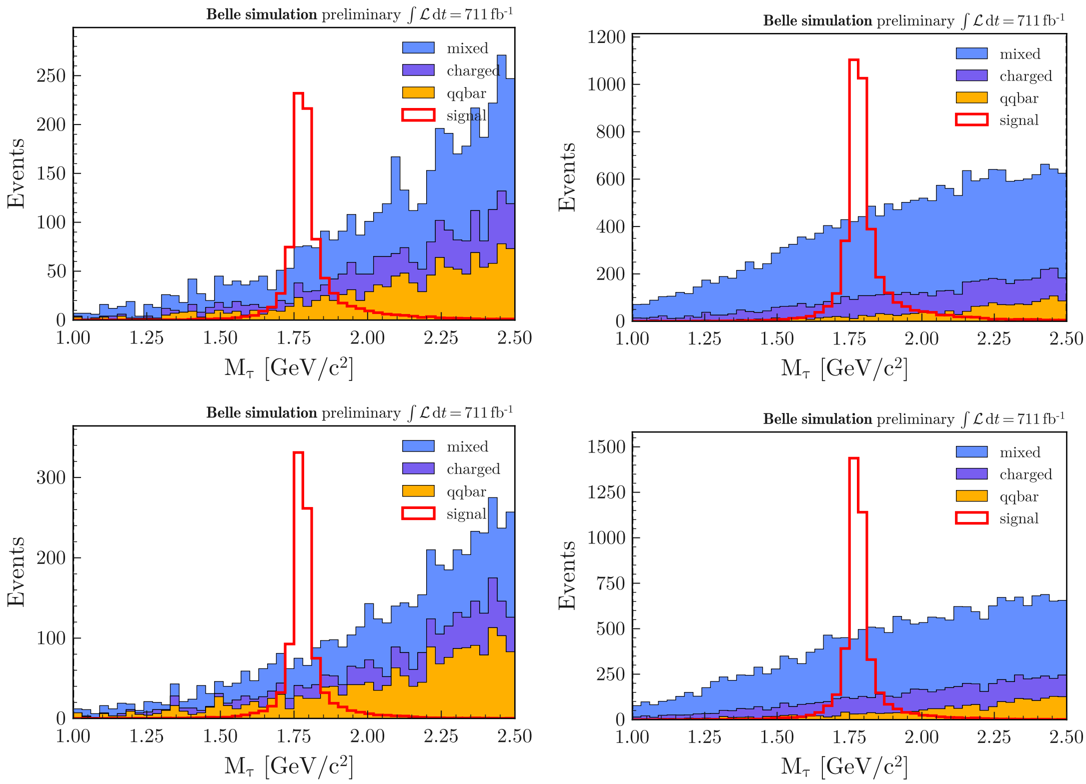


Figure 4.18: τ recoil mass after the cut-based selection for Belle (from top left to bottom right to OSe, SSe, $OS\mu$, $SS\mu$). Background is scaled to the data luminosity. The signal MC is truthmatched for the $K^{*0}\ell$ system and arbitrarily scaled for better visualisation and shape comparison.

Belle	OSe	SSe	$OS\mu$	$SS\mu$
$N_{q\bar{q}}$	610	920	1130	1350
	(86)	(130)	(160)	(189)
$N_{B\bar{B}}$	1920	11450	1720	11390
	(271)	(1610)	(242)	(1600)
$N_{\text{bkg}}^{\text{tot}}$	2530	12370	2850	12740
	(357)	(1740)	(402)	(1790)
ε_{sig}	0.098%	0.101%	0.079%	0.085%

Table 4.10: Signal MC efficiency and number of background events left in generic MC scaled to the data luminosity after the cut-based selection for Belle. The values in parentheses are scaled to 100 fb^{-1}.

4.4.4. Boosted Decision Tree classification

After the selection targeting specific physics backgrounds, a Boosted Decision Tree (BDT) [129] is trained to discriminate signal from the remaining background. A BDT classifier is a machine learning algorithm using supervised learning on an ensemble of decision trees to classify events, here in signal and background categories. A single BDT is used to target both the $B\overline{B}$ and $q\overline{q}$ backgrounds.

For each of the Belle and Belle II MC sets, the equivalent of the data luminosity is kept for the validation of the BDT, while the remaining $3\times\, L_{\text{data}}$ are separated into the train and test samples with a ratio 80%-20%. A subset of the available signal MC with the same number of events as in the train-test background sample is picked and used for the training and testing of the BDT. The test sample is used for performance estimation and the optimisation of the hyperparameters, while the validation sample, completely independent of the events used to train and optimise the BDT performance, is used to optimise the BDT selection. A schematic view of the separation of the samples is shown in Figure 4.19.

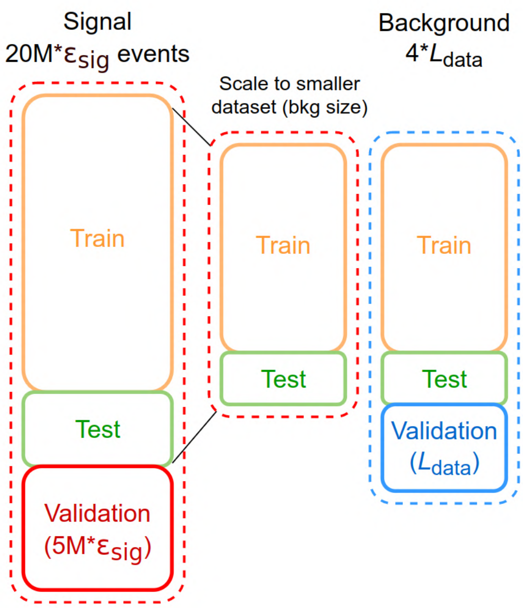


Figure 4.19: Schematic view of the BDT strategy.

We train one BDT for each of the four modes $(OS, SS)\times(e, \mu)$ and experiment (Belle, Belle II), resulting in a total of 8 BDTs. The BDTs are based on the `fastBDT` library [130] and the hyperparameters are optimised for each BDT with the `Optuna` library [131].

BDT training

The hyperparameters considered for optimisation are the number of decision trees to train (`nTrees`), the maximum depth of the trees (`nLevels`), the fraction of events to use to train each tree (`randRatio`) and the learning rate controlling the contribution of each tree at every model update (`shrinkage`). These parameters are optimised in order to maximise an objective function F depending on the receiver operating characteristic (ROC) curve and the loss function of the train and test samples:

$$F = \frac{AUC_{\text{test}}}{L_{\text{test}}} \cdot \left(\frac{AUC_{\text{test}}}{AUC_{\text{train}}}\right)^2 \cdot \frac{L_{\text{train}}}{L_{\text{test}}}, \tag{4.3}$$

where $AUC_{\text{train/test}}$ is the area under the receiver operating characteristic (ROC) curve of the train/test sample and $L_{\text{train/test}}$ is the log loss function on the train/test sample. The AUC and L ratio terms help to control the overtraining while the $AUC_{\text{test}}/L_{\text{test}}$ term favours better performance. The optimised values of the hyperparameters are displayed in Table 4.11 for each of the 8 BDTs.

		OSe		*SSe*		*OSμ*		*SSμ*	
Parameter	Range	Belle II	Belle	Belle II	Belle	Belle II	Belle	Belle II	Belle
nTrees	(10, 300)	220	220	290	180	230	260	135	280
nLevels	(1, 5)	2	2	2	2	1	2	2	2
randRatio	(0.1, 0.8)	0.1	0.3	0.2	0.1	0.4	0.6	0.7	0.3
shrinkage	(0.005, 0.5)	0.05	0.07	0.09	0.11	0.26	0.12	0.38	0.11

Table 4.11: Values of the hyperparameters from fastBDT for each of the 8 BDTs (OS, SS) × (e, μ) × (Belle II, Belle).

To train the BDTs, a set of 12 to 14 variables chosen from a common larger set is used for each of the BDTs, where the set of chosen variables can vary from one mode to another, but are the same for Belle and Belle II in each mode. A list of about 25 kinematic and event-shape variables expected to have different behaviour between signal and background was used as primary training. Then, for each mode independently, the variables with very low or no contribution to improve the model's performance were removed, according to the ranking visible for the final set of input variables in Figures 4.20 and 4.21. Performance was verified to be equivalent or better when removing input variables.

Among the variables used, different categories can be distinguished. First, we use variables related to the signal kinematics, with the invariant masses $M(K^{*0}\ell)$ and $M(K^{*0}t_\tau)$ discussed in the previous section, the energies of the prompt lepton and the 1 – prong track, and information on the $K^{*0}\ell$ vertex. The second category consists of variables related to the ROE, such as the number of neutral ROE clusters in the ECL and their energy, as well as the ROE total momentum (tracks and clusters) and the number of charged tracks found. Finally, the last category comprises the event shape variables, such as the sphericity and the modified Super Fox-Wolfram moments (KSFW) [127]. The Fox-Wolfram (FW) moments are a set of rotationally invariant observables H_k such that for N particles with momenta p_i, the k-th order FW moment is defined as $H_k = \sum_{i,j=1}^{N} |\vec{p}_i||\vec{p}_j|P_k(\cos\theta_{ij})$, where P_k are the Legendre polynomials and θ_{ij} is the angle between the momenta $\vec{p}_i$ and $\vec{p}_j$. At Belle and Belle II the FW moments are modified to take into account the origin of the particles on which the sum is run (s for particles from $B_{\rm tag}$, o for ROE particles, leading to modified FW moments H_k^{ss}, H_k^{oo} and H_k^{so}). The KSFW moments are defined as linear combinations of these modified FW moments. The last category is particularly useful to separate the $q\overline{q}$ background from the signal, as the $q\overline{q}$ events are more jet-like than the $B\overline{B}$ events. The variables used in the BDT training are listed in Table 4.12 for each of the four modes, and the most important features are shown for Belle II in Figure 4.20 and for Belle in Figure 4.21.

Variable	Description	OSe	SSe	$OS\mu$	$SS\mu$
$M(K^{*0}\ell)$	$K^{*0}\ell$ invariant mass	✓	✓	✓	✓
$M(K^{*0}t_\tau)$	$K^{*0}t_\tau$ invariant mass	–	✓	–	✓
E_ℓ	prompt lepton energy	✓	✓	✓	✓
E_{t_τ}	1-prong track energy	✓	✓	✓	✓
$dr(V_{K^{*0}\ell})$	transverse distance in respect to IP of $K^{*0}\ell$ vertex	✓	✓	✓	✓
$chiProb(V_{K^{*0}\ell})$	χ^2 probability of the $K^{*0}\ell$ vertex	✓	✓	✓	✓
$E_{\rm ROE}^{\rm neutral}$(ECL)	energy from neutral ROE clusters in ECL	✓	✓	–	–
$N_{\rm ROE}^{\rm neutral}$(ECL)	number of neutral ROE clusters in ECL	✓	✓	✓	✓
$N_{\rm ROE}^{\rm charged}$	number of charged tracks in the ROE	✓	✓	✓	✓
$p_{\rm ROE}$	ROE momentum	✓	✓	✓	✓
$hso00$	KSFW moment	✓	–	✓	–
$hso20$	KSFW moment	✓	✓	✓	✓
$hoo0$	KSFW moment	✓	✓	✓	✓
$mm2$	KSFW moment	–	✓	–	✓
$sphericity$	event sphericity	✓	✓	✓	✓

Table 4.12: Input variables in BDT (for KSFW moment, see [127]).

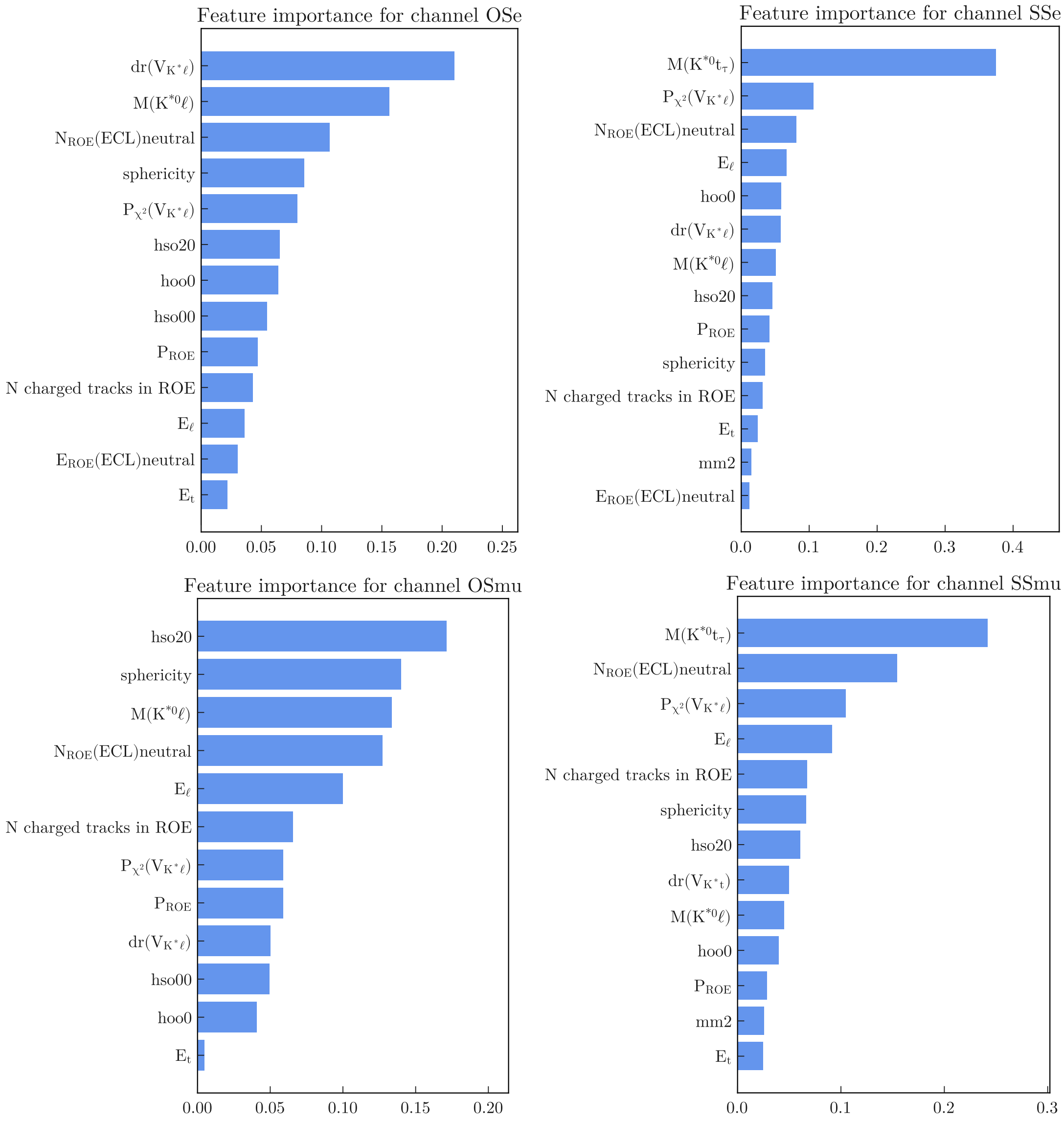


Figure 4.20: Importance of the BDT features in Belle II (top left to bottom right : *OSe*, *SSe*, *OSμ*, *SSμ*).

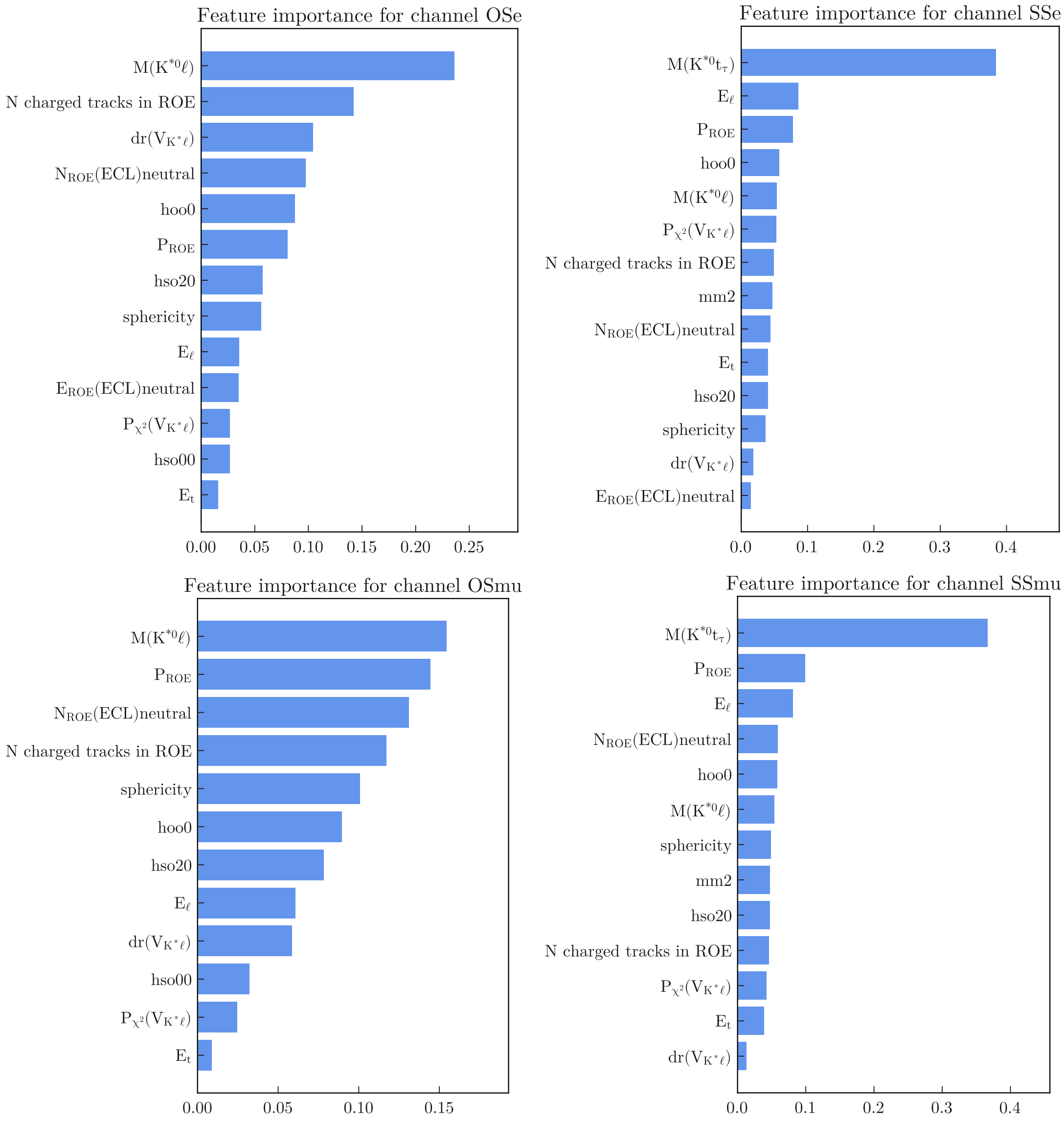


Figure 4.21: Importance of the BDT features in Belle (top left to bottom right : *OSe*, *SSe*, $OS\mu$, $SS\mu$).

BDT performance

To evaluate the performance of the BDTs, we can use the ROC curves and BDT score distributions for the train and test samples, which are shown on Figure 4.22 for Belle II and 4.23 for Belle. The BDT score distributions of $B\overline{B}$ and $q\overline{q}$ events separately in the validation sample are also presented.

A small overtraining is visible in SSe and $OS\mu$ Belle modes. However, as mentioned previously, the BDTs are optimised in order to minimise the overtraining and maximise the performance according to the function (4.3) so the overtraining should not be an issue as long as the performance is still improving in the test sample as well. In addition, a selection cut will be applied on the BDT score, with respectively 40% and 72% efficiency for SSe and $OS\mu$. One can see in Figure 4.24 that in the region corresponding to such BDT cut efficiency, the train and test samples ROC curves are similar.

The BDTs perform well for the $q\overline{q}$ rejection and are slightly worse for the $B\overline{B}$ rejection, in particular for the OS modes. This is expected as $q\overline{q}$ events differ more from the signal due to their more jet-like geometry. The difference between OS and SS modes can be partially explained by the very different distributions in signal and background of the $K^{*0}\ell$ invariant mass in OS modes while although different too, the signal and background $K^{*0}t_\tau$ invariant masses are both distributed towards low mass. An attempt at treating separately the $q\overline{q}$ and the $B\overline{B}$ backgrounds was tried, but resulted in lower performance due to the low amount of MC simulation available to train the BDTs. One can also see in Figures 4.25 and 4.26 that the BDT score is globally homogeneous along the τ recoil mass range in signal and background, and that the BDTs do not sculpt a peaking structure in background in the signal region in M_τ.

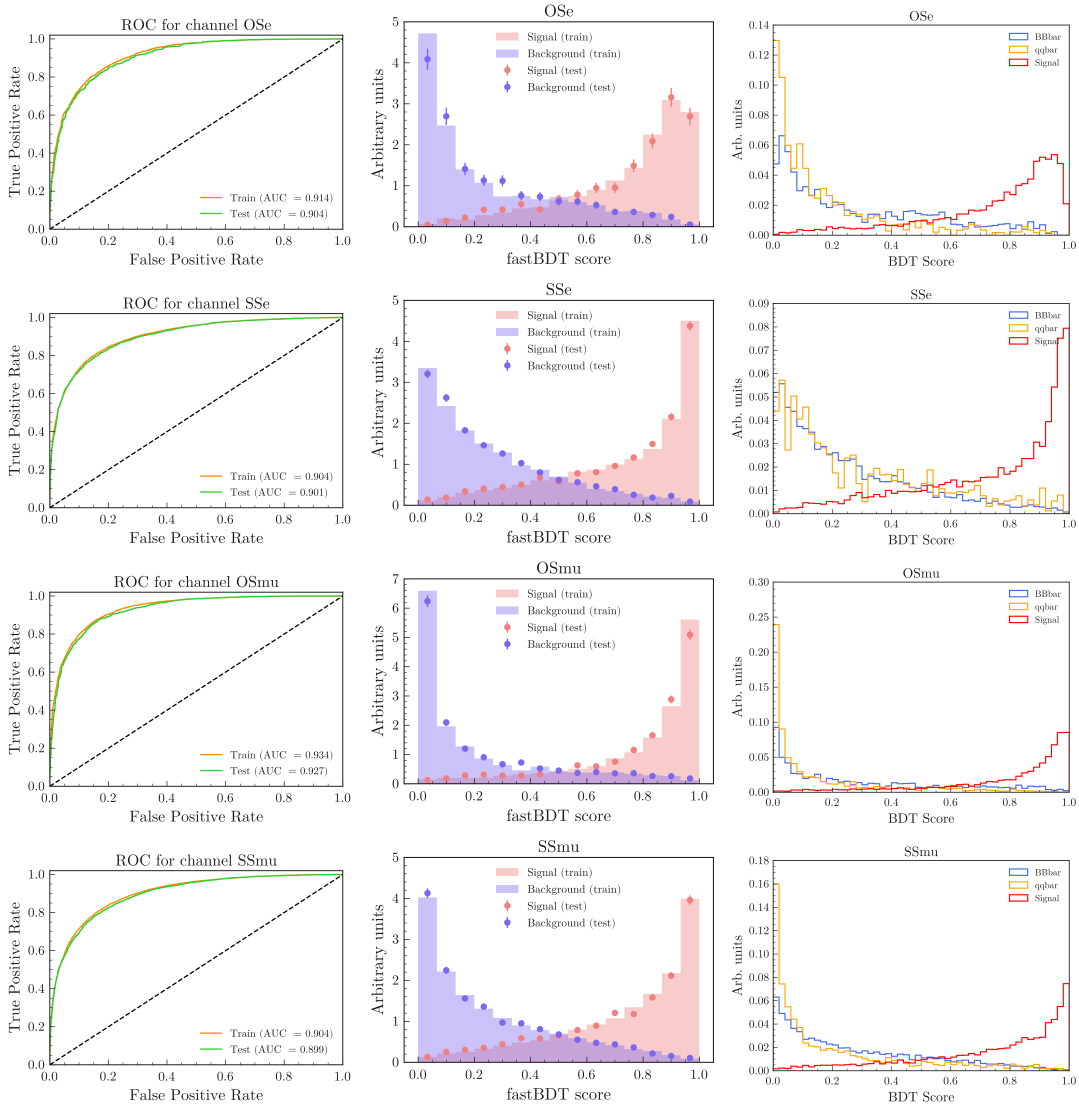


Figure 4.22: BDT ROC curves (left), scores (centre) and breakdown for $B\overline{B}$ and $q\overline{q}$ (right) for Belle II MC. From top to bottom : OSe, SSe, $OS\mu$, $SS\mu$.

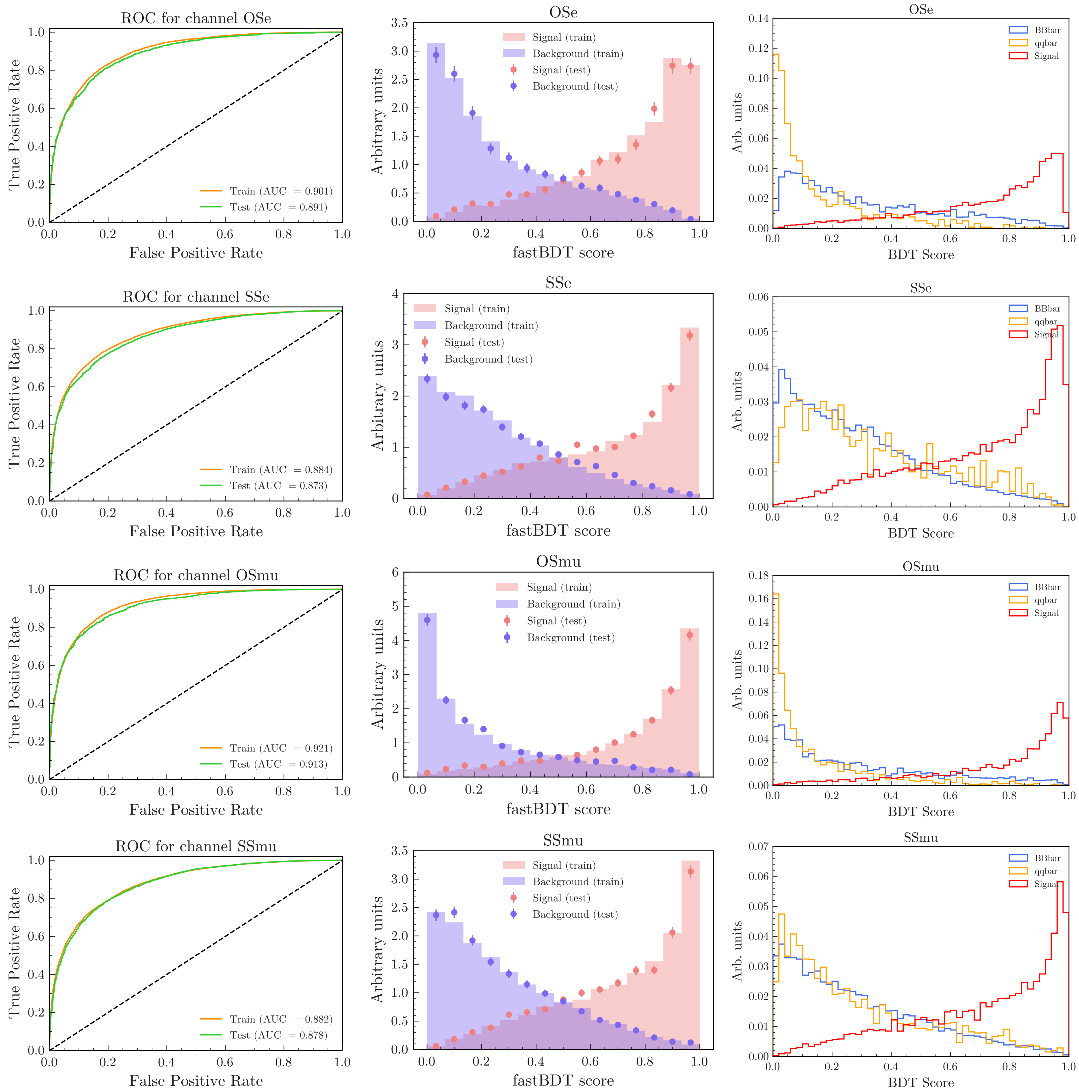


Figure 4.23: BDT ROC curves (left), scores (centre) and breakdown for $B\overline{B}$ and $q\overline{q}$ (right) for Belle MC. From top to bottom : OSe, SSe, $OS\mu$, $SS\mu$.

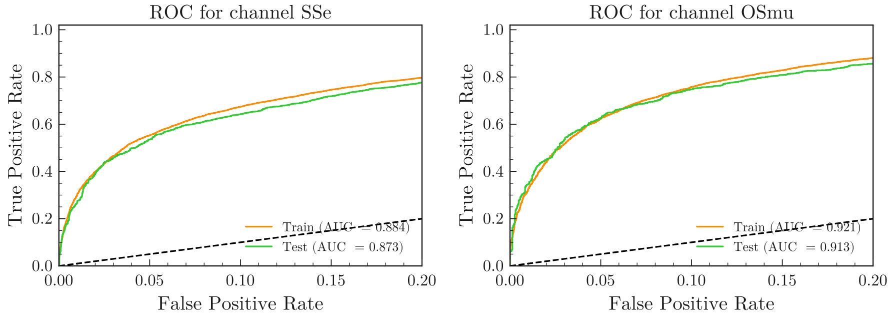


Figure 4.24: Zoom of the ROC curve of the Belle SSe (left) and $OS\mu$ (right) modes. The BDT cut applied corresponds to a signal efficiency of 40% for SSe and 72% for $OS\mu$. In these regions, no overtraining is visible. The black dashed line corresponds to a random classification.

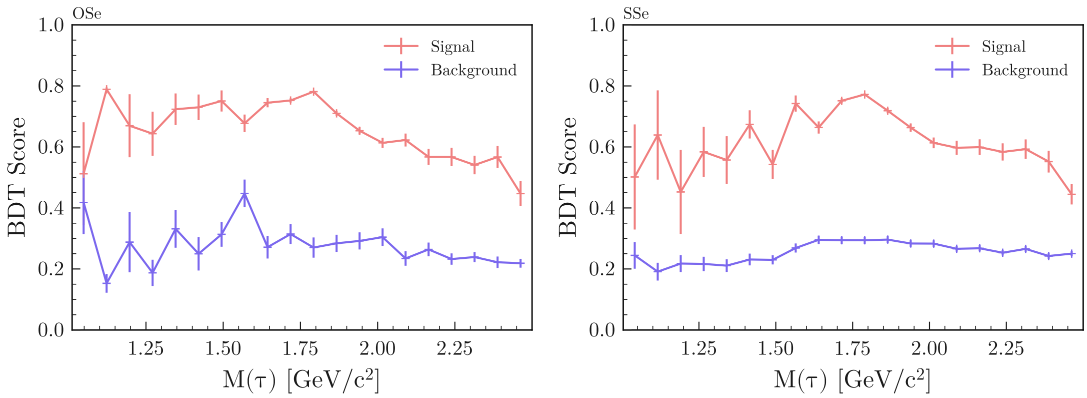


Figure 4.25: Profile plots of the BDT score with respect to the τ recoil mass for OSe (left) and SSe (right) modes in Belle II MC.

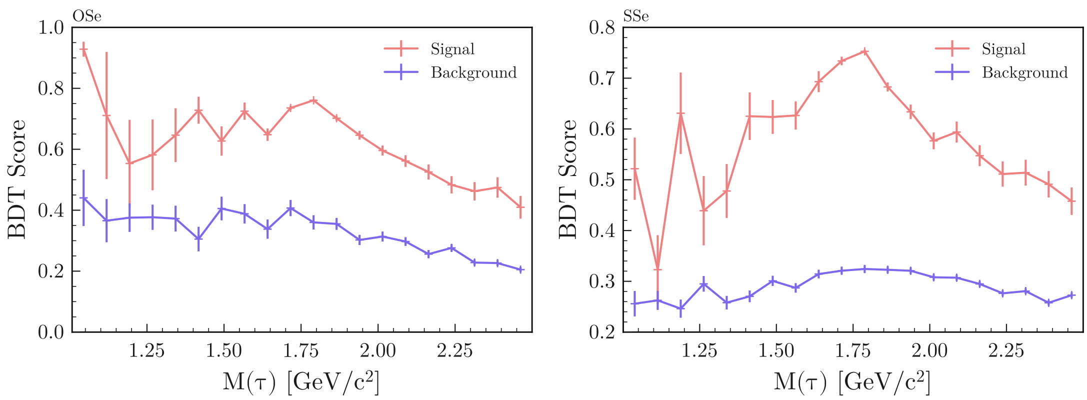


Figure 4.26: Profile plots of the BDT score with respect to the τ recoil mass for OSe (left) and SSe (right) modes in Belle MC.

BDT selection

The BDT selection is optimised in the [1.7, 1.85] GeV/c^2 region, which corresponds to approximately 85 to 90% signal coverage depending on the modes, with the Punzi figure of merit $\frac{\varepsilon_{\text{sig}}}{3/2+\sqrt{N_{\text{bkg}}}}$ [124]. The Punzi figure of merit (fom) value according to the BDT cut value is shown in Figures 4.27 and 4.28 for Belle II and Belle, respectively. The resulting BDT cut values, BDT efficiencies and number of background events left in the fitting region [1.0, 2.5]GeV/c^2 are displayed in Table 4.13 for Belle II and 4.14 for Belle.

In the next steps of the analysis, the τ recoil mass distribution of the events left after the BDT selection will be fitted. In some particular cases, the optimised BDT selection leads to a very low number of background events left in the fitting region, which can result in unstable fits. Specifically, since we allow the signal yield to be negative in the fit, a very low number of background events coupled to a negative signal can lead to a negative total probability density function (PDF). In these cases, we artificially loosen the BDT selection criterion to allow more events to be fitted. This is the case of the Belle II *SSe* mode, where the BDT cut is lowered from the optimised value of 0.95 to 0.8, which corresponds to the previous local maximum of the Punzi fom. Similarly, for the Belle II *OSe* mode, the cut applied on the BDT score is lowered to 0.65. For these two modes, the efficiency and number of background events with the optimised selection are shown in parentheses in Table 4.13.

