NNT/NL : 2023AIXM0358/048ED352

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 31 octobre 2023 par

Robin LEBOUCHER

Search for $\tau^- \rightarrow \mu^-\mu^+\mu^-$ lepton flavour violating decays and measurement of the SVD cluster position resolution at Belle II experiment

Discipline
Physique et Sciences de la Matière

Spécialité
Physique des Particules et Astroparticules

École doctorale
Physique et Sciences de la matière (ED 352)

Laboratoire/Partenaires de recherche
Centre de Physique des Particules
de Marseille (CPPM) UMR 7346
Centre National de la Recherche Scientifique (CNRS)

Composition du jury

Sasha GLAZOV DESY, Hamburg, Germany	Rapporteur
Patrick ROBBE IJCLab (CNRS/IN2P3), Orsay, France	Rapporteur
Victor COCO CERN, Geneva, Switzerland	Examinateur
Cristinel DIACONU CPPM (CNRS/IN2P3), Marseille, France	Président du jury
Justine SERRANO CPPM (CNRS/IN2P3), Marseille, France	Directrice de thèse

Affidavit

Je soussigné, Robin Leboucher, 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 d'Aix-Marseille Université 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 28 août 2023

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et participation aux conférences

Liste des publications réalisées dans le cadre du projet de thèse :

- R. Leboucher, A. Martini, A. Rostomyan, and J. Serrano, Search for lepton-flavor-violating $\tau^- \to \mu^-\mu^+\mu^-$ decays in Belle II data, Belle II Internal Note, Jul. 2023, BELLE2-NOTE-PH-2023-034
- R. Leboucher, K. Adamczyk, L. Aggarwal, et al., Measurement of the cluster position resolution of the Belle II Silicon Vertex Detector, NIMA, vol. 1033, Jun. 2022, p. 166 746. doi : 10.1016/j.nima.2022.166746
- G. Casarosa, G. Dujany, C. Finck, et al., Measurement of the SVD Cluster Position Resolution, Belle II Internal Note, Mar. 2022, BELLE2-NOTE-TE-2022-005
- R. Leboucher, J. Serrano, L. Zani, and F. Tenchini, Good track selection for tau events, Belle II Internal Note, Jun. 2020, BELLE2-NOTE-PH-2020-029

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

Conférences :

- **The 21st International Conference on B-Physics at Frontier Machines, "BEAUTY 2023"**, Jul. 6th 2023, *Clermont-Ferrand, France*, Title : *Dark matter and tau results at Belle II.*
- **45^{th} Belle II General Meeting (B2GM)**, Jun. 7th 2023, *Nagoya (online), Japan*, Shared talk with A. Martini (Desy, Hamburg, Germany), Title : *Status of the $\tau \to \mu\mu\mu$ lepton flavour violation decays searches at Belle II.*
- **Annual workshop of Groupement de Recherche "Intensity Frontier" (GdR-InF)**, Nov. 4th 2022, *Lyon, France*, Shared talk with L. Polat (CPPM, Marseille, France), Title : *Overview of tau lepton flavour violating decays at Belle II.*
- **Annual workshop of Groupement de Recherche "Intensity Frontier" (GdR-InF)**, Nov. 17th 2021 *Orsay, France*, Shared talk with L. Martel (IPHC, Strasbourg, France), Title : *Measurements of the cluster position resolution of the Belle II Silicon Vertex Detector.*
- **The 30th International Workshop on Vertex Detectors**, Sep. 27th 2021, *Online Organised by University of Oxford, United Kingdom*, Title : *Measurements of the*

cluster position resolution of the Belle II Silicon Vertex Detector.

Ecoles d'été :

- **IN2P3 School of Statistics 2021**, May 2022, *Carry-le-Rouet, France* by members of Institut national de physique nucléaire et de physique des particles.
- **2021 Belle II Physics Week : Starter Kit**, Nov. 2021 *Rome, Italy* by members of the Belle II experiment.

Remerciements

La rédaction de cette thèse conclue un long voyage et une longue partie de ma vie. En débutant ce voyage en même temps que mes études je n'imaginais pas pouvoir en arriver jusqu'ici. En réalité je refusais même d'y songer ne m'estimant pas à ma place pour y parvenir. C'est ainsi que tel *Frodon* amenant l'anneau jusqu'à *Fondcombe*, ou dans mon cas jusqu'au master, je ne voulais pas songer à la suite de l'aventure. Les passages m'y ayant mené m'ont fait douter de mes capacités et de ma légitimité à poursuivre dans ce que j'aimais faire, apprendre et comprendre le monde. C'est dans ces moments que ma Communauté, si je file la métaphore fantaisiste, fut d'un support indéfectible.

En recherchant un stage de fin de Master, j'ai rencontré ma future directrice de thèse, Justine, qui a fait confiance à mon parcours atypique, provenant d'un Master de Physique numérique de Besançon (coucou les clochards de la 313C). Elle a continué à me faire confiance pour poursuivre mon travail en thèse malgré cette fois-ci d'excellents candidats (coucou Arthur). Et ainsi j'ai repris le voyage depuis *Fondcombe* vers l'accomplissement de la thèse. Pour cette confiance mais aussi ta patience, ta disponibilité, je voudrais très sincèrement remercier Justine.

Durant ces années de thèse, Justine a formé une formidable équipe : la *Justine League*, que je souhaite remercier pour leur bienveillance, leur patience et leur enthousiasme communicatif. A Giampi, Laura, Klemens et Valerio, d'avoir toujours été là pour éclaircir les mystères de Belle II. En particulier je voudrais du plus profond de mon cœur dire merci à Laura, tu as toujours répondu présente pour nous les doctorants, tant bien du point de vue d'experte en physique, que d'un point de vue humain. Avec qui l'on se sent en sécurité pour discuter quand ça ne va pas. J'ai aussi une pensée pour mes collègues doctorants, Leonard qui a ouvert la voie que l'on a tous suivi, Arthur pour nos longues discussions et ton réel optimisme, Clothilde et Isaac. Ce fut un plaisir et un honneur de travailler avec vous et c'est avec un certain pincement que je quitte Belle 2 Marseille.

Une salutation aux différentes communautés de doctorant s'impose. Qu'elles soient sur Discord, avec *"Bien Dans Ma Thèse"*, ou dans Belle 2, avec *les étoiles* de Strasbourg, elles sont un des lieux ou l'on trouvera de l'aide, du soutien et une oreille pour nos tracas malgré la distance physique, sanitaire et de domaine scientifique.

J'aimerais remercier les membres de la collaboration de Belle 2 et du CPPM, et plus particulièrement l'administration et la direction, pour offrir un accueil chaleureux et bienveillant, et de faire vivre les lieux de sciences.

Aux membres du jury de thèse, pour avoir répondu présent et avoir donné du temps pour valider toutes ces années de travail et m'accueillir dans la communauté des physiciens.

Enfin, ce périple, que dis-je cette quête, aurait été impossible sans le support indéfectible, l'aide et leur confiance quand j'en manquais, de ma famille. Malgré la distance, qui se fait plus dure aujourd'hui, j'ai depuis toujours ressenti votre amour. A mes parents, pour nous avoir transmis leur amour, leurs valeurs d'humanité et de solidarité. À mon frère et ma sœur qui furent et demeurent mes modèles à suivre dans la vie. A toi Timothé je souhaite que le monde puisse s'inspirer un peu plus de nos collaborations scientifiques pour bâtir une société plus solidaire face aux enjeux à venir.

Pour finir, cette thèse n'aurait pas existé sans la personne avec qui je partage ma vie depuis huit ans, Laura. La personne qui a été présente dans les meilleurs mais aussi les plus durs moments de doute. Pour m'avoir si souvent ramassé en larmes lors de mes crises d'angoisse et de fatigue, armée de ta patience, et ton réconfort. Tu es un phare dans ma vie, ma lumière d'*Eärendil* qui luit d'une douce et chaleureuse lumière dans les ténèbres pour que je trouve ma voie auprès de toi. Cette aventure aurait été insurmontable sans toi et je t'en suis infiniment reconnaissant. Je suis conscient qu'une thèse peut être éprouvante pour un couple avec les sacrifices qui vont avec mais je suis fière d'avoir traversé cette aventure avec toi et je suis prêt pour les prochaines.

Merci profondément d'avoir partager cette aventure!

Résumé

Mot-clefs : Physique des Particules; Recherche de nouvelle physique; Violation de la saveur leptonique; Désintégration du lepton tau; Résolution spatiale; Détecteur de vertex

Les désintégrations violant la saveur des leptons sont considérées comme l'un des moyens les plus efficaces de rechercher de la physique au-delà du modèle standard, car elles ne sont pas autorisées dans le modèle standard de la physique des particules. Un certain nombre de modèles de nouvelle physique prédisent que les fractions d'embranchement de $\tau^- \to \mu^-\mu^+\mu^-$ sont juste en dessous des limites expérimentales actuelles.

L'expérience Belle II, qui opère au laboratoire KEK au Japon, a déjà enregistré une luminosité de 424 fb^{-1} entre 2019 et 2022 à l'énergie de résonance des mésons $\Upsilon(4S)$ $\sqrt{s} = 10,58\,\mathrm{GeV}$ et $\Upsilon(5S)$. De plus Belle II fourni un environnement idéal pour étudier les désintégrations de tau en raison de son environnement propre et de la section efficace élevée de $\tau^-\tau^+$.

Cette thèse présente une recherche de désintégration $\tau^- \to \mu^-\mu^+\mu^-$ dans les événements $e^+e^- \to \tau^+\tau^-$. La stratégie est basée sur une reconstruction "non étiquetée" d'un tau en trois muons, tandis que le tau de charge opposée n'est pas contraint afin de maximiser l'efficacité de la sélection du signal. Pour rejeter le bruit de fond, une stratégie en trois étapes est adoptée, basée sur des variables d'identification des muons, des sélections préliminaires et des arbres de décision boostés, qui prennent comme entrées des variables cinématiques, topologiques et d'autres variables liées à l'événement. Le rejet du bruit de fond est optimisé à l'aide d'échantillons simulés par Monte-Carlo afin de minimiser la figure de mérite de Punzi. Après avoir déterminé le nombre attendu de données et l'incertitude systématique, la limite supérieure attendue des rapports d'embranchement $\tau^- \to \mu^-\mu^+\mu^-$ est estimée à 1.56×10^{-8} à 90% de niveau de confiance en utilisant 424 fb^{-1}. On s'attend, donc à ce que la limite de Belle établie à 2.1×10^{-8} puisse être améliorée avec la moitié de sa luminosité.

En outre, la thèse comprend également une mesure de la résolution spatiale du détecteur de vertex en utilisant des capteurs qui se chevauchent. Les événements de type $e^+e^- \to \mu^-\mu^+$ sont sélectionnés pour ne conserver que le cas où une particule a laissé deux impacts dans la même couche du détecteur. La méthode estime que la résolution spatiale est approximativement comprise entre 15 μm et 32 μm en fonction de la couche et du côté du détecteur.

Abstract

Keywords: Particle physics; Search for New Physics; Lepton Flavour Violation; Tau lepton decays; Spatial resolution; Vertex detector

Lepton flavour violating decays are considered one of the most effective ways to search for physics beyond the standard model, as they are not allowed in the Standard Model of particle physics. A number of new physics models predict that the branching fractions of $\tau^- \to \mu^-\mu^+\mu^-$ are just below the current experimental limits.

The Belle II experiment, which operates at the KEK laboratory in Japan, has already collected a luminosity of $424\,\mathrm{fb}^{-1}$ between 2019 and 2022 at the $\Upsilon(4S)$ $\sqrt{s} = 10.58\,\mathrm{GeV}$ and $\Upsilon(5S)$ mesons resonance energies. In addition, Belle II provides an ideal environment to study tau decays due to its clean environment and high $\tau^-\tau^+$ cross-section.

This thesis presents a search for $\tau^- \to \mu^-\mu^+\mu^-$ decays in $e^+e^- \to \tau^+\tau^-$ events. The strategy is based on an "untagged" reconstruction of one tau going into three muons, while the oppositely charged tau is left unconstrained to maximize the signal selection efficiency. A three-step strategy is adopted to reject background based on muon identification variables, cut-based selections, and boosted decision trees, which take kinematic, topological, and other event-related variables as inputs. The background rejection is optimized using Monte-Carlo simulated samples to minimize Punzi's figure of merit. After determining the expected number of data and the systematic uncertainty, the expected upper limit on $\tau^- \to \mu^-\mu^+\mu^-$ branching fractions is estimated to be 1.56×10^{-8} at 90% confidence level using $424\,\mathrm{fb}^{-1}$. Thus, the limit of Belle that reached 2.1×10^{-8} is expected to be improved with half of the luminosity.

Moreover, the thesis also includes measuring the vertex detector spatial resolution using overlapping sensors. The $e^+e^- \to \mu^-\mu^+$ data event is selected to keep only the case where a particle has left two hits in the same detector layer. The method estimates the spatial resolution approximately between $15\,\mu\mathrm{m}$ and $32\,\mu\mathrm{m}$, depending on the layer and the sensor side.

Résumé Long

Le modèle standard et la violation de la saveur leptonique

Le modèle standard de la physique des particules est une théorie développée au 20e siècle pour expliquer les constituants fondamentaux de la nature et la manière dont ils interagissent les uns avec les autres [1]. Au fil du temps, de nombreux résultats expérimentaux ont confirmé l'exactitude des prédictions du modèle standard concernant les mécanismes à l'échelle subatomique. Cette théorie repose sur la description des particules élémentaires par des champs répondant aux propriétés de la relativité restreinte et de la mécanique quantique. Deux types de particules se distinguent : les fermions et les bosons. Les fermions se répartissent en quarks (up, down, strange, charm, top et beauty) qui composent le noyaux atomique et en leptons chargés (électrons, muons et tauons) et neutres (neutrinos électronique, muonic et tauic). Les bosons comprenennt les bosons jauge, médiateurs des trois forces fondamentales, et enfin le boson de Higgs responsable de la masse intrinsèque des particules. Il existe aussi des antiparticules qui ont les mêmes propriétés que les particules usuelles mais avec des nombres quantiques opposées.

Toutefois le modèle standard faillit dans la description de plusieurs phénomènes observés, avec entre autres :

- l'existence de la matière noire et de l'énergie sombre, prouvée par des observations en astrophysiques et cosmologie,
- l'unification de l'interaction gravitationnelle avec les autres interactions décrites par la théorie quantique des champs,
- l'asymétrie d'abondance entre la matière et l'antimatière dans l'univers,
- la faible masse des neutrinos introduite par le phénomène d'oscillation des neutrinos entre les saveurs électroniques, muoniques et tauiques.

Ces phénomènes conduisant les physiciens à suggérer que le modèle standard est une "théorie effective", limitée à un certain domaine d'énergie, tandis qu'une théorie plus générale l'engloberait. Dans ce cadre sont développés de nombreux modèles allant au-delà du modèle standard, aussi appelés "nouvelle physique", afin de pallier ces lacunes. Ces modèles de nouvelle physique introduisent des champs quantiques additionnels, avec éventuellement de nouvelles symétries de jauge, dans le lagrangien du modèle standard qui décrit les interactions entre particules. Ces nouvelles particules existeraient à des énergies supérieures à celles explorées de nos jours mais joueraient un rôle dans des phénomènes observés aux énergies plus faibles.

Nous pouvons ranger les leptons en trois différentes saveurs, électronique, muonique et tauique, chacune composée d'un lepton chargé et d'un neutrino. En associant un nombre leptonique à chacune des trois saveurs, une propriété de conservation de ces nombres apparaît accidentellement. Il existe toutefois une source de violation de cette saveur leptonique dans le secteur des neutrinos avec leurs oscillations. Les oscillations de neutrinos sont aussi responsables d'une violation de la saveur leptonique dans le secteur des leptons chargés, cependant de tels phénomènes ont des taux d'apparition inférieurs à 10^{-50} [2, 3]. Ce taux est loin des sensibilités des expériences actuelles et même futures. L'observation de telles désintégrations violant la saveur leptonique dans les collisionneurs de particule serait une preuve indéniable de "nouvelle physique". De plus certains modèles comme ceux de la supersymmétrie, ou du boson de Higgs léger, ou de nouveaux bosons de jauges [4, 5] joueraient un role dans la désintégration $\tau^- \to \mu^-\mu^+\mu^-$ et augmenteraient le taux d'apparition en desous de 10^{-8}. A ces niveaux les expériences actuelles et futures seraient en capacité d'observer de telles désintégrations.

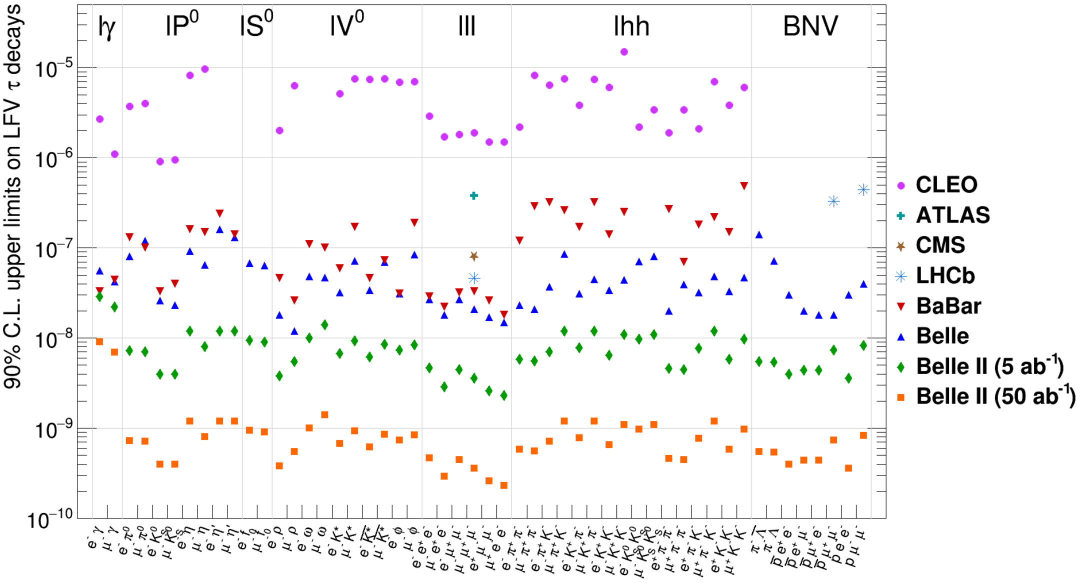


FIGURE 1. – Limites supérieures sur les rapports d'embranchements de diverses désintégrations du tauon violant la saveur leptonique à un niveau de confiance de 90%, établies par les expériences CLEO, ATLAS, CMS, LHCb, BABAR et Belle, et estimées dans Belle II pour des luminosités intégrées de 5 (losanges verts) et 50 ab^{-1} (carrés oranges). Extrait de [6].

Parmis les canaux de désintégration violant la saveur leptonique étudiés dans l'expérience Belle II, certains sont qualifiés de "canaux d'or" par leur implication dans de nouveaux modèles ou par leur faible contamination par le bruit de fond. Le canal $\tau^- \to \mu^-\mu^+\mu^-$ étudié dans cette thèse en fait partie grâce à son état final composé exclusivement de leptons, le rendant facilement identifiable par rapport au bruit de fond. L'expérience Belle II promet, sur le long terme, d'améliorer d'un à deux ordres de grandeur les limites supérieures actuelles sur les rapports d'embranchement de désintégrations du leptons tau violants la saveur (voir Figure 1). Les meilleurs mesures à l'heure actuelle ont été établies par l'expérience Belle à 2.1×10^{-8} à 90% de niveau de confiance comme présenté dans le Tableau 1.

TABLE 1. – Valeurs expérimentales sur la limite supérieure du taux de désintégration du canal du tauon en trois leptons à 90% de niveau de confiance établies par les expériences BABAR [7], Belle [8] et celles du Large Hadron Collider [9, 10, 11]. Les valeurs présentées sont des multiples de 10^8.

Mode	Belle	Babar	LHCb	ATLAS	CMS
$\mu^-\mu^+\mu^-$	2.1	3.3	4.6	3.8	2.9
$e^-e^+e^-$	2.7	2.9	-	-	-
$e^-\mu^+\mu^-$	2.7	3.2	-	-	-
$e^-e^+\mu^-$	1.8	2.2	-	-	-
$e^+\mu^-\mu^-$	1.7	2.6	-	-	-
$\mu^+e^-e^-$	1.5	1.8	-	-	-

L'expérience Belle II

L'expérience Belle II [12] est menée au laboratoire KEK de Tsukuba, au Japon. Elle se concentre sur l'étude des propriétés des mésons, qui sont des particules composées d'un quark et d'un antiquark, notamment les mésons beaux, charme et les leptons tau. L'objectif est de mesurer avec précision les propriétés du modèle standard et de rechercher de nouvelles formes de physique. Les installations se composent de deux éléments principaux : le collisionneur d'électrons et de positrons SuperKEKB et le détecteur Belle II, qui sont des mises à niveau des installations précédentes utilisées dans l'expérience Belle de 1998 à 2010.

L'accélérateur de particules SuperKEKB produit des collisions entre électron et positron à une énergie de $\sqrt{s} = 10.58\,\mathrm{GeV}$ [13] dans le référentiel du centre de masse. Cette énergie correspond à l'énergie de résonance du méson $\Upsilon(4S)$ constitué d'une paire de quark-antiquark beaux. Les propriétés du mésons B, produit lors des désintégrations du $\Upsilon(4S)$, notamment la recherche de violation de Charge et Parité dans ses désintégrations, sont le sujet d'étude privilégié de Belle II. Cependant les collisions réalisées à SuperKEKB produisent aussi d'autres particules en grande quantité comme des paires de quarks plus légers ($u\overline{u}$, $d\overline{d}$, $c\overline{c}$ et $s\overline{s}$) mais aussi des paires de leptons : $e^+e^- \to e^+e^-$ (diffusion Bhabha), $\mu^-\mu^+$ ou $\tau^-\tau^+$. Ainsi Belle II est une expérience parfaitement adaptée à l'étude des leptons tau avec une section efficace de production de $\tau^-\tau^+$ à $0.919\,\mathrm{nb}$ proche de celle du méson $\Upsilon(4S)$ à $1.110\,\mathrm{nb}$. Pour atteindre l'énergie de résonance $\Upsilon(4S)$, les électrons et positrons sont accélérés dans deux anneaux circulaires avant leur collision, le Low Energy Ring (LER) où les positrons atteignent une énergie de 4 GeV et le High Energy Ring (HER) où les électrons sont accélérés à 7 GeV. La différence d'énergie crée un "boost" pour les particules produites durant la collision et permet de les propulser dans une direction privilégiée (l'axe du faisceau dans le sens des électrons) dans le référentiel du laboratoire. Cette direction privilégiée permet d'obtenir une meilleure séparation spatiale de leurs désintégrations.

SuperKEKB se distingue des autres accélérateurs dans le monde par sa luminosité instantanée record et dont l'objectif est d'atteindre $\mathscr{L} = 6 \times 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$ soit 30 fois celle enregistrée par KEKB. Cette objectif est possible grâce à une avancée technologique récente réduisant la taille du faisceau dans lequel les particules sont concentrées, tout en imposant une intensité électrique environ 1.5 fois plus grande. A terme l'expérience Belle II a pour objectif de proposer un échantillon de données d'une luminosité intégrée de $50\,\text{ab}^{-1}$, permettant des mesures d'une précision jamais obtenue.