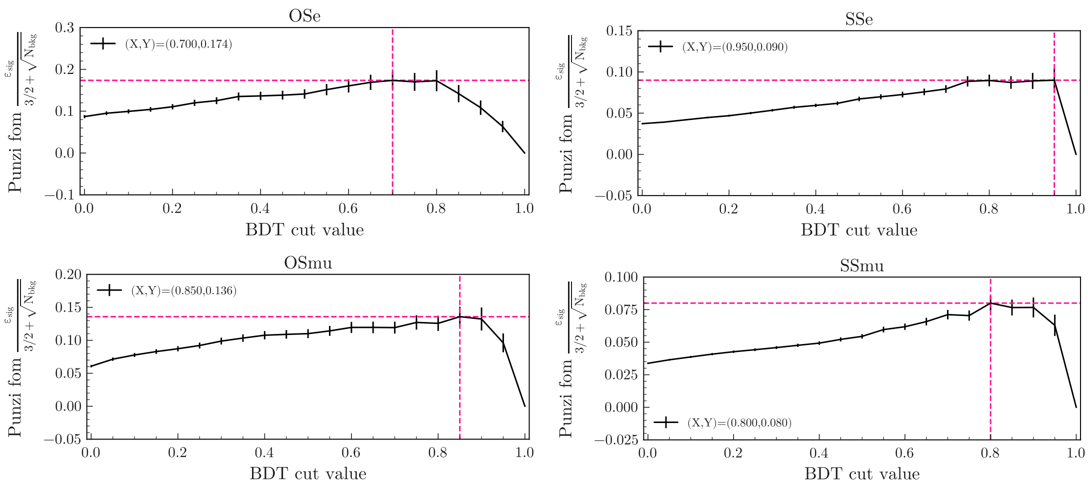


Figure 4.27: Punzi figure of merit and optimised BDT selection for each mode (top left to bottom right: *OSe*,*SSe*, *OSμ* *SSμ*) in Belle II.

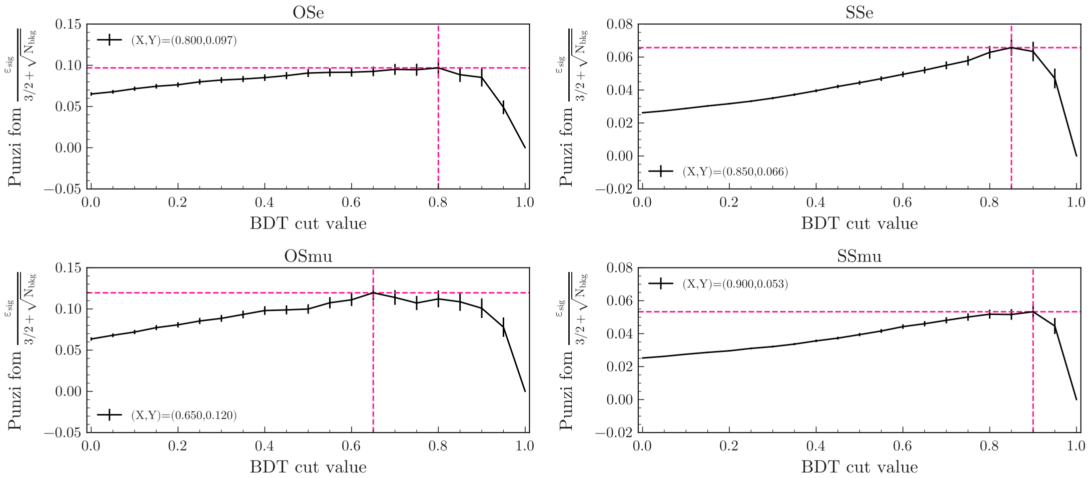


Figure 4.28: Punzi figure of merit and optimised BDT selection for each mode (top left to bottom right: *OSe*,*SSe*, *OSμ* *SSμ*) in Belle.

	OSe	*SSe*	*OSμ*	*SSμ*
BDT cut value	0.65 (0.7)	0.8 (0.95)	0.85	0.8
$\varepsilon_{\text{sig}}^{\text{BDT}}$	72.6% (67.5%)	53.0% (24.3%)	64.1%	52.4%
$\varepsilon_{\text{sig}}^{\text{tot}}$	0.085% (0.079%)	0.061% (0.028%)	0.066%	0.055%
N_{bkg}	116 (91)	183 (26)	131	260

Table 4.13: BDT cut value, corresponding efficiency, resulting total signal efficiency and number of background events left in the [1.0, 2.5]GeV/c^2 $M(\tau)$ region scaled to the data luminosity for Belle II. For relevant modes, the values in parentheses are for the optimised selection while the others are for the actually applied cut.

	OSe	*SSe*	*OSμ*	*SSμ*
BDT cut value	0.8	0.85	0.65	0.9
$\varepsilon_{\text{sig}}^{\text{BDT}}$	49.0%	39.6%	72.2%	29.4%
$\varepsilon_{\text{sig}}^{\text{tot}}$	0.048%	0.040%	0.057%	0.025%
N_{bkg}	91	248	258	142

Table 4.14: BDT cut value, corresponding efficiency, resulting total signal efficiency and number of background events left in the [1.0, 2.5]GeV/c^2 $M(\tau)$ region scaled to the data luminosity for Belle.

Finally, Figures 4.29 and 4.30 show the impact of the BDT selection on the τ recoil mass, where the signal is represented with a branching fraction on 3×10^{-5} for Belle II and Belle.

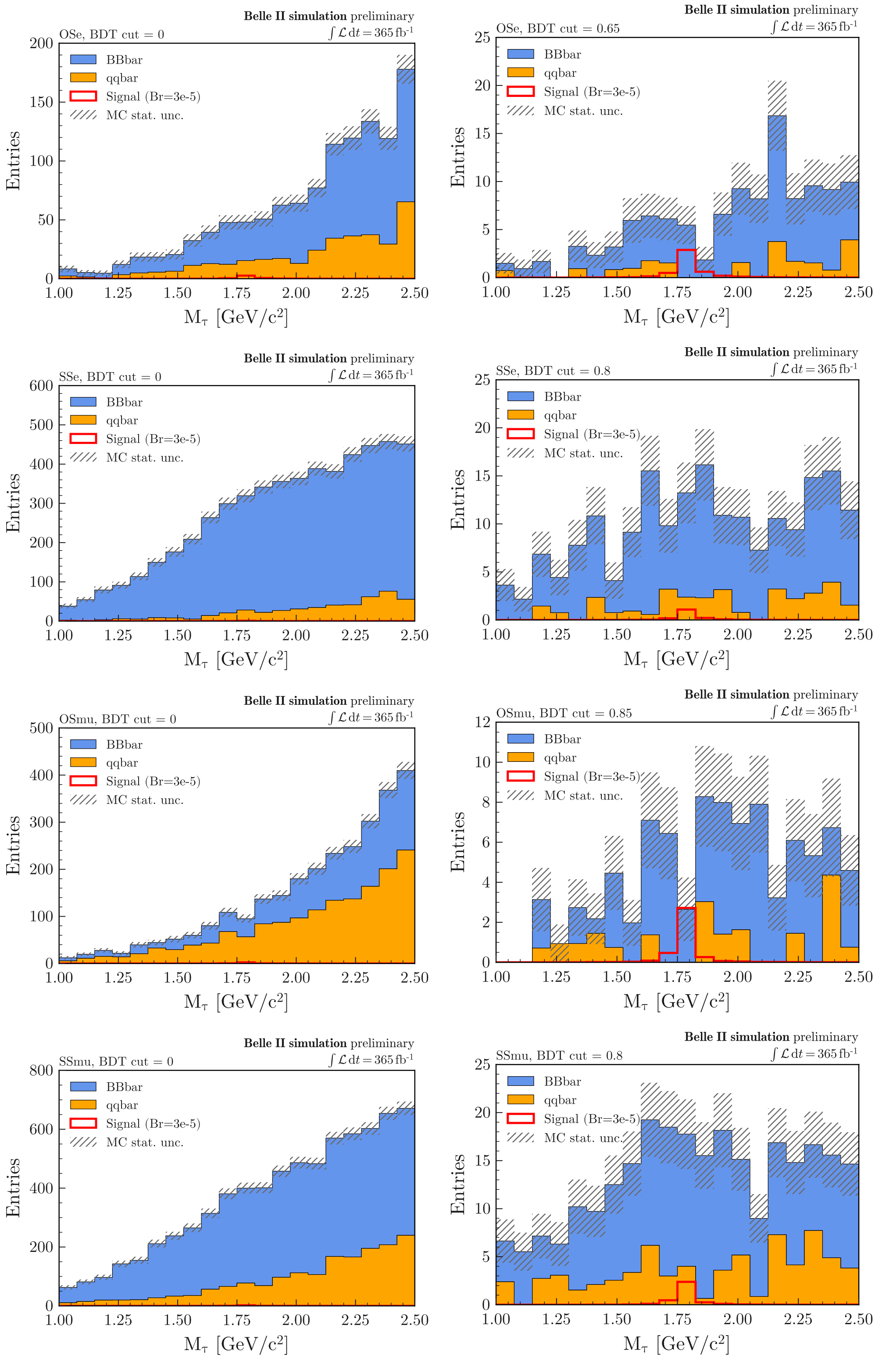


Figure 4.29: Effect of BDT selection on τ recoil mass for Belle II. Left is before the BDT selection, right is after (from top to bottom to *OSe*, *SSe*, *OSμ*, *SSμ*). The signal is plotted assuming a branching fraction of 3×10^{-5}.

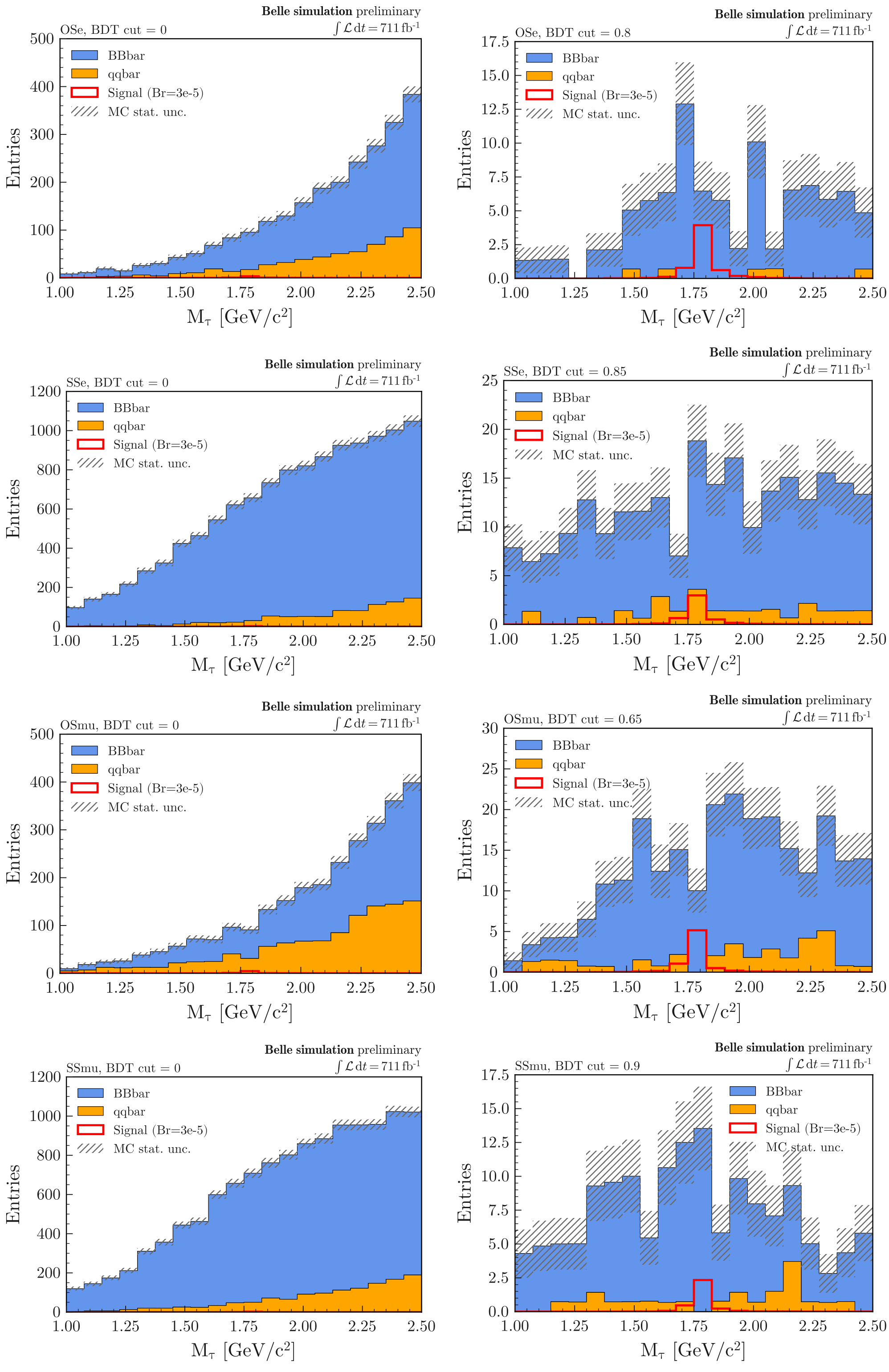


Figure 4.30: Effect of BDT selection on τ recoil mass for Belle. Left is before the BDT selection, right is after (from top to bottom to OSe, SSe, $OS\mu$, $SS\mu$). The signal is plotted assuming a branching fraction of 3×10^{-5}.

4.4.5. Summary

In general, more than one candidate per event may pass the selection criteria, resulting in so-called multiple candidates. The very last step of the selection is therefore to perform a best candidate selection within the remaining candidates.

At this stage, the total average multiplicity is between 1 and 1.2. Events with multiple candidates represent from 5 to 15% of all events for OS modes and 1 to 6% for SS modes. In both cases, the average multiplicity of these events is between 2 and 2.1. We choose a random candidate among the candidates left in order to not create any bias. The average multiplicity $\hat{M}$ in each sample is shown in Table 4.15 together with the proportion of events with multiple candidates R_M.

		OSe		SSe		$OS\mu$		$SS\mu$	
		Belle II	Belle	Belle II	Belle	Belle II	Belle	Belle II	Belle
Signal MC	$\hat{M}$	1.04	1.12	1.04	1.07	1.09	1.10	1.02	1.06
	R_M	4%	10%	4%	6%	7%	9%	2%	5%
Generic MC	$\hat{M}$	1.04	1.20	1.03	1.07	1.12	1.18	1.01	1.04
	R_M	4%	16%	3%	6%	11%	14%	1%	4%

Table 4.15: Average multiplicity $\hat{M}$ and proportion of events with multiple candidates R_M for each mode in signal and generic MC at the end of the selection process.

The final efficiencies and background yields for Belle II and Belle are presented in Table 4.16 where the total number of background events in parentheses is in the fit region $[1.3, 2.3]\mathrm{GeV}/c^2$.

The remaining background is as expected dominated by $B^0\overline{B}^0$ events. Using the TopoAna tool [132], allowing to match the reconstructed events to the MC generated decays, on Belle II mixed samples, we observe that the dominant modes left are $B^0 \to \ell\nu_\ell D^{(\prime,*)}_{(1,2)}$ modes for SS channels. They represent approximately 60% of the remaining $B^0\overline{B}^0$ decays, and the rest of the decays appear only once and cannot be grouped into specific background categories. For OS modes, the background left is dominated by $B^0 \to t(\nu)D^{(*)}(\to \ell\nu_\ell K^{*0})$ modes (where t is π, e or μ), but they represent a much smaller fraction of what is left (20% of remaining $B^0\overline{B}^0$ events for $OS\mu$, 10% for OSe). There is also a small component of $B^0 \to \ell\nu_\ell D^{(\prime,*)}_{(1,2)}$ decays in both OSe and $OS\mu$ cases and some $B^0 \to \pi^+\pi^-K^{*0}J/\psi(\to e^+e^-)$ for the OSe mode. Then, the large majority of the decay modes left appear only once and cannot be grouped into specific background categories.

		OSe		*SSe*		*OSμ*		*SSμ*	
		Belle II	Belle	Belle II	Belle	Belle II	Belle	Belle II	Belle
	$N_{q\bar{q}}$	14	4	30	24	27	31	58	15
	$N_{B^+B^-}$	26	22	30	39	25	61	45	23
	$N_{B^0\bar{B}^0}$	58	61	111	178	60	138	146	102
Final	N_{bkg}	97 (70)	87 (70)	171 (122)	241 (179)	111 (89)	230 (186)	249 (192)	140 (106)
	$\varepsilon_{\mathrm{sig}}^{\mathrm{tot}}$	0.075%	0.046%	0.056%	0.038%	0.060%	0.053%	0.051%	0.024%

Table 4.16: Signal MC efficiency and number of background events left in the [1.0, 2.5] GeV/c^2 ([1.3, 2.3] GeV/c^2 for values in parentheses) M_τ region in generic MC scaled to the data luminosity for Belle II and Belle.

4.4.6. Data-MC comparison

Both the cut-based selection and the BDT training are based on MC simulations. Although the final measurement relies only on data and not on generic MC, it is still important to check the data/MC agreement and ensure that the MC simulation reproduces reliably the data. Indeed, even if in this analysis, a potential yield difference between data and MC is not a problem, we still want to make sure that the data behaves as expected by the MC simulations and that no suspicious event distribution trend is observed. However, to avoid introducing any bias in the selection, a region around the expected signal location, also referred to as *blind region* is hidden and the data/MC comparison is performed in the so-called sideband region. The sidebands are defined such that $M_\tau \in [1.0, 1.65[\cup]1.9, 2.5]\mathrm{GeV}/c^2$. The central region $[1.65,\ 1.9]\mathrm{GeV}/c^2$ is therefore blinded, and corresponds to approximately 90% coverage of the signal in electron modes and 95% in muon modes.

The data/MC yields ratios are presented in Table 4.17 for each mode before and after the application of the BDT for Belle II and Belle. The BDT score distributions for data and MC are shown for the four modes in Figures 4.31, 4.32 for Belle II and Belle respectively. The distributions of the τ recoil mass in data and MC of all the modes are also shown in Figures 4.33 for Belle II and 4.34 for Belle before and after the BDT application.

For Belle II we see a data excess in the *OS* modes both before and after the BDT selection, that is also visible in the distributions of M_τ and the BDT variables. For the *SS* modes, the data/MC ratios are close to 1 both before and after the BDT selection. Similarly, we see a data excess in the *OS* modes for Belle, where the excess is more pronounced for *OSe*. The *SS* modes show data/MC ratios close to 1 before the BDT selection. After it, a data deficit is clearly visible.

The overall data/MC agreement is not perfect in terms of yields but no suspicious structure is observed neither in the BDT score and τ recoil mass distributions nor in the BDT input variables, which are available in Appendix A.1. Moreover, the generic MC information is used only to train and optimise the BDTs. The only impact of such data/MC discrepancy would therefore be a suboptimal choice for the BDT selection criterion. Otherwise, the PDFs and yields used for the upper limits estimation are extracted from the data sidebands. Hence, we make no use of the generic MC and thus do not apply any correction factor or additional systematic to take this discrepancy into account.

	OSe	*SSe*	*OSμ*	*SSμ*
Belle II before BDT	1.25±0.04	1.06±0.02	1.23±0.03	1.08±0.02
Belle II after BDT	1.29±0.13	0.97±0.09	1.19±0.12	1.02±0.08
Belle before BDT	1.27±0.03	1.06±0.01	1.12±0.02	1.03±0.01
Belle after BDT	1.53±0.18	0.85±0.07	1.10±0.08	0.70±0.09

Table 4.17: Data/MC ratio before and after the BDT application for Belle II and Belle.

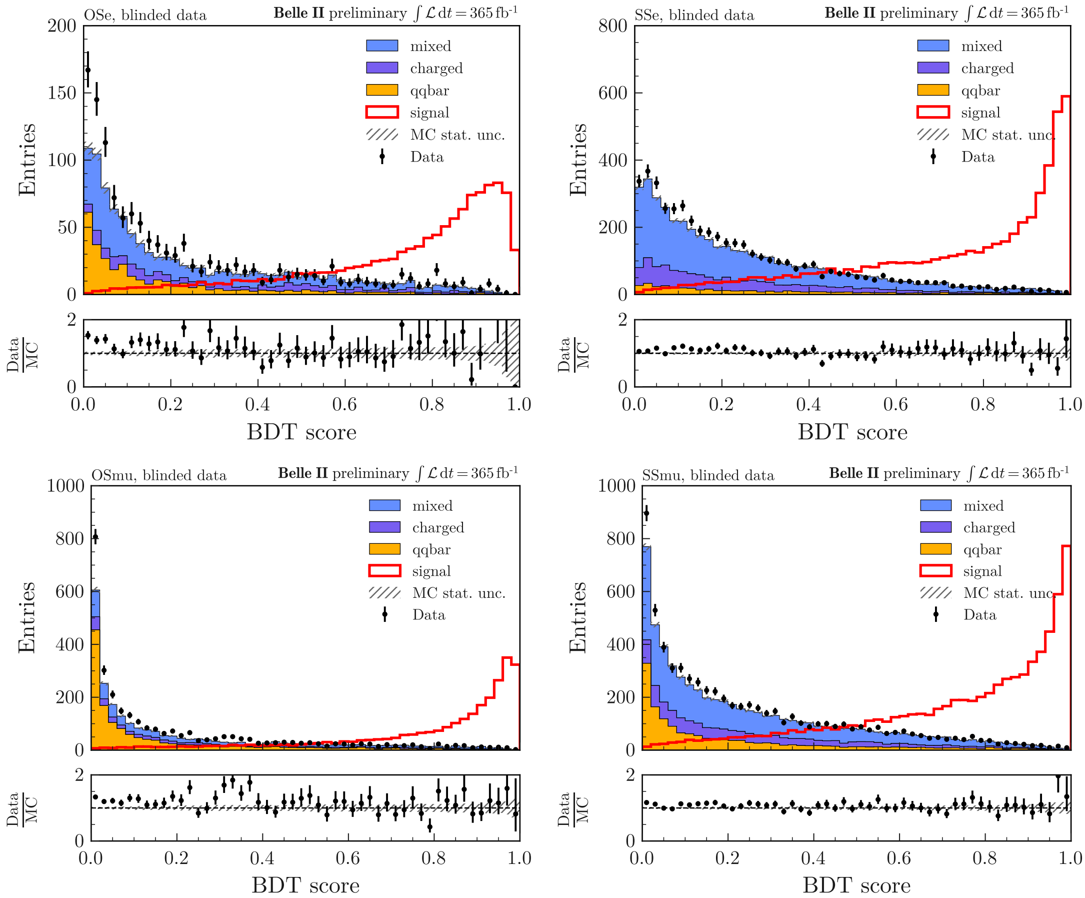


Figure 4.31: Data/MC comparison of the BDT score in M_τ sidebands for Belle II. Top left to bottom right : *OSe*, *SSe*, *OSμ*, *SSμ*.

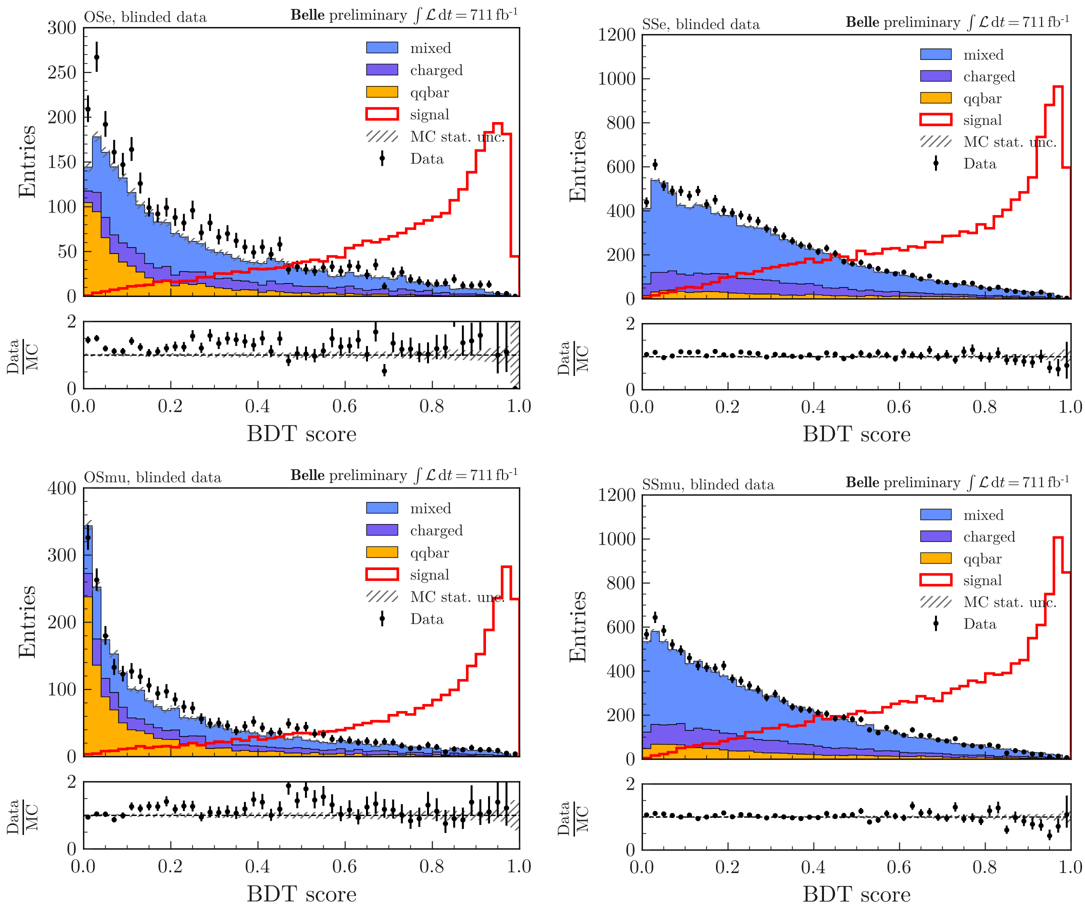


Figure 4.32: Data/MC comparison of the BDT score in M_τ sidebands for Belle. Top left to bottom right : OSe, SSe, $OS\mu$, $SS\mu$.

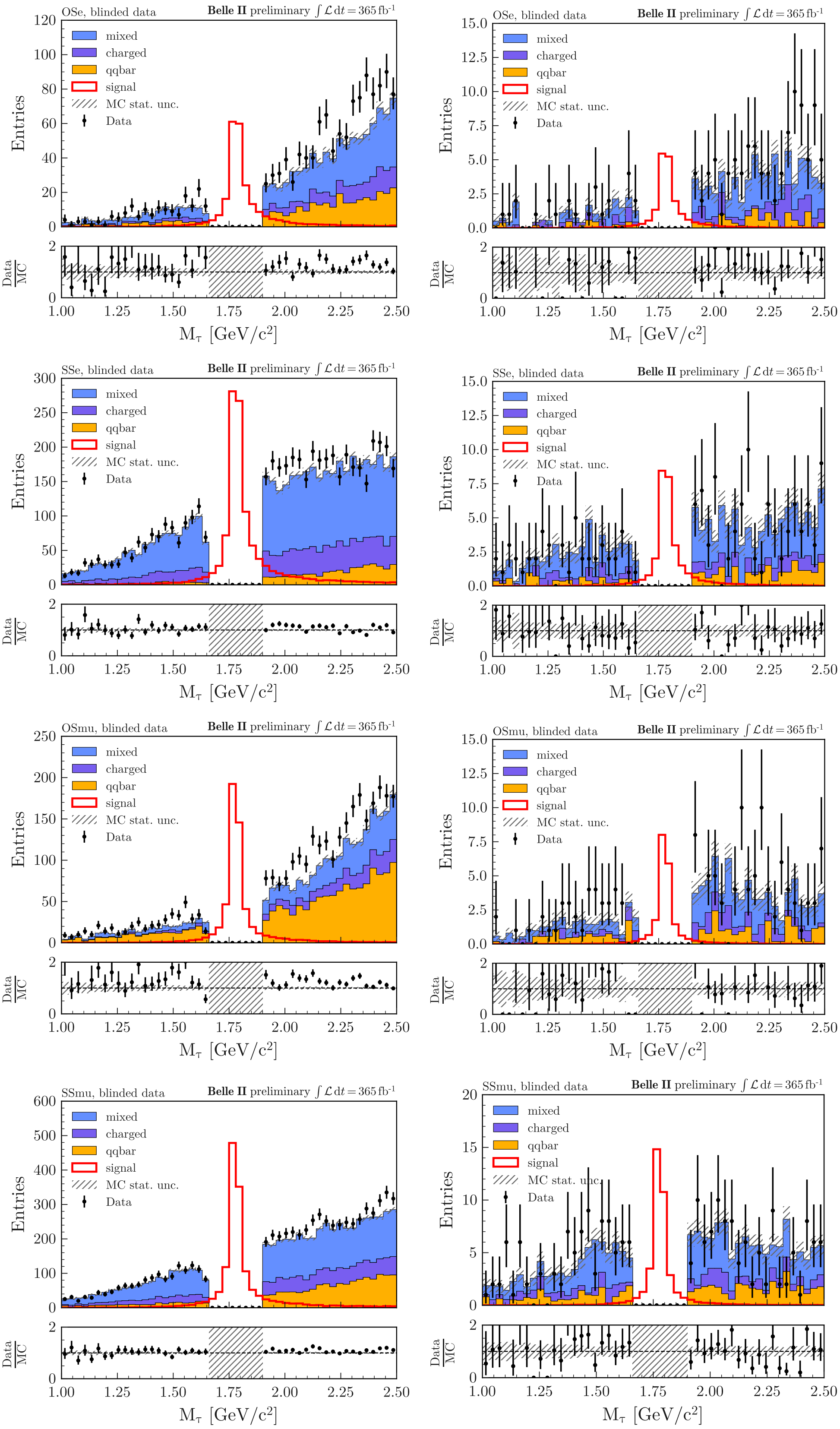


Figure 4.33: Data/MC comparison of the τ recoil mass before (left) and after (right) the BDT selection in M_τ sidebands for Belle II. Top to bottom: *OSe*, *SSe*, *OS*μ, *SS*μ.

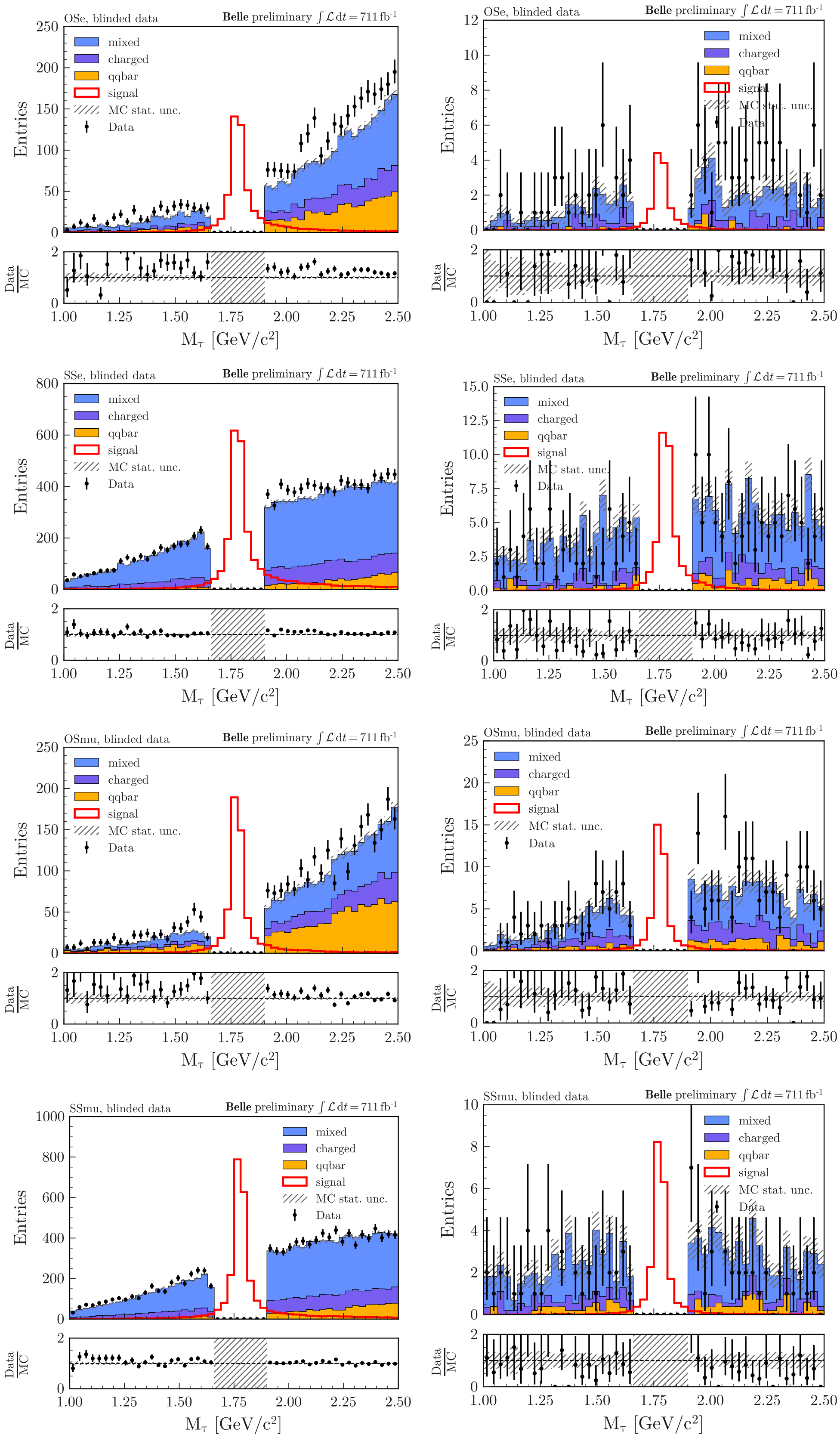


Figure 4.34: Data/MC comparison of the τ recoil mass before (left) and after (right) the BDT selection in M_τ sidebands for Belle. Top to bottom: *OSe*, *SSe*, *OSμ*, *SSμ*.

4.5. Signal extraction

The events remaining in the signal region after the event selection are fitted simultaneously in the Belle II and Belle datasets to extract the signal yield and the branching fraction. The fitting region is chosen to be $M_\tau \in [1.3, 2.3]\mathrm{GeV}/c^2$. Indeed, the M_τ distribution showing low yields at the edges of the $[1.0, 2.5]\mathrm{GeV}/c^2$ region, in particular at low mass, the range is reduced for fitting in order to avoid having negative PDFs at the edges of the range. The Belle II and Belle data are fitted with the sum of a signal component and a background component.

4.5.1. Signal modelling

The Belle II and Belle signal MC are considered separately.