Le détecteur Belle II se présente comme une amélioration des composants et technologies éprouvés par l'expérience Belle. Les différentes améliorations visent à faire face à l'augmentation de la luminosité instantanée et ainsi à l'augmentation du bruit de fond et du taux d'occupation attendus. Le détecteur Belle II est composé de multiples sous-détecteurs disposés en cylindres imbriqués couvrants un large angle solide allant de 17 à 150° pour les couches les plus internes. Belle II se distingue grâce à son architecture, à ses déclencheurs spécifiques et par la connaissance précise des états initiaux et finaux des collisions e^+e^-, permettant ainsi d'avoir une très bonne reconstruction des processus à faible multiplicité ou impliquants de l'énergie manquante. Cela permet à Belle II de proposer un environnement idéal pour l'étude des propriétés du lepton τ^- ou des désintégrations avec neutrinos. Pour mener à bien son rôle, le detecteur possède différents constituants qui permettent de reconstruire la trajectoire des particules par combinaison des points d'impact en "traces" lors du tracking, de restituer leur quadri vecteur impulsion et enfin d'identifier la particule. Les sous-détecteurs de Belle II des couches les plus internes aux plus externes sont :

- le détecteur de vertex en silicium (VerteX Detector (VXD)), divisé en deux couches de pixels (PiXel Detector (PXD)) et quatre couches de capteurs à bande de silicium double face (Silicon Vertex Detector (SVD)). La fonction du VXD est de reconstruire le point de production des particules secondaires [1], mais aussi de participer à la reconstruction des traces, et à l'identification des particules (avec la mesure du dépôt d'énergie).
- la chambre à dérive centrale (Central Drift Chamber (CDC)), contenant un réseau de fil sous tension dans une atmosphère gazeuse. Ce réseau divise le volume en "cellules". Lorsqu'une particule chargée traverse une cellule, elle ionise le gaz et les électrons arrachés créent un signal dans les fils. La CDC permet ainsi de reconstruire les traces et impulsions [2] des particules, ou encore de donner des

---

[1] Les particules secondaires sont issues de la désintégration des particules produites lors de la collision.

[2] Les particules soumises au champ magnétique produit par un solénoïde (1.5T) décrivent une trajectoire parabolique permettant de mesurer leur impulsion.

informations sur leur identité en mesurant la perte d'énergie dans le gaz.

- le système d'identification des particules, est divisé en un compteur de temps de propagation (Time-Of-Propagation (TOP)) pour la partie cylindrique, ainsi qu'un détecteur Tcherenkov (Aerogel Ring-Imaging Cherenkov (ARICH)) dans le couvercle avant du détecteur (pour couvrir les particules boostées). En traversant le milieu de diffusion les particules émettent des cônes de photons par effet Tcherenkov. Les deux détecteurs déterminent la vitesse de la particule à l'aide du temps de diffusion des photons émis pour le TOP et de l'angle d'ouverture du cône de photon pour l'ARICH. En combinant cette vitesse avec l'impulsion on détermine la masse invariante de la particule émettrice.
- le calorimètre électromagnétique (Electromagnetic CaLorimeter (ECL)), collecte et mesure l'énergie déposée dans ses cristaux par les photons qui peuvent représenter une part importante de l'énergie dégagée dans les collisions, ou encore par les électrons, qu'il permettra donc de distinguer des autres hadrons chargés.
- le détecteur de K-long et de muons (K_L and Muon detector (KLM)), constitué de scintillateurs permettant d'identifier ces particules.

Les programmes contrôlant le fonctionnement des différents détecteurs, des simulations, de la prise de données et de la reconstruction des événements voulus sont centralisés dans le logiciel *Belle II* Analysis Software Framework (*basf2*), le logiciel d'analyse spécifique à Belle II. Une production officielle des échantillons de simulations Monte-Carlo (MC) et des données collectées est mise en place. Pour cette analyse les données traitées correspondent à des échantillons à l'énergie de la résonance $\Upsilon(4S)$ ($362\,\text{fb}^{-1}$), off résonance (60 MeV en dessous de l'$\Upsilon(4S)$), et à la résonance $\Upsilon(5S)$ pour un total de $424\,\text{fb}^{-1}$ enregistrés entre 2019 et 2022, avant le premier long arrêt visant à mettre à jour les installations.

Mesure de la résolution spatiale du détecteur de vertex de Belle II

Le détecteur SVD joue un rôle important dans la reconstruction de la trajectoire d'une particule. Afin de fournir une reconstruction des traces de la meilleure qualité possible, il est crucial d'estimer avec précision la résolution spatiale du détecteur. En effet elle intervient dans l'ajustement de la trace, mais elle permet aussi de propager les incertitudes de la position d'impact dans le calcul des paramètres de la trace.

Les quatres couches du SVD [14] sont composées de capteurs à bandes de silicium regroupés en modules disposés en couches cylindriques grâce à une disposition en moulin, Figure 2. Afin de capter les particules boostées, des modules trapézoïdaux sont disposés en bout de couches sauf pour la 3e couche. Les deux bandes de silicium sont disposées orthogonalement le long des capteurs de manière à mesurer la position selon la coordonnée $r\phi$ pour les bandes u/P et z pour les bandes v/N.

Lorsqu'une particule chargée traverse un capteur, plusieurs bandes sont activées et regroupées en grappe. Les propriétés de chacune des grappes, Figure 2 regroupent :

- la position de la grappe m obtenue comme le centre de gravité de toutes les positions X_i des bandes la constituant pondérées par leur charge collectée S_i, $m = \frac{\sum_i X_i S_i}{\sum_i S_i}$;
- la position t et son erreur σ_t sont définies à partir de l'intersection entre le capteur et la trace. Elles sont déterminées de manière non-biaisée, signifiant que la grappe considérée est ignorée pour reconstruire la trace considérée ;
- la vraie position x est l'intersection entre le capteur et la trajectoire réelle de la particule. La trajectoire réelle est seulement disponible dans les simulations.

La résolution spatiale des grappes est définie comme étant la variance du vrai résidu $zeps = m - xpos$. Cependant cette définition est applicable seulement dans les simulations ou la vraie position de la particule est connue. Après certaines approximations la résolution spatiale des grappes peut être déterminée dans les données en utilisant le résidu $R = m - t$ entre les positions de la grappe et celle de la trace.

Dans cette thèse, une mesure de la résolution spatiale des grappes a été effectuée en utilisant une nouvelle méthode à Belle II, dite de recouvrement ou de paire, basée sur le recouvrement des modules d'une même couche, initialement proposée par CMS [15]. Cette méthode est réalisable grâce à l'arrangement en moulin des modules pour former une couche.

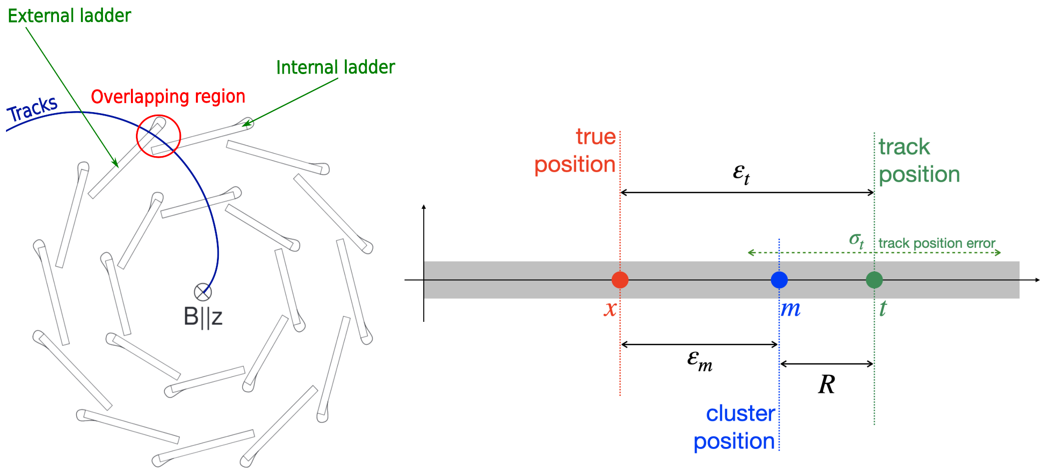


FIGURE 2. – Vue transversale (direction $r\phi$) du SVD illustrant la structure en moulin et le recouvrement des modules dans une même couche (Gauche). Schémas des différentes positions et résidus liés à une activation d'un capteur (Droite).

Dans la méthode des recouvrements, les traces sont reconstruites, en n'acceptant que celles qui intersectent deux capteurs de deux modules consécutifs d'une même couche (donc dans la région de recouvrements des modules), Figure 2. Le résidu $m - t$ est déterminé sur chacun des deux modules (internes et externes) qui se chevauchent. Puis les résidus du module interne et externe sont soustraits pour définir un double résidu ΔR. Une correction est appliquée au double résidu pour tenir compte de la géométrie des modules. La résolution par la méthode des recouvrements $\sigma_{cl}^{\text{recouvrement}}$

est obtenue par l'ajustement du double résidu par une distribution t-Student T définie par les paramètres suivant : le nombre de degrés de liberté ν; la moyenne de la distribution μ, et la variance σ^2. La variance de cette distribution est déterminée comme étant le carré de la demi-distance `sigma-68` entre les 16e et 84e quantiles :

$$\sigma_{cl}^{\text{recouvrement}} = \texttt{sigma-68}\big(T(X,\nu,\mu,\sigma)\big). \tag{0.1}$$

L'exploitation des modules se recouvrant a l'avantage d'annuler la contribution de l'erreur sur la position des traces, et d'être peu sensible à l'effet de Coulomb. Mais aussi de dissocier la contribution de la précision de l'ajustement des traces et la contribution de la résolution spatiale des grappes. Cependant elle limite les angles d'incidence de la trajectoire atteignable.

Les résultats obtenus par la méthode des recouvrements sont reportés dans la Figure 3, et dans le Tableau 2 [16]. En comparaison avec les autres méthodes essayées à Belle II ("Event-by-event" et "Global"), la méthode des recouvrements donne généralement une estimation moins bonne de la résolution spatiale à l'exception des bandes v/N des couches les plus externes.

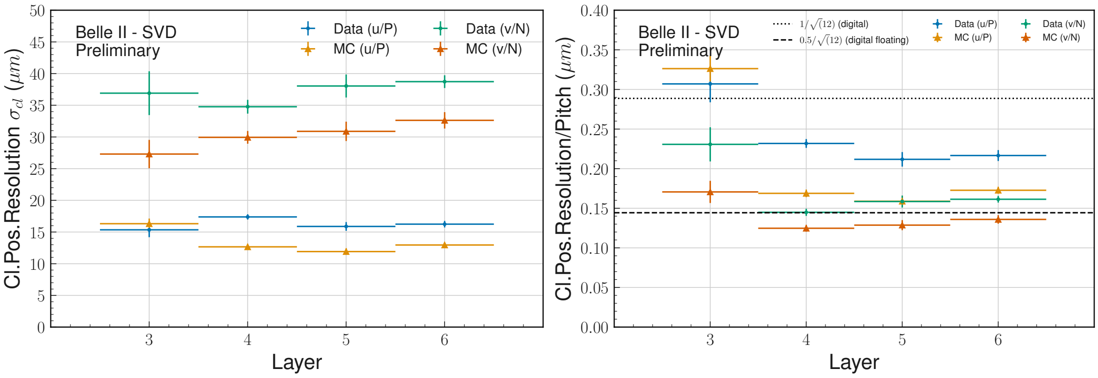


FIGURE 3. – La résolution spatiale des grappes (Gauche) et la résolution normalisée par l'espacement des bandes (Droite) sont calculées pour les côtés u/P et v/N de chaque couche du détecteur à partir de la méthode des recouvrements.

	Digital	EBE	Global	Pair
Layer 3 u/P (μm)	7	7	9	15
Layer 456 u/P (μm)	11	10	11	16-17
Layer 3 v/N (μm)	23	24	23	33
Layer 456 v/N (μm)	35	32	35	29-36

TABLE 2. – Résumé des résolutions spatiales obtenue avec les différentes méthodes essayées à Belle II [16].

Recherche de désintégration $\tau^- \to \mu^-\mu^+\mu^-$ violant la saveur leptonique

Dans l'étude de la désintégration du $\tau^- \to \mu^-\mu^+\mu^-$ dans les données de Belle II, une nouvelle méthode de reconstruction non étiquetée est proposée pour les événements de paires de τ^-, qui a été précédemment testée dans l'étude du canal de désintégration Lepton Flavour Violation (LFV) $\tau \to \ell\phi$ ($\ell = e, \mu$). Dans cette méthode, dite inclusive, les événements $e^+e^- \to \tau^+\tau^-$ sont reconstruits avec un lepton tau se désintégrant en trois muons (ce tau sera désigné comme signal) alors que le second tau de charge opposée (étiquetté) n'est pas restreint sur son mode de désintégration. Ceci se distingue des méthodes de reconstruction présentées par Belle [8] et *BABAR* [7], où le tau signal est reconstruit avec une désintégration en trois muons et le tau étiqueté est reconstruit avec une désintégration du modèle standard en un électron, ou muon ou pion [17]. Cette approche inclusive vise à améliorer l'efficacité de la sélection des signaux en tenant compte de tous les modes possibles de désintégration du second tau. Cependant, cette inclusivité introduit un bruit de fond supplémentaire qui doit être compensé par des critères de sélection efficaces pour maintenir des gains significatifs en matière de pureté des échantillons.