The signal PDFs are obtained from signal MC and the best shape is found to be a Johnson PDF (4.4) [32]:

$$\mathscr{P}_{\mathrm{Johnson}}(M_\tau,\mu,\lambda,\gamma,\delta) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+(\frac{M_\tau-\mu}{\lambda})^2}} \exp\left[-\frac{1}{2}\left(\gamma+\delta\sinh^{-1}\left(\frac{M_\tau-\mu}{\lambda}\right)\right)^2\right], \tag{4.4}$$

where μ represents the mean of the Gaussian component, λ its width, γ the distortion of the distribution to the left/right, and δ the strength of the Gaussian-like component.

This choice is made over more common PDFs such as CrystalBall and Gaussian functions because it allows a better description of the tails and a better stability in the final fit.

The fits to the signal MC can be seen in Figure 4.35 for the *OSe* mode.

4.5.2. Background modelling

The background shape is fitted with a 2nd order Chebychev polynomial (4.5). Order 1 and 3 Chebychev polynomials as well as other types of polynomials were also considered, but fits were more stable with order 2 Chebychev polynomials:

$$\mathscr{P}_{\mathrm{Chebychev}}(M_\tau,c_1,c_2) = T_0(M_\tau) + c_1 T_1(M_\tau) + c_2 T_2(M_\tau), \tag{4.5}$$

$$T_0(M_\tau) = 1\,; \quad T_1(M_\tau) = M_\tau\,; \quad T_2(M_\tau) = 2M_\tau T_1(M_\tau) - T_0(M_\tau) = 2M_\tau^2 - 1.$$

The choice of the background PDF is first made on generic MC, as illustrated in Figure 4.36 for the *OSe* mode. Then, once the PDF is chosen, the background shape is refitted on the data sidebands $M_\tau \in [1.3, 1.65[\cup]1.9, 2.3]$ with free polynomial coefficients to estimate properly the shape and number of background events in the fitting region. This PDF and yield are later used to generate a toy dataset for the expected upper limit estimation. The data sidebands fits are shown in Figure 4.37 for Belle II and Belle for the four mode.

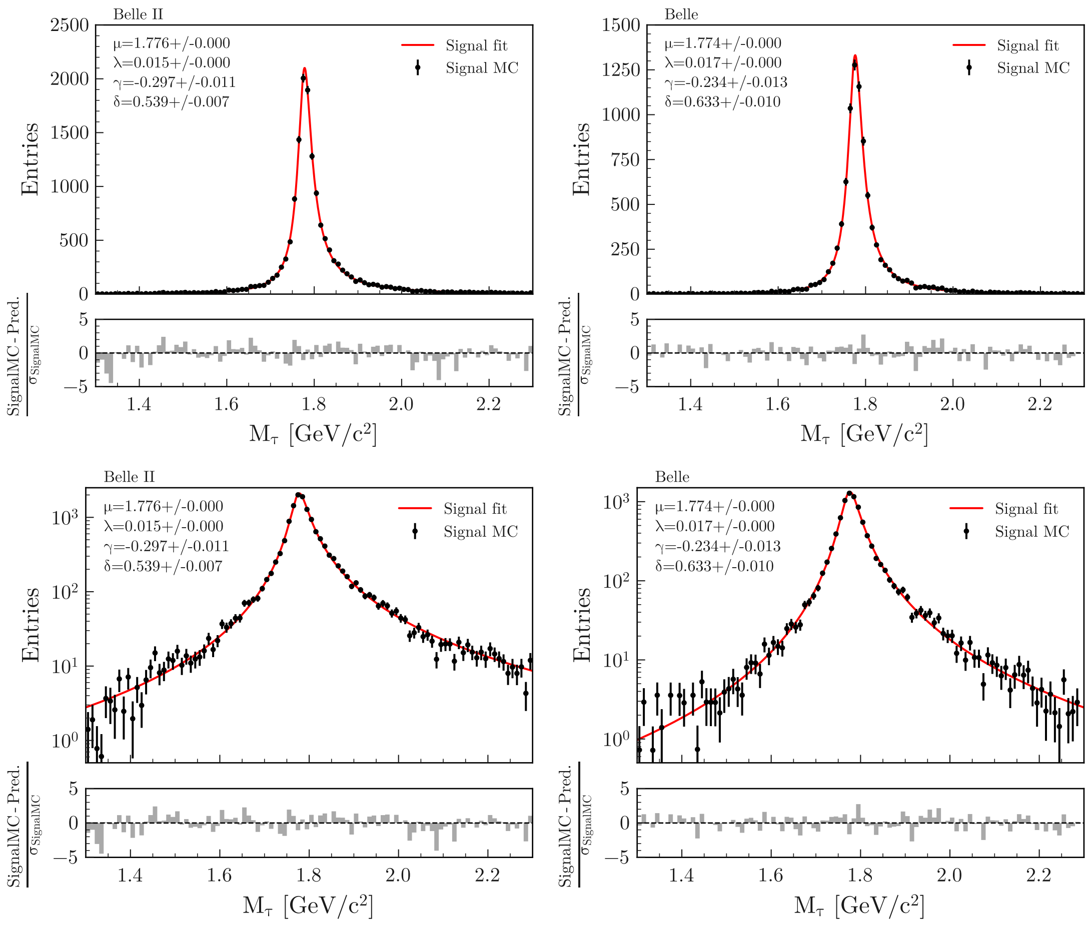


Figure 4.35: Signal fit for the *OSe* mode in linear (top) and logarithmic (bottom) scale for Belle II (left) and Belle (right) signal MC.

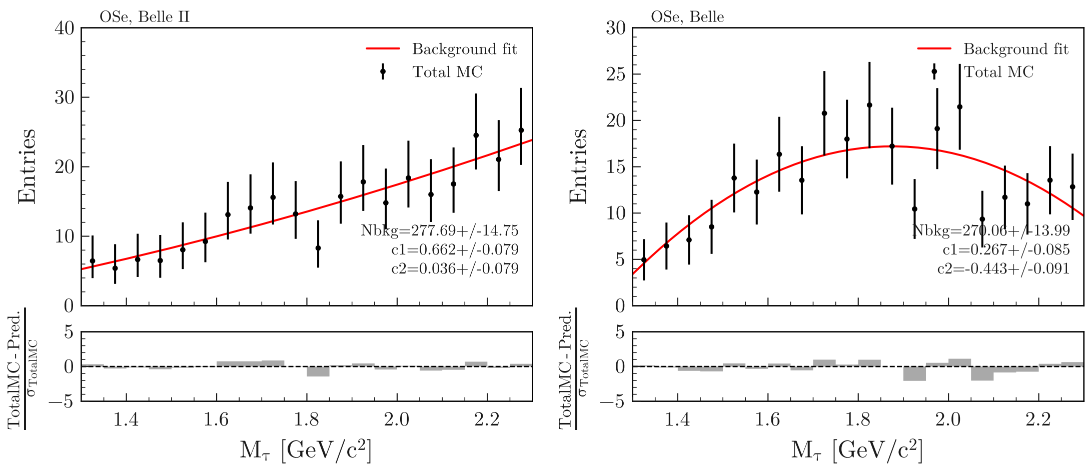


Figure 4.36: Background fit for the *OSe* mode in Belle II (left) and Belle (right) generic MC.

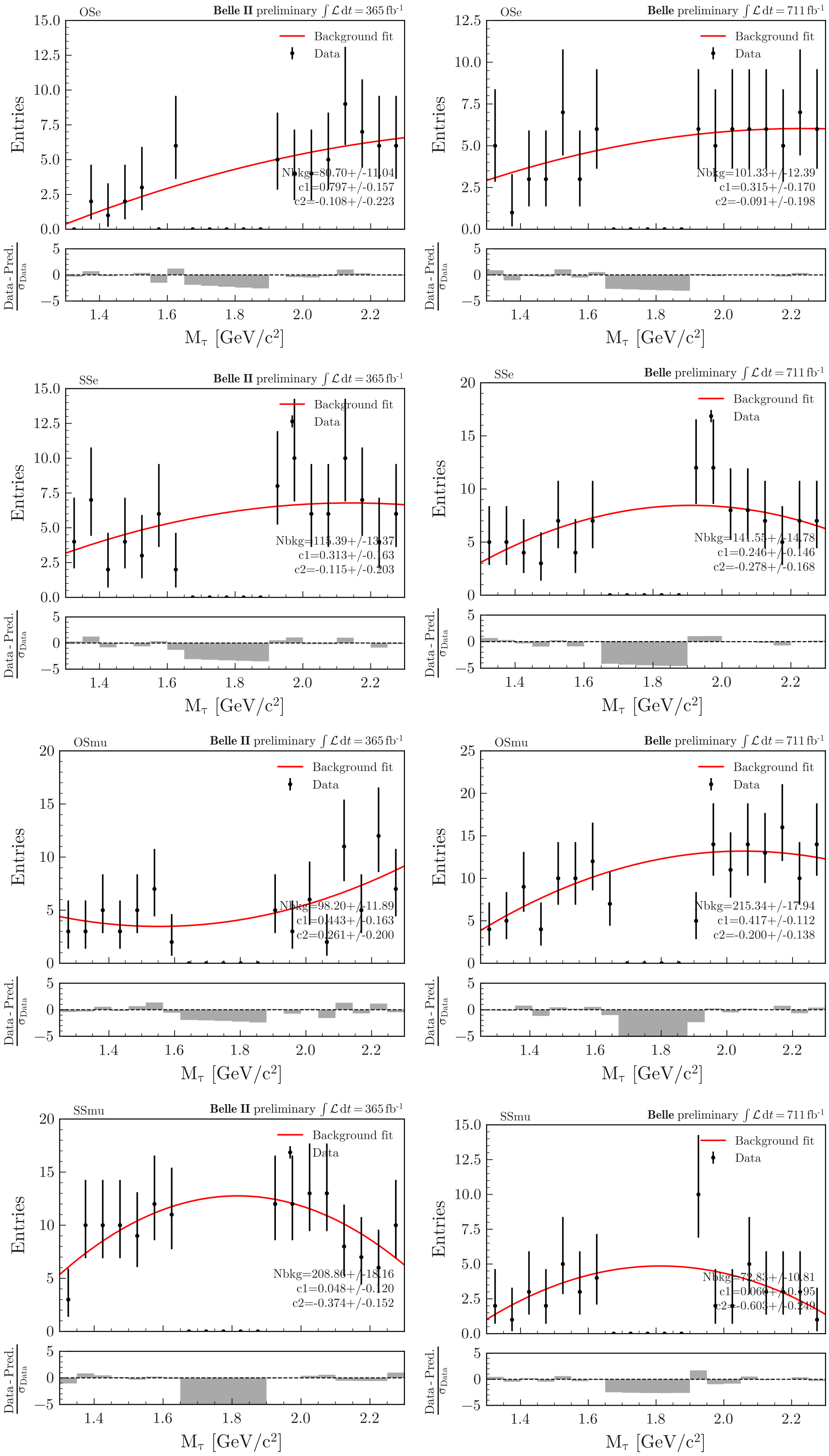


Figure 4.37: Background fit of the sidebands data for Belle II (left) and Belle (right). From top to bottom : OSe, SSe, $OS\mu$, $SS\mu$.

4.5.3. Total fit

To perform the total fit, the Belle II and Belle datasets are fitted simultaneously with an unbinned maximum likelihood fit. Each of the Belle II and Belle total PDFs is built from the sum of their corresponding signal and background components with free yields (4.6). The background PDFs coefficients are also free in the total fit, while the signal parameters are fixed to the values obtained in the signal MC fits:

$$\mathscr{P}_{\text{total}}(M_\tau) = N_{\text{sig}} \cdot \mathscr{P}_{\text{Johnson}}(M_\tau, \mu, \lambda, \gamma, \delta) + N_{\text{bkg}} \cdot \mathscr{P}_{\text{Chebychev}}(M_\tau, c_1, c_2), \tag{4.6}$$

where N_{sig}, N_{bkg}, c_1 and c_2 are free in the final fit while the other parameters are fixed.

For each of the Belle II and Belle total PDFs, the N_{sig} parameter is expressed as a function of the corresponding Belle II or Belle efficiency $\varepsilon_{\text{sig}}^{\text{exp}}$ and a free common branching fraction parameter $Br = \mathscr{B}(B^0 \to K^{*0}\tau^\pm\ell^\mp)$ (4.7):

$$N_{\text{sig}}^{\text{exp}} = Br \cdot \varepsilon_{\text{sig}}^{\text{exp}} \cdot 2 \cdot f_{00} \cdot N_{\Upsilon(4S)}^{\text{exp}} \cdot \mathscr{B}(K^{*0} \to K^+\pi^-), \tag{4.7}$$

where exp = Belle II, Belle, and all parameters except Br are fixed. The other parameters are described in Eq. (4.2).

At this stage, the total fit is performed on Belle II and Belle pseudo-experiments (also referred to as toy datasets) generated from the background PDFs fitted in data sidebands, under the assumption that no signal is present in data. This allows to validate the fit procedure as well as estimate the expected upper limits under the hypothesis of no signal. An example of the fits on toy datasets are shown for the four modes in Figure 4.38, and the fitted branching fractions are shown in Table 4.18.

	OSe	SSe	$OS\mu$	$SS\mu$
Br ($\times 10^{-5}$)	-1.45 ± 1.24	2.77 ± 2.51	0.89 ± 1.81	-0.23 ± 2.81

Table 4.18: Branching fraction of $B^0 \to K^{*0}\tau^\pm\ell^\mp$ obtained from τ recoil mass fit in pseudo-experiments generated from the data sidebands PDF.

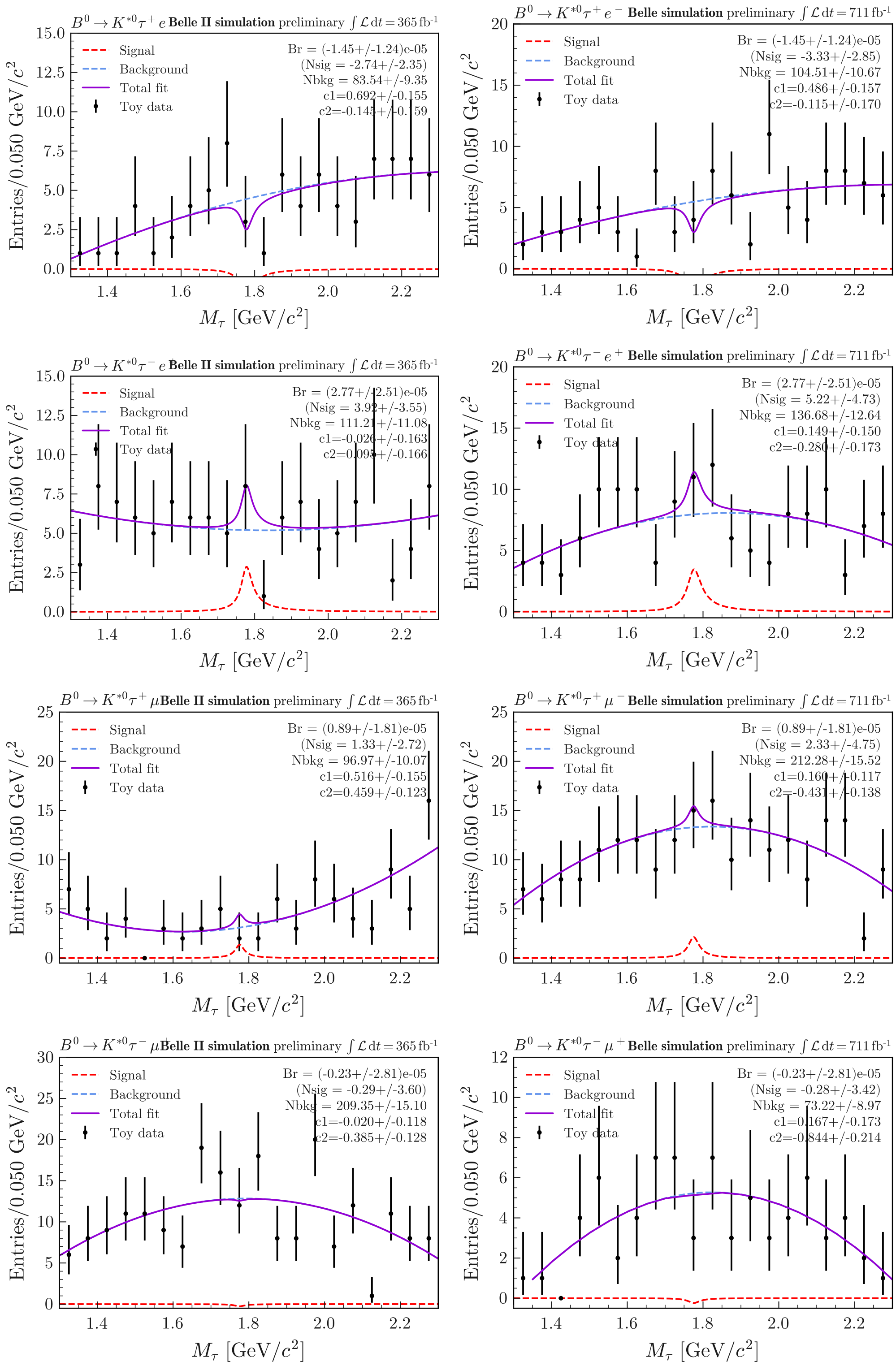


Figure 4.38: Total Belle II (left) and Belle (right) simultaneous fits of the τ recoil mass in toy datasets generated from data sideband. From top to bottom: OSe, SSe, $OS\mu$, $SS\mu$.

4.5.4. Fit validation

To validate the fits, 1000 pairs of Belle II and Belle toy datasets are generated for input branching fractions varying from 0 to 10×10^{-5} and fitted using the procedure described in the previous section.

Linearity tests presented in Figure 4.39 are performed to validate the fitting procedure. For each input branching fraction point, the average fitted branching fraction is measured. If the fitting procedure is unbiased, the average fitted branching fraction should be equal to the input branching fraction, resulting in a linear relation with slope 1 and intercept 0 between the input and fitted branching fractions. Here, the slopes are all compatible with 1 and the intercept is slightly incompatible with 0. However, the difference to 0 being one order of magnitude smaller than the statistical uncertainty on the branching fraction, we consider this light underestimation of the branching fraction to be absorbed in the statistical uncertainty.

In addition, we present the pull distributions (4.8) of the branching fraction and the background yields in Figure 4.40 with an input branching fraction of 0, as is expected in data. The pull distributions are defined as:

$$\mathrm{Pull}(X) = \frac{X_{\mathrm{fit}} - X_{\mathrm{expected}}}{\sigma(X_{\mathrm{fit}})}, \tag{4.8}$$

where for a variable X, X_{fit} is the value of the variable extracted from the fit, X_{expected} is its expected value, and $\sigma(X_{\mathrm{fit}})$ is the statistical uncertainty on the fit value. A non biased fitting procedure should result in pull distributions corresponding to standard Gaussian distributions. The pulls are indeed compatible with a mean of 0 and a width of 1. The pull distributions for higher input signal branching fractions are also Gaussian and no bias is observed. They are presented for an input branching fraction of 4×10^{-5} in Figure 4.41.

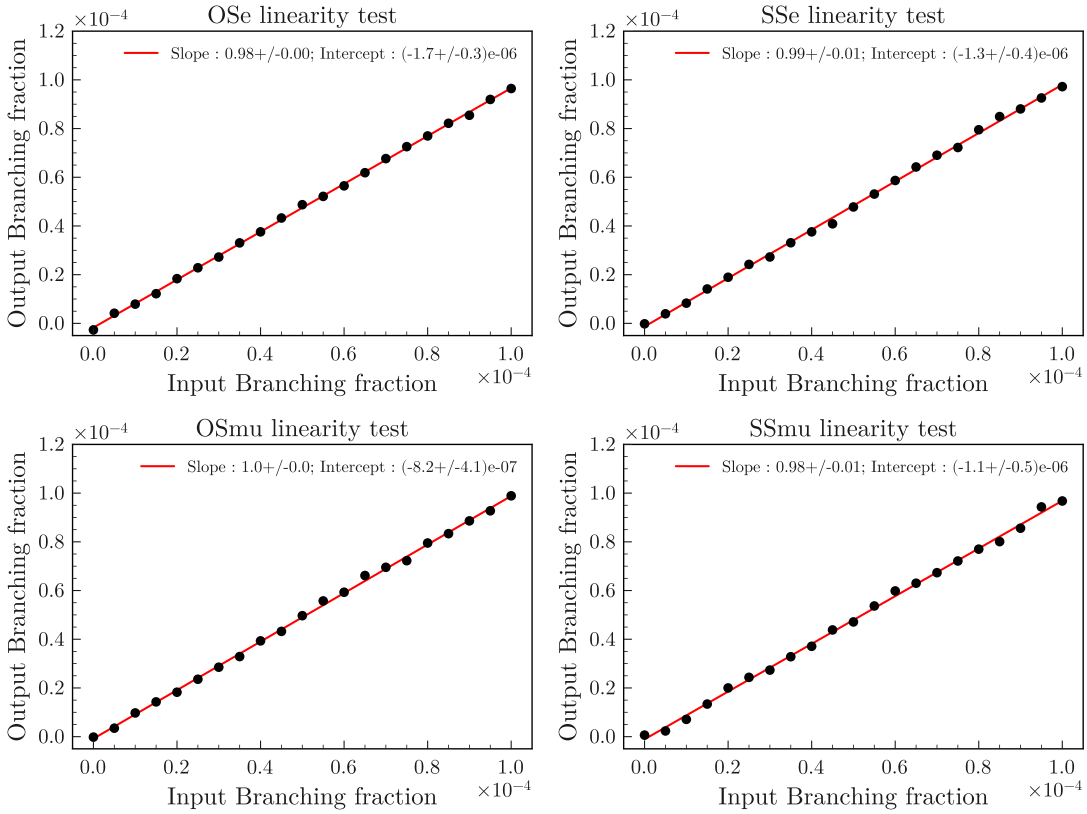


Figure 4.39: Linearity tests of measured branching fraction as function of the input one. From top left to bottom right : *OSe*, *SSe*, *OSμ*, *SSμ*.

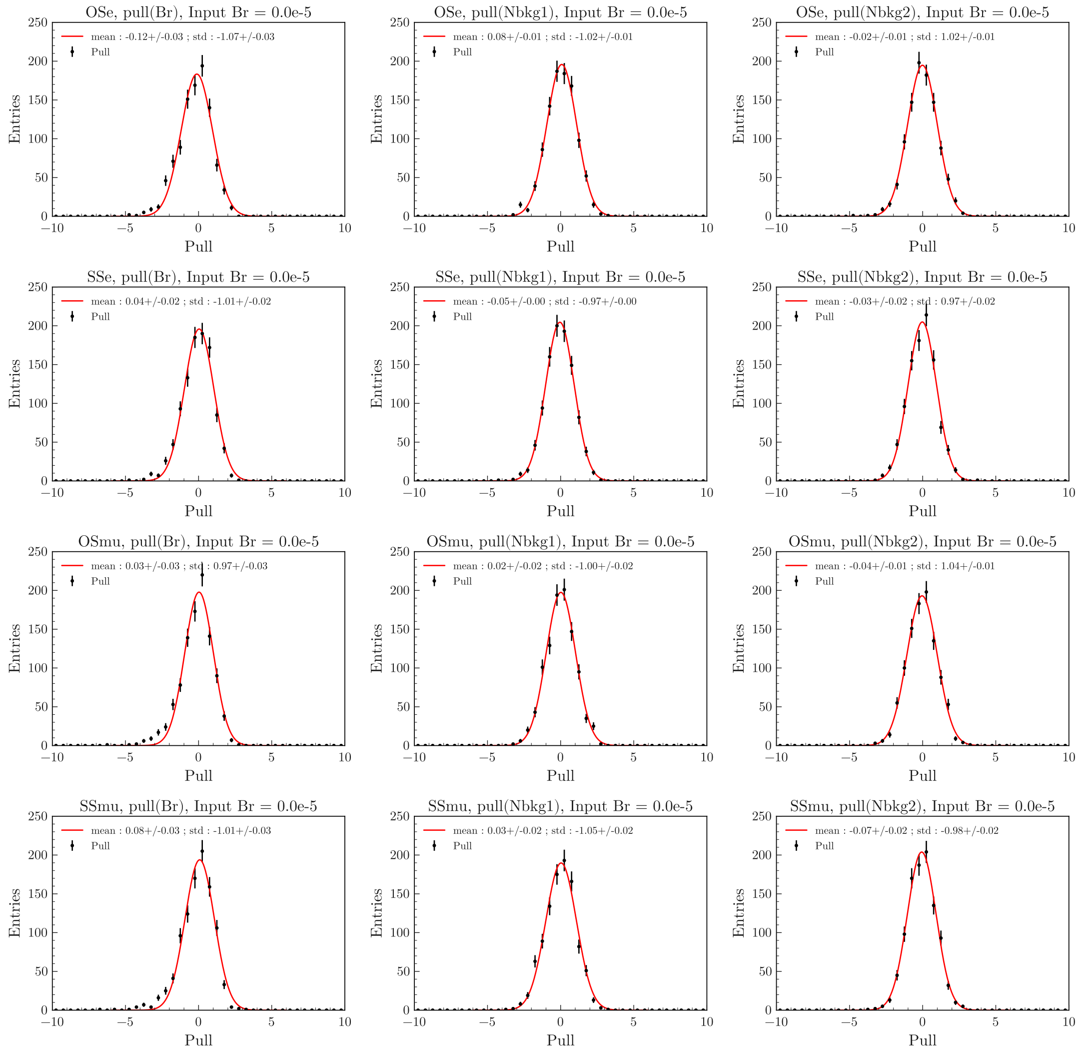


Figure 4.40: Pull distributions for the signal branching fraction (left) and background yields (middle for Belle II, right for Belle) of the fit for the 0 input signal point. From top to bottom : *OSe*, *SSe*, *OSμ*, *SSμ*.

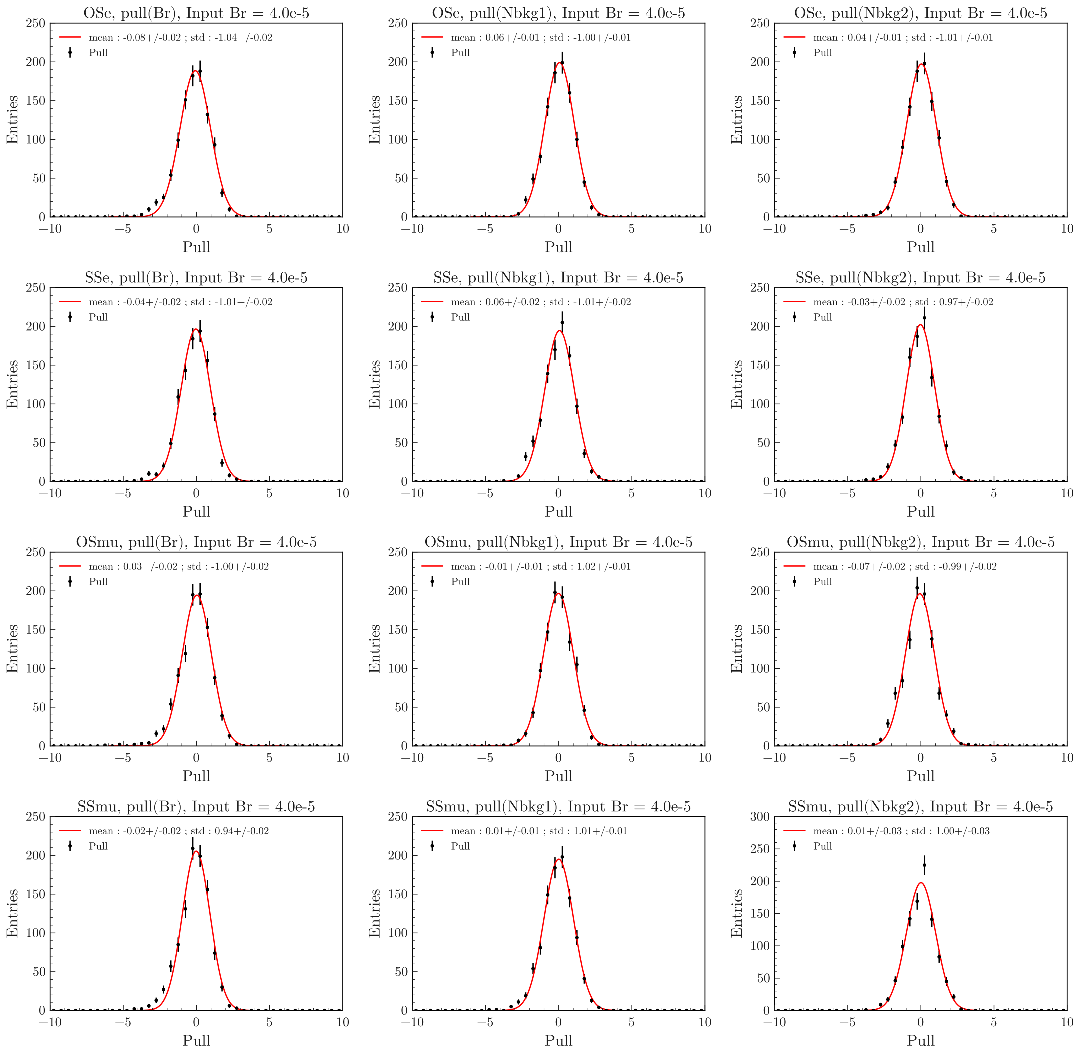


Figure 4.41: Pull distributions for the signal branching fraction (left) and background yields (middle for Belle II, right for Belle) of the fit for an input branching fraction of 4×10^{-5}. From top to bottom : OSe, SSe, $OS\mu$, $SS\mu$.

4.6. Control channels

Some steps of the analysis, such as the determination of signal efficiency, fully rely on MC simulations. To make sure that we are not introducing biases in these steps, it is common practice to exploit a control channel. The idea is to use a channel that is similar to the analysis channel, in particular for the final state, but for which a signal can be observed. Thereby, the selection or any analysis procedure can be applied to the control channel in both data and MC, and the differences can be used to assign systematic uncertainties to the analysis results.

In order to validate the BDT procedure and the signal PDF, two control channels $B^0 \to D^- D_s^+ (\to \overline{K}^{*0} K^+)$ and $B^0 \to D^- D_s^+ (\to \phi\pi^+)$ where the D decay is inclusive are used. Both channels are used together in order to increase the size of the data sample and allow a lower statistical uncertainty on the performed measurements. The D_s is either reconstructed into $\overline{K}^{*0} K^+$ or $\phi\pi^+$ to mimic the $K^{*0}\ell$ system, and one random track with the correct charge from the D is reconstructed as our track from the τ. The tracks left from the D are considered as the missing energy from the neutrino(s) and the additional τ tracks of $\tau \to 3-$prongs which are not reconstructed. However, since these tracks still end up in the Rest Of Event, its characteristics will differ between nominal and control channels. In particular, more tracks are expected in the ROE for the control channel on average, which leads to higher energy and more activity in the ECL compared to the nominal channels. A schematic view of the reconstruction is shown in Figure 4.42. The recoil mass of D can then be extracted in the same way as the recoil mass of the τ in the nominal channels.

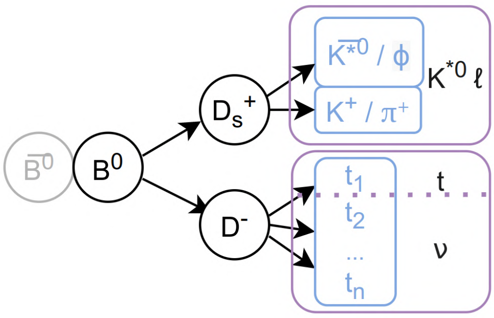


Figure 4.42: Schematic view of the $B^0 \to D^- D_s^+$ sample reconstruction in comparison to the nominal modes.

For the reconstruction, the same requirements as for the nominal channels are asked when relevant (tracking, FEI selection). The PID requirements are adapted such that pions (π^+ from $\overline{K}^{*0}$ in case of $\overline{K}^{*0}K^+$ reconstruction, pion fromD_s^+ in case of $\phi\pi^+$ reconstruction) have pionID ≥ 0.6, and kaons (K^+ from D_s^+ and K^- from $\overline{K}^{*0}$ in case of $\overline{K}^{*0}K^+$ reconstruction, kaons from ϕ in case of $\phi\pi^+$ reconstruction) have kaonID ≥ 0.6. A selection around the K^{*0} mass at $\pm$50MeV is applied in case of $\overline{K}^{*0}K^+$ reconstruction, and a selection around the ϕ mass at $\pm$20MeV is applied in $\phi\pi^+$ case. In both cases, a selection around the D_s^+ mass is applied at $\pm$20MeV. A single candidate is chosen randomly among the multiple candidates due to the different possibilities of reconstruction of a single track from the D^-. Similarly to the nominal modes, a selection is also applied on the number of tracks in the ROE such that this number is strictly below 3. Since there is no lepton in these modes, the J/ψ veto is not applied, neither is the $K\pi\pi$ veto, as there is no large component of $D_s^+ \to K\pi\pi$.

For both the validation of the BDT and the signal PDF, the distribution of the D recoil mass needs to be fitted. To do so, the two samples corresponding to the $\overline{K}^{*0}K^+$ and the $\phi\pi^+$ reconstructions are added, and the signal is fitted with a Johnson PDF (4.4) obtained from signal MC. Since the measured observable is the recoil mass extracted from the kinematics of the B_{tag} and the D_s^+ decay, a peak at the D^* mass corresponding to the $B^0 \to D^{*-}D_s^+$ decay is also visible in the recoil mass distribution. To model it correctly in generic MC and data, the D^* peak is fitted as well with a Johnson PDF obtained from corresponding signal MC, as shown in Figure 4.43. Then, the generic MC and the data are fitted with the two signal components corresponding to the D and D^* peaks and a Chebychev polynomial of order 2 for the background. The fitted MC and data for the Belle II dataset are shown in Figure 4.44. In these plots, the signal parameters and the D^* peak parameters are fixed to the ones determined in the signal MC fits shown in Figure 4.43 and the background parameters are free.