L'analyse repose sur les données expérimentales ainsi que des échantillons simulés de signal $\tau^- \to \mu^-\mu^+\mu^-$ et de bruit de fond. Afin d'être le plus général possible dans les modèles de nouvelle physique explorés, aucune contrainte sur le paramétrage de l'espace de phase du signal simulé n'est imposée.

Il es possible de définir une "région de signal" dans le plan bidimensionnel ($M_{3\mu}, \Delta E_{3\mu}$), où $M_{3\mu}$ est la masse mesurée du tau reconstruit et $\Delta E_{3\mu}$ est la différence d'énergie entre le tau et le faisceau ($\sqrt{s}$) dans le référentiel du centre de masse. En raison de l'absence de neutrinos (qui entraîne une perte d'énergie détectée) dans l'état final de $\tau^- \to \mu^-\mu^+\mu^-$, la distribution du signal présente un pic autour de $M_{3\mu} \approx 1{,}777\,\mathrm{GeV}/c^2$ [13] (la masse du lepton tau) et $\Delta E_{3\mu} \approx 0\,\mathrm{GeV}$. Ainsi, une région autour de ce pic est définie où il est plus probable de détecter une désintégration $\tau^- \to \mu^-\mu^+\mu^-$ dans les données. Les données dans cette région sont cachées jusqu'à la fin de l'analyse pour éviter tout biais humain dans l'obtention des résultats.

Afin d'assurer la reconstruction de trois muons provenant d'un lepton tau, les traces laissées par les particules dans le détecteur doivent satisfaire plusieurs prérequis :

- elles doivent provenir d'une zone proche de la région où les électrons et positrons se collisionnent,
- les particules chargées ayant une probabilité d'être un muon ($muonID$) supérieure à 0.5 sont identifiées comme muon,

les valeurs des coupures sur les variables d'identification sont volontairement faibles afin de permettre une optimisation de celle-ci par la suite.

Les éventuelles particules non chargées de l'événement, comme les photons et les pions neutres (π^0), sont également reconstituées pour déterminer leur quantité, qui peut varier en fonction de la nature des particules produites après la collision, ainsi que pour évaluer précisément les propriétés cinétiques et géométriques de l'événement.

Afin de limiter la quantité de bruit de fond traitée, les données subissent un prétraitement pour identifier les événements d'intérêt pour la recherche désintégrations du τ^- violant la saveur leptonique. Cette sélection couvre une région du plan $(M_{3\mu};\Delta E_{3\mu})$ avec une excellente rétention du signal. A la reconstruvtion, des exigences sur la topologie de l'événement sont imposées : un maximum de 6 traces doit être présent, et les trois muons doivent être situés dans un seul hémisphère, dont l'axe est la direction de la poussée totale de l'événement (thrust). De plus les événements doivent activer les lignes de déclencheurs du détecteur ECL correspondant aux basses multiplicités ou du détecteur CDC. À cette étape la composition des échantillons simulés est résumée dans le Tableau 3.

TABLE 3. – Taux de survie du signal et composition du bruit de fond pour 362 fb^{-1} dans les échantillons générés par simulation après la reconstruction des événements $e^+e^- \to \tau^-\tau^+$ avec un lepton tau se désintégrant selon $\tau^- \to \mu^-\mu^+\mu^-$.

	ε_{sig}^{abs}	N_{bkg}	$N_{\tau\tau}$	$N_{q\bar{q}}$	$N_{B\bar{B}}$	N_{lowm}
train	34.30%	1803.70	7.98	830.10	0.70	964.92
test	34.35%	1819.67	6.15	842.73	1.23	969.56

Parmi les propriétés importantes afin de distinguer le signal du bruit de fond se trouvent celles provenant du reste de l'événement (Rest-of-Event (ROE)). Ces propriétés sont obtenues en utilisant toutes les particules chargées et neutres non utilisées dans la reconstruction du $\tau^- \to \mu^-\mu^+\mu^-$. Celles-ci sont nettoyées à l'aide de masque, basé sur leur point d'origine, leur impulsion pour les particules chargées et sur leur énergie pour les particules neutres.

Avant de procéder à l'optimisation du rejet du bruit de fond plusieurs régions du plan $(M_{3\mu};\Delta E_{3\mu})$ sont définies. Les différentes régions sont définies à partir des résolutions δ des variables $M_{3\mu}$ et $\Delta E_{3\mu}$ obtenues par un ajustement de courbe à l'aide d'une gaussienne asymétrique. De cet ajustement sont extraits la moyenne $\bar{\mu}$, la résolution gauche $\delta^{\downarrow}$ et droite $\delta^{\uparrow}$ des distributions des échantillons simulés de signal, les valeurs sont résumées dans le Tableau 4. À partir de ces résolutions, sont définies

TABLE 4. – Résolutions obtenues avec l'ajustement des distributions $M_{3\mu}$ et $\Delta E_{3\mu}$ et utilisées comme unités pour définir les régions du signal.

Mode	Variable	Moyenne $\bar{\mu}$	Résolution à gauche δ^{down}	Résolution à droite δ^{up}
$\tau^- \to \mu^-\mu^+\mu^-$	$M_{3\mu}$ (MeV/c^2)	1777.35±0.07	4.80±0.07	4.44±0.06
	$\Delta E_{3\mu}$ (MeV)	0.7±0.3	14.9±0.3	10.0±0.5

différentes régions de taille et de forme différentes : rectangulaires ou elliptiques (avec une rotation d'angle θ extrait à partir de la corrélation des deux variables).

- la région $\pm 20\delta$ rectangulaire sert de référence pour l'optimisation des présélections et de l'apprentissage de l'algorithme de classification signal/bruit de fond,
- la région $\pm 20\delta$ en $M_{3\mu}$ et $\pm 10\delta$ en $\Delta E_{3\mu}$ rectangulaire est employée comme bande pour confirmer que les variables simulées représentent correctement les données, dans ce cas-là les données sont dissimulées selon la région ci-dessous,
- la région $\pm 5\delta$ rectangulaire est dissimulée dans les données pour éviter tout risque de biais,
- les régions $\pm 3\delta$ elliptique et rectangulaire, où sont évalués le signal et le bruit de fond restant.

La première étape du rejet du bruit de fond consiste à affiner les valeurs de coupures sur la variable d'identification du muon $muonID$ afin de garder une haute efficacité sur le signal. L'idée pour permettre cela est de trier les trois muons provenant du signal τ^- selon leur valeur de $muonID$ et de permettre une coupure différente selon leur rang (premier, deuxième et troisième) :

$$\mu ID_{premier} > 0.95, \tag{0.2}$$

$$\mu ID_{deuxieme} > 0.95, \tag{0.3}$$

$$\mu ID_{troisieme} > 0.5. \tag{0.4}$$

Une telle sélection permet de conserver une efficacité relative de 97% pour le signal et de 32% pour le bruit de fond contre respectivement 68% et 24% dans le cas où la coupure à 0.95 est imposée pour les trois muons. Le surplus de bruit de fond obtenu est aisément rejeté à l'aide de l'algorithme de classification.

La deuxième étape de la stratégie de rejet du bruit de fond repose sur les variables cinématiques et topologiques liées aux différentes particules (muons, tau, photons et π^0), à l'énergie manquante, au ROE, ou à l'événement en général. Ces variables sont exploitées en définissant des sélections préliminaires, des présélections, et des variables d'entrée pour un algorithme d'apprentissage automatique [18] qui classe les événements en fonction de leur "probabilité" de ressembler à un signal. Cependant, les présélections ont un impact sur les performances de l'apprentissage de l'algorithme, l'objectif est donc de trouver la meilleure combinaison de présélections pour entraîner l'algorithme. Pour y parvenir, la méthode consiste à :

- Choisir un set de sélections préliminaires parmi un ensemble de présélection défini antérieurement. Les différents sets de présélection sont définis en cherchant à éliminer les contributions de bruit de fond à basse multiplicité facilement distinguable du signal sur un nombre réduit de variables. Les différents sets de sélections préliminaires ainsi que la composition des échantillons après application sont donnés dans le Tableau 5.
- Appliquer le set de présélection choisi puis effectuer l'apprentissage de l'algorithme. Nous utilisons comme algorithme de classification un arbre de décision

boosté (Boosted Decision Tree (BDT)) issue de la bibliothèque `XGBoost` [18]. Celui-ci est entraîné sur les événements simulés de signal et de bruit de fond $q\overline{q}$ et $\tau^-\tau^+$. En sortie du BDT nous obtenons pour chaque événement une variable de vraisemblance de correspondre à du signal, illustré en Figure 4. Une coupure est optimisée sur cette variable de sortie en maximisant la figure de mérite de Punzi [19] dans la région du signal à 5δ.

Afin de valider les performances du BDT, les échantillons sont divisés en trois : un échantillon d'entraînement, un de validation et un de test qui retranscrit le rôle des données expérimentales dans le cadre de la simulation, c'est-à-dire qu'il n'est consulté que pour l'obtention des résultats dans le MC. Cependant pour conserver un nombre significatif d'événements de bruit de fond dans les échantillons d'entraînement et de test, nous avons fait le choix de diviser les $8\,\mathrm{ab}^{-1}$ d'événements simulées en deux échantillons d'entraînement et de test égaux ($4\,\mathrm{ab}^{-1}$). Puis l'échantillon d'entraînement est sous échantillonné à l'aide d'une méthode de "k-folding" [20] pour obtenir des "folds" d'entraînement et de validation.

Afin d'éviter les biais survenants par un surapprentissage du BDT, les hyperparamètres sont optimisés à l'aide de la bibliothèque `Optuna` [21] en minimisant la fonction logarithmique de perte des "folds" de validation.

Le processus est répété pour tout les différents sets de présélection et la combinaison obtenant la figure de mérite de Punzi maximale, Figure 4 (droite) est conservée. Suivant cette méthode la sélection permettant le rejet du bruit de fond est donnée dans le Tableau 6.

L'efficacité du signal de la méthode de réjection du bruit de fond est de $19.70^{+0.06}_{-0.06} stat\%$ alors qu'il subsiste $0.08^{+0.21}_{-0.07} stat$ événements de bruit de fond estimé sur $4\,\mathrm{ab}^{-1}$ d'échantillon de `Test` normalisé à la luminosité des données $362\,\mathrm{fb}^{-1}$.

Afin de valider notre stratégie de rejet du bruit de fond des vérifications sont nécessaires en comparant les échantillons de données et de simulation. En effet nos simulations doivent toujours correctement représenter les données. Pour cela nous réalisons une première comparaison, en utilisant la région des bandes latérales $\pm 20\delta$ selon $M_{3\mu}$ et $\pm 10\delta$ selon $\Delta E_{3\mu}$ en prenant soin d'exclure la région à 5δ après avoir appliqué les présélections pour vérifier que les variables utilisées dans le BDT représentent correctement les données. Le nombre d'événement après les présélections dans les bandes latérales est de $118.1^{+9.8}_{-5.1} stat$ dans les simulations et de $94.0^{+10.7}_{-9.7} stat$ dans les données. Après le BDT le nombre d'événement tombe à $3.3^{+1.2}_{-0.7} stat$ dans les simulations et $7.0^{+3.8}_{-2.6} stat$ dans les données. Les données et les simulations sont en accord selon les erreurs statistiques.

Plusieurs méthodes pour estimer le nombre de bruit de fond dans les données ont été testées en utilisant les bandes latérales. Nous avons retenu la méthode "ABCD" [22] basé uniquement sur les données, celle-ci n'est pas sensible aux problèmes de description des données par les simulations. Dans cette méthode quatre régions, Figure 5, sont définie par :

- **A**, en dehors de la région du signal à $\pm 5\delta$ et probabilité de sortie du BDT entre

TABLE 5. – Définition des différents sets de sélection préliminaires ainsi que leur efficacité sur le signal et la composition du bruit de fond survivant dans la région du signal ±20δ après avoir appliqué la sélection sur le muonID définit plus tôt. Le nombre d'événements de bruit de fond est normalisé à la luminosité des données $\Upsilon(4S)$, 362 fb^{-1}. Les variables considérées pour les sélections préliminaires sont : θ^{CM}_{miss} l'angle polaire de l'impulsion manquante, l'exposant CM dénote qu'une quantité est mesuré dans le référentiel de centre de masse ; $p^{T,CM}_{miss}$ la composante trasverse à l'axe du faisceau de l'impulsion manquante ; E^{CM}_{vis} énergie visible ; ΔE_{ROE} la différence d'énergie entre le ROE et la collision.

Name	Preselection	ε^{rel}_{sig} (%)	ε^{abs}_{sig} (%)	ε^{rel}_{bkg} (%)	N_{bkg}	$N_{\tau-pair}$	$N_{q\bar{q}}$	$N_{B\bar{B}}$	N_{lowm}
Reference	$0.3 < \theta^{CM}_{miss} < 2.7$	96.88	31.11	89.99	938.82	3.08	287.52	0	648.22
Set 1	$0.3 < \theta^{CM}_{miss} < 2.7$ $0.89 < Thrust < 0.97$	95.48	30.67	30.83	321.64	2.96	270.87	0	47.82
Set 2	$0.3 < \theta^{CM}_{miss} < 2.7$ $0.935 < Thrust_{\tau ROE}$ $0.95 > Thrust_{\tau ROE}$	96.35	30.94	61.78	644.50	2.58	244.94	0	396.97
Set 3	$0.3 < \theta^{CM}_{miss} < 2.7$ $E^{CM}_{vis} < 10.$	90.54	29.08	14.89	155.30	2.98	127.81	0	24.52
Set 4	$0.3 < \theta^{CM}_{miss} < 2.7$ $E^{CM}_{miss} > 0.6$	90.22	28.98	14.69	153.29	2.91	125.85	0	24.52
Set 5	$0.3 < \theta^{CM}_{miss} < 2.7$ $p^{T,CM}_{miss} > 0.4$	91.12	29.26	15.89	165.74	2.77	135.08	0	27.90
Set 6	$0.3 < \theta^{CM}_{miss} < 2.7$ $M_{ROE} < 2.2$ $-5. < \Delta E_{ROE} < -0.2$	90.76	29.15	16.49	172.08	2.62	106.08	0	63.39

TABLE 6. – Liste de sélections appliquées pour rejeter le bruit de fond après optimisation.

$\tau^- \to \mu^-\mu^+\mu^-$
$muID_{lead} > 0.95$
$muID_{sub} > 0.95$
$muID_{third} > 0.5$
$0.3 < \theta^{CM}_{miss} < 2.7$
$0.89 < thrust < 0.97$
$p^{BDT}(Signal) > 0.89$
3δ ellipse region

0.2 et 0.5,

- **B**, dans la région du signal à ±3δ et probabilité de sortie du BDT entre 0.2 et 0.5,

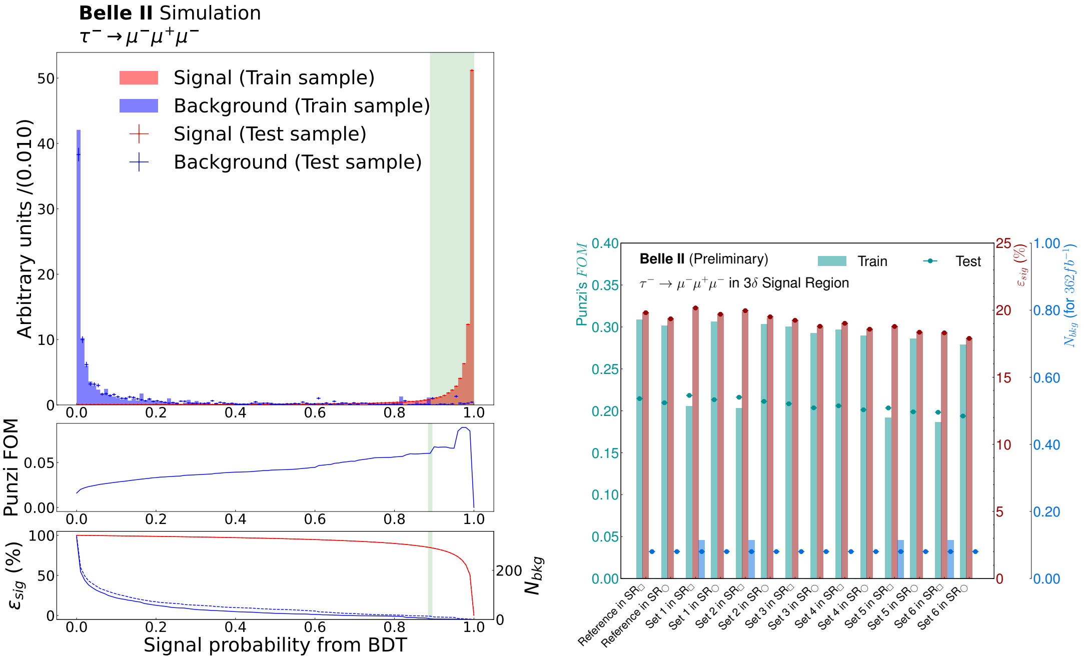


FIGURE 4. – À gauche : Probabilité de vraisemblance au signal obtenue en sortie du BDT estimée sur les échantillons d'`Entrainement` et de `Test` dans une région du signal à $\pm 20\delta$. Les deux graphiques du dessous représentent respectivement la figure de mérite de Punzi et les efficacités du signal (rouge) et de bruit de fond (bleue) en fonction de la valeur de coupure. La région verte représente la région gardée après coupure sur le BDT, cette région est optimisée pour maximiser la figure de mérite de Punzi dans l'échantillon de validation dans une région à $\pm 5\delta$.
À droite : Figure de mérite de Punzi (cyan), efficacité du signal (rouge) et nombre de bruits de fond survivant (bleue) pour les échantillons d'`Entrainement` et de `Test` pour les différentes combinaisons de présélection et de BDT.

- **C**, en dehors de la région du signal à $\pm 5\delta$ et probabilité de sortie du BDT supérieur à 0.89,
- **D**, dans la région du signal à $\pm 3\delta$ et probabilité de sortie du BDT supérieur à 0.89.

Le but est d'estimer le nombre de donnée dans la région **D** par extrapolation depuis la région **C** et un facteur de transfert **B/A**, $N_D^{attendue} = N_C \times (N_B / N_A)$. Nous obtenons ainsi $0.50^{+1.38}_{-0.50} stat$ événement attendu.

La limite supérieure attendue sur le rapport d'embranchement de la désintégration $\tau^- \to \mu^-\mu^+\mu^-$ selon :

$$\mathscr{B}^{LS}(\tau^- \to \mu^-\mu^+\mu^-) = \frac{N^{attendu} - N^{observe}}{2L\sigma_{\tau\tau}\varepsilon_{3\mu}}, \tag{0.5}$$

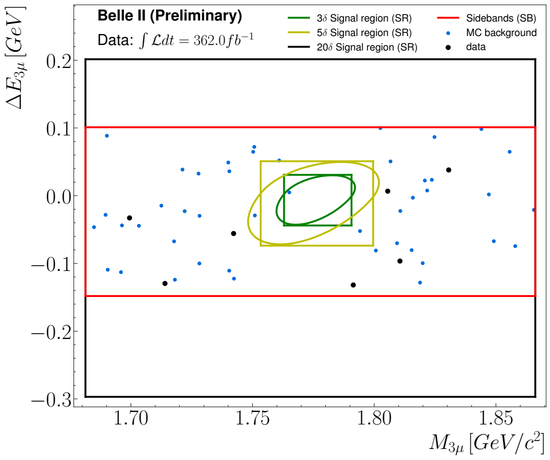


FIGURE 5. – Distribution des Événements de données (noir) et de bruit de fond simulés (bleu) survivants aux sélections de rejet du bruit de fond dans le plan de la distance elliptique au pic ($M_{3\mu}, \Delta E_{3\mu}$) et de la variable de sortie du BDT. Le lignage rouge délimite les différentes régions ABCD de la méthode. Les données restent dissimulées dans la région D.

où L est la luminosité intégrée, $\sigma_{\tau\tau}$ la section efficace de production des paires $\tau^-\tau^+$, $\varepsilon_{3\mu}$ l'efficacité de sélection du signal et respectivement $N^{attendu}$ et $N^{observe}$ les nombres d'événements dans les données attendues avec la méthode "ABCD" et observés. Les données étant toujours dissimulées, nous faisons l'hypothèse que $N^{observe} = 0$. La limite supérieure est calculée en utilisant la méthode CL_s [23] implémentée dans le langage `RooStats`. Les incertitudes systématiques sur les différentes quantités du rapport d'embranchement sont résumées dans le Tableau 7. Les résultats de la méthode CL_s sont représentés en Figure 6. La limite supérieure attendue à un niveau de confiance de 90% est de 1.77×10^{-8} en utilisant les 362 fb^{-1} de données à $\Upsilon(4S)$ et de 1.51×10^{-8} en utilisant toute la statistique de Belle II, 424 fb^{-1}. Si une telle limite supérieure est conservée après dévoilement des données, nous serions en capacité d'établir une nouvelle référence jusque-là établie par Belle à 2.1×10^{-8} à 90%

TABLE 7. – Incertitudes systématiques relatives sur les différentes quantités de la formule du rapport d'embranchement.

Quantité affectée	Source	Valeur
$\varepsilon_{\ell\ell\ell}$	Indentification des leptons	2.39%
	Efficacité du tracking	0.72%
	Efficacité des tiggers	1.0%
N_{exp}	Correction de l'impulsion	5.0%
L	Luminosité	0.6%
$\sigma_{\tau\tau}$	Section efficace de production de $\tau^-\tau^+$	0.3%

de niveau de confiance.

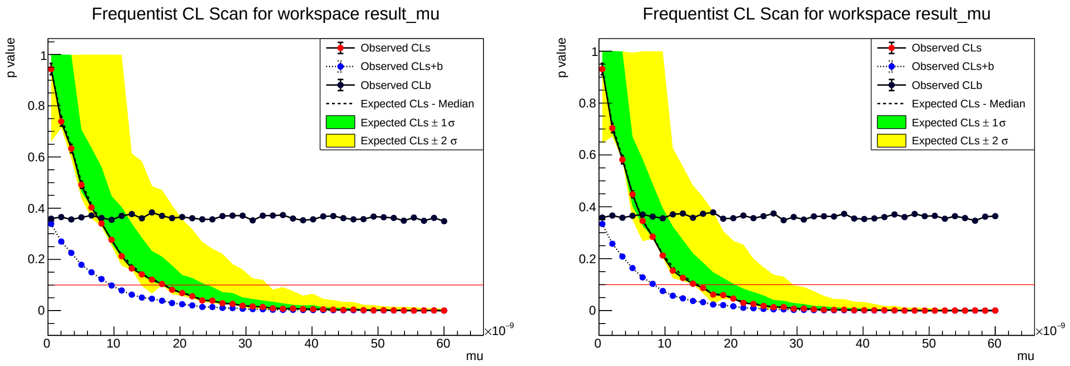


FIGURE 6. – CL_s attendu en fonction de la limite supérieure sur les rapports d'embranchement de $\tau^- \to \mu^-\mu^+\mu^-$ pour une luminosité de 362 fb^{-1} à droite et 424 fb^{-1} à gauche. La droite horizontale rouge dénote un niveau de confiance à 90%.

Contents

Introduction

The Standard Model (SM) of particle physics is a theory developed in the 20th century to explain the fundamental constituents of nature and how they interact with each other. Over time, numerous experimental results have confirmed the accuracy of the SM predictions regarding mechanisms at the subatomic scale. Through the utilization of quantum fields in accordance with the principles of special relativity, the mathematical structure of the SM permits the description of the behaviour of fundamental particles, including leptons, quarks, gauge bosons, and hadrons through three fundamental interactions: the electromagnetism, weak, and strong interactions. However, the SM presents several limitations in explaining some observed phenomena, such as dark matter, matter-antimatter asymmetry, hierarchy problems, and many others. Therefore, new physics models are being investigated by particle physics experiments. Among them, the Belle II experiment is a second-generation experiment at a B meson factory, which exploits the SuperKEKB electron-positron collider and its detector at the KEK laboratory in Japan. The Belle II experiment provides the perfect environment for precision measurements.