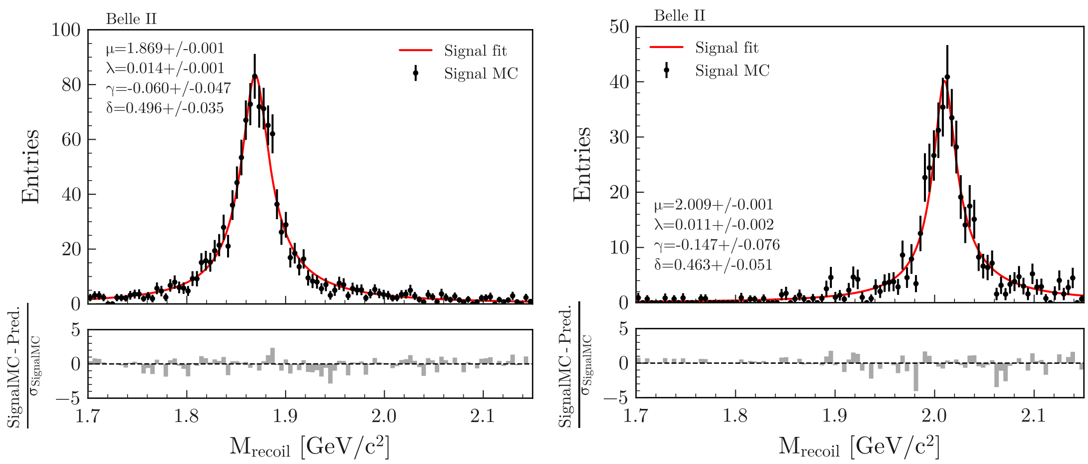


Figure 4.43: Fits of the D recoil mass (left) and of the D^* peak (right) of the control channel $B^0 \to D^-D_s^+$ in Belle II signal MC.

The validation of the BDT efficiency and the signal PDF performed with these control channels are described in the next sections on systematic uncertainty estimations, 4.7.4 and 4.7.5, respectively.

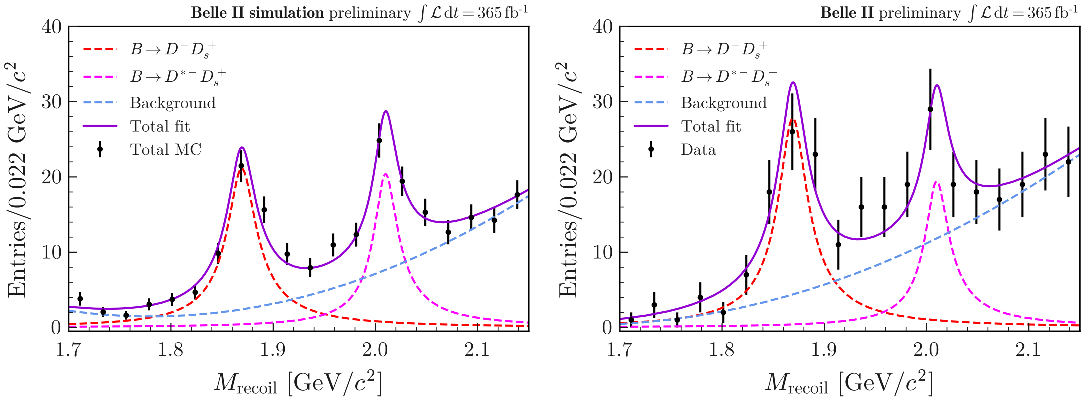


Figure 4.44: Fits of the D recoil mass of the control channel $B^0 \to D^- D_s^+$ in generic MC (left) and data (right) before the BDT selection for Belle II.

4.7. Systematic uncertainties

The systematic uncertainties are estimated separately for the Belle and Belle II samples.

For the systematic uncertainties on signal efficiency coming from data/MC corrections provided by the performance group, that's to say FEI and PID systematics (see Sections 4.7.1 and 4.7.2), we use the following procedure. For each systematic source independently, we vary the corresponding weights with the up and down errors provided. When both a systematic and a statistic error are provided, we define the total error on the weight $\sigma_{w,\text{source}}^{\uparrow,\downarrow}$ as the sum in quadrature of the systematic and statistical uncertainties:

$$\sigma_{w,\text{source}}^{\uparrow,\downarrow} = \sqrt{\left(\sigma_{w,\text{source}}^{\uparrow,\downarrow,\text{stat}}\right)^2 + \left(\sigma_{w,\text{source}}^{\uparrow,\downarrow,\text{syst}}\right)^2}.$$

Then we estimate the up and down relative efficiency variations and take as systematic uncertainty the largest error between the up and down variations:

$$\sigma_{\text{source}} = \max(\sigma_{\text{source}}^{\uparrow}, \sigma_{\text{source}}^{\downarrow}), \quad \sigma_{\text{source}}^{\uparrow,\downarrow} = \frac{|\varepsilon_{\text{sig}} - \varepsilon_{\text{sig}}^{\text{source},\uparrow,\downarrow}|}{\varepsilon_{\text{sig}}}, \tag{4.9}$$

where $\varepsilon_{\text{sig}}^{\text{source},\uparrow,\downarrow}$ is the signal efficiency measured varying the systematic source weights.

4.7.1. FEI efficiency

FEI calibration factors are measured by the Belle II performance group and provided to correct MC simulation efficiencies. For Belle II, the calibration factors are extracted mode by mode using the factors measured in the $B \to X\ell\nu$ and $B \to D\pi$ samples to which the hadronic FEI is applied [110]. Each event is weighted accordingly to the FEI mode reconstructed with the adequate calibration factor. For Belle, we use the global calibration factor measured using $B \to D\pi$ [111] as no covariance tables are provided for the application mode by mode.

To estimate the systematic uncertainty, the weights are varied up and down according to their uncertainty, and the maximum error between the nominal efficiency and the up/down one is taken as systematic uncertainty following Eq. (4.9).

The systematic uncertainties are presented for each mode, Belle and Belle II in Table 4.19.

	OSe	SSe	$OS\mu$	$SS\mu$
Belle II	6.2%	6.1%	6.1%	6.2%
Belle	4.9%	4.9%	4.9%	4.9%

Table 4.19: Systematic uncertainty assigned for each mode for Belle II and Belle to account for the B_{tag} reconstruction efficiency with the FEI.

4.7.2. PID efficiency

In the same way as for FEI, the simulations are corrected for the PID requirements efficiency. For Belle, correction tables developed in [133] using $e^+e^- \to \ell^+\ell^-(\gamma)$ and $B^0 \to J/\psi(\to \ell^+\ell^-)X$ ($\ell = e, \mu$, $X =$ anything) are used for lepton ID corrections. For hadron ID, corrections are extracted from the study of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decays developed in [134]. For Belle II, lepton ID corrections developed in [135] using various channels such as $J/\psi \to \ell^+\ell^-$ and $e^+e^- \to \ell^+\ell^-(\gamma)$ are provided to calibrate the simulation. For hadron ID, the Systematic Framework [136] is used to build the efficiency correction tables, thanks to the studies in [137] on two-body decays of D^0, K^0 and Λ to charged particles.

Similarly to the FEI systematics, the weights are varied up and down according to their uncertainty and the maximum efficiency variation is taken as systematic.

The systematic uncertainties are presented for each mode, Belle and Belle II in Table 4.20 for lepton identification and hadron identification.

		OSe	SSe	$OS\mu$	$SS\mu$
Lepton ID	Belle II	0.7%	1.1%	0.7%	0.6%
	Belle	2.0%	2.4%	2.2%	2.2%
Hadron ID	Belle II	3.7%	3.7%	3.6%	3.7%
	Belle	1.9%	2.0%	1.9%	2.0%

Table 4.20: Systematic uncertainty assigned for each mode for Belle II and Belle to account for the particle identification efficiency.

4.7.3. Tracking efficiency

For Belle II, tracking efficiency studies performed by the tracking group defined a 0.27% systematic uncertainty per track to be applied to MC [123]. For Belle, tracking efficiency studies determined a 0.34% systematic uncertainty per track [138]. The number of reconstructed tracks in the signal side is 4, we thus apply a systematic uncertainty of 1.1% to each mode for Belle II, and 1.4% for Belle to account for the tracking efficiency.

4.7.4. BDT selection efficiency

The BDT procedure is validated using $B^0 \to D^- D_s^+ (\to \overline{K}^{*0} K^+ / \phi\pi^+)$ described in Section 4.6 as a control channel.

For each LFV channel, the corresponding BDT is applied to generic MC reconstructed as the control channel, and the selection on the BDT score applied on the control sample ("Control cut" in Table 4.21) is set such that its efficiency ("Control ε_{BDT}" in Table 4.21) is the same as the efficiency of the applied BDT selection on the nominal channel ("Nominal ε_{BDT}" in Table 4.21). The BDT is also applied on data and the events are selected according to the newly set cut on the BDT score. The distribution of each BDT score (corresponding to each nominal channel BDT) applied to the control channel is shown in Figure 4.45 in the D peak region [1.815, 1.930] GeV/c^2 for Belle II (4.46 for Belle) together with the nominal and control BDT selections.

To estimate the systematic uncertainty on the BDT selection efficiency, the data and MC efficiencies ratio $\mathscr{R}^{\text{data/MC}}_{\varepsilon_{\text{BDT}}} = \varepsilon^{\text{Control, data}}_{\text{BDT}} / \varepsilon^{\text{Control, MC}}_{\text{BDT}}$ is measured, and the assigned uncertainty is symmetrised such that it covers 68.3% of the area of a Gaussian with mean $1 - \mathscr{R}^{\text{data/MC}}_{\varepsilon_{\text{BDT}}}$ and standard deviation the statistical uncertainty on the ratio, according to [139]. To measure efficiencies in generic MC and data, fitted signal yields of the D recoil mass before and after the BDT selection are used.

The cut values, the corresponding efficiencies in MC and data and the data/MC ratios are shown in Tables 4.21 for Belle II and in Table 4.22 for Belle. The fits of generic MC and data after the BDT selection are shown in Figures 4.47 for Belle II and 4.48 for Belle for the OSe mode, as example.

The resulting assigned systematic uncertainties are 29% (OSe), 31% (SSe), 34% ($OS\mu$), and 33% ($SS\mu$) for Belle II, and 27% (OSe), 21% (SSe), 18% ($OS\mu$), 23% ($SS\mu$) for Belle.

	Nominal cut	Nominal ε_{BDT}	Control cut	Control ε_{BDT}	Data ε_{BDT}	$\mathscr{R}^{\text{data/MC}}_{\varepsilon_{\text{BDT}}}$
OSe	0.65	0.73±0.01	0.18	0.74±0.10	0.72±0.19	0.97±0.29
SSe	0.8	0.53±0.01	0.68	0.54±0.08	0.54±0.12	1.00±0.31
$OS\mu$	0.85	0.64±0.01	0.14	0.65±0.09	0.70±0.18	1.07±0.33
$SS\mu$	0.8	0.52±0.01	0.58	0.52±0.08	0.54±0.15	1.03±0.33

Table 4.21: BDT selection cuts, corresponding efficiencies and data/MC ratio for Belle II.

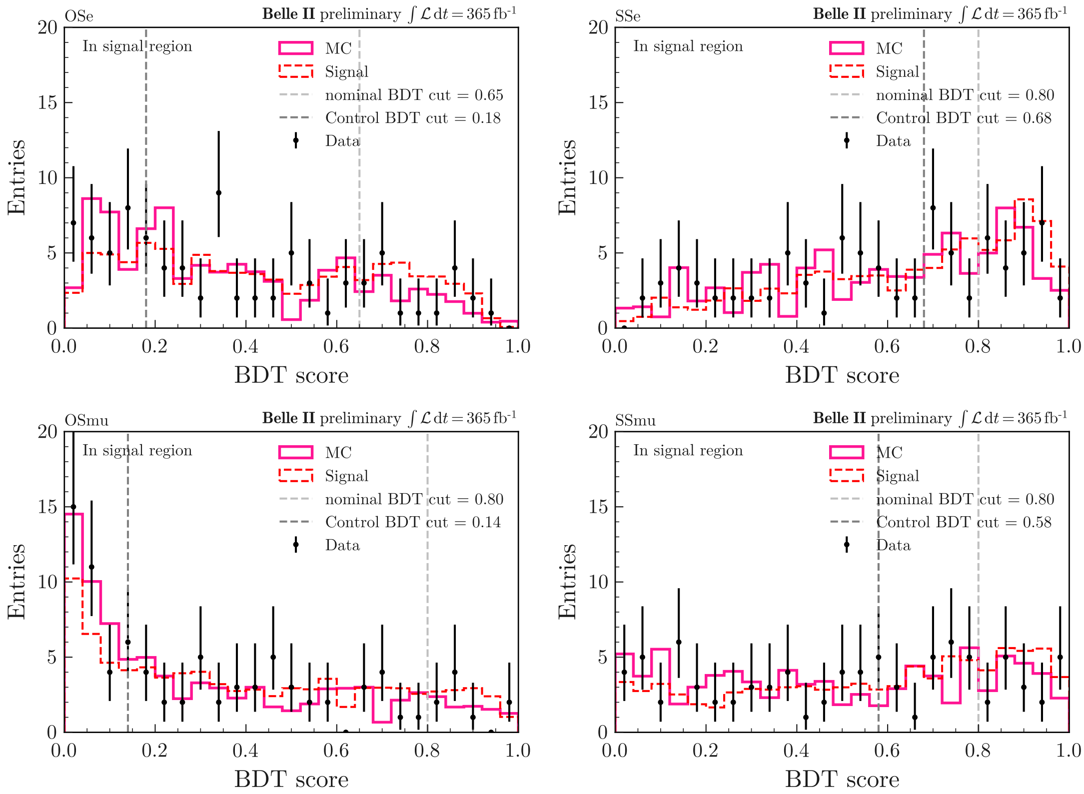


Figure 4.45: Distribution of the BDT score in MC and data in the D peak region $[1.815, 1.930]\mathrm{GeV/c}^2$ for the control channel in the four modes in Belle II (from top left to bottom right : OSe, SSe, $OS\mu$, $SS\mu$). The generic and signal MC are normalised to the same area as data for shape visualisation. The signal MC shown is $B^0 \to D^- D_s^+ (\to K^{*0}K/\phi\pi)$ signal MC.

	Nominal cut	Nominal $\varepsilon_{\mathrm{BDT}}$	Control cut	Control $\varepsilon_{\mathrm{BDT}}$	Data $\varepsilon_{\mathrm{BDT}}$	$\mathscr{R}^{\mathrm{data/MC}}_{\varepsilon_{\mathrm{BDT}}}$
OSe	0.8	0.49±0.01	0.6	0.49±0.05	0.39±0.06	0.80±0.15
SSe	0.85	0.39±0.05	0.84	0.40±0.04	0.36±0.06	0.89±0.17
$OS\mu$	0.65	0.73±0.01	0.28	0.74±0.07	0.69±0.10	0.93±0.16
$SS\mu$	0.9	0.30±0.01	0.86	0.29±0.03	0.25±0.05	0.87±0.18

Table 4.22: BDT selection cuts, corresponding efficiencies and data/MC ratio for Belle.

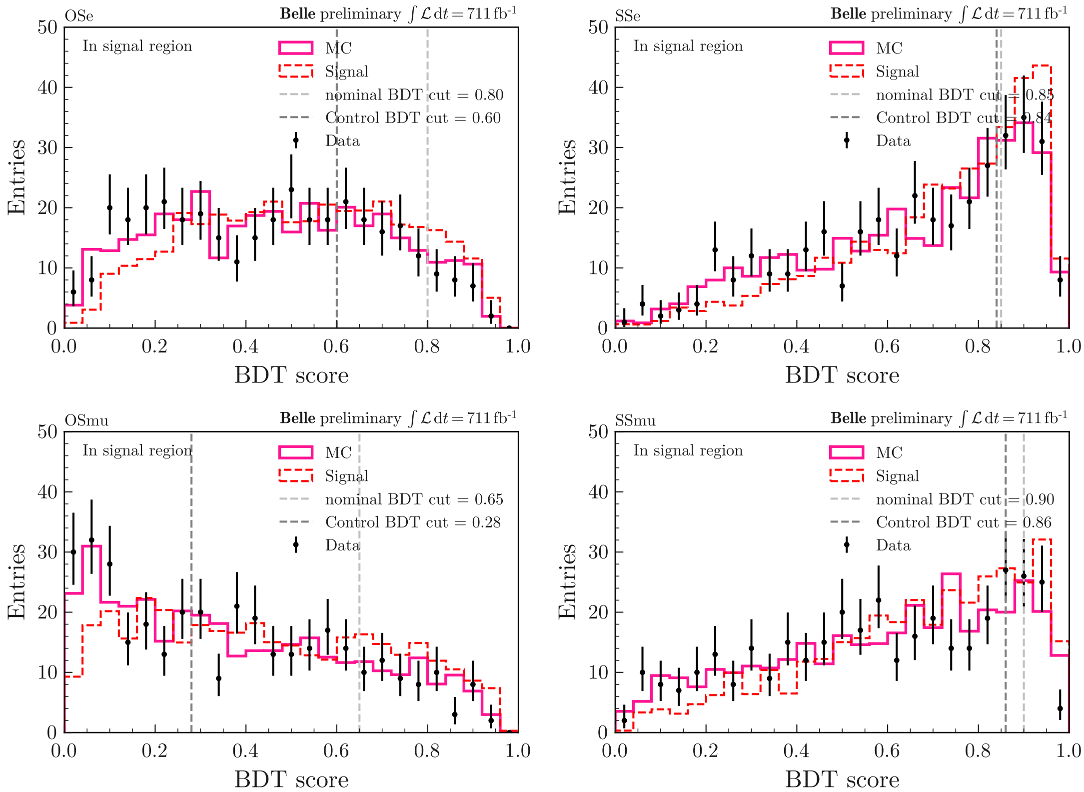


Figure 4.46: Distribution of the BDT score in MC and data in the D peak region [1.815, 1.930]GeV/c^2 for the control channel in the four modes in Belle (from top left to bottom right : OSe, SSe, $OS\mu$, $SS\mu$). The generic and signal MC are normalised to the same area as data for shape visualisation. The signal MC shown is $B^0 \to D^- D_s^+ (\to K^{*0}K/\phi\pi)$ signal MC.

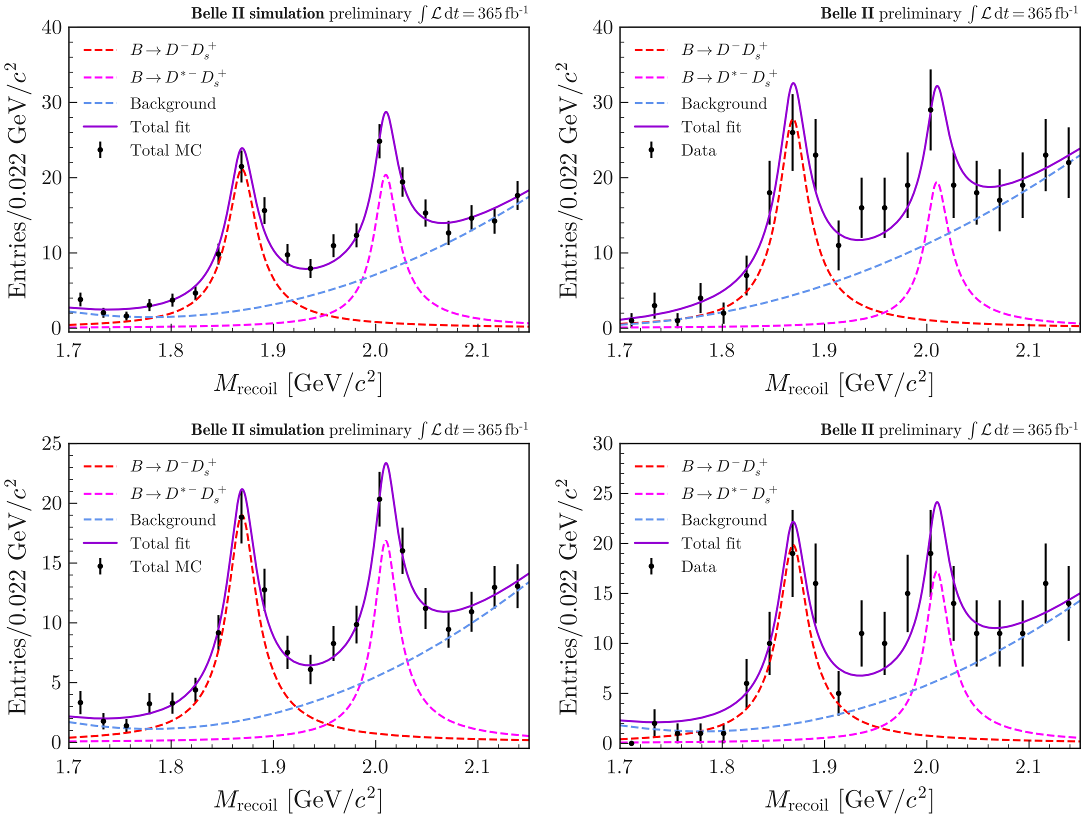


Figure 4.47: Fits to the control channel $B^0 \to D^- D_s^+$ in MC (left) and data (right) before (top) and after (bottom) the BDT selection for the *OSe* mode in Belle II.

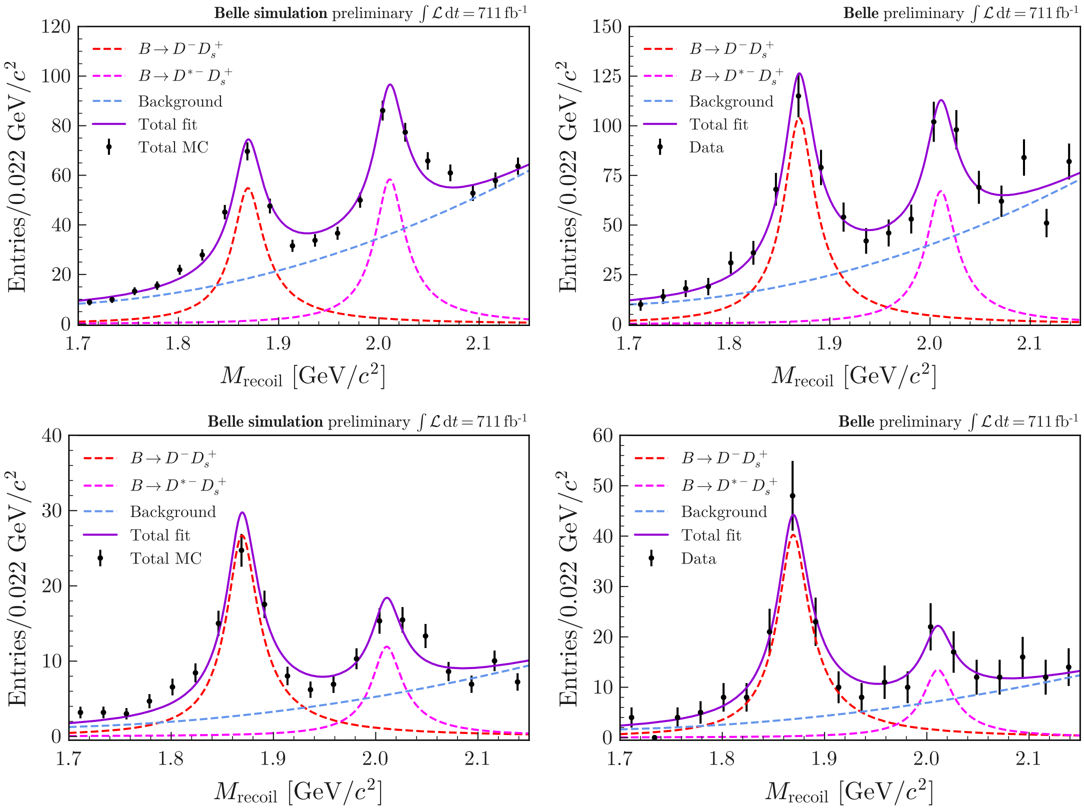


Figure 4.48: Fits to the control channel $B^0 \to D^- D_s^+$ in MC (left) and data (right) before (top) and after (bottom) the BDT selection for the *OSe* mode in Belle.

The BDT score distributions shown in Figures 4.45 and 4.46 are not necessarily peaking at 1, as expected for signal, in particular for the *OS* modes. This is mainly due to the fact that the dominant variables in the *OS* BDTs are related to the $K^{*0}\ell$ system, which is the D_s in the control channel. The variables distributions are thus very different from the nominal channel. Also in both *OS* and *SS* modes, the ROE variables are playing an important role and have different distributions between nominal and control channels as the additional tracks of the D decay are counted in the ROE. In addition, most of the variables have a very similar distribution between signal and background in the control channel, which can explain a worse separation of the two. The BDT input variables distributions for the control channel, as well as comparisons of the nominal and control channel signal distributions are available in Appendix A.2.

4.7.5. Signal PDF validation

The signal PDF is validated using the same control channels $B^0 \to D^- D_s^+ (\to \overline{K}^{*0} K^+ / \phi\pi^+)$, first fitting the signal MC in the same way as for the nominal channels as illustrated in Figure 4.49.

For the parameter corresponding to the mean, we define $\mu' = \mu + \delta m$ where μ is

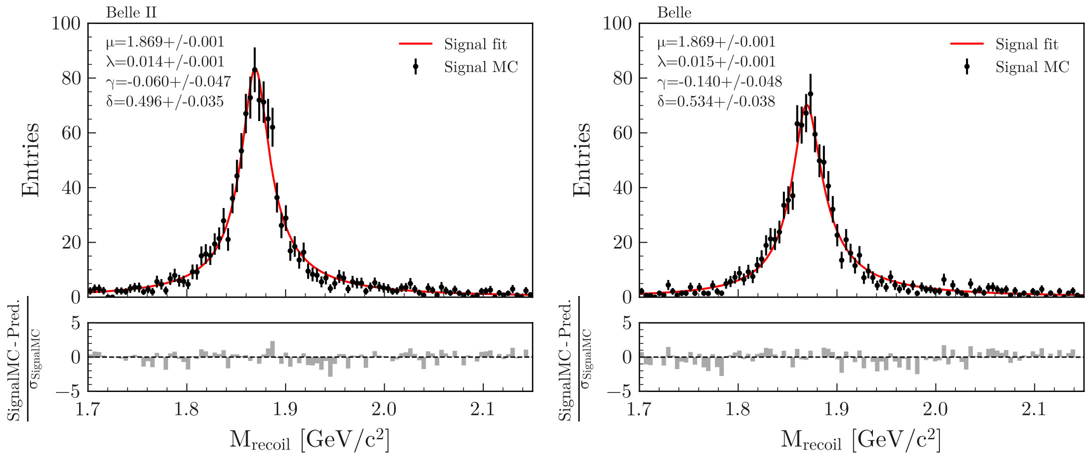


Figure 4.49: Fit to the signal MC D recoil mass for the control channel $B^0 \to D^- D_s^+ (\to \overline{K}^{*0} K^+ / \phi\pi^+)$ for Belle II (left) and Belle (right).

fixed to the value fitted in signal MC and δm is free to vary in the fit to data. Then, we fit the data with μ' and we take as systematic uncertainty on the mean the largest value between the difference of δm and its error. For the parameter λ, corresponding to the width of the signal distribution, we define $\lambda' = \lambda \cdot f$ where λ is fixed to the value fitted in signal MC and f is a free fudge factor. Since the value of the resolution parameter λ is the same for the D and the D^* peaks in signal MC and as both peaks represent the same recoil mass, there is no reason for the resolution of the two peaks to be different and to behave differently in data and MC. Hence, we constrain the resolutions of the two peaks to vary identically and we fit both resolutions with λ' in data. The maximum between $1 - f$ and the error on f is taken as systematic uncertainty on this parameter. The fits to data are the same as the ones shown in the top right plots of Figures 4.47 and 4.48 for Belle II and Belle, respectively.

The tails parameters γ and δ being very sensitive to the background, estimating a systematic uncertainty in the same way would lead to very large values, in particular since we are limited by the sample size of the control channel. We do not assess any systematic on these parameters, and a dedicated systematic on the background description will be estimated.

The δm and f parameters are shown in Table 4.23. The systematic uncertainty assigned to the parameters are 0.2% for μ and 59% for λ for Belle II and 0.1% for μ and 21% for λ for Belle.

4.7.6. Background shape validation

In the analysis, the background distribution is fitted with a second order Chebychev polynomial (4.5) with free coefficients. In order to validate the choice of the PDF, we refit for each mode the background shape with an order 3 Chebychev polynomial.

To do so, we generate a Belle II and a Belle toy dataset from the fit to the data

	Parameter	δm; f
Belle II	μ [GeV/c^2]	-0.002±0.004
	λ	1.31±0.59
Belle	μ [GeV/c^2]	0.001±0.002
	λ	0.98±0.21

Table 4.23: Variation of the signal PDF parameters fitted in data.

sidebands. One could fit directly this pair of datasets with a total PDF comprising once a 2nd order background polynomial and once a 3rd order background polynomial, and measure the difference in the obtained branching fraction. However, the statistical uncertainty on the branching fraction is very large (same order as the branching fraction itself) as seen in Section 4.5.3. To make sure that the difference observed between the two fits (order 2 and order 3 background polynomials) is not due to the statistical uncertainty, we generate pseudo-experiments. Since the object we want to validate is the background PDF, it would not be relevant to generate toy datasets from a PDF fitted on data sidebands. To bypass this issue, we bootstrap the toy dataset extracted from data sidebands (after unblinding, the generation of a toy dataset from data sidebands is not needed any more and data is directly bootstrapped). Bootstrapping has the advantage of not relying on a PDF to generate the pseudo-experiments, but rather on the data itself.

500 bootstrapped pseudo-experiments are thus fitted with the total PDF, comprising the signal and background components. This step is done once with a 2nd order polynomial and once with a 3rd order polynomial for the background PDF. The fits with the two different polynomials of one Belle and Belle II set are visible in Figure 4.50 for OSe.

The difference between the average branching fraction obtained from fit with order 3 and order 2 polynomials is taken as systematic uncertainty on the branching fraction. For the expected upper limits, it corresponds to 0.1×10^{-5} for OSe, 0.1×10^{-5} for SSe, 0.1×10^{-5} for $OS\mu$ and 0.6×10^{-5} for $SS\mu$. The detailed values measured are presented in Table 4.24.

	Average $\mathscr{B}_{\text{2ndOrder}}$	Average $\mathscr{B}_{\text{3rdOrder}}$	Difference (3rd−2nd)
OSe	-1.66±0.06	-1.79±0.06	-0.13±0.08
SSe	2.62±0.11	2.66±0.11	0.05±0.16
$OS\mu$	0.74±0.08	0.71±0.08	-0.02±0.12
$SS\mu$	-0.63±0.13	-1.30±0.13	-0.64±0.18

Table 4.24: Average fitted branching fraction ($\times10^{-5}$) with a 2nd order and a 3rd order polynomial background PDF in Belle and Belle II toy datasets generated from the data sidebands.

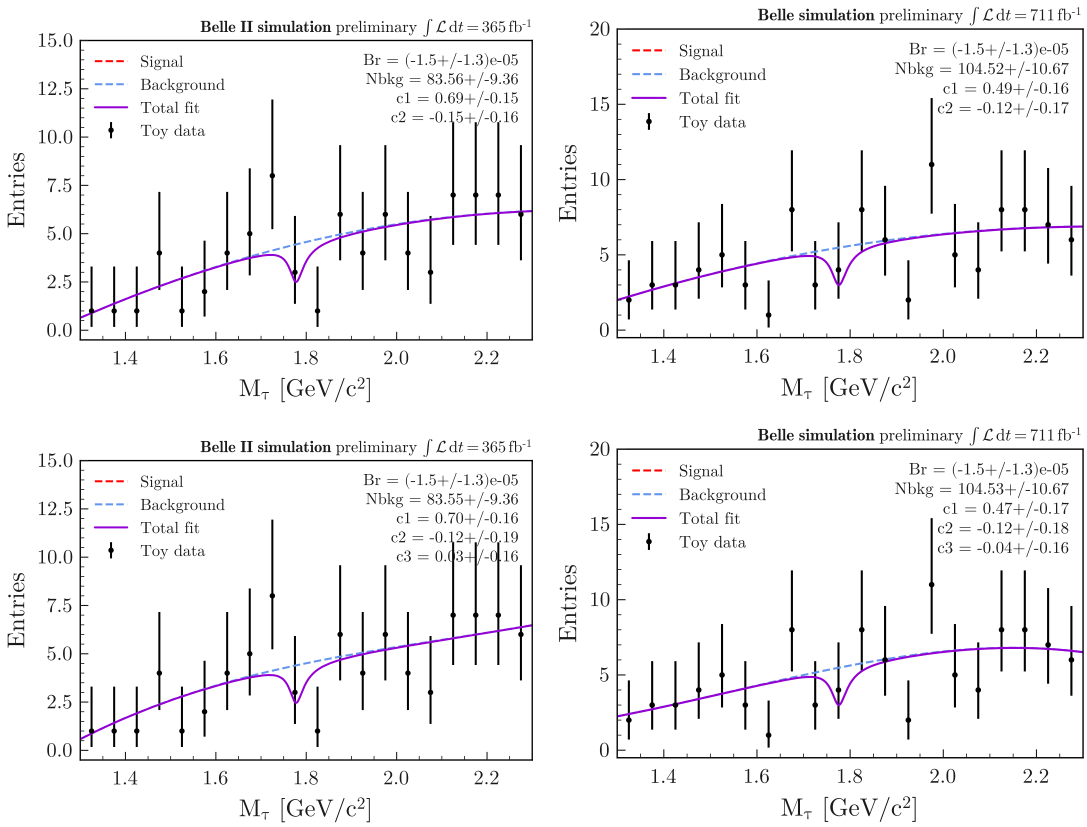


Figure 4.50: Total simultaneous fit of the τ recoil mass in OSe toy dataset generated from data sidebands with a second order polynomial (top) and third order polynomial (bottom) background PDF in Belle II (left) and Belle (right).

After unblinding, the systematic uncertainty on the background modelling is reevaluated from fits to data. The corresponding systematic uncertainties are 0.1×10^{-5} for OSe, 0.3×10^{-5} for SSe, 0.1×10^{-5} for $OS\mu$ and 0.1×10^{-5} for $SS\mu$.

4.7.7. Other sources

$\Upsilon(4S)$

For Belle, the number of $\Upsilon(4S)$ produced is $(771.58\pm10.57)\cdot10^6$ [122], while for Belle II, it is $(387\pm6)\cdot10^6$ [123]. In both cases, the uncertainty is taken as systematic uncertainty, i.e. 1.4% for Belle and 1.6% for Belle II.

f_{00}

The value of the fraction of $\Upsilon(4S)$ decaying into $B^0\overline{B}^0$ pairs f_{00} is (48.61 ± 0.80)% [15]. The uncertainty is taken as systematic uncertainty, that is 0.80%.