Particles possess unique properties, including mass, electric and colour charges, spin, and other quantum properties that influence their interactions with one another. Among these properties, the three lepton numbers (electronic, muonic, tauic) are associated with the lepton flavour and represent the leptonic nature and mass family. In the SM, lepton flavour is typically conserved by chance in most cases. However, neutrinos can spontaneously oscillate between the three neutrino flavours, which constitutes a violation of the lepton flavour in the neutrino sector. This phenomenon can also participate in charged Lepton Flavour Violation (cLFV) decays through charged currents of the weak interaction. The SM predicts rates below 10^{-50}, which is far below current and future experiment sensitivity. So the observation of cLFV decays would provide indisputable evidence of "new physics" beyond the Standard Model. In particular various models of "new physics" can increase the rates of τ LFV decays just below the current experimental sensitivity. Some LFV decay channels studied at Belle II are classified as "Golden channels" due to their physics interest or low background contamination. The $\tau^- \to \mu^-\mu^+\mu^-$ decay, which is studied in this thesis, stands in the second type of "Golden channels". Indeed $\tau^- \to \mu^-\mu^+\mu^-$ presents a low background contamination due to its purely leptonic final state. A new method is proposed in this document aiming a better signal retention with an untagged events reconstruction coupled thanks to a machine learning background rejection.

In Chapter 1, we will present the SM of particle physics with its LFV mechanism and theories behind the possible new physics that motivate τ LFV searches. A presentation of the Belle II experiment and how it operates is done in Chapter 2. Chapter 3 presents a side project consisting of a measurement of the spatial resolution of the vertex subdetectors using the overlapping sensors. Finally, we report in Chapter 4 the search for $\tau^- \to \mu^-\mu^+\mu^-$ decays based on an untagged and machine learning-based selection.

1. The Standard Model and Beyond: Focus on the Lepton Flavour Violation

Table of contents

1.1. Introduction to the Standard Model of particle physics

1.1.1. Elementary particles

The Standard Model (SM) is a phenological description of the propagation and the interaction of elementary particles through three fundamental forces: electromagnetic, weak, and strong interactions. At present, the SM provides the most precise description of the subatomic scale [1].

In the SM, the notion of elementary particles emerges from the Quantum Field Theory (QFT) mathematical framework. Such framework is based on special relativity to describe space, using three spatial dimensions and adding time as the fourth dimension. The initial concept of the QFT involves populating the spacetime with mathematical entities called fields. A field is a quantity attached to each point of the spacetime. In QFT, fields have to follow the special relativity's symmetries linked to the geometry of our universe, *e.g.* symmetries of translation and rotation. Following these requirements, the fields are composed of a few kinds of mathematical objects; three of them stand apart: scalar, vector and spinor [2]. Moreover, the symmetries of special relativity also imply the conservation of some physics quantities, like energy, momentum, angular momentum and velocity of the centre of mass, according to the Noether's theorem.

When describing the scale of elementary particles, the QFT must not only adhere to the requirements of special relativity but also incorporate quantum physics. The fields experience perturbations in the QFT. Mathematical operators, known as creation and annihilation operators, are defined to increment those perturbations by an integer. These perturbations are the particles [3]. Additionally, the fields evolve based on a superposition of all possible states with varying probabilities of occurrence.

From the quantum fields, emerges the definition of elementary particles. The universe is not limited to being filled by only one field but by several in the meantime. Each field represents one elementary particle of the standard model, represented in Figure 1.1 with their properties. These coexisting fields are in self-interaction and/or interactions with others. These interacting fields are traducing the fundamental forces of physics: electromagnetic, weak and strong interactions.

Particles are distinguished into two families according to their spin. If the spin quantum number is a half-odd-integer, the particle is described by a spinor field and named Fermions from the Fermi-Dirac statistic they follow. On the contrary, if the spin is an integer, the particle is described by a scalar (spin 0) or vector (spin 1) field and called Boson as they follow the Bose-Einstein statistic. Fermions are the components of matter, and they are classified into two subclasses: leptons and quarks.

Leptons exist under three mass generations composed of one charged and one neutral leptons: electrons e^- and electronic neutrinos ν_e, muons μ^- and muonic neutrinos ν_μ, and tau τ^- and tauic neutrinos ν_τ. These three doublets correspond to the three leptons flavours (electronic, muonic and tauic). The three charged leptons, electron, muon and tau, interact through electromagnetic and weak interactions, while neutral leptons, the neutrinos, interact only through weak interactions.

---

[1] The description of the Standard Model is based on different works: literature [24, 25, 26, 1, 27], thesis [28, 29, 30], online resources [31] and lectures.

[2] To these objects is defined a number, 0 for scalar, 1/2 for spinors and 1 for vectors which is called the spin of the particle.

[3] The representation of particles as perturbations of unique fields which fill the universe allows us to understand why particles have the same properties everywhere and everytime.

Quarks exist in six flavours: three *up-type* up u, charm c, and top t with a charge $q = +2/3$ and three *down-type* down d, strange s, and bottom b with a charge $q = -1/3$. In addition to the electromagnetic and weak forces, quarks interact through the strong force described by the Quantum ChromoDynamics (QCD). Quarks can only be found under bound states, the hadrons: two quarks (quarks-antiquarks) states such as B ($q\bar{b}$) are called mesons, and three quarks states such as proton (uud) and neutron (udd) are called baryons. Under the QCD, quarks carry a colour charge: red, green, and blue for quarks and anti-red, anti-green, and anti-blue for anti-quarks (also called sometimes magenta, yellow and cyan). Hadrons must have zero total colour charge (being white by following additive colour properties) because of a phenomenon called colour confinement. Only lighter fermions are stable and make the atoms with a nucleus made up of protons ($u\ u\ d$) and neutrons ($u\ d\ d$) and a cloud of electrons arranged in atomical orbits. A whole variety of atoms can be created depending on the number of protons, while the stability of the atom is ensured by the number of neutrons. The electrons are responsible for the organisation of atoms in molecules.

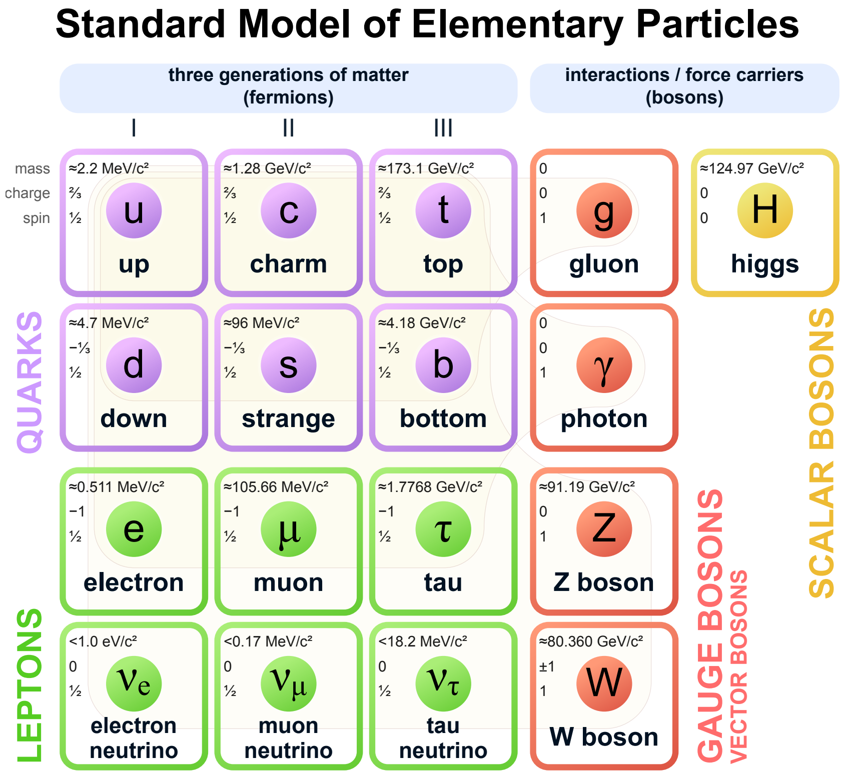


Figure 1.1. – Summary of the elementary particles featured in the Standard Model, along with their respective properties such as mass, charge, and spin. Three generations of fermions have been identified, ordered by their respective masses. The first generation is stable and comprises all of the matter that surrounds us. Credits [32]

In the SM, the three fundamental interactions are seen as an exchange of vector bosons (spin 1) between fermions:

- Photons γ are massless and neutral and carry the electromagnetic interaction between electrically charged particles,
- $W^{\pm}$ and Z^0 bosons are the mediators for the weak interaction between all fermions, such as β[4] radioactivity,
- Gluons g are massless and neutral and mediate the strong interaction. They exist in eight colours to bind quarks into hadrons, *e.g.* protons and neutrons in atoms. The Strong interaction is also responsible for the atomic nucleus stability.

Aside from this, the complex conjugate of these particles, which have opposite complex numbers, are also part of the SM and form the anti-matter.

In order to finalize the SM, it is necessary to incorporate the Higgs boson, one scalar boson which is the manifestation of the Higgs field. Despite being neutral, this massive boson plays a crucial role in determining the mass of leptons, quarks, and $W^{\pm}$ and Z^0 bosons. The particle's mass increases in proportion to the strength of its interaction with the Higgs field.

1.1.2. Evolution and interactions of the fields under the Standard Model

In the previous section, we defined the space-time comprising the three dimensions of space with the one time. In addition, the space-time is filled by different fields, one per type of elementary particle. Now let's see how these different fields evolve in time and interact with each other.

Only four fundamental forces are needed to describe all the interactions between objects: gravitation, electromagnetism, weak, and strong forces. The SM is able to gather three of them thanks to QFT under the gauge group product $SU(3)_c \times SU(2)_L \times U(1)_Y$ summarizing symmetries of the model. The colour charge c, the chirality (left/right-handed) and the weak hypercharge Y arise from the corresponding indexed group. Under such formalism, the lagrangian[5] [1] density of the SM can be written as:

$$\mathscr{L}_{SM} = \mathscr{L}_{gauge} + \mathscr{L}_{EW} + \mathscr{L}_{QCD} + \mathscr{L}_{Higgs} + \mathscr{L}_{Yukawa}. \tag{1.1}$$

The first term of the lagrangian $\mathscr{L}_{gauge}$ describes the self-interaction of the different gauge bosons with:

$$\mathscr{L}_{gauge} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\mathbf{W}^a_{\mu\nu}\mathbf{W}_a^{\mu\nu} - \frac{1}{4}\mathbf{G}_{\mu\nu}\mathbf{G}^{\mu\nu}, \tag{1.2}$$

where $B_{\mu\nu}$, $W^a_{\mu\nu}$, $G_{\mu\nu}$ are respectively the strength tensor of the gauge boson fields B_μ under $U(1)_Y$, W^a_μ under $SU(2)_L$, both being the bosons associated to the ElectroWeak

---

[4] β radioactivity is the emission of an electron (positron) and an associated neutrino (anti-neutrino) during the proton/neutron transition in the atomic nucleus.

[5] The Lagrangian is a quantity which describes the dynamic of the system as a function of a set of coordinates dependant on time. This quantity is convenient because the evolution of a system can be viewed as the sum of the free evolution of each subsystem and the interaction between each of them.

(EW) interaction; and the gluons fields G_μ under $SU(3)_c$ associated to the strong interaction.

The spinor fields, like fermions, have two components - a left-handed L component and a right-handed R component, except for neutrinos[6]. In addition, neutral currents[7] involve both chiralities, but only left-handed fermions and right-handed antifermions interact through charged currents[8]. Doublets can be defined for left-handed fermions, while singlets can be defined for right-handed fermions as:

$$\ell_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \; q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \qquad e_R,\; u_R,\; d_R, \tag{1.3}$$

where the three generation are summarized such as: $e = \left(e, \mu, \tau\right)$ and $\nu = \left(\nu_e, \nu_\mu, \nu_\tau\right)$ for the leptons and $u = (u, c, t)$ and $d = (d, s, b)$ for the quarks.