Intermediate branching fraction

In the computation of the upper limit on the branching fraction (4.2), the branching fraction $\mathscr{B}(K^{*0} \to K^+\pi^-) = 2/3$ in the isospin limit appears. Its uncertainty is negligible and thus not taken into account.

MC statistics

To account for the limited size of the available signal MC used to measure the efficiency, we consider the corresponding uncertainty to be $\sqrt{\varepsilon(1-\varepsilon)/N_{\text{tot}}}$ where ε is the signal efficiency and N_{tot} is the number of generated signal events. The final efficiencies being between 0.024% and 0.058%, and the number of generated signal MC events being 20 millions for each mode and experiment, the uncertainty on the signal MC statistics is below 0.0005% (< 1% relative uncertainty) in all cases and thus ignored.

4.7.8. Systematic uncertainties summary

A summary of the systematic uncertainties is presented in Table 4.25. All the systematic uncertainties applied to the efficiency are combined with a sum in quadrature and are displayed in the "Total efficiency" line. The uncertainties expressed as percentages are multiplicative and applied to the corresponding parameter via a Gaussian constraint in the fit. The background PDF uncertainties are additive and directly applied to the branching fraction and upper limit. The "Background shape" uncertainty differs between the expected upper limits (exp) and the observed ones (obs) as it was reevaluated after unblinding.

The systematic uncertainties are taken into account into the upper limit measurement by applying a Gaussian constraint to the parameters of the branching fraction. For each parameter, the Gaussian constraint has for mean the nominal value of the parameter and for standard deviation the systematic uncertainty quoted in Table 4.25. For the background shape systematic, the systematic uncertainty is added on the final upper limit as the uncertainty is directly assigned to the branching fraction. Indeed, since the branching fraction is the parameter of interest of the fit, it would not make sense to apply a Gaussian constraint on it.

The effect of each systematic uncertainty source on the fitted branching fraction and on the observed upper limit is shown in Table 4.26.

Source	Belle				Belle II			
	OSe	*SSe*	*OSμ*	*SSμ*	*OSe*	*SSe*	*OSμ*	*SSμ*
FEI efficiency [%]	4.9	4.9	4.9	4.9	6.2	6.1	6.1	6.2
Lepton ID efficiency [%]	2.0	2.4	2.2	2.2	0.7	1.1	0.7	0.6
Hadron ID efficiency [%]	1.9	2.0	1.9	2.0	3.7	3.7	3.6	3.7
BDT efficiency [%]	27	21	18	23	29	31	34	31
Tracking efficiency [%]	1.4				1.1			
Total efficiency [%]	27.6	21.8	18.9	23.7	29.8	31.8	34.7	31.7
Signal PDF μ [%]	0.1				0.2			
Signal PDF λ [%]	21				59			
$N_{\Upsilon(4S)}$ [%]	1.4				1.6			
f_{00} [%]	0.8							
Background PDF (exp) ($\times 10^{-5}$)	0.11	0.28	0.09	0.02	0.11	0.28	0.09	0.02
Background PDF (obs) ($\times 10^{-5}$)	0.13	0.05	0.02	0.64	0.13	0.05	0.02	0.64
Total impact on UL (obs) ($\times 10^{-5}$)	0.3	1.0	0.4	0.5	0.3	1.0	0.4	0.5

Table 4.25: Summary of the systematic uncertainties. The upper part of the table are systematic uncertainties applied to the efficiency, which are combined in the “Total efficiency" line.

Source	Variation of the fitted branching fraction (obs. upper limit) ($\times 10^{-5}$)			
	OSe	*SSe*	*OSμ*	*SSμ*
Total efficiency	0 (+0.1)	0 (+0.3)	+0.01 (+0.2)	0 (+0.3)
Signal PDF μ	+0.04 (+0)	0 (+0.1)	+0.01 (+0.1)	+0.01 (+0.1)
Signal PDF λ	+0.11 (+0.1)	+0.01 (+0.3)	+0.04 (+0.1)	+0.01 (+0.1)
Background PDF (exp)	–0.1 (+0.1)	–0.1 (+0.1)	–0.1 (+0.1)	–0.6 (+0.6)
Background PDF (obs)	–0.1 (+0.1)	–0.3 (+0.3)	–0.1 (+0.1)	–0.1 (+0.1)
Other sources	0 (+0)	0 (+0)	0 (+0)	0 (+0)
Total (obs)	0 (+0.3)	+0.1 (+1.0)	+0.1 (+0.4)	0 (+0.5)

Table 4.26: Impact of the systematic uncertainties on the branching fraction and the upper limit. The values quoted correspond to the absolute change in the fitted branching fraction when we consider the systematic only on the given source. The value in parentheses is the change in the observed upper limit when considering only the given source. The "Total" line gives the difference between the branching fraction (upper limit) with and without systematic uncertainties.

4.8. Results

Once the analysis strategy is completely fixed, the hidden signal region is unblinded, and we can perform the final fit to the data.

4.8.1. Total fit to data

The Belle and Belle II unblinded datasets are fitted simultaneously with an unbinned extended maximum likelihood fit. The results are shown in Figure 4.51 and the measured branching fractions as well as the other free parameters of the fit are displayed in Table 4.27. The fits are validated with 1000 toy datasets and no bias is observed.

		Br ($\times10^{-5}$)	N_{sig}	N_{bkg}	c_1	c_2
OSe	Belle II	-0.24 ± 1.46	-0.5 ± 2.8	86.5 ± 9.8	0.76 ± 0.15	-0.22 ± 0.19
	Belle		-0.6 ± 3.3	99.5 ± 10.4	0.33 ± 0.17	-0.03 ± 0.17
SSe	Belle II	1.17 ± 2.77	1.7 ± 3.9	120.1 ± 11.8	0.32 ± 0.17	-0.18 ± 0.17
	Belle		2.2 ± 5.2	140.0 ± 12.7	0.26 ± 0.14	-0.26 ± 0.15
$OS\mu$	Belle II	1.07 ± 1.80	1.6 ± 2.7	109.1 ± 10.9	0.45 ± 0.16	0.07 ± 0.16
	Belle		2.8 ± 4.7	207.4 ± 15.0	0.43 ± 0.11	-0.14 ± 0.12
$SS\mu$	Belle II	0.48 ± 2.61	0.6 ± 3.3	198.4 ± 14.5	0.07 ± 0.12	-0.31 ± 0.14
	Belle		0.6 ± 3.2	76.4 ± 9.3	0.06 ± 0.19	-0.66 ± 0.22

Table 4.27: Values of the free parameters of the total τ recoil mass fit in data. The uncertainties include both statistical and systematic uncertainties.

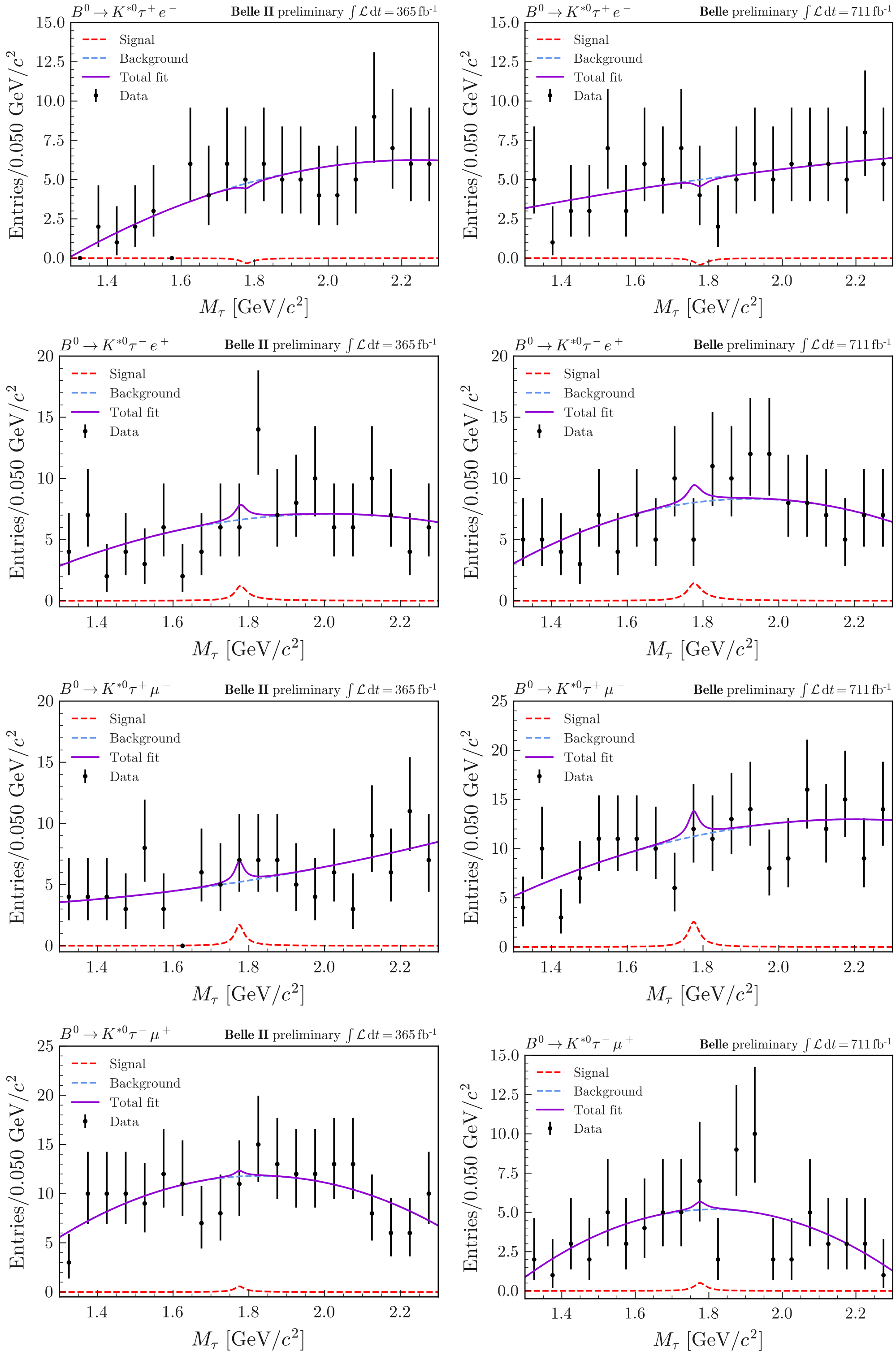


Figure 4.51: Total Belle II (left) and Belle (right) simultaneous fits of the τ recoil mass in data. From top to bottom: *OSe*, *SSe*, *OS*μ, *SS*μ.

4.8.2. Upper limits on the branching fractions

Since no significant signal is observed, we derive upper limits (UL) at 90% confidence level (CL) on the branching fractions of the four decay modes. To do so, we use the so-called CLs method [33], with an asymptotic approach [34]: for each scanned values of input branching fractions varying from 0 to 1×10^{-4}, confidence level (or p-value) is calculated for a background only hypothesis $\mathrm{CL_b}$ and for a signal+background hypothesis $\mathrm{CL_{s+b}}$, and the upper limit is defined such that the corresponding confidence level for signal $\mathrm{CL_s} = \mathrm{CL_{s+b}}/\mathrm{CL_b}$ is equal to 0.1.

The expected and observed upper limits on the branching fractions are given in Table 4.28 together with the corresponding efficiencies for Belle II and Belle. The expected upper limits are derived using an extrapolation of the data sidebands with a 0 signal hypothesis, while the observed ones are derived from the fit to Belle and Belle II data. The CLs scans with expected and observed limits are also shown in Figure 4.52.

The limits are given with (and without) the systematic uncertainties detailed in Section 4.7.

The dominant systematic uncertainties are the ones on the BDT efficiency (~30% for Belle II, ~20% for Belle), on the signal PDF resolution (59% for Belle II, 21% for Belle) and on the background shape.

	$\varepsilon_{\rm sig}^{\rm Belle\ II}$ [%]	$\varepsilon_{\rm sig}^{\rm Belle}$ [%]	$\mathscr{B}_{\rm exp}^{UL}$ ($\times10^{-5}$)	$\mathscr{B}_{\rm obs}^{UL}$ ($\times10^{-5}$)
OSe, $B^0 \to K^{*0}\tau^+e^-$	0.075	0.046	2.8 (2.7)	2.9 (2.6)
SSe, $B^0 \to K^{*0}\tau^-e^+$	0.056	0.038	4.4 (4.3)	6.4 (5.4)
$OS\mu$, $B^0 \to K^{*0}\tau^+\mu^-$	0.060	0.053	3.0 (2.9)	4.2 (3.8)
$SS\mu$, $B^0 \to K^{*0}\tau^-\mu^+$	0.051	0.024	5.5 (4.9)	5.6 (5.1)

Table 4.28: Signal efficiencies, expected and observed upper limits on the branching fractions. The value in parentheses is the value without systematic uncertainties.

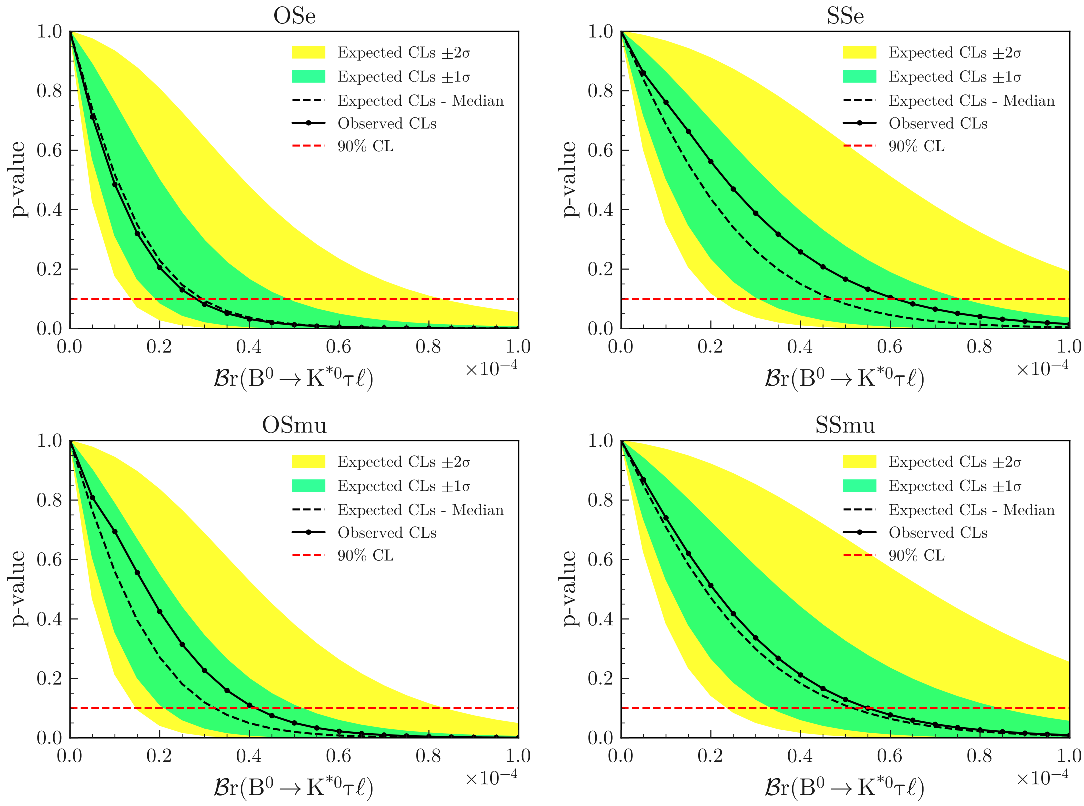


Figure 4.52: Confidence level as function of the branching fraction. The limits include the systematic uncertainties. The observed (expected) upper limit at 90% CL corresponds to the intersection of the solid (dashed) black line and the dashed red one. From top left to bottom right : *OSe*, *SSe*, *OSμ*, *SSμ*.

4.8.3. Upper limits recasting for different NP scenarios

For this analysis, the upper limits are estimated based on signal MC generated from a phase space distribution hypothesis. However, nothing guarantees that in case of NP, the signal will be distributed as such. Indeed, alternative NP scenarios may have very different q^2 or angular distributions, leading to a different signal efficiency. For this reason we propose in this section to recast the upper limits according to different NP scenarios.

For this, we consider two model-independant scenarios discussed in Section 1.3: The case where only left-handed operators contribute to the decay ($\Delta C_9 = -\Delta C_{10} = -1$), and the case where only scalar operator does ($\Delta C_S = 1$).

We define as q^2 the squared invariant mass of the dilepton system, i.e. $q^2 = M(\tau\ell)^2$. For the angular variables, we consider the angles θ_ℓ, defined as the angle between the lepton and the B-meson flight direction in the $\tau\ell$ rest frame, θ_K, defined as the angle between the K from K^{*0} and the B-meson flight direction in the K^{*0} rest frame, and φ, defined as the angle between the decay plane of the K^{*0} and the decay plane of the $\tau\ell$ system in the B rest frame.

The q^2 and angular distributions of the differential branching fraction for the phase space hypothesis and the two NP scenarios are shown in Figure 4.53 for the *OSe* decay mode as example. .

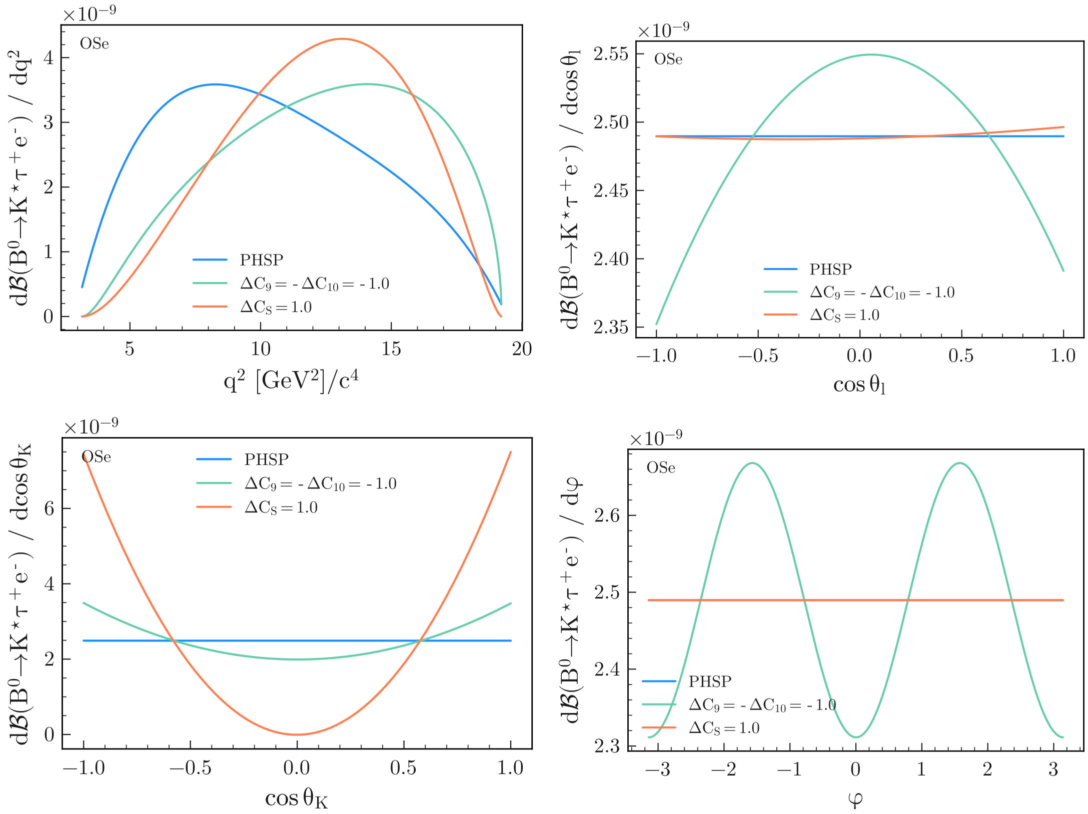


Figure 4.53: Differential branching fraction for the *OSe* decay mode as function of q^2 (top left), $\cos\theta_\ell$ (top right), $\cos\theta_K$ (bottom left) and φ (bottom right) for the phase space hypothesis (blue), for the left-handed NP hypothesis (green) and for the scalar NP hypothesis (orange) before any selection.

For each mode and experiment, the final efficiency is evaluated as a function of q^2 and angular variables, and is reweighted by the differential branching fraction of the considered NP scenario. The alternative upper limits are then remeasured using the same CLs method as described in Section 4.8.2, but with the reweighted efficiencies. The efficiency as function of q^2 and angular variables is shown in Figure 4.54 (4.55) for the Belle II (Belle) *OSe* decay mode as example. The reweighted efficiencies are given in Table 4.29 for all modes, Belle and Belle II, together with the remeasured upper limits. The results are similar or higher than the phase space model by up to 19%.

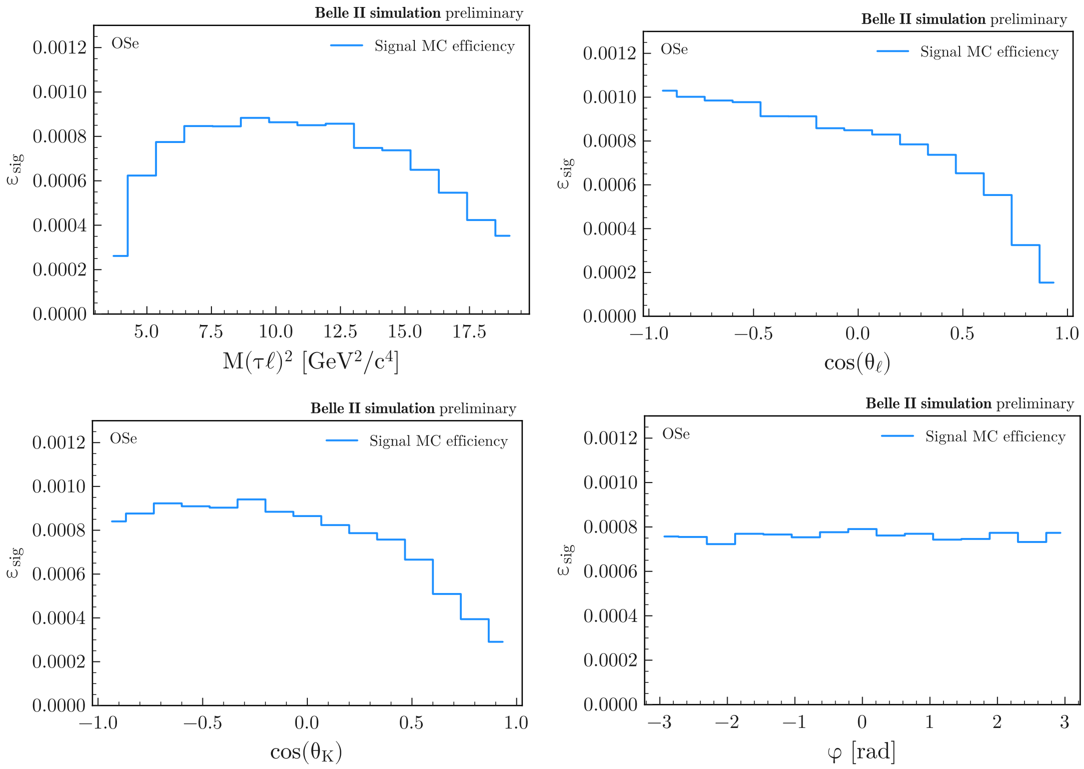


Figure 4.54: Final efficiency as function of $q^2 = M(\tau\ell)^2$ (top left), $\cos\theta_\ell$ (top right), $\cos\theta_K$ (bottom left) and φ (bottom right) for the Belle II *OSe* mode.

		$\varepsilon_{sig}^{\text{Belle II}}$ [%]	$\varepsilon_{sig}^{\text{Belle}}$ [%]	$\mathcal{B}_{obs}^{UL}$ ($\times 10^{-5}$)
	PHSP	0.075	0.046	2.9
OSe, $B^0 \to K^{*0}\tau^+e^-$	Left-handed	0.072	0.045	3.0
	Scalar	0.066	0.042	3.2
	PHSP	0.056	0.038	6.4
SSe, $B^0 \to K^{*0}\tau^-e^+$	Left-handed	0.048	0.033	7.3
	Scalar	0.045	0.032	7.6
	PHSP	0.060	0.053	4.2
$OS\mu$, $B^0 \to K^{*0}\tau^+\mu^-$	Left-handed	0.060	0.055	4.1
	Scalar	0.056	0.053	4.3
	PHSP	0.051	0.024	5.6
$SS\mu$, $B^0 \to K^{*0}\tau^-\mu^+$	Left-handed	0.047	0.023	6.0
	Scalar	0.043	0.022	6.4

Table 4.29: Reweighted efficiencies and upper limits on the branching fractions for the different left-handed and scalar NP scenarios. The "PHSP" line corresponds to the phase space hypothesis used in the analysis.

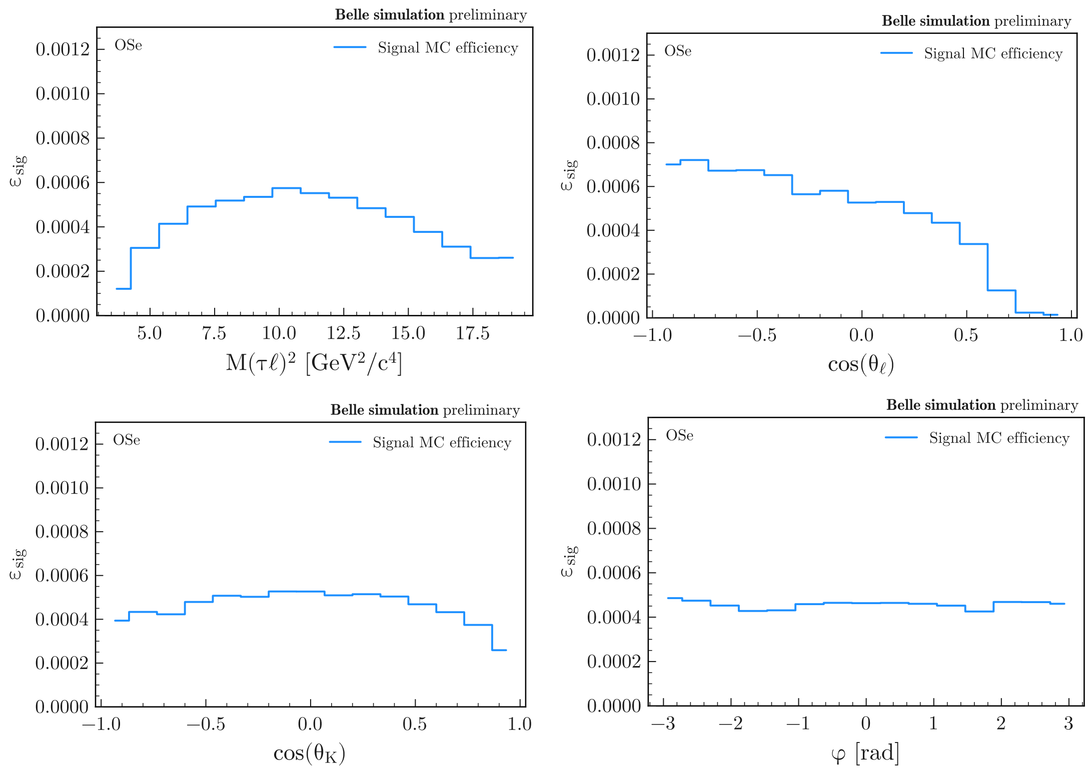


Figure 4.55: Final efficiency as function of $q^2 = M(\tau\ell)^2$ (top left), $\cos\theta_\ell$ (top right), $\cos\theta_K$ (bottom left) and φ (bottom right) for the Belle *OSe* mode.

Conclusion

The first part of the work presented in this thesis consisted in tracking performance studies for the Belle II experiment.

To begin with, the optimisation of the DAF (Deterministic Annealing Filter) algorithm was presented. Facing the higher event rates due to the luminosity increase, improvements in the computing time of the tracking chain are necessary. In this context, some parameters of the DAF algorithm, used to improve the accuracy of track reconstruction, were tuned to allow a more optimal behaviour of the algorithm. Computing power time improvement between 10% and 15% was achieved for simulated $B\overline{B}$ events in the detector configuration used for the reconstruction on HLT. In addition, the DAF parameters were implemented in a payload to ease the access and tuning of these parameters in the future.

The second tracking performance study was dedicated to the fake tracks rate at Belle II. The fake tracks rate was measured from $e^+e^- \to \tau^+\tau^-$ MC and data events from the last data processing campaign, and was found to be around 4%. This value is compatible with the previous measurements carried out with older data and MC samples. In addition, the evolution of the fake tracks rate was studied under different reconstruction conditions based on the SVD timing. It was shown that exploiting the SVD timing information to select tracks plays an important role in reducing the fake tracks rate, in particular in high beam background conditions.

The second and main part of this manuscript was the search for $B^0 \to K^{*0}\tau^\pm\ell^\mp$ lepton flavour violating decays. The analysis was performed using the full Belle II dataset recorded during the first data taking period (2019-2022) as well as the full Belle dataset, consisting in respectively 365+711 fb^{-1} of data collected at the $\Upsilon(4S)$ resonance. The study relies on the hadronic B-tagging technique, which allows to fully reconstruct the tagged B meson in hadronic decays. The signal can then be searched for in the system recoiling against the tagged B meson and the $K^{*0}\ell$ system. Such technique, despite a very low efficiency, allows a strong constraint on the τ recoil mass resolution, providing a clean signal if any.

The background rejection process was performed first with a cut-based selection, targeting specific and localised background sources, and then with machine learning techniques using a BDT to reject the remaining background. The branching fractions $\mathcal{B}(B^0 \to K^{*0}\tau^\pm\ell^\mp)$ were then extracted from a maximum likelihood fit to the τ recoil mass distribution in the Belle and Belle II datasets simultaneously. Since no signal was observed, the upper limits on the branching fractions were derived at 90% confidence level using an asymptotic $\mathrm{CL_s}$ method and set in the range $2.9 - 6.4 \times 10^{-5}$.

The results obtained in this thesis for the muon modes are not competitive with the pre-existing results from the LHCb experiment. For the electron modes, a similar analysis was performed by the LHCb experiment in parallel of the work done for this thesis, which preliminary results are the most stringent upper limits in $b \to s\tau\ell$ transitions to date. The comparison of the results obtained in this thesis with the LHCb results is shown in Table 4.30.

	$\mathcal{B}^{UL}_{\text{obs}}$ ($\times 10^{-5}$)	$\mathcal{B}^{UL}_{\text{LHCb}}$ ($\times 10^{-5}$) [19, 20]
OSe, $B^0 \to K^{*0}\tau^+e^-$	2.9	0.49 *
SSe, $B^0 \to K^{*0}\tau^-e^+$	6.4	0.59 *
$OS\mu$, $B^0 \to K^{*0}\tau^+\mu^-$	4.2	1.0
$SS\mu$, $B^0 \to K^{*0}\tau^-\mu^+$	5.6	0.8

Table 4.30: Comparison of the upper limits on the branching fractions of $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ obtained in this thesis with other existing results. The asterisk indicates preliminary results.

The limits obtained in the analysis presented in thesis are nonetheless the first results of lepton flavour violation search in the $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ decays performed at a B factory. Moreover, the results are obtained with a totally different analysis method and independent data from the LHCb experiment, which allows to corroborate the results obtained by LHCb.

In the coming years, the integrated luminosity of data collected by the Belle II experiment will increase significantly, allowing a better sensitivity to the $B^0 \to K^{*0}\tau^{\pm}\ell^{\mp}$ decays. When rescaling naively the Belle II background yield to 10 and 50 ab^{-1}, the expected upper limits can reach $0.9 - 1.7 \times 10^{-5}$ and $0.5 - 0.8 \times 10^{-5}$ respectively. By also considering a reduction of the systematic uncertainties thanks to the larger data samples, and the expected efficiency gain from the FEI algorithm (estimated at +30% for the new training of the algorithm with respect to the version used in the analysis presented in this thesis), the expected upper limits reach $0.6 - 1.1 \times 10^{-5}$ for 10 ab^{-1} and $0.2 - 0.5 \times 10^{-5}$ for 50 ab^{-1}. Even better limits could be achieved in the future by optimising the analysis for a larger dataset. Apart from the size of the data sample, the other limit of this analysis is the extremely low efficiency, mainly due to the hadronic B-tagging technique. Using a semi-leptonic tagging instead of hadronic could help by increasing the efficiency by a factor four, but the missing energy in the tag side leads to a highly degraded τ recoil mass resolution, resulting in probably lower performance for the final upper limits. However, the combination of hadronic and semi-leptonic tagging techniques could be a good compromise to increase the sensitivity. A dedicated analysis for semi-leptonic tagging would thus be necessary to know if better performance can be achieved with this technique. Other strategies to develop B tagging performance at Belle II are also being studied, such as the use of graph neural networks, but the results are not yet conclusive.

Appendices

A. BDT training variables

A.1. Data-MC comparison in $M(\tau)$ sidebands

A.1.1. Belle II

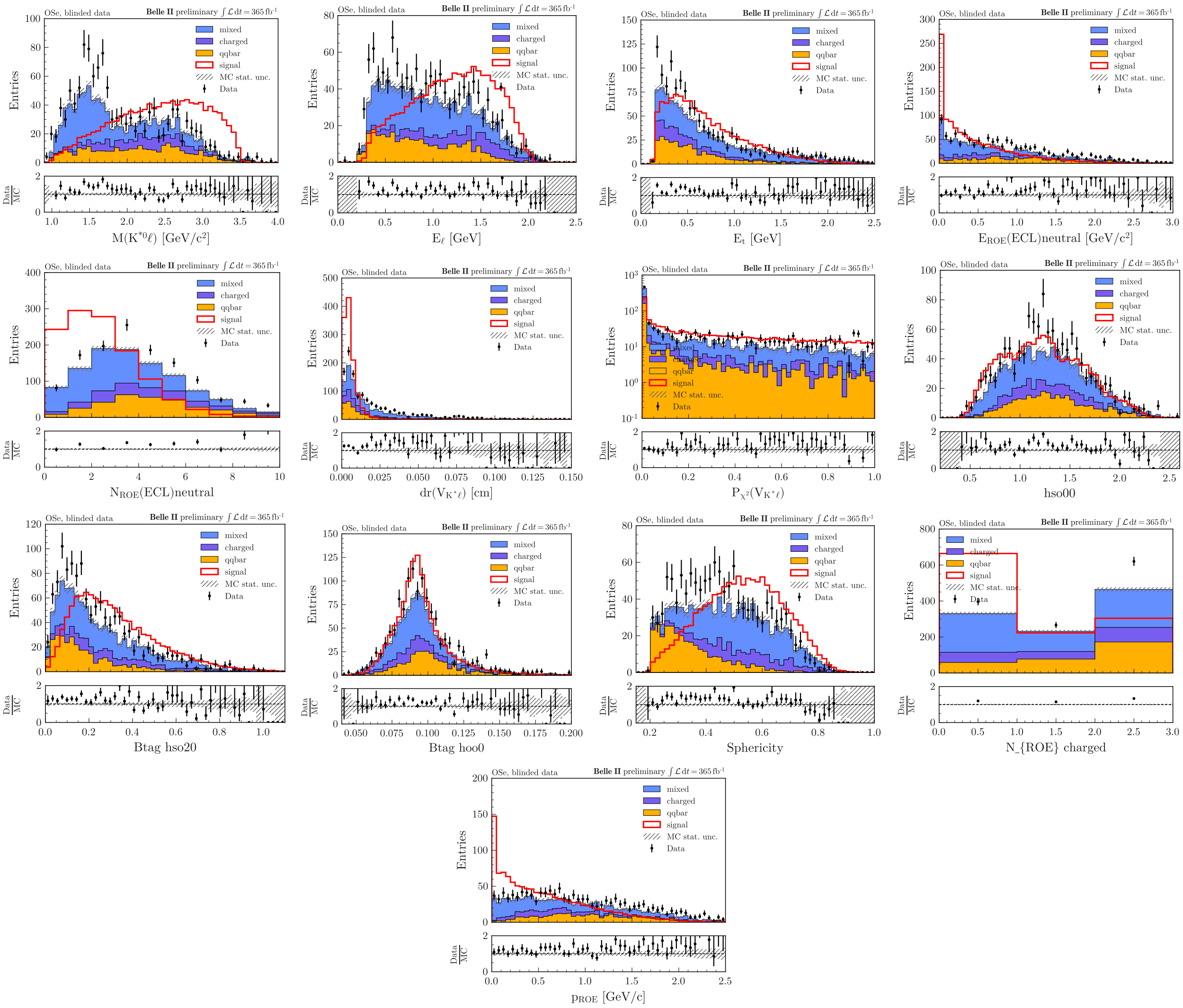


Figure 1: data/MC comparison for the BDT training variables in $M(\tau)$ sidebands for the *OSe* mode, Belle II.

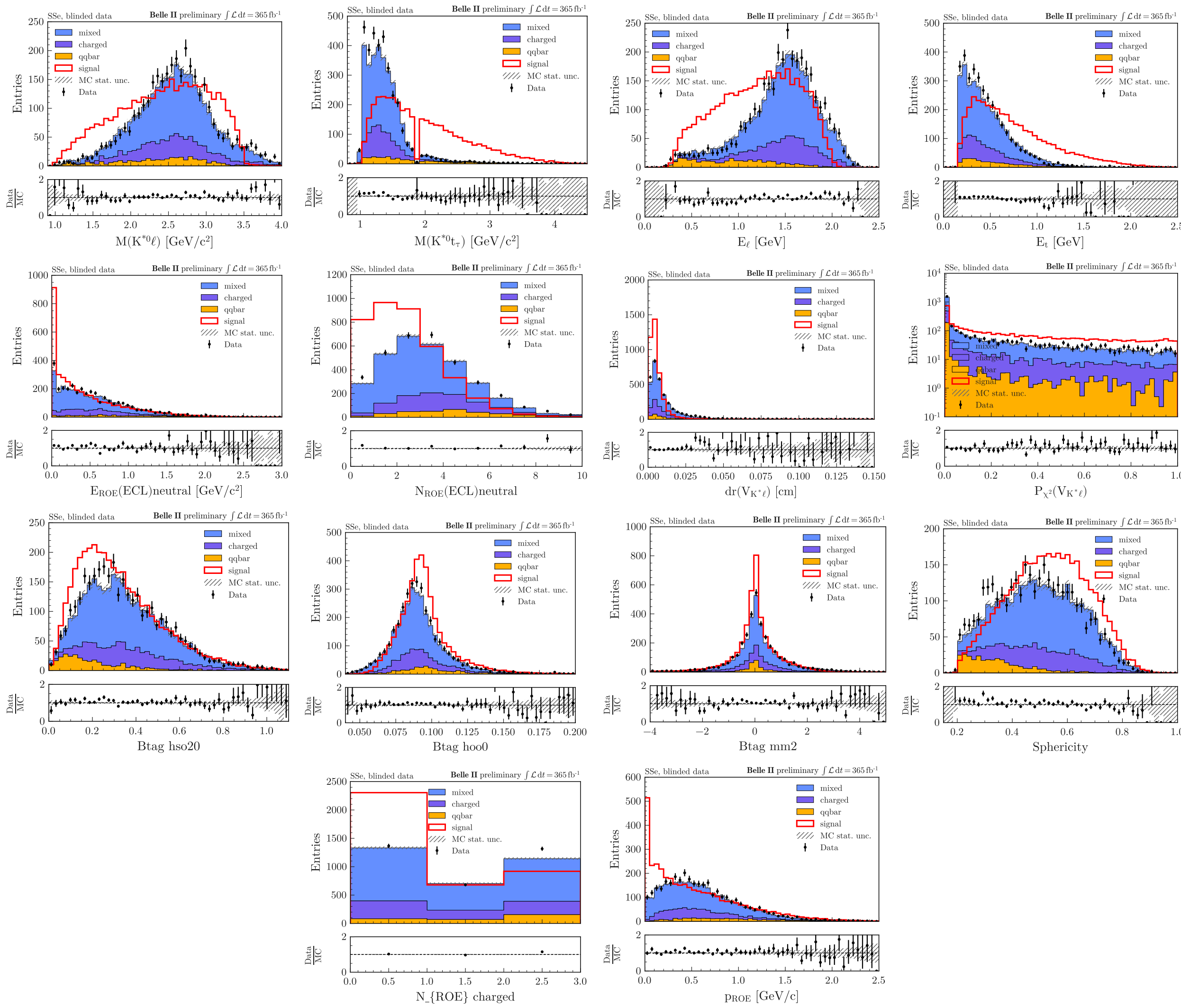


Figure 2: data/MC comparison in $M(\tau)$ sidebands for the BDT training variables for the SSe mode, Belle II.

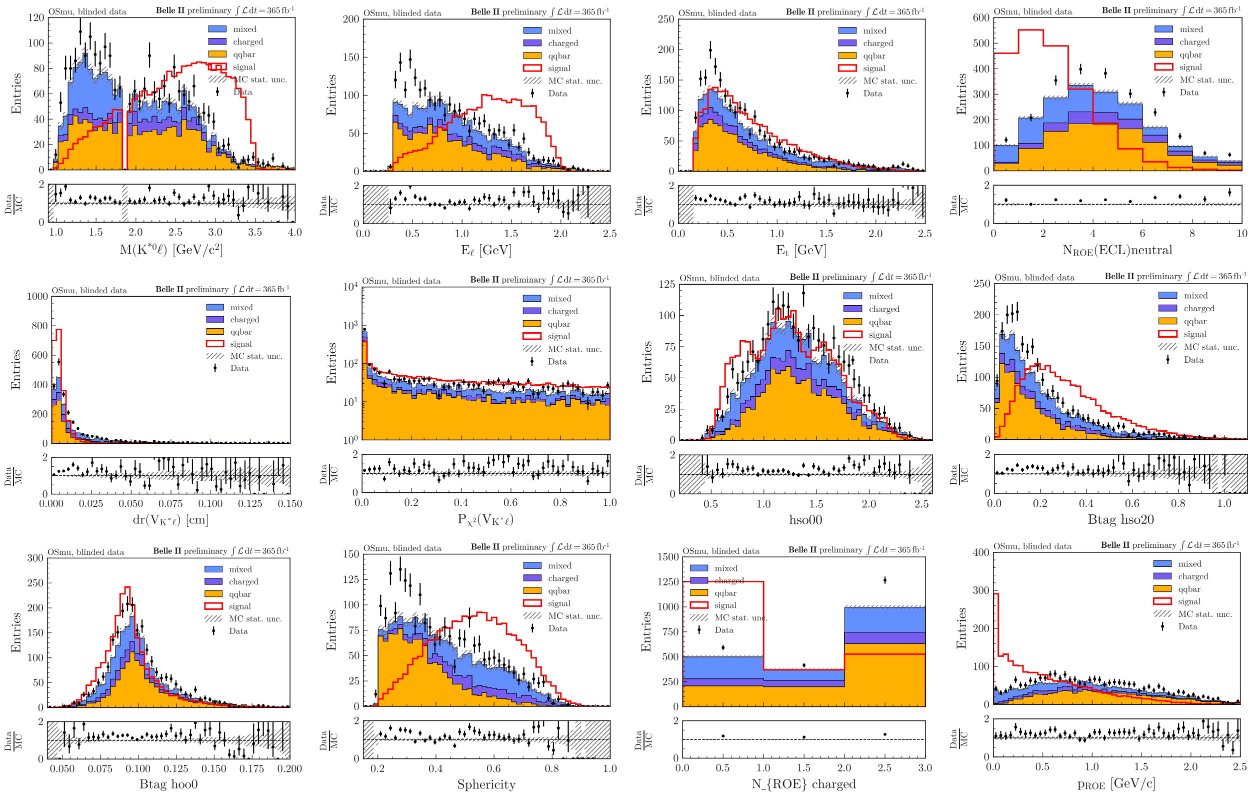


Figure 3: data/MC comparison in $M(\tau)$ sidebands for the BDT training variables for the $OS\mu$ mode, Belle II.

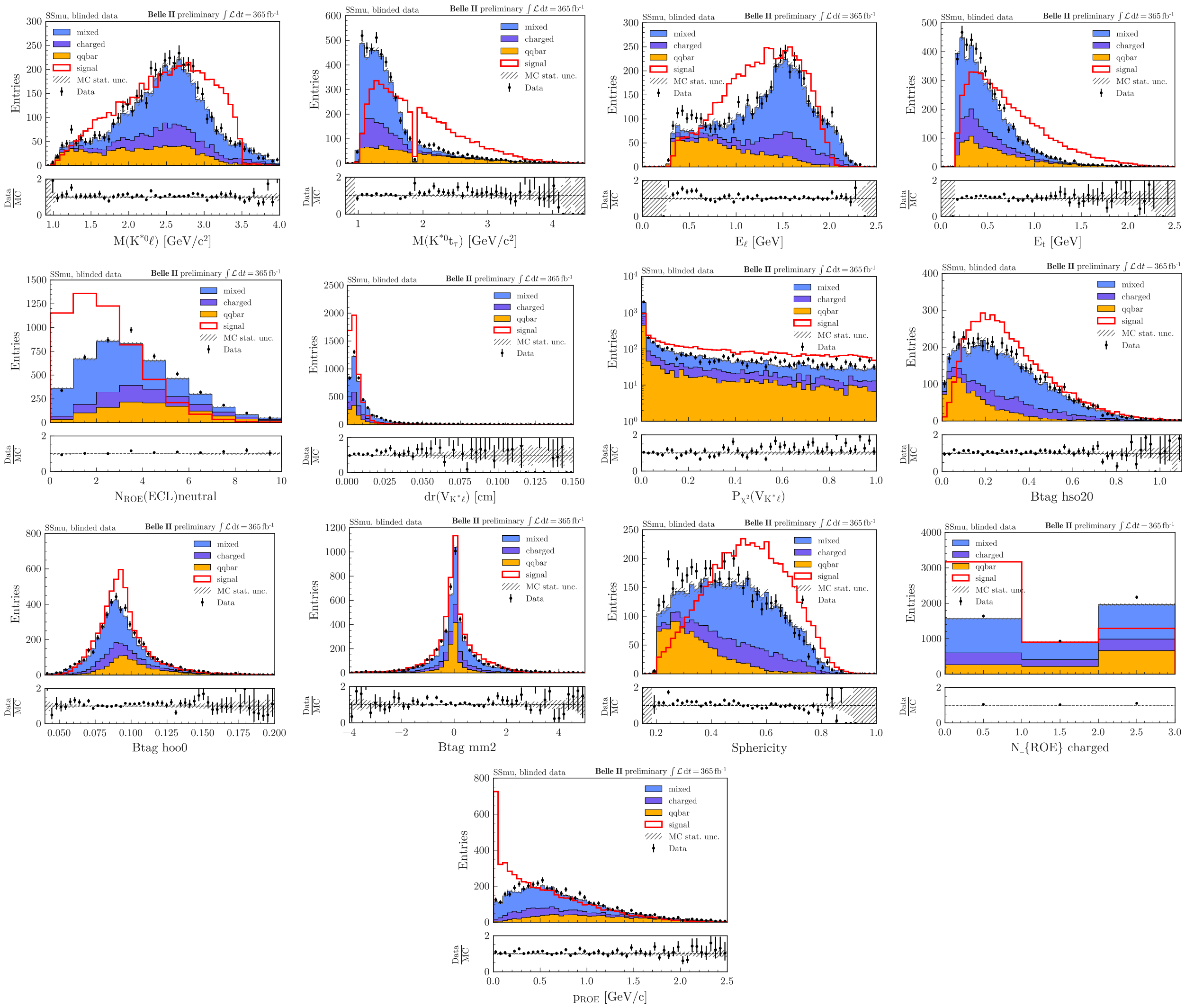


Figure 4: data/MC comparison in $M(\tau)$ sidebands for the BDT training variables for the $SS\mu$ mode, Belle II.

A.1.2. Belle

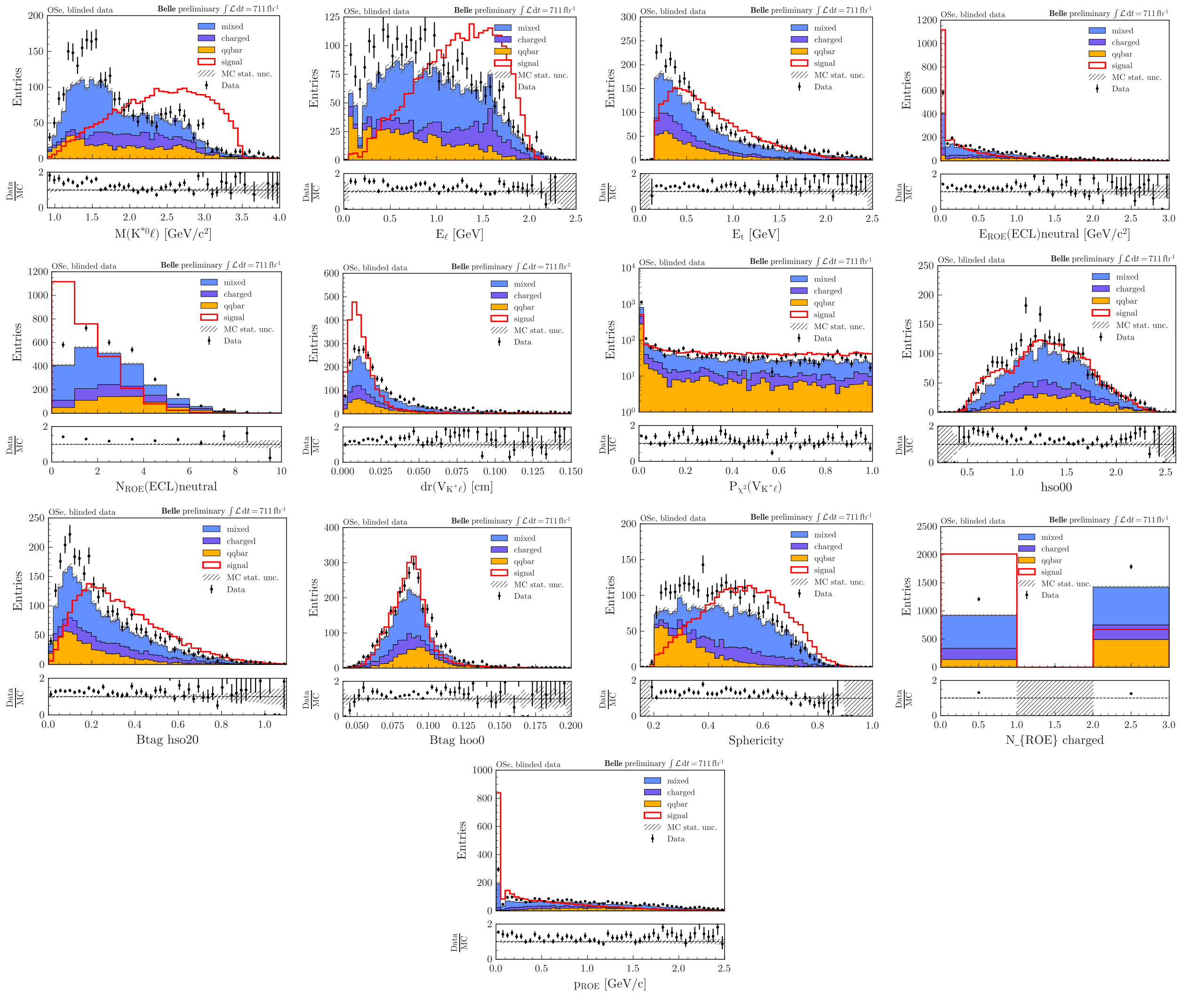


Figure 5: data/MC comparison for the BDT training variables in $M(\tau)$ sidebands for the OSe mode, Belle.

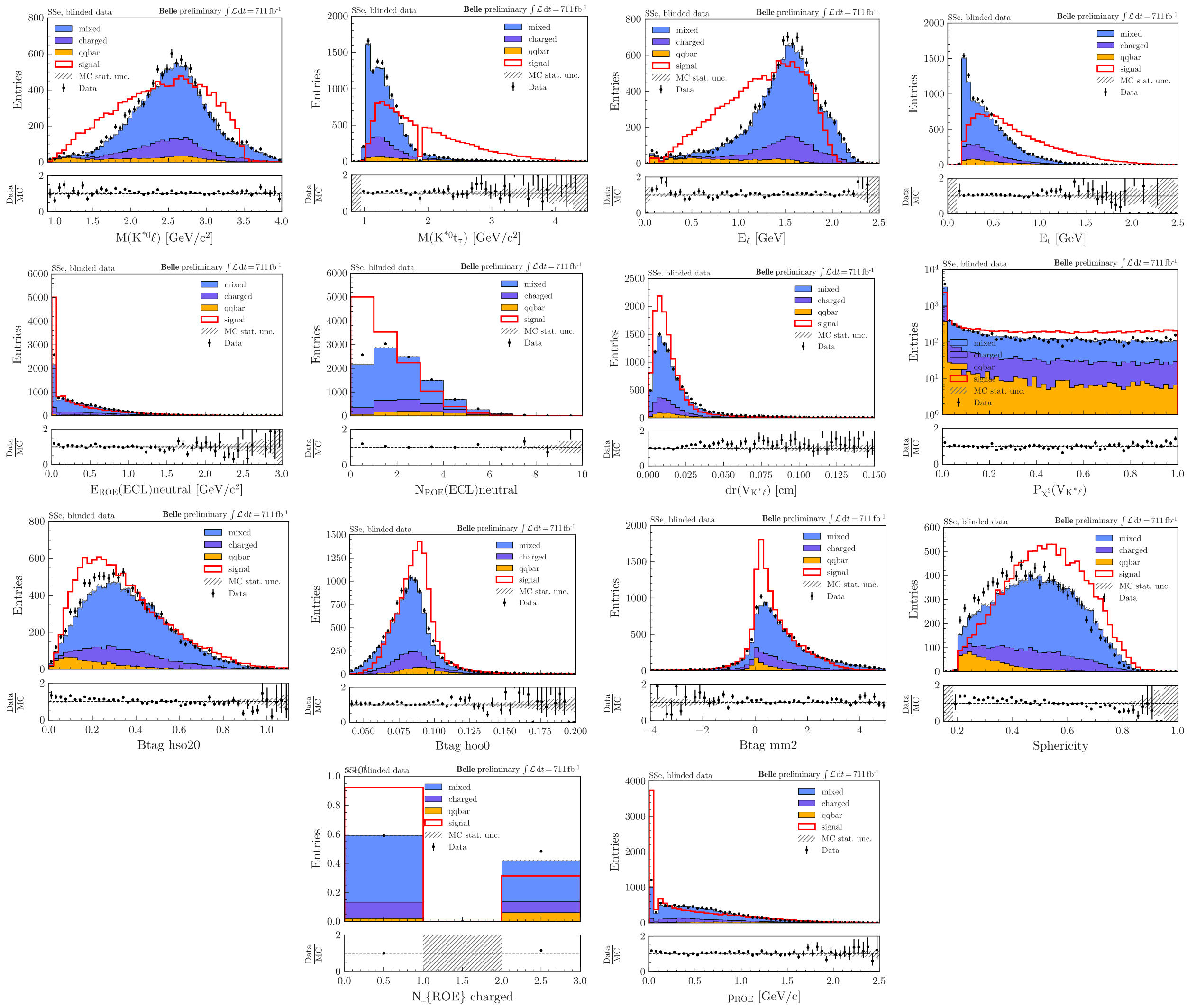


Figure 6: data/MC comparison in $M(\tau)$ sidebands for the BDT training variables for the *SSe* mode, Belle.

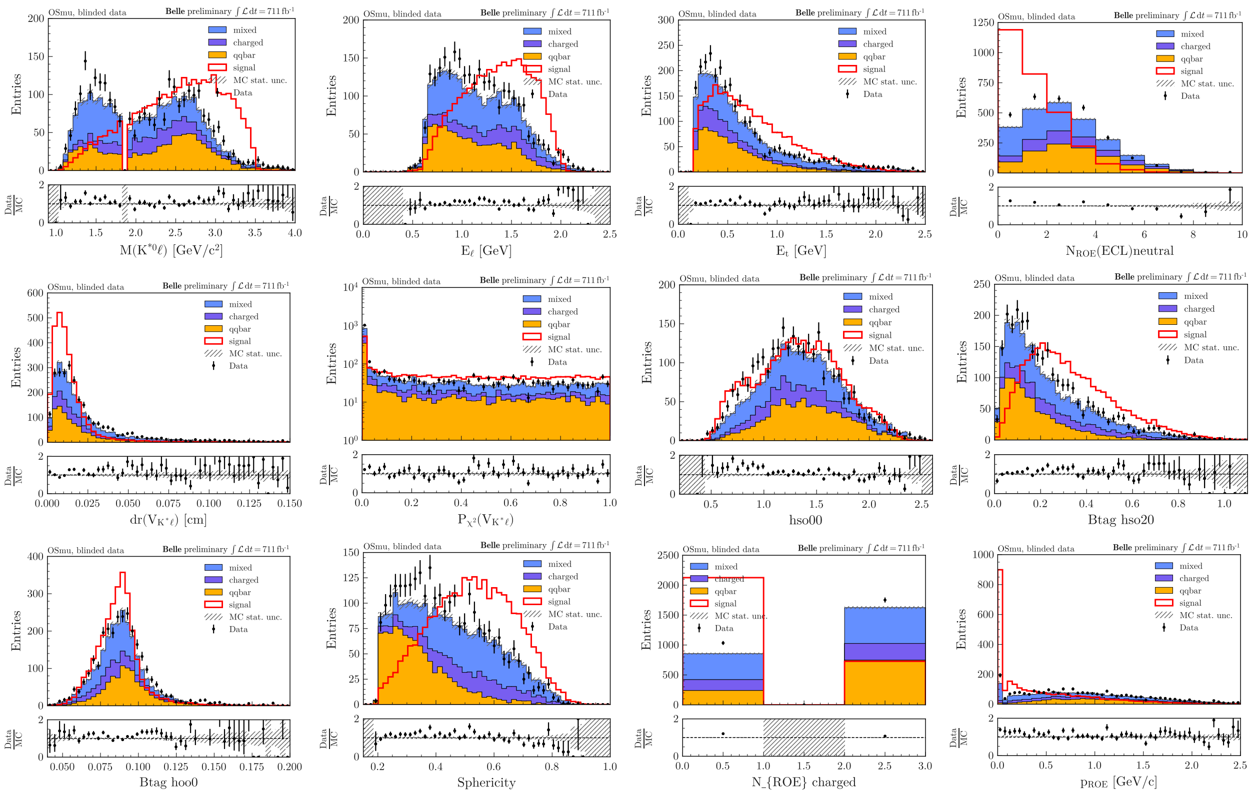


Figure 7: data/MC comparison in $M(\tau)$ sidebands for the BDT training variables for the $OS\mu$ mode, Belle.

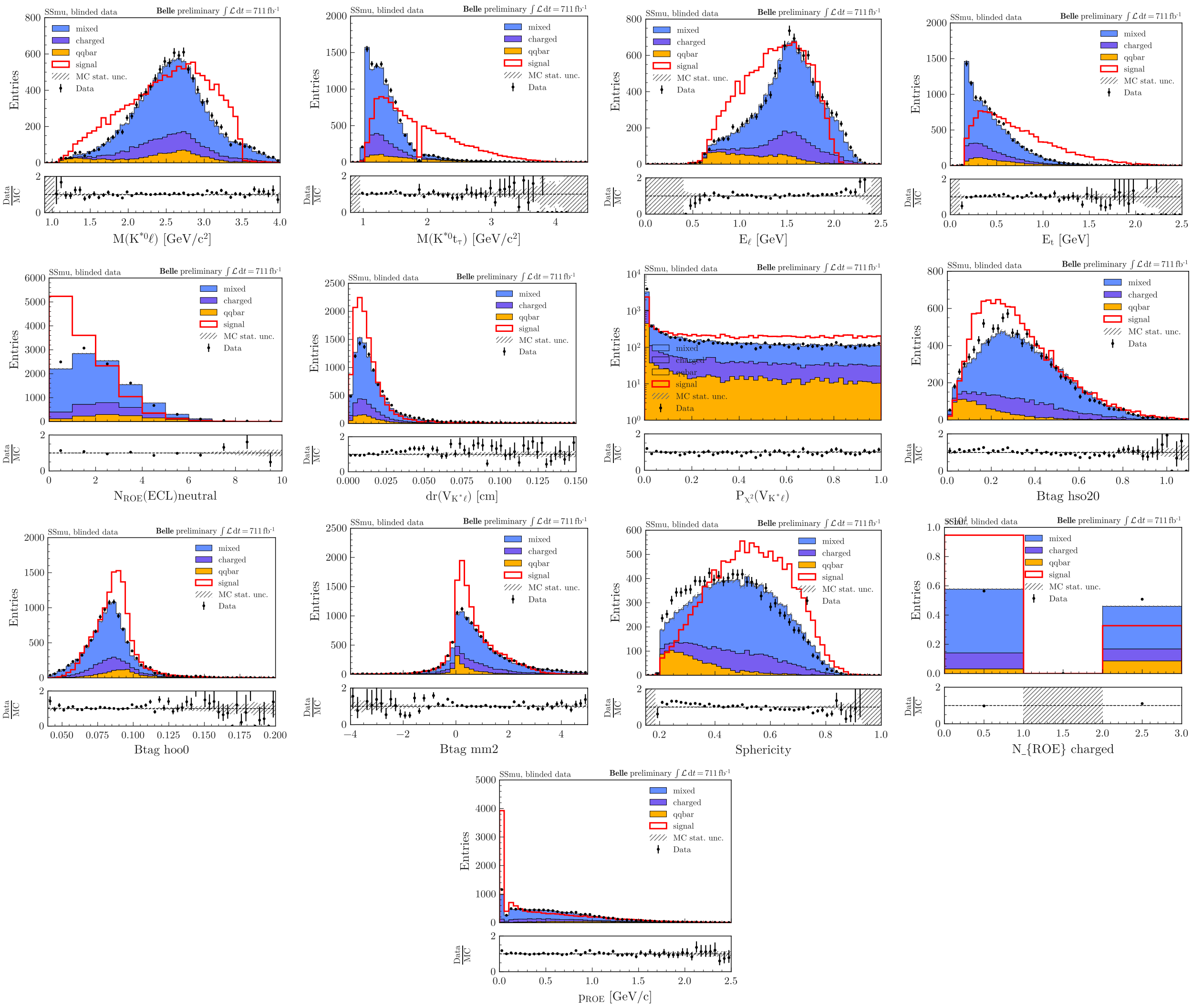


Figure 8: data/MC comparison in $M(\tau)$ sidebands for the BDT training variables for the $SS\mu$ mode, Belle.

A.2. Control channel

A.2.1. Belle II

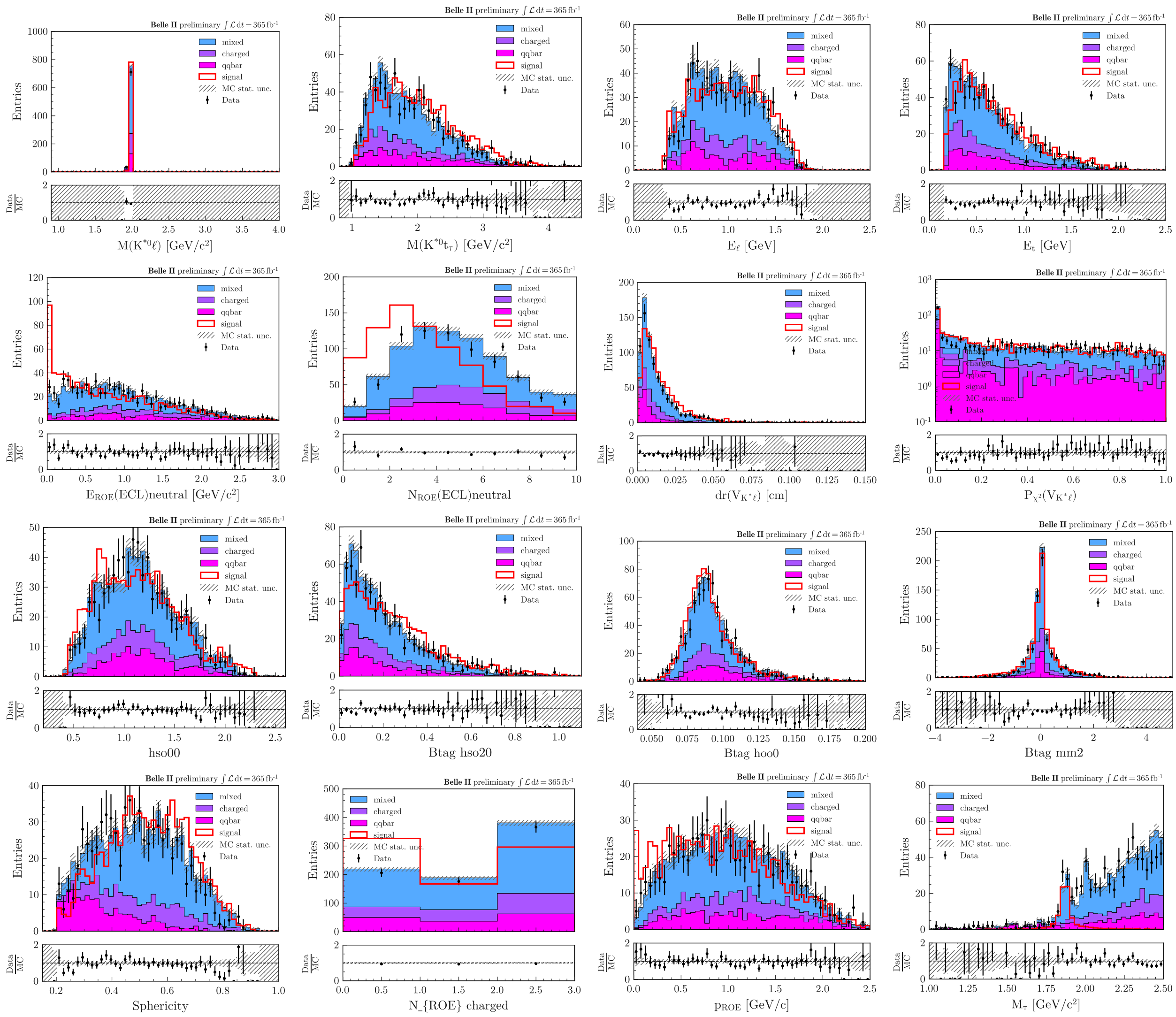


Figure 9: data/MC comparison for the BDT training variables and the recoil mass for the control channel $B^0 \to D^- D_s^+ (\to K^{*0} K^+ / \phi\pi^+)$ mode, Belle II. For the BDT variables, the signal MC is scaled to the same area as generic MC, for the recoil mass it is scaled the the expected number of events.

A.2.2. Belle

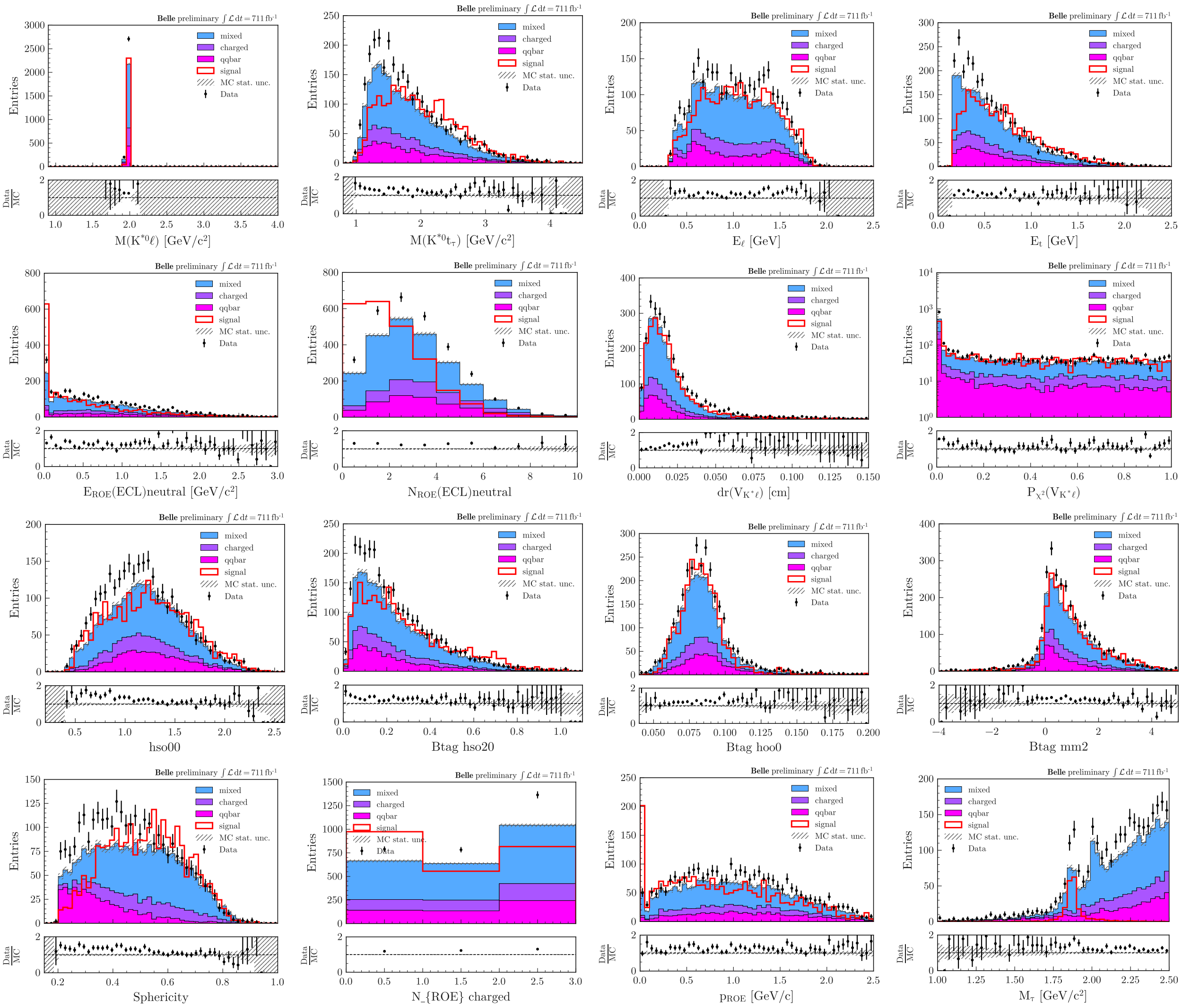


Figure 10: data/MC comparison for the BDT training variables for the control channel $B^0 \to D^- D_s^+ (\to K^{*0} K^+ / \phi \pi^+)$ mode, Belle. For the BDT variables, the signal MC is scaled to the same area as generic MC, for the recoil mass it is scaled the the expected number of events.