The EW lagrangian $\mathscr{L}_{EW}$ is built from the representation in left/right-handed fermions such as:

$$\begin{aligned} \mathscr{L}_{EW} = {} & \bar{\ell}_L i\gamma^\mu D_\mu \ell_L + \bar{e}_R i\gamma^\mu D_\mu e_R \\ & + \bar{q}_L i\gamma^\mu D_\mu q_L + \bar{u}_R i\gamma^\mu D_\mu u_R + \bar{d}_R i\gamma^\mu D_\mu d_R, \end{aligned} \tag{1.4}$$

with γ_μ the gamma matrices used for the chiral representation and:

$$D_\mu = \partial_\mu - \frac{ig}{2} T^a W^a_\mu - \frac{ig'}{2} Y B_\mu, \tag{1.5}$$

where the derivative term ∂_μ gives the fermion kinetic term. T^a is the weak isospin operator, while Y is the weak hypercharge operator, the respective coupling constant being written g and g'. Isospin and weak hypercharge are linked to the electrical charge with $Q = T^3 + Y/2$.

In the world of quarks, colour confinement is a crucial principle that keeps them bounded together as hadrons. This is made possible by the action of gluons and it follows specific guidelines corresponding to the quark's colour charge under the QCD theory. $\mathscr{L}_{QCD}$ is the term associated with the strong interaction between gluons and quarks under the QCD theory and is written as:

$$\mathscr{L}_{QCD} = \bar{q}_{L,i} i\gamma^\mu D_\mu q_{L,j} + \bar{u}_R i\gamma^\mu D_\mu u_R + \bar{d}_R i\gamma^\mu D_\mu d_R, \tag{1.6}$$

with:

$$D_\mu = i g_s T_{A,ij} G_{A,\mu}, \tag{1.7}$$

---

[6] Neutrinos have been experimentally proven to be entirely left-handed, whereas anti-neutrinos are right-handed.

[7] In the field of particle physics, there are two currents that describe the ways in which particles can interact via the weak interaction, each mediated by a different boson. If the interaction involves a Z^0 boson we are talking about the neutral current of the weak interaction. The charged current involves W bosons.

[8] The weak interaction is violating the parity symmetry, which is the symmetry under the inversion of the space coordinates $\begin{pmatrix} x & y & z \end{pmatrix} \rightarrow \begin{pmatrix} -x & -y & -z \end{pmatrix}$.

where $T_{A,ij}$ are the eight generator matrices of the $SU(3)_c$ gauge group with i, j indexing the three colours, g_s is the strong coupling.

At this point, the lagrangian of the SM does not contain a term to assign mass to the fields. Fields' masses are rising from the interaction with a complex field ϕ, the Higgs field. The interaction of the Higgs field with gauge bosons is included in $\mathscr{L}_{Higgs}$ term:

$$\mathscr{L}_{Higgs} = \left|D_\mu\phi\right|^2 + \mu^2\left|\phi\right|^2 - \lambda\left|\phi\right|^4, \tag{1.8}$$

with $\mu \in \mathbb{C}$ and $\lambda \in \mathbb{R}^*$. The potential $V(\left|\phi\right|^2) = -\mu^2\left|\phi\right|^2 + \lambda\left|\phi\right|^4$ has two local minimums at $\phi_0 = 0$ for $\mu^2 \leq 0$ and $\phi_0 = \sqrt{\frac{\mu^2}{2\lambda}}e^{i\theta_W} \equiv \frac{v}{\sqrt{2}}e^{i\theta_W}$ if $\mu^2 > 0$, where v is the vacuum expectation value of the Higgs field. In the event that the field has a non-zero value, the global $U(1)$ symmetry is spontaneously broken [33] through the selection of a phase value for $\theta_W = cos^{-1}\frac{g}{\sqrt{g^2+g'^2}}$. Using the phase θ_W, the gauge bosons for EW can be rotated as:

$$W^\pm_\mu = \frac{W^1_\mu \mp W^2_\mu}{\sqrt{2}}, \tag{1.9}$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} cos\theta_W & -sin\theta_W \\ sin\theta_W & cos\theta_W \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}, \tag{1.10}$$

and the $\mathscr{L}_{Higgs}$ can be developed to define mass terms depending on v and the coupling constants g and g' for the EW bosons. We can identify the two charged gauge bosons fields for the weak interaction $W^\pm_\mu$ with a mass $\sqrt{g^2v^2/4}$, one neutral Z_μ with a mass $\sqrt{(g^2+g'^2)v^2/4}$ and finally one gauge boson for the electromagnetism A_μ which remains massless.

The interaction of the quarks and leptons with the Higgs field is described by the last part $\mathscr{L}_{Yukawa}$, where:

$$\mathscr{L}_{Yukawa} = y^d_{ij}\bar{q}^i_L\phi d^j_R + y^u_{ij}\bar{q}^i_L\phi^\dagger u^j_R + y^e_{ij}\bar{e}^i_L\phi e^j_R + h.c. \tag{1.11}$$

with the indices i, j referring to the three generations of leptons or quarks. The interaction is defined by the coupling constants y^u_{ij}. The symbol $\phi^\dagger$ represents the conjugate of ϕ in the $SU(2)$ group, and $h.c.$ denotes the Hermitian conjugate. The conjugate is necessary for up-type quarks but not for leptons, as neutrinos are not right-handed and do not interact with the Higgs field and are therefore assumed to be massless.

To summarize, the Lagrangian density for the SM is linked to the following: the development and interplay of the gauge bosons, the EW interaction between fermions utilizing the mediation of $W^\pm$, Z^0 and γ bosons, the strong interaction between quarks where they exchange gluons carrying a colour charge within the Quantum Chromodynamics (QCD) and the definition of the masses of gauge bosons and fermions through their interaction with the Higgs field.

1.2. The Standard Model limitations

1.2.1. An incomplete theory

The Standard Model has been rigorously tested for several decades and the discovery of the Higgs boson by the Large Hadron Colliders experiments in 2012 (which led to the Nobel Prize for François Englert and Peter Higgs in 2013) was the final missing piece of the puzzle. At present, the SM provides the most precise explanation of physics at the subatomic level. However, some theoretical issues have not been resolved and various experimental observations still require explanation. One of the most compelling theoretical topics is the inability of the SM to incorporate the fourth fundamental interaction: gravitation. The SM's QFT framework is incompatible with the mathematical description of general relativity.

From a phenomenological viewpoint, the existence of dark energy and dark matter, which constitute most of the universe energy content, is unanticipated by the SM. Dark matter is perhaps the most well-known example of the SM's shortcomings. Additionally, the fact that the universe is made up of matter, while matter and antimatter were expected to be created in equal quantities in the early universe, cannot be explained by the Charged Parity (CP) violation in the weak interaction. The origin of three generations in the fermion sector and the difference in masses between charged leptons and neutrinos are other facts that are not properly explained and demonstrate the SM's incomplete nature.

The weaknesses of the SM have prompted physicists to explore evidence of New Physics (NP) beyond the standard model, which would include new particles and interactions that could address the aforementioned problems and others.

1.2.2. How do we search for beyond the Standard Model physics?

In experimental high-energy physics, the search for new physics beyond the SM follows two main strategies: the most direct one is to achieve the highest collision energy to produce new heavier particles directly; this is known as the *energy frontier*, Figure 1.2-left. This strategy is adopted by the experiments done at the CERN with the Large Hadron Collider, such as the ATLAS and CMS experiments, which recorded the highest collision energy in the world at 13 TeV.

The second way, indirect, consists of performing precision measurements at lower energies to measure deviations from the SM, a sign of new physics, Figure 1.2-right; the method is called *intensity frontier*. The B factories, like KEKB and PEP-II, or the new generation SuperKEKB as well as LHCb, adopt this method.

1.3. Lepton flavour and its violation in the Standard Model

1.3.1. The flavour of a fermion

In the SM, some accidental symmetries can arise, which are reflected in mathematical description by the conservation of additional quantum numbers.

In the quarks sector, there are six distinct flavours: up, down, charm, strange, top, and bottom. One flavour for each existing quark. The unique way to change the flavour of a quark is under the weak interaction that involves W bosons mediators. Indeed, interaction with W bosons allows the transition from *up* to *down* quark types. As seen in Section 1.1.2, the Yukawa Lagrangian Equation (1.11), defines the quarks' masses through mass matrices. The mass eigenstates do not coincide with flavour eigenstates. From this postulate, a matrix V_{CKM}, called Cabibbo-Kobayashi-Maskawa (CKM) matrix [34, 35], can be built to describe the flavour transitions between quarks via weak interaction charged currents (with W bosons as mediators):

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{1.12}$$

where the elements V_{ud} is for example the amplitude of transition between u and d quarks, as shown in Figure 1.3 (left). The standard parametrisation of the CKM matrix is:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.13}$$

where $c_{ij} = cos\theta_{ij}$ and $s_{ij} = sin\theta_{ij}$ for $i < j = 1,2,3$, θ_{ij} are the rotation angles between the mass generation (up/down, charm/strange and top/bottom) and δ a complex phase. The complex phase was introduced when the third generation (top/bottom) was discovered and is the unique source for CP symmetry violation in the SM. From the experimental measurements of the elements [13], we know that

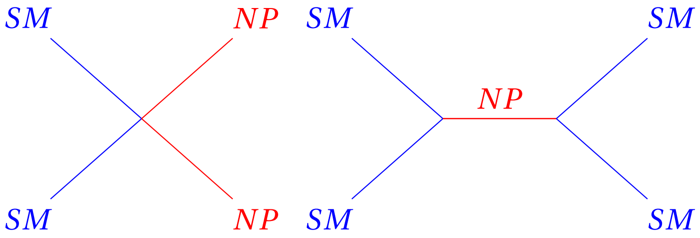


Figure 1.2. – Schemes of the strategies to search for new physics beyond the standard model.

diagonal elements dominate the matrix:

$$V_{CKM} \approx \begin{pmatrix} 0.974 & 0.224 & 0.004 \\ 0.221 & 0.975 & 0.041 \\ 0.009 & 0.042 & 1.014 \end{pmatrix}. \tag{1.14}$$

Transitions among quarks from the same generation are far more likely than transitions among quarks from different generations.

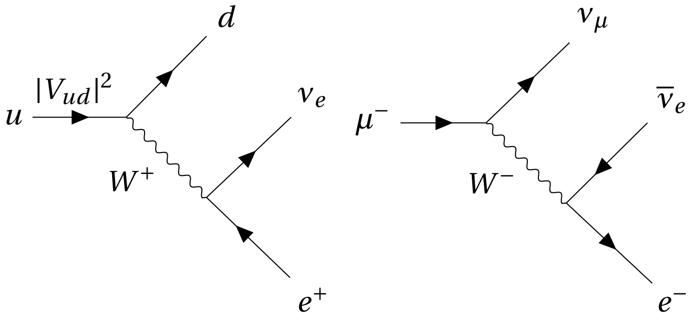


Figure 1.3. – Feynman diagram the for u to d transition (left) in the weak interaction charged currents, responsible for the β radioactive decay in atoms' nuclei. μ^- decays to e^- with the emission of neutrinos through the weak interaction (right).

In the context of the lepton sector, the Standard Model maintains a global quantum number known as the lepton number (L), calculated by subtracting the number of antileptons from the number of leptons. Along with the lepton number, the Standard Model also preserves three lepton flavour numbers. Indeed the leptons are gathered into three flavour doublets composed of one charged (electron e^-, muon μ^- and tau τ^- leptons) and one neutral lepton (neutrino electronic ν_e, muonic ν_μ, tauic ν_τ). The three lepton flavour numbers are:

- The electron number L_e,
- The muon number L_μ,
- The tau number L_τ,

where, for example, L_e is equal to 1 for e^- and ν_e, -1 for e^+ and $\overline{\nu}_e$ and 0 otherwise. The lepton number L and the lepton flavour numbers $L_{e,\mu,\tau}$ have to be the same between the initial and the final states of a SM interaction. Unless the lepton flavour numbers are not conserved in which case we talk of Lepton Flavour Violation (LFV), Figure 1.4. The weak interaction involving W boson allows the transition between Leptons, Figure 1.3 (right), but only in the same flavour doublets ($e^- \leftrightarrow \nu_e$), inducing the conservation of the lepton flavours.