A.2.3. Signal comparison between nominal and control channels

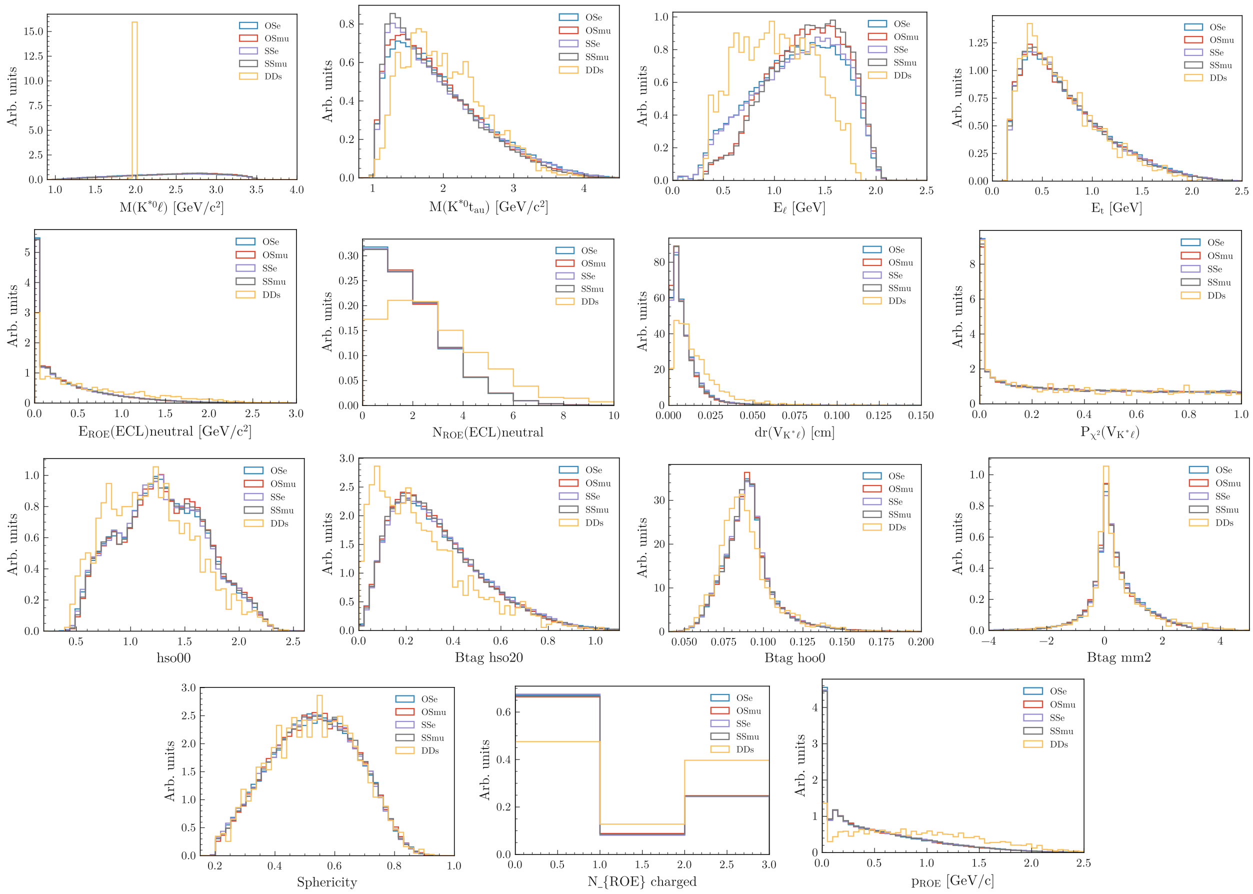


Figure 11: Signal MC comparison between the nominal channels and the control channel for the BDT training variables. The Belle and Belle II signal MC are added together. All signal are normalised to the same area for the shape comparison.

A.3. Correlation matrices

A.3.1. Belle II

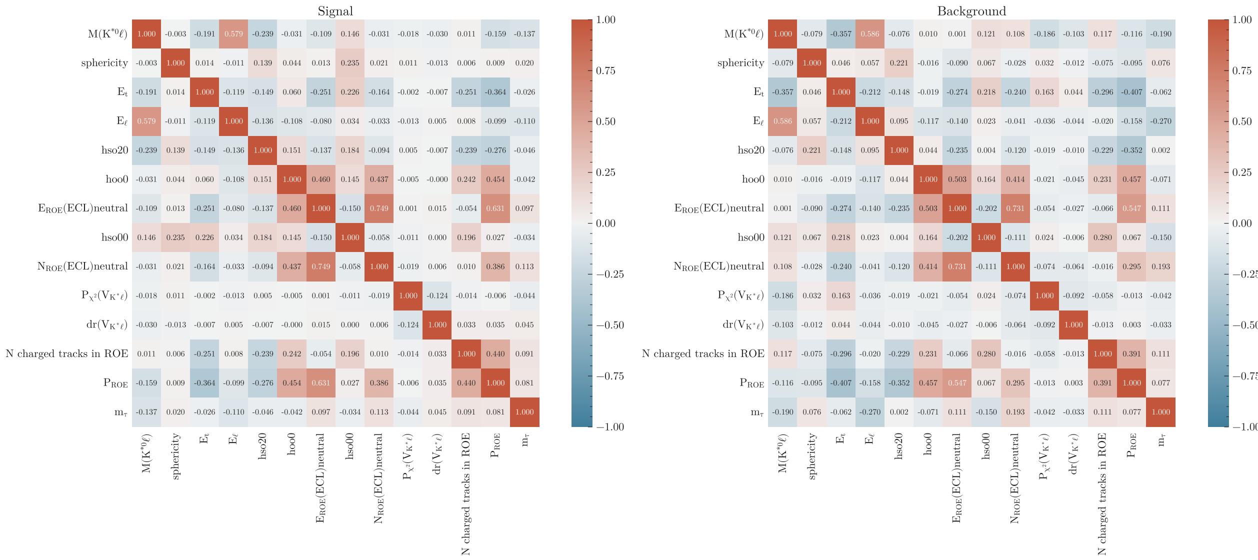


Figure 12: Correlation matrices of the BDT variables for signal (left) and background (right) for the *OSe* mode, Belle II.

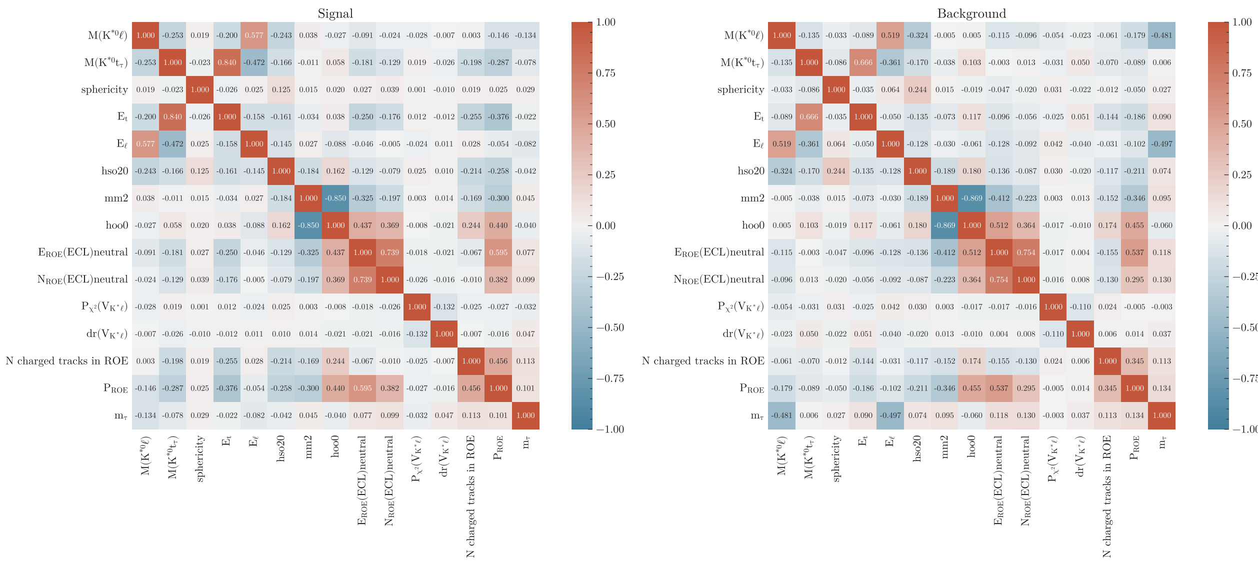
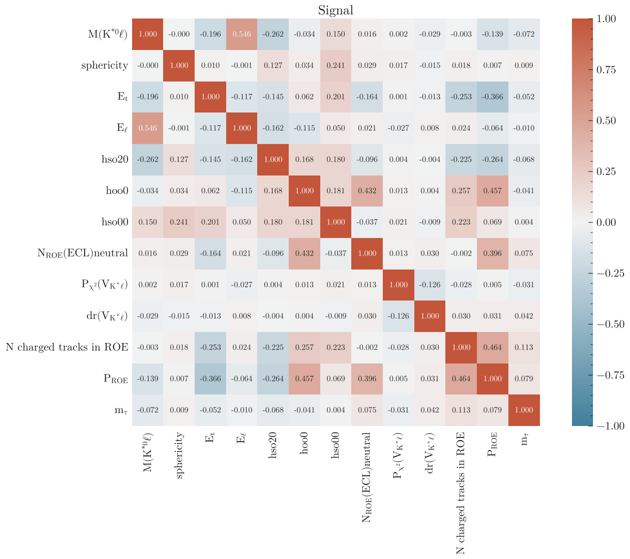


Figure 13: Correlation matrices of the BDT variables for signal (left) and background (right) for the *SSe* mode, Belle II.



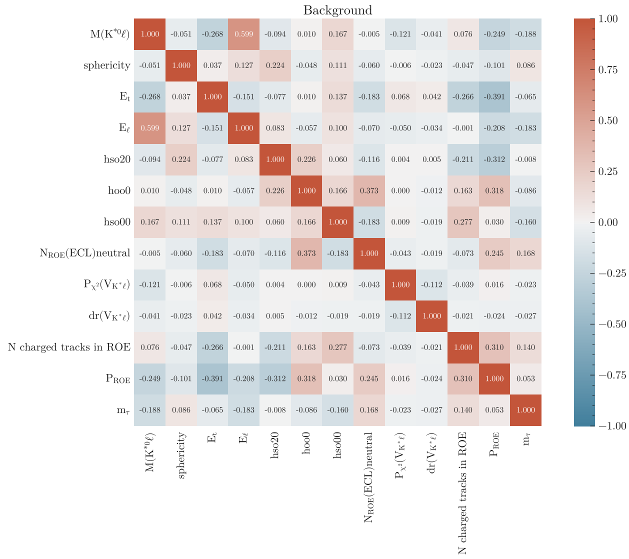
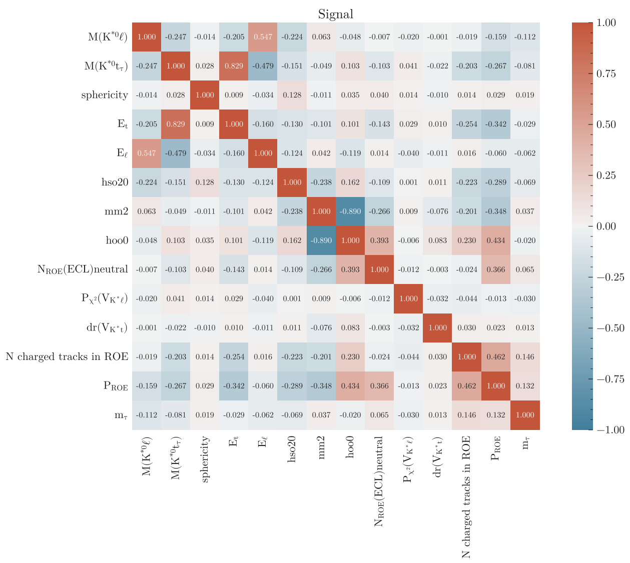


Figure 14: Correlation matrices of the BDT variables for signal (left) and background (right) for the $OS\mu$ mode, Belle II.



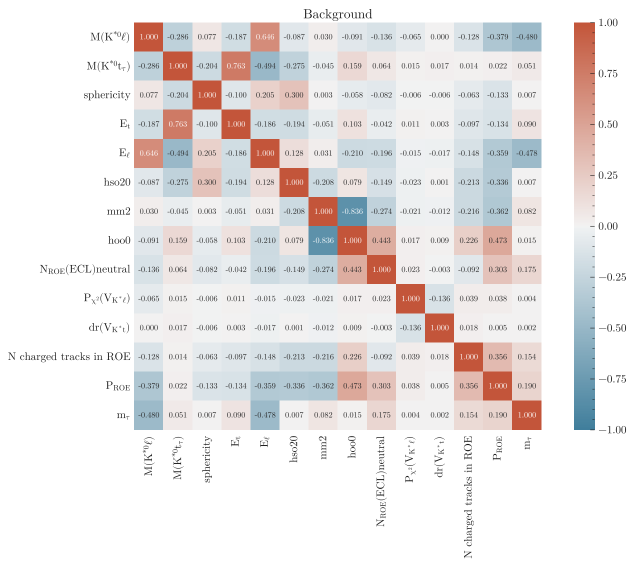


Figure 15: Correlation matrices of the BDT variables for signal (left) and background (right) for the $SS\mu$ mode, Belle II.

A.3.2. Belle

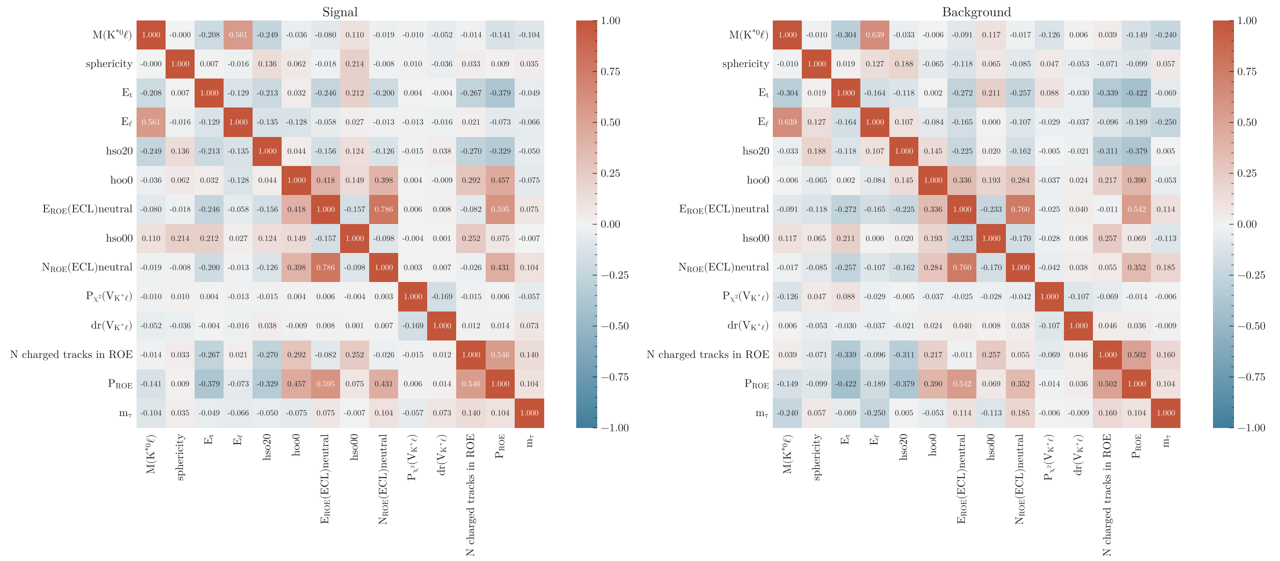


Figure 16: Correlation matrices of the BDT variables for signal (left) and background (right) for the *OSe* mode, Belle.

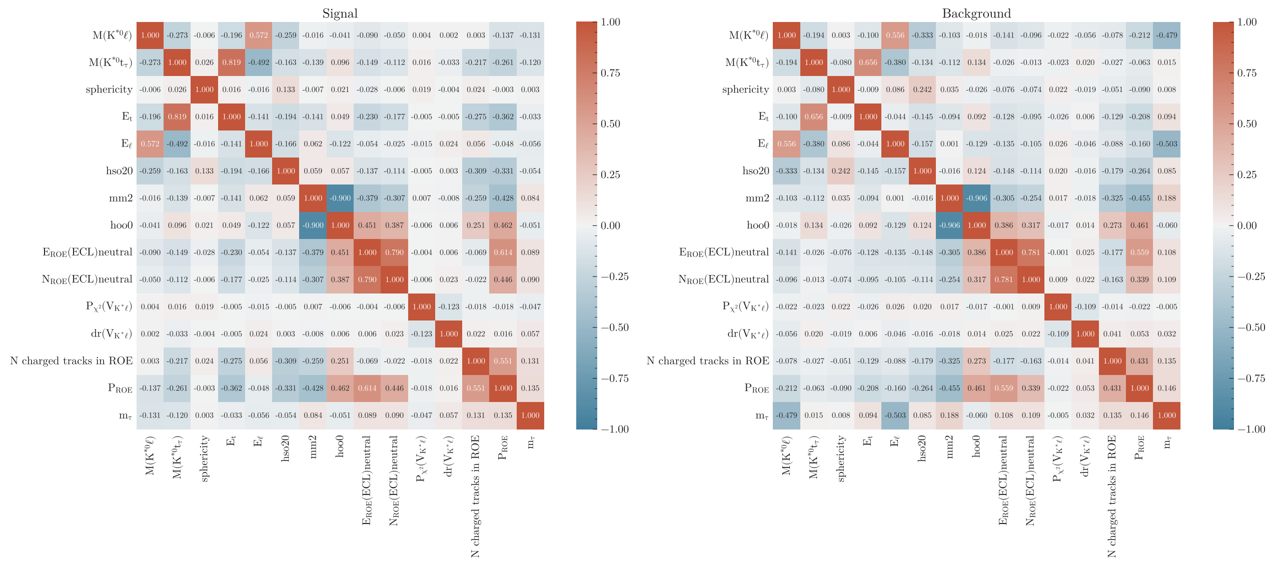


Figure 17: Correlation matrices of the BDT variables for signal (left) and background (right) for the *SSe* mode, Belle.

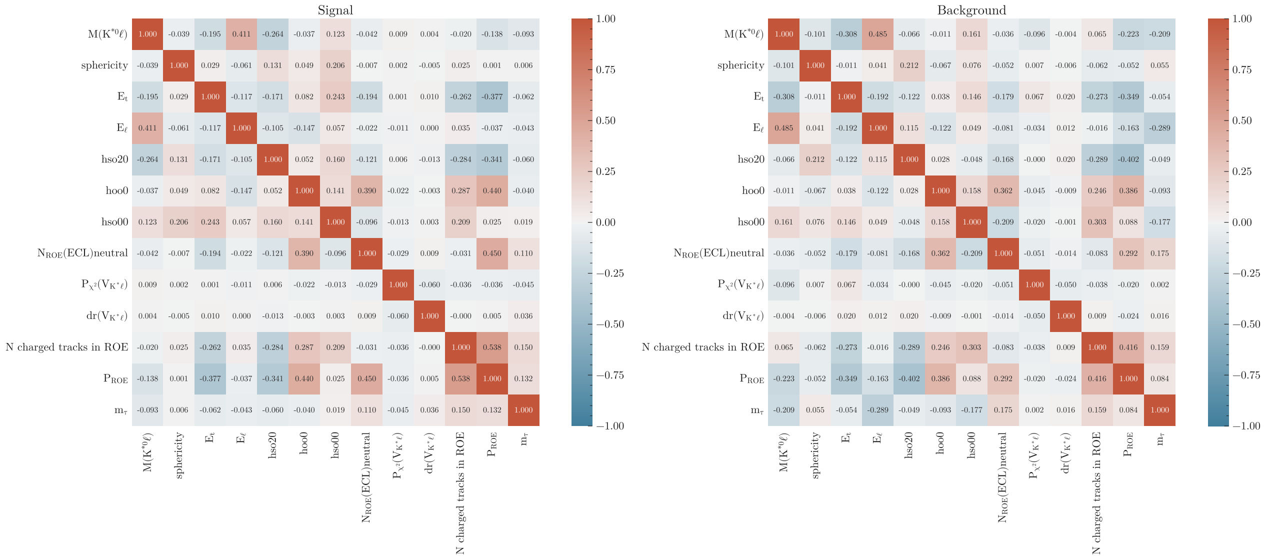


Figure 18: Correlation matrices of the BDT variables for signal (left) and background (right) for the $OS\mu$ mode, Belle.

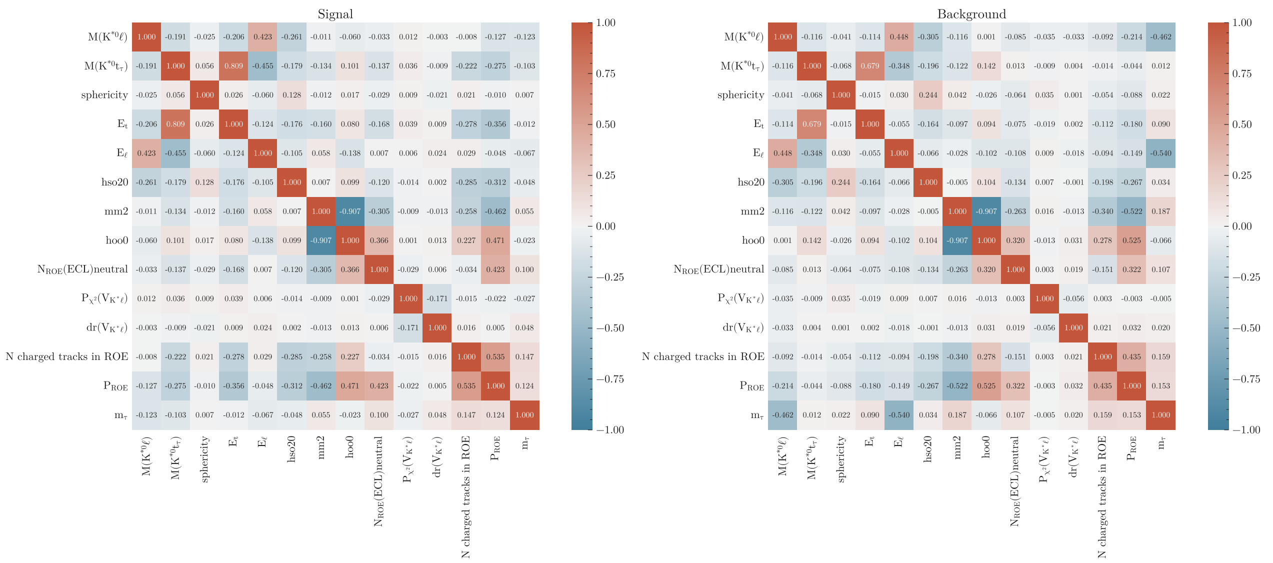


Figure 19: Correlation matrices of the BDT variables for signal (left) and background (right) for the $SS\mu$ mode, Belle.

B. Efficiency maps

B.1. Belle II

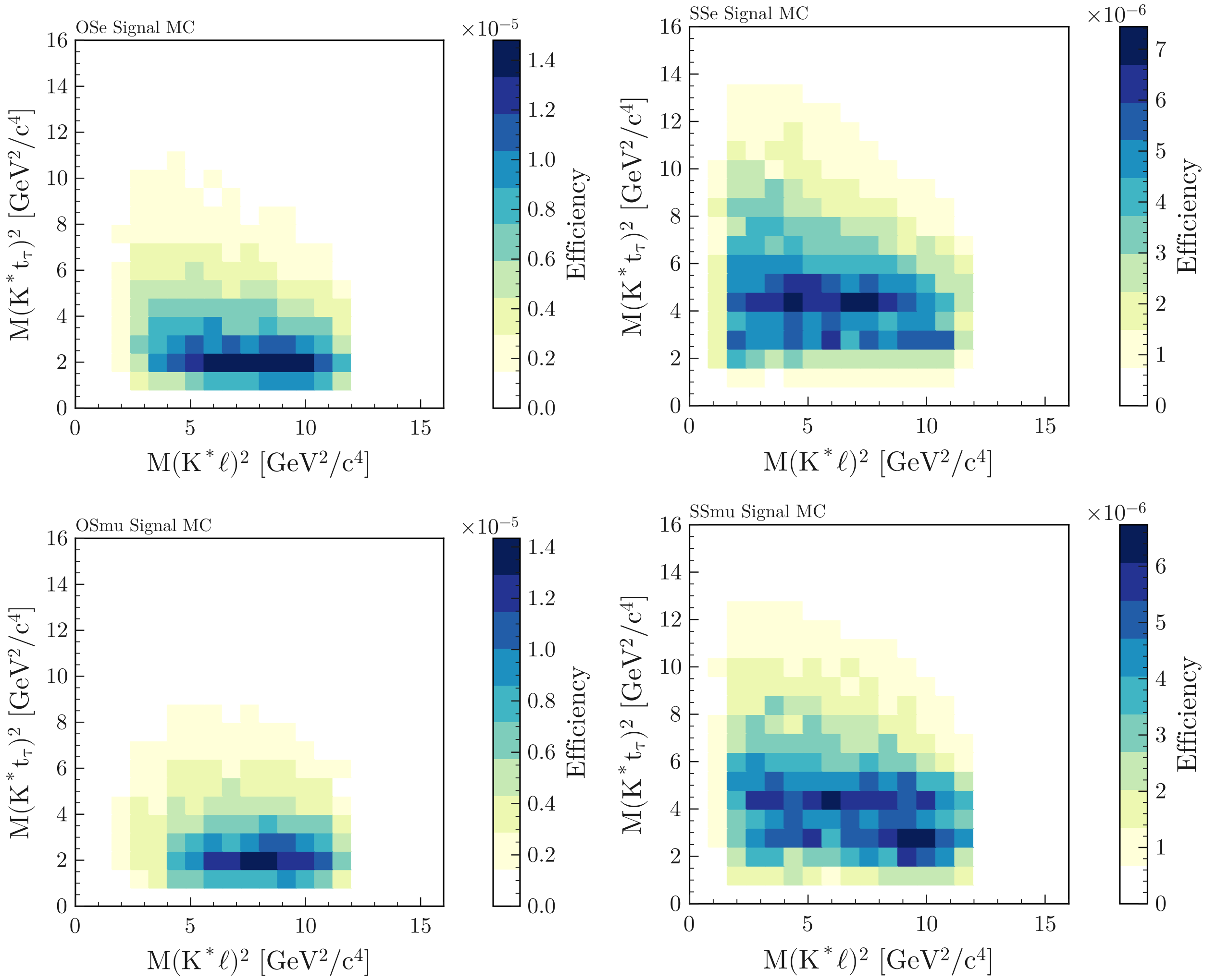


Figure 20: Efficiency maps of $M(K^{*0}\ell)$ vs $M(K^{*0}t_\tau)$ for Belle II signal MC. From top left to bottom right: *OSe*, *SSe*, *OSμ*, *SSμ*.

B.2. Belle

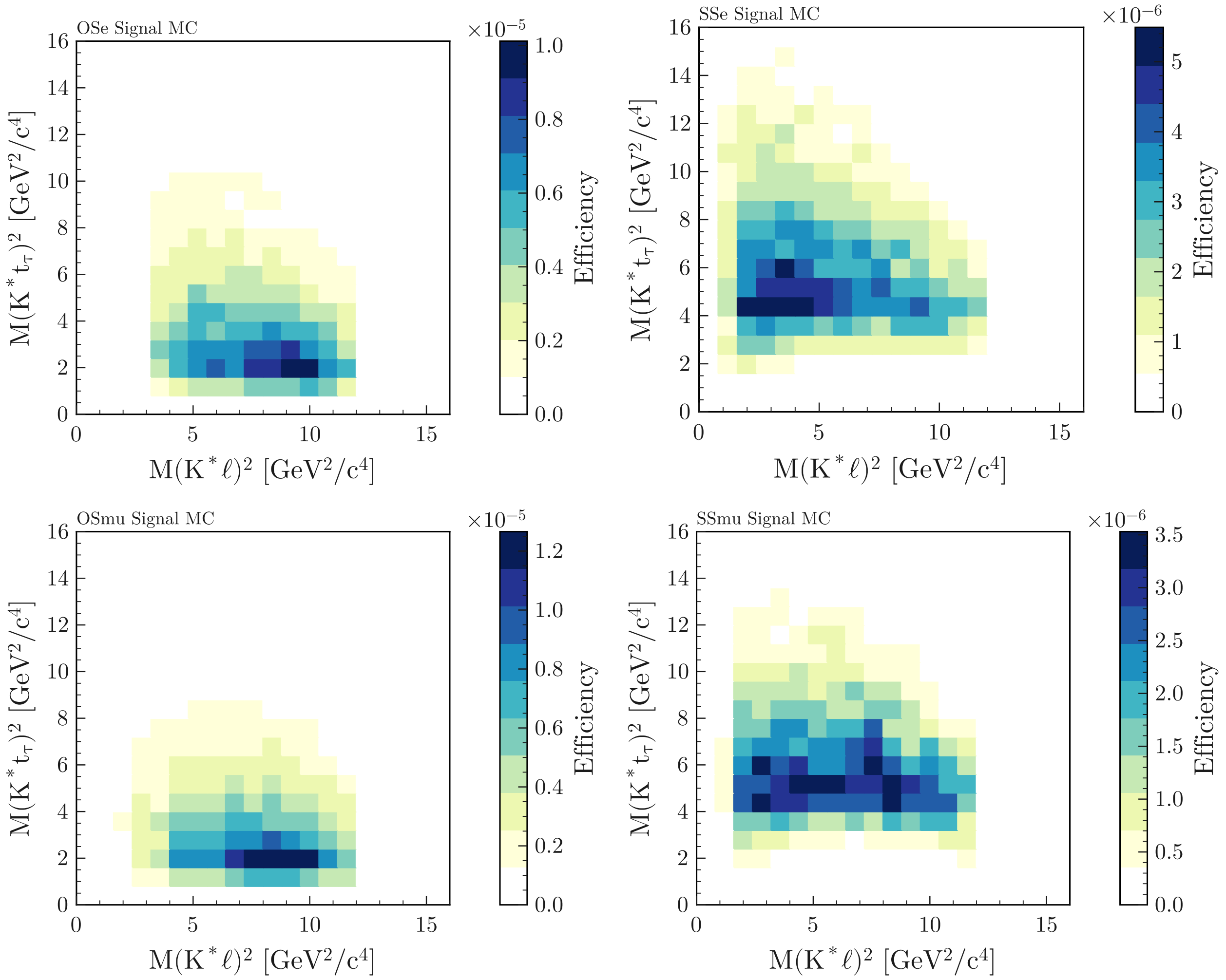


Figure 21: Efficiency maps of $M(K^{*0}\ell)$ vs $M(K^{*0}t_\tau)$ for Belle signal MC. From top left to bottom right : *OSe*, *SSe*, *OS*μ, *SS*μ.

List of Figures

List of Tables

List of acronyms

ACC

Aerogel Cherenkov Counter.

ARICH

Aerogel Ring-Imaging CHerenkov detector.

basf

Belle Analysis Software.

basf2

Belle II Analysis Software.

BDT

Boosted Decision Tree.

BSM

Beyond the Standard Model.

CDC

Central Drift Chamber.

CKF

Combinatorial Kalman Filter.

CKM

Cabibbo-Kobayashi-Maskawa.

CL

Confidence Level.

CMS

Centre-of-mass frame.

CP

Charge Parity.

CPT
: Charge Parity Time reversal.

DAF
: Deterministic Annealing Filter.

DAQ
: Data Acquisition.

DSSD
: Double-Sided Silicon Strip Detector.

ECL
: Electro-magnetic CaLorimeter.

EFT
: Effective Field Theory.

EW
: Electro-Weak.

FCNC
: Flavour Changing Neutral Current.

FEI
: Full Event Interpretation.

FR
: Full Reconstruction.

FW
: Fox-Wolfram.

HER
: High Energy Ring.

HLT
: High Level Trigger.

ID
: Identification.

IP
: Interaction Point.

KLM
: K Long and Muon detector.

KSFW
: modified Super Fox-Wolfram.

L1
: Level 1 trigger.

LER
: Low Energy Ring.

LFU
: Lepton Flavour Universality.

LFV
: Lepton Flavour Violation.

LH
: Left-Handed.

LHC
: Large Hadron Collider.

LINAC
: Linear Accelerator.

LQ
: Leptoquark.

LS1
: Long Shutdown 1.

MC
: Monte-Carlo.

MCP-PMT
: Micro-Channel Plate Photo-Multiplier Tube.

MCrd
Monte-Carlo run-dependent.

MCri
Monte-Carlo run-independent.

mdst
Mini Data-Summary Table.

NP
New Physics.

OS
Opposite Sign.

PID
Particle IDentification.

PMNS
Pontecorvo-Maki-Nakagawa-Sakata.

PMT
Photo-Multiplier Tube.

POCA
Point Of Closest Approach.

PXD
PiXel Detector.

QCD
Quantum Chromodynamics.

QED
Quantum Electrodynamics.

RH
Right-Handed.

ROE
Rest Of Event.

ROI
: Region Of Interest.

RPC
: Resistive Plate Chamber.

SiPM
: Silicon Photo-Multiplier.

SM
: Standard Model.

SS
: Same Sign.

SSB
: Spontaneous Symmetry Breaking.

SVD
: Silicon Vertex Detecor.

TOF
: Time Of Flight counter.

TOP
: Time Of Propagation detector.

TSC
: Trigger Scintillation Counter.

UL
: Upper Limit.

vev
: Vacuum Expectation Value.

VXD
: VerteX Detector.

Bibliography

[1] M. Maggiore. *A modern introduction to quantum field theory*. en. Oxford master series in physics. Oxford University Press, 2005. ISBN: 978-0-19-852073-3 (cit. on pp. 9, 28).

[2] M. E. Peskin and D. V. Schroeder. *An introduction to quantum field theory*. en. The advanced book program. CRC Press, Taylor & Francis Group, 2019. ISBN: 978-0-201-50397-5 (cit. on pp. 9, 28).

[3] M. Thomson. *Modern Particle Physics*. en. Cambridge University Press, Sept. 2013. ISBN: 9781139525367. DOI: 10.1017/CBO9781139525367 (cit. on pp. 9, 28).

[4] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. en. Cambridge University Press, 2007. ISBN: 978-1-009-40168-5. DOI: 10.1017/9781009401685 (cit. on pp. 9, 28).

[5] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020 (cit. on pp. 9, 37).

[6] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021 (cit. on pp. 9, 37).

[7] Super-Kamiokande Collaboration. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Physical Review Letters* 81.8 (Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562 (cit. on pp. 10, 39).

[8] SNO Collaboration. "Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory". In: *Physical Review Letters* 87.7 (July 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301 (cit. on pp. 10, 39).

[9] KamLAND Collaboration. "First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance". In: *Physical Review Letters* 90.2 (Jan. 2003), p. 021802. DOI: 10.1103/PhysRevLett.90.021802 (cit. on pp. 10, 39).

[10] B. W. Lee and R. E. Shrock. "Natural suppression of symmetry violation in gauge theories: Muon- and electron-lepton-number nonconservation". In: *Physical Review D* 16.5 (Sept. 1977), pp. 1444–1473. DOI: 10.1103/PhysRevD.16.1444 (cit. on pp. 10, 40).

[11] BABAR Collaboration. “Evidence for an Excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ Decays”. In: *Physical Review Letters* 109.10 (Sept. 2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802 (cit. on pp. 10, 41).

[12] BABAR Collaboration. “Measurement of an excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays and implications for charged Higgs bosons”. In: *Physical Review D* 88.7 (Oct. 2013), p. 072012. DOI: 10.1103/PhysRevD.88.072012 (cit. on pp. 10, 41).

[13] Belle Collaboration. “Measurement of $\mathscr{R}(D)$ and $\mathscr{R}(D^*)$ with a Semileptonic Tagging Method”. In: *Physical Review Letters* 124.16 (Apr. 2020), p. 161803. DOI: 10.1103/PhysRevLett.124.161803 (cit. on pp. 10, 41).

[14] LHCb Collaboration. “Measurement of the Ratios of Branching Fractions $\mathscr{R}(D^*)$ and $\mathscr{R}(D^0)$”. In: *Physical Review Letters* 131.11 (Sept. 2023). ISSN: 1079-7114. DOI: 10.1103/physrevlett.131.111802 (cit. on pp. 10, 41).

[15] HFLAV Collaboration. “Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023”. In: arXiv:2411.18639 (Nov. 2024). arXiv:2411.18639 [hep-ex]. DOI: 10.48550/arXiv.2411.18639 (cit. on pp. 10, 41, 111, 171).

[16] Belle II Collaboration. “Evidence for $B^+ \to K^+\nu\overline{\nu}$ decays”. In: *Physical Review D* 109.11 (June 2024), p. 112006. DOI: 10.1103/PhysRevD.109.112006 (cit. on pp. 10, 41).

[17] Belle Collaboration. “Search for the Lepton Flavor Violating Decays $B^+ \to K^+\tau^\pm\ell^\mp$ ($\ell = e, \mu$) at Belle”. In: *Phys. Rev. Lett.* 130 (26 June 2023), p. 261802. DOI: 10.1103/PhysRevLett.130.261802 (cit. on pp. 10, 46).

[18] Belle and Belle II Collaborations. “Search for lepton flavor-violating decay modes $B^0 \to K^0_S\tau^\pm\ell^\mp$ ($\ell = \mu, e$) with hadronic B-tagging at Belle and Belle II”. In: arXiv:2412.16470 (Dec. 2024). DOI: 10.48550/arXiv.2412.16470 (cit. on pp. 10, 46).

[19] LHCb Collaboration. “Search for the lepton-flavour violating decays $B^0 \to K^{*0}\tau^\pm\mu^\mp$ ”. In: *Journal of High Energy Physics* 2023.6 (June 2023). ISSN: 1029-8479. DOI: 10.1007/jhep06(2023)143 (cit. on pp. 10, 20, 46, 182).

[20] LHCb Collaboration. *Search for the lepton-flavour-violating decays $B^0 \to K^{*0}\tau^\pm e^\mp$*. June 2025. URL: https://arxiv.org/abs/2506.15347v1 (cit. on pp. 10, 20, 46, 182).

[21] W. Altmannshofer et al. “The Belle II Physics Book”. In: *PTEP* 2019.12 (2019). Ed. by E. Kou and P. Urquijo. [Erratum: PTEP 2020, 029201 (2020)], p. 123C01. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex] (cit. on pp. 11, 49).

[22] T. Abe, I. Adachi, K. Adamczyk, et al. *Belle II Technical Design Report*. 2010. arXiv: 1011.0352 [physics.ins-det] (cit. on pp. 11, 49, 63, 64).

[23] K. Akai, K. Furukawa, and H. Koiso. “SuperKEKB collider”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (Nov. 2018), pp. 188–199. ISSN: 0168-9002. DOI: 10.1016/j.nima.2018.08.017 (cit. on pp. 11, 49).

[24] Y. Ohnishi et al. "Accelerator design at SuperKEKB". In: *PTEP* 2013 (2013), 03A011. DOI: 10.1093/ptep/pts083 (cit. on pp. 11, 49).

[25] A. Abashian, K. Gotow, N. Morgan, et al. "The Belle detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 479.1 (2002). Detectors for Asymmetric B-factories, pp. 117–232. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)02013-7 (cit. on pp. 11, 49, 61, 64).

[26] S. Kurokawa and E. Kikutani. "Overview of the KEKB accelerators". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment.* KEK-B: The KEK B-factory 499.1 (Feb. 2003), pp. 1–7. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(02)01771-0 (cit. on pp. 11, 49).

[27] T. Abe, K. Akai, N. Akasaka, et al. "Achievements of KEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 2013), 03A001. ISSN: 2050-3911. DOI: 10.1093/ptep/pts102 (cit. on pp. 11, 49).

[28] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097 (cit. on pp. 11, 17, 50, 110, 111).

[29] N. Katayama, R. Itoh, A. Manabe, et al. "Belle computing model". In: *Computer Physics Communications* 110.1 (May 1998), pp. 22–25. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(97)00148-3 (cit. on pp. 13, 69, 70).

[30] T. Kuhr, C. Pulvermacher, M. Ritter, et al. "The Belle II Core Software: Belle II Framework Software Group". In: *Computing and Software for Big Science* 3.1 (Nov. 2018). ISSN: 2510-2044. DOI: 10.1007/s41781-018-0017-9 (cit. on pp. 13, 69).

[31] T. Keck et al. "The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 6. DOI: 10.1007/s41781-019-0021-8. arXiv: 1807.08680 [hep-ex] (cit. on pp. 13, 17, 71–73, 110, 121).

[32] N. L. Johnson. "Systems of frequency curves generated by methods of translation". In: *Biometrika* 36 (1949), pp. 149–176. DOI: 10.1093/biomet/36.1-2.149 (cit. on pp. 20, 151).

[33] A. L. Read. "Presentation of search results: The CL_s technique". In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313 (cit. on pp. 20, 111, 176).

[34] G. Cowan, K. Cranmer, E. Gross, et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0 (cit. on pp. 20, 111, 176).

[35] R. Leboucher. “Search for $\tau^- \to \mu^-\mu^+\mu^-$ lepton flavour violating decays and measurement of the SVD cluster position resolution at Belle II experiment”. en. PhD thesis. Aix-Marseille Université, Oct. 2023. URL: https://hal.science/tel-04448902 (cit. on p. 28).

[36] L. Polat. “Search for $\tau \to \ell\phi$ lepton flavour violating decays and fake tracks rate measurement with $e^+e^- \to \tau^+\tau^-$ events at the Belle II experiment”. en. PhD thesis. Aix-Marseille université, Dec. 2022. URL: https://hal.science/tel-04071458 (cit. on p. 28).

[37] E. Majorana and L. Maiani. “A symmetric theory of electrons and positrons”. In: *Ettore Majorana Scientific Papers: On occasion of the centenary of his birth*. Ed. by Giuseppe Franco Bassani. Berlin, Heidelberg: Springer, 2006, pp. 201–233. ISBN: 978-3-540-48095-2. DOI: 10.1007/978-3-540-48095-2_10 (cit. on p. 28).

[38] Wikipedia contributors. *Standard Model*. 2025. URL: https://en.wikipedia.org/wiki/Standard_Model (cit. on p. 29).

[39] CMS Collaboration. “Observation of magic states of top quark pairs produced in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: (2025). URL: https://cds.cern.ch/record/2926751 (cit. on p. 29).

[40] LHCb Collaboration. “Observation of J/ψ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays”. In: *Physical Review Letters* 115.7 (Aug. 2015), p. 072001. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.115.072001 (cit. on p. 29).

[41] LHCb Collaboration. “Observation of a Narrow Pentaquark State, $P_c(4312)^+$, and of the Two-Peak Structure of the $P_c(4450)^+$”. In: *Physical Review Letters* 122.22 (June 2019), p. 222001. DOI: 10.1103/PhysRevLett.122.222001 (cit. on p. 29).

[42] BESIII Collaboration. “Observation of a Charged Charmoniumlike Structure in $e^+e^- \to \pi^+\pi^- J/\psi$ at $\sqrt{s}$=4.26 GeV”. In: *Physical Review Letters* 110.25 (June 2013), p. 252001. DOI: 10.1103/PhysRevLett.110.252001 (cit. on p. 29).

[43] Belle Collaboration. “Study of $e^+e^- \to \pi^+\pi^- J/\psi$ and Observation of a Charged Charmoniumlike State at Belle”. In: *Physical Review Letters* 110.25 (June 2013), p. 252002. DOI: 10.1103/PhysRevLett.110.252002 (cit. on p. 29).

[44] LHCb Collaboration. “Observation of the Resonant Character of the $Z(4430)^-$ State”. In: *Physical Review Letters* 112.22 (June 2014), p. 222002. DOI: 10.1103/PhysRevLett.112.222002 (cit. on p. 29).

[45] Planck Collaboration. “Planck 2018 results - VI. Cosmological parameters”. In: *Astronomy & Astrophysics* 641 (Sept. 2020), A6. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201833910 (cit. on p. 37).

[46] N. Bozorgnia, J. Bramante, J. M. Cline, et al. “Dark Matter Candidates and Searches”. In: arXiv:2410.23454 (Oct. 2024). DOI: 10.48550/arXiv.2410.23454 (cit. on p. 37).

[47] R. Davis, D. S. Harmer, and K. C. Hoffman. "Search for Neutrinos from the Sun". In: *Physical Review Letters* 20.21 (May 1968), pp. 1205–1209. DOI: 10.1103/PhysRevLett.20.1205 (cit. on p. 39).

[48] LHCb Collaboration. "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *Journal of High Energy Physics* 2017.8 (Aug. 2017), p. 55. ISSN: 1029-8479. DOI: 10.1007/JHEP08(2017)055 (cit. on p. 41).

[49] LHCb Collaboration. "Test of lepton universality in beauty-quark decays". In: *Nature Physics* 18.3 (Mar. 2022), pp. 277–282. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01478-8 (cit. on p. 41).

[50] LHCb Collaboration. "Test of Lepton Universality in $b \to s\ell^+\ell^-$ Decays". In: *Phys. Rev. Lett.* 131 (5 Aug. 2023), p. 051803. DOI: 10.1103/PhysRevLett.131.051803 (cit. on p. 41).

[51] LHCb Collaboration. "Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *Physical Review D* 108.3 (Aug. 2023), p. 032002. DOI: 10.1103/PhysRevD.108.032002 (cit. on p. 41).

[52] HFLAV Collaboration. *Average of R(D) and R(D*)*. June 2025. URL: https://hflav-eos.web.cern.ch/hflav-eos/semi/spring25/html/RDsDsstar/RDRDs.html (cit. on p. 41).

[53] LHCb Collaboration. "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$". In: *Physical Review Letters* 111.19 (Nov. 2013), p. 191801. DOI: 10.1103/PhysRevLett.111.191801 (cit. on p. 41).

[54] LHCb collaboration. "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity". en. In: *Journal of High Energy Physics* 2016.2 (Feb. 2016), p. 104. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2016)104 (cit. on p. 41).

[55] Belle Collaboration. "Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$". In: *Physical Review Letters* 118.11 (Mar. 2017), p. 111801. DOI: 10.1103/PhysRevLett.118.111801 (cit. on p. 41).

[56] L. Allwicher, D. Bečirević, G. Piazza, et al. "Understanding the first measurement of $\mathcal{B}(B \to K\nu\bar{\nu})$". In: *Physics Letters B* 848 (Jan. 2024), p. 138411. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2023.138411 (cit. on pp. 41, 44).

[57] I. Plakias and O. Sumensari. "Lepton Flavor Violation in Semileptonic Observables". In: arXiv:2312.14070 (July 2024). DOI: 10.48550/arXiv.2312.14070 (cit. on p. 42).

[58] D. Bečirević, F. Jaffredo, J. P. Pinheiro, et al. "Lepton flavor violation in exclusive $b \to d\ell_i\ell_j$ and $b \to s\ell_i\ell_j$ decay modes". In: arXiv:2407.19060 (July 2024). DOI: 10.48550/arXiv.2407.19060 (cit. on pp. 42, 43, 45).

[59] N. Gubernari, A. Kokulu, and D. van Dyk. “$B \to P$ and $B \to V$ Form Factors from B-Meson Light-Cone Sum Rules beyond Leading Twist”. In: *Journal of High Energy Physics* 2019.1 (Jan. 2019), p. 150. ISSN: 1029-8479. DOI: 10.1007/JHEP01(2019)150 (cit. on p. 43).

[60] A. Bharucha, D. M. Straub, and R. Zwicky. “$B \to V\ell^+\ell^-$ in the Standard Model from Light-Cone Sum Rules”. In: *Journal of High Energy Physics* 2016.8 (Aug. 2016), p. 98. ISSN: 1029-8479. DOI: 10.1007/JHEP08(2016)098 (cit. on p. 43).

[61] D. M. Straub. “flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond”. In: arXiv:1810.08132 (Oct. 2018). arXiv:1810.08132 [hep-ph]. DOI: 10.48550/arXiv.1810.08132 (cit. on p. 43).

[62] I. Doršner, S. Fajfer, A. Greljo, et al. “Physics of leptoquarks in precision experiments and at particle colliders”. In: *Physics Reports*. Physics of leptoquarks in precision experiments and at particle colliders 641 (June 2016), pp. 1–68. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2016.06.001 (cit. on p. 44).

[63] L. Calibbi, A. Crivellin, and T. Li. “Model of vector leptoquarks in view of the B-physics anomalies”. In: *Phys. Rev. D* 98 (11 Dec. 2018), p. 115002. DOI: 10.1103/PhysRevD.98.115002 (cit. on p. 44).

[64] D. Bečirević, I. Doršner, S. Fajfer, et al. “Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies”. In: *Physical Review D* 98.5 (Sept. 2018), p. 055003. DOI: 10.1103/PhysRevD.98.055003 (cit. on p. 44).

[65] J. Aebischer, G. Isidori, M. Pesut, et al. “Confronting the vector leptoquark hypothesis with new low- and high-energy data”. In: *The European Physical Journal C* 83.2 (Feb. 2023), p. 153. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11304-5 (cit. on p. 44).

[66] G. Hiller, D. Loose, and K. Schönwald. “Leptoquark flavor patterns and B decay anomalies”. In: *Journal of High Energy Physics* 2016.12 (Dec. 2016), p. 27. ISSN: 1029-8479. DOI: 10.1007/jhep12(2016)027 (cit. on p. 44).

[67] C. Hati, J. Kriewald, J. Orloff, et al. “The fate of $\mathbf{V_1}$ vector leptoquarks: the impact of future flavour data”. In: *The European Physical Journal C* 81.12 (Dec. 2021), p. 1066. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09824-z (cit. on p. 44).

[68] J. Fuentes-Martín, G. Isidori, M. König, et al. “Vector leptoquarks beyond tree level. III. Vectorlike fermions and flavor-changing transitions”. In: *Physical Review D* 102.11 (Dec. 2020), p. 115015. DOI: 10.1103/PhysRevD.102.115015 (cit. on p. 44).

[69] A. Angelescu, D. Be čirevi ć, D. A. Faroughy, et al. “Single leptoquark solutions to the B-physics anomalies”. In: *Phys. Rev. D* 104 (5 Sept. 2021), p. 055017. DOI: 10.1103/PhysRevD.104.055017 (cit. on p. 44).

[70] D. Bečirević, I. Doršner, S. Fajfer, et al. “Model with two scalar leptoquarks: R_2 and S_3”. In: *Physical Review D* 106.7 (Oct. 2022), p. 075023. DOI: 10.1103/PhysRevD.106.075023 (cit. on p. 44).

[71] A. Angelescu, D. Bečirević, D. A. Faroughy, et al. "Closing the window on single leptoquark solutions to the B-physics anomalies". en. In: *Journal of High Energy Physics* 2018.10 (Oct. 2018), p. 183. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2018)183 (cit. on p. 44).

[72] D. Bečirević, N. Košnik, O. Sumensari, et al. "Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive $b \to s\ell_1\ell_2$ modes". In: arXiv:1608.07583 (May 2019). DOI: 10.48550/arXiv.1608.07583 (cit. on p. 44).

[73] A. Crivellin, L. Hofer, J. Matias, et al. "Lepton-flavor violating B decays in generic $Z^{'}$ models". In: *Phys. Rev. D* 92 (5 Sept. 2015), p. 054013. DOI: 10.1103/PhysRevD.92.054013 (cit. on p. 45).

[74] D. Bečirević, O. Sumensari, and R. Z. Funchal. "Lepton flavor violation in exclusive $b \to s$ decays". In: *The European Physical Journal C* 76.3 (Mar. 2016), p. 134. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-3985-0 (cit. on p. 45).

[75] MEG Collaboration. "Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment". In: *Eur. Phys. J. C* 76.8 (2016), p. 434. DOI: 10.1140/epjc/s10052-016-4271-x (cit. on p. 45).

[76] SINDRUM II Collaboration. "A Search for muon to electron conversion in muonic gold". In: *Eur. Phys. J. C* 47 (2006), pp. 337–346. DOI: 10.1140/epjc/s2006-02582-x (cit. on p. 45).

[77] COMET Collaboration. *COMET Phase-I Technical Design Report.* Dec. 2018. DOI: 10.1093/ptep/ptz125 (cit. on p. 45).

[78] N. Teshima. *Status of the DeeMe Experiment, an Experimental Search for μ-e Conversion at J-PARC MLF.* 2019. arXiv: 1911.07143 [physics.ins-det] (cit. on p. 45).

[79] L. Bartoszek, E. Barnes, J. P. Miller, et al. *Mu2e Technical Design Report.* 2015. arXiv: 1501.05241 [physics.ins-det] (cit. on p. 45).

[80] P.M. Lewis, I. Jaegle, H. Nakayama, et al. "First measurements of beam backgrounds at SuperKEKB". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 914 (2019), pp. 69–144. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.05.071 (cit. on p. 49).

[81] SuperB Collaboration. *SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report.* 2007. arXiv: 0709.0451 [hep-ex] (cit. on p. 51).

[82] Belle-ECL, V Aulchenko, A Bobrov, et al. "Electromagnetic calorimeter for Belle II". In: *Journal of Physics: Conference Series* 587.1 (Feb. 2015), p. 012045. DOI: 10.1088/1742-6596/587/1/012045 (cit. on pp. 55, 63).

[83] D. Matvienko. “The Belle II experiment: status and physics program”. In: *EPJ Web of Conferences* 191 (Jan. 2018), p. 02010. DOI: 10.1051/epjconf/201819102010 (cit. on p. 56).

[84] S. Mondal, K. Adamczyk, L. Aggarwal, et al. “Silicon vertex detector of the Belle II experiment”. In: *Journal of Instrumentation* 19.02 (Feb. 2024), p. C02038. ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/02/c02038 (cit. on pp. 56, 57).

[85] H. Ye, F. Abudinen, K. Ackermann, et al. “Commissioning and performance of the Belle II pixel detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 987 (2021), p. 164875. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2020.164875 (cit. on pp. 56, 57).

[86] Z. Natkaniec, H. Aihara, Y. Asano, et al. “Status of the Belle silicon vertex detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 560.1 (2006). Proceedings of the 13th International Workshop on Vertex Detectors, pp. 1–4. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2005.11.228 (cit. on pp. 57, 58).

[87] K. Adamczyk, L. Aggarwal, H. Aihara, et al. “The design, construction, operation and performance of the Belle II silicon vertex detector”. In: *Journal of Instrumentation* 17.11 (Nov. 2022), P11042. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/11/p11042 (cit. on pp. 57, 58).

[88] T.V. Dong, S. Uno, M. Uchida, et al. “Calibration and alignment of the Belle II central drift chamber”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 930 (2019), pp. 132–141. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2019.03.072 (cit. on pp. 58, 60).

[89] H. Hirano, M. Akatsu, Y. Fujita, et al. “A high-resolution cylindrical drift chamber for the KEK B-factory”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 455.2 (2000), pp. 294–304. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(00)00513-1 (cit. on p. 58).

[90] V. Bertacchi, T. Bilka, N. Braun, et al. “Track finding at Belle II”. In: *Computer Physics Communications* 259 (Feb. 2021), p. 107610. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2020.107610 (cit. on pp. 60, 66, 68).

[91] T Iijima, I Adachi, M Amami, et al. “Aerogel Cherenkov counter for the BELLE experiment”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment.* Proceedings of the Sixth International Conference on Instrumentation for Experiments at e+ e- Colliders 379.3 (Sept. 1996), pp. 457–459. ISSN: 0168-9002. DOI: 10.1016/0168-9002(96)00609-2 (cit. on p. 60).

[92] H. Kichimi, Y. Yoshimura, T. Browder, et al. “The BELLE TOF system”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics 453.1 (Oct. 2000), pp. 315–320. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(00)00651-3 (cit. on p. 61).

[93] R. Pestotnik, I. Adachi, R. Dolenec, et al. “The aerogel Ring Imaging Cherenkov system at the Belle II spectrometer”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016) 876 (Dec. 2017), pp. 265–268. ISSN: 0168-9002. DOI: 10.1016/j.nima.2017.04.043 (cit. on pp. 61, 62).

[94] S. Iwata, I. Adachi, K. Hara, et al. “Particle identification performance of the prototype aerogel RICH counter for the Belle II experiment”. In: *Progress of Theoretical and Experimental Physics* 2016.3 (Mar. 2016), 033H01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptw005 (cit. on p. 62).

[95] U. Tamponi. “The TOP counter of Belle II: Status and first results”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 10th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2018) 952 (Feb. 2020), p. 162208. ISSN: 0168-9002. DOI: 10.1016/j.nima.2019.05.049 (cit. on p. 62).

[96] K. Miyabayashi. “Belle electromagnetic calorimeter”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Proceedings of the 8th International Conference on Instrumentatio n for Colliding Beam Physics 494.1 (Nov. 2002), pp. 298–302. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(02)01483-3 (cit. on p. 63).

[97] M. Nakao, T. Higuchi, R. Itoh, et al. “Data acquisition system for Belle II”. en. In: *Journal of Instrumentation* 5.12 (Dec. 2010), p. C12004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/5/12/C12004 (cit. on p. 64).

[98] Y. Iwasaki, B. Cheon, E. Won, et al. “Level 1 trigger system for the Belle II experiment”. In: *2010 17th IEEE-NPSS Real Time Conference*. May 2010, pp. 1–9. DOI: 10.1109/RTC.2010.5750454 (cit. on p. 65).

[99] S Lee, R Itoh, T Higuchi, et al. “Belle-II High Level Trigger at SuperKEKB”. en. In: *Journal of Physics: Conference Series* 396.1 (Dec. 2012), p. 012029. ISSN: 1742-6596. DOI: 10.1088/1742-6596/396/1/012029 (cit. on p. 65).

[100] A. Glazov, P. Rados, A. Rostomyan, et al. “Measurement of the tracking efficiency in Phase 3 data using tau-pair events. ” In: (Feb. 2020). BELLE2-NOTE-PH-2020-006. URL: https://docs.belle2.org/record/1867? (cit. on p. 67).

[101] P. Rados and L. Zani. “Measurement of the Tracking Efficiency with Tau-pair Events in Run1 Data and MC15rd”. In: (July 2024). BELLE2-NOTE-PH-2024-037. URL: https://docs.belle2.org/pub_data/documents/24/ (cit. on p. 67).

[102] The Belle II Collaboration Charged PID Group. "Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2022". In: (Mar. 2022). BELLE2-CONF-PH-2022-003. URL: https://docs.belle2.org/pub_data/documents/3274/ (cit. on p. 68).

[103] Vismaya V. S. S. Maharana S. Sandilya. "Study of Kaon and Pion Identification Performances in Phase III data using the decay $D^{*+} \to D^0[K^-\pi^+\pi^0]\pi^+$". In: (Feb. 2023). BELLE2-NOTE-PH-2022-051. URL: https://docs.belle2.org/pub_data/documents/148/ (cit. on p. 68).

[104] K. Hanagaki, H. Kakuno, H. Ikeda, et al. "Electron identification in Belle". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 485.3 (June 2002), pp. 490–503. ISSN: 0168-9002. DOI: 10.1016/s0168-9002(01)02113-1 (cit. on p. 69).

[105] A. Abashian, K. Abe, P. K. Behera, et al. "Muon identification in the Belle experiment at KEKB". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 491.1 (Sept. 2002), pp. 69–82. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(02)01164-6 (cit. on p. 69).

[106] Belle II Software Group. *Belle II Software documentation*. URL: https://software.belle2.org/development/sphinx/index.html (cit. on p. 69).

[107] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. New Computing Techniques in Physics Research V 389.1 (Apr. 1997), pp. 81–86. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00048-X (cit. on p. 70).

[108] M. Gelb, T. Keck, M. Prim, et al. "B2BII: Data Conversion from Belle to Belle II". In: *Computing and Software for Big Science* 2.1 (Nov. 2018). ISSN: 2510-2044. DOI: 10.1007/s41781-018-0016-x (cit. on p. 70).

[109] M. Feindt, F. Keller, M. Kreps, et al. "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (Oct. 2011), pp. 432–440. ISSN: 0168-9002. DOI: 10.1016/j.nima.2011.06.008 (cit. on p. 71).

[110] W. Sutcliffe, N. Rout, M. Liu, et al. "Combination of calibration factors for hadronic tagging". In: *BELLE2-NOTE-PH-2023-029* (June 2023) (cit. on pp. 71, 162).

[111] M. Liu, N. Rout, K. Trabelsi, et al. "Hadronic FEI calibration with $D^{(*)}\pi$ samples". In: *BELLE2-NOTE-PH-2023-004* (Jan. 2023) (cit. on pp. 71, 163).

[112] A. Ryd, D. Lange, N. Kuznetsova, et al. "EvtGen: A Monte Carlo Generator for B-Physics". In: (May 2005). URL: https://phab.hepforge.org/source/evtgen/ (cit. on pp. 73, 112).

[113] C. Bierlich, S. Chakraborty, N. Desai, et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: *SciPost Phys. Codebases* (2022), p. 8. DOI: 10.21468/SciPostPhysCodeb.8 (cit. on pp. 73, 112).

[114] S. Agostinelli, J. Allison, K. Amako, et al. "Geant4 - a simulation toolkit". In: *Nucl. Instrum. Methods Phys. Res. A* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8 (cit. on pp. 73, 112).

[115] R. Brun, F. Bruyant, F. Carminati, et al. "GEANT Detector Description and Simulation Tool". In: (Oct. 1994). DOI: 10.17181/CERN.MUHF.DMJ1 (cit. on pp. 73, 112).

[116] P. Billoir. "Progressive track recognition with a Kalman-like fitting procedure". In: *Computer Physics Communications* 57.1 (Dec. 1989), pp. 390–394. ISSN: 0010-4655. DOI: 10.1016/0010-4655(89)90249-X (cit. on p. 76).

[117] K. Trabelsi. *Belle II High Level Trigger*. Jan. 2023. URL: https://indico.belle2.org/event/8043/contributions/51065/attachments/20558/30432/HLT_20230130.pdf (cit. on pp. 79, 80).

[118] V. Bertacchi. *Status and Plan for DAF optimisation*. May 2022. URL: https://indico.belle2.org/event/6621/contributions/35146/attachments/16583/24806/bertacchi%20-%20F2F%20tracking%20Marseille%20-%20DAF%20optimization.pdf (cit. on p. 80).

[119] D. Pitzl. *CDC tracks in release 8*. Feb. 2025. URL: https://indico.belle2.org/event/14521/contributions/89077/attachments/32908/48612/Pitzl-2025-02-cdc.pdf (cit. on p. 89).

[120] The Belle II Tracking Group. "Tracking Performance in Early Belle II Data". In: (Apr. 2022). BELLE2-NOTE-TE-2022-006 (cit. on pp. 91, 94).

[121] S. Banerjee and A. Pathak. "Update of TAUOLA Monte Carlo for Belle II software". In: (July 2020). BELLE2-NOTE-PH-2020-055 (cit. on p. 94).

[122] Belle Collaboration. *Number of B events in Hadron B(J)*. URL: https://belle.kek.jp/secured/nbb/nbb.html (cit. on pp. 111, 112, 171).

[123] Belle II Collaboration. *Conference readiness*. URL: https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Conference+readiness (cit. on pp. 111, 113, 164, 171).

[124] G. Punzi. "Sensitivity of searches for new signals and its optimization". In: *eConf* C030908 (2003). Ed. by L. Lyons, R. P. Mount, and R. Reitmeyer, MODT002 (cit. on pp. 116, 141).

[125] J. D. Bjorken and S. J. Brodsky. "Statistical Model for Electron-Positron Annihilation into Hadrons". In: *Physical Review D* 1.5 (Mar. 1970), pp. 1416–1420. DOI: 10.1103/PhysRevD.1.1416 (cit. on p. 118).

[126] G. C. Fox and S. Wolfram. "Event shapes in e^+e^- annihilation". In: *Nuclear Physics B* 149.3 (Mar. 1979), pp. 413–496. ISSN: 0550-3213. DOI: 10.1016/0550-3213(79)90003-8 (cit. on p. 118).

[127] G. C. Fox and S. Wolfram. "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes". In: *Phys. Rev. Lett.* 41 (23 Dec. 1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581 (cit. on pp. 118, 133, 134).

[128] M. Dasgupta and G. P. Salam. "Event shapes in e^+e^- annihilation and deep inelastic scattering". In: *Journal of Physics G: Nuclear and Particle Physics* 30.5 (Feb. 2004), R143–R181. ISSN: 1361-6471. DOI: 10.1088/0954-3899/30/5/r01 (cit. on p. 118).

[129] Y. Coadou. "Boosted Decision Trees". In: *Artificial Intelligence for High Energy Physics*. WORLD SCIENTIFIC, Feb. 2022, pp. 9–58. ISBN: 9789811234033. DOI: 10.1142/9789811234033_0002 (cit. on p. 132).

[130] T. Keck. "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification". In: *Computing and Software for Big Science* 1 (Sept. 2017). ISSN: 2510-2044. DOI: 10.1007/s41781-017-0002-8 (cit. on p. 132).

[131] T. Akiba, S. Sano, T. Yanase, et al. *Optuna: A Next-generation Hyperparameter Optimization Framework*. 2019. arXiv: 1907.10902 [cs.LG] (cit. on p. 132).

[132] X. Zhou, S. Du, G. Li, et al. "TopoAna: A generic tool for the event type analysis of inclusive Monte-Carlo samples in high energy physics experiments". In: *Computer Physics Communications* 258 (2021), p. 107540. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2020.107540 (cit. on p. 145).

[133] L. Hinz. "Lepton ID effiency correction and systematic error". In: *BELLE-NOTE-954* (July 2006) (cit. on p. 163).

[134] S. Nishida. "Study of Kaon and Pion Identifiction Using Inclusive D^* Sample". In: *BELLE-NOTE-779* (Jan. 2005) (cit. on p. 163).

[135] M. Milesi, R. Dhamija, P. Feichtinger, et al. "Evaluation of electron- and muon-identification performance with Belle II data from 2019 to 2021". In: *BELLE2-PUB-DRAFT-2022-010* (Oct. 2022) (cit. on p. 163).

[136] Belle II Collaboration. *Systematic Corrections Framework*. URL: https://syscorrfw.readthedocs.io/en/latest/ (cit. on p. 163).

[137] Belle II Collaboration. "Charged-hadron identification at Belle II". en. In: *arXiv.org* (June 2025). URL: https://arxiv.org/abs/2506.04355v2 (cit. on p. 163).

[138] Belle Collaboration. *Systematics*. URL: https://belle.kek.jp/secured/wiki/doku.php?id=software:systematics (cit. on p. 164).

[139] A. Alavi-Harati et al. “Measurements of Direct CP Violation, CPT Symmetry, and Other Parameters in the Neutral Kaon System”. In: *Phys. Rev. D* 67 (2003). [Erratum: Phys.Rev.D 70, 079904 (2004)], p. 012005. DOI: 10.1103/PhysRevD.70.079904 (cit. on p. 164).