1.3.2. Neutral lepton flavour oscillation

In the SM, quarks and charged leptons acquire a mass from the Yukawa coupling with the Higgs fields by spontaneous symmetry breaking. The neutrinos' masses are not defined since the Yukawa coupling is not applicable to neutrinos without right-handed neutrinos. Nevertheless, if the SM does not define the mass of neutrinos, it allows neutrinos to get masses below 1 eV.

Bruno Pontecorvo has proposed that under the assumption that neutrinos have mass, the weak interaction (flavours) eigenstates $(\nu_e, \nu_\mu, \nu_\tau)$ do not coincide with mass eigenstates (ν_1, ν_2, ν_3) analogously to the quarks sector. This property implies that the neutrino can oscillate: one neutrino's flavour can spontaneously change as it propagates (*e.g.* $\nu_e \to \nu_\mu$). Several experiments have observed the neutrino oscillation from different sources. Multiple experiments using different sources of neutrinos (solar [36], atmospheric [37], accelerator [38] and reactor [39]), have established a deficit in the observed number of neutrinos. The first evidence was given by the Super-Kamiokande [37] experiment in 1998 using atmospheric neutrinos.

The rotation between the mass $\begin{pmatrix}\nu_1 & \nu_2 & \nu_3\end{pmatrix}^T$ and flavour $\begin{pmatrix}\nu_e & \nu_\mu & \nu_\tau\end{pmatrix}^T$ eigenstates is described by:

$$\begin{pmatrix}\nu_e \\ \nu_\mu \\ \nu_\tau\end{pmatrix} = \begin{pmatrix}U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3}\end{pmatrix} \cdot \begin{pmatrix}\nu_1 \\ \nu_2 \\ \nu_3\end{pmatrix} = U \cdot \begin{pmatrix}\nu_1 \\ \nu_2 \\ \nu_3\end{pmatrix}, \tag{1.15}$$

where U is the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [40, 41] similar to the CKM matrix. The element matrix U_{ei} represent for example the amplitude of the transition $|\nu_e\rangle = \sum_{i=1}^{3} U_{ei}\, |\nu_i\rangle$. The PMNS matrix can be represented using three rotation angles θ_{12} θ_{13} θ_{23}, and one complex phase δ. If the neutrinos are from

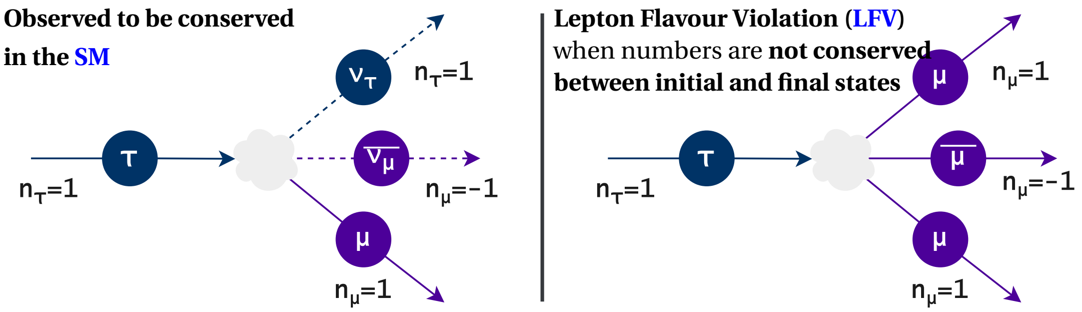


Figure 1.4. – Examples of τ^- leptons decay with the conservation of the lepton numbers between initial and final states (left), and in the case where the lepton numbers are not conserved, called Lepton Flavour Violation (right).

Majorana [9] ($\nu = \bar{\nu}$), two more phase α_2, α_3 are needed:

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{i\alpha_2/2} & 0 \\ 0 & 0 & e^{i\alpha_3/2} \end{pmatrix}, \tag{1.16}$$

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

Giving the PMNS matrix elements U_{ij} and neglecting the complex phases, the probability of oscillation, using three families, between a flavour state $\nu_{i=e,\mu,\tau}$ to a different state ν_j is:

$$P_{\nu_i \to \nu_j}(x) = \sum_{k=1}^{3} U_{jk}^2 U_{ik}^2 + 2\sum_{k>l} U_{jk}U_{ik}U_{jl}U_{il}\left(1 + 2sin^2\frac{\delta m_{kl}^2 x}{4E}\right). \tag{1.17}$$

The oscillation frequency is given by the mass difference δm_{ij}^2. As shown in Figure 1.5, the neutrino oscillations follow two regimes because we observed $\delta m_{12}^2 \ll \delta m_{23}^2, \delta m_{13}^2$. A slow oscillation dominated by $\nu_1 \to \nu_2$ transition and a fast one. $\nu_1 \to \nu_3$ amplitude is relatively low. Thanks to the oscillation in the two regimes, the system can be

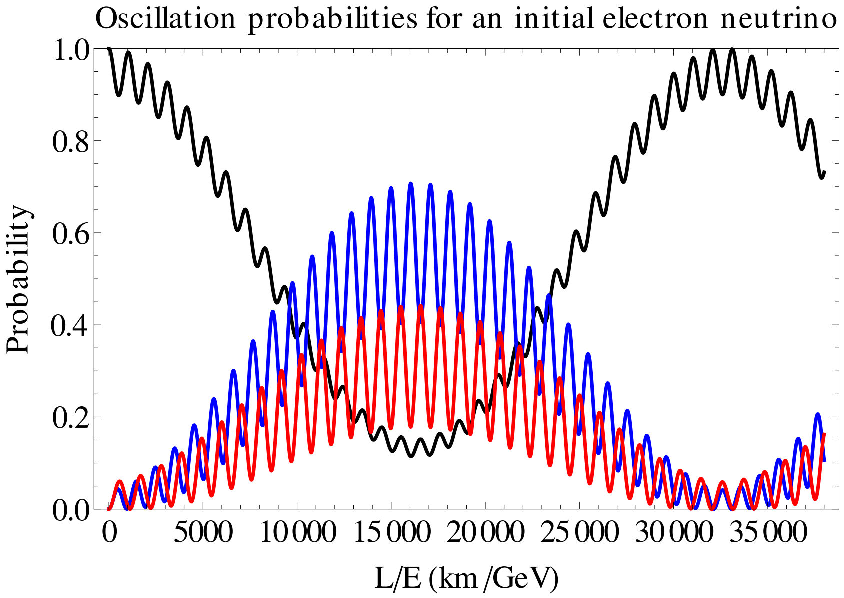


Figure 1.5. – Proportion of neutrino in the function of the distance of propagation. The black, blue and red curves are, respectively the proportion of electron neutrino, muon neutrino and tau neutrino. Credits [42]

---

[9] SM particles can be of Dirac type or Majorana type, depending if they are different or equal to their anti-particle. These types lead to two different mass terms in the SM Lagrangian. Dirac describes masses with a single number, and Majorana describes masses by a complex matrix.

simplified to a two neutrinos oscillation, giving an oscillation length L:

$$L = \frac{4\pi E}{\delta m^2}, \tag{1.18}$$

which depend on the energy E of the initial state and the difference of mass δm^2. The oscillation length is of the order of several kilometres by GeV.

The discovery of neutrino oscillations demonstrates that processes involving LFV can occur within the SM through neutral currents. This discovery also suggests that at least two out of the three neutrinos have mass. Due to the lack of observation of right-handed neutrinos, it is impossible for neutrinos to interact with Higgs fields via Yukawa couplings. However, more complex mechanisms, such as the seesaw mechanism, can explain the differences in behaviour between charged and neutral lepton.

1.3.3. Neutrino oscillations in charged lepton flavour mixing

In the SM, heavy charged leptons can naturally decay into lighter ones through the weak interaction. However, this decay is always accompanied by the emission of a neutrino or anti-neutrino, which preserves the lepton flavour. Nonetheless, the mechanism of neutrino oscillation makes possible transitions between two charged leptons without neutrinos in initial and final states, so a mechanism for charged Lepton Flavour Violation (cLFV) in the SM exists.

cLFV processes, *e.g.* $\mu \to e\gamma$, $\tau \to \ell\gamma$, and $\tau \to \ell\ell\ell$ involve W boson loop or more complex processes like boxes [2, 43, 44]. In the $\tau^- \to \mu^-\mu^+\mu^-$ decay, Figure 1.6, when the τ^- lepton decays, it produces a neutrino ν_τ through the charged weak current. This neutrino oscillates and reabsorbs the W boson, ultimately leading to the creation of a μ^- lepton. Throughout this process, leptons (charged or neutrinos), or W bosons may emit a photon or a Z^0 boson, which can result in the production of a pair of muons. In the resulting process, the initial and final states do not present any neutrinos and the lepton flavour is violated.

The contribution of penguin diagrams in the $\tau \to \mu$ transition has an amplitude $A_{\tau\mu}$ proportional to the PMNS matrix elements $U_{\tau i}$ and $U^*_{\mu i}$. The amplitude also depends on the ratio between the neutrino mass $m^2_{\nu_i}$ and the W boson mass m^2_W such as:

$$A_{\tau\mu} \propto \sum_{i=1}^{3} U_{\tau i} U^*_{\mu i} \left(1 + f \frac{m^2_{\nu_i}}{m^2_W}\right), \tag{1.19}$$

with a factor f. So, the cLFV processes are highly suppressed since the neutrinos masses are small with respect to the W one. Using the current knowledge on the PMNS matrix elements, the branching fraction $\mathscr{B}$ [10] for the $\tau^- \to \ell^-\ell^+\ell^-$ and other cLFV processes is of the order of 10^{-55} [2, 3]. The contributions from the penguin

---

[10] The branching fraction can be viewed as the probability to decays into a given final state.

diagram dominate the branching fraction, while box diagram contributions are of the order of 10^{-57}.

The inclusion of the neutrino oscillation mechanism in the SM allows for the cLFV. However, due to the very low branching fraction for these decays, as explained above, it is unattainable for current and future experiments to observe them. Thus, the observation of cLFV decays can only be attributed to unknown Beyond Standard Model (BSM) processes.

1.4. Lepton flavour violation beyond the Standard Model and implications

The concept of cLFV emerged from new theoretical frameworks attempting to explain the limitations of the SM. A way to constrain the parameter space of NP models or exclude them is by measuring the upper limit[11] of the branching fraction of various LFV decays [45, 46]. When searching for a cLFV process, it is preferable to focus on the third lepton generation rather than the lighter generations due to the mass dependence of several NP model couplings[12]. Additionally, unlike the muon and the electron, the tau lepton can decay into quarks and leptons, allowing access to a larger number of decay modes that can be enhanced by NP models [5, 47, 48, 49, 50] and studied experimentally. The predictions of several BSM models presented in Table 1.1, can be investigated using the current and future experiment sensitivities, including Belle II.

---

[11] The upper limit represents the highest likely value of the parameter. For example, an experiment that sets an upper limit on a parameter at 0.5 can invalidate a model that predicts a value for the parameter between 0.7 and 0.9.

[12] Coupling refers to the interaction between two particles by one of the four fundamental forces. The intensity of the force exerted in this interaction is established by a numerical value known as the coupling constant or gauge coupling parameter.

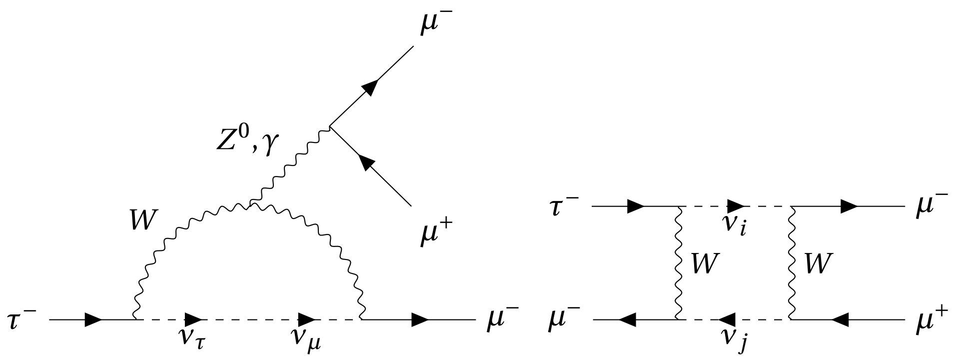


Figure 1.6. – Feynman diagrams for $\tau^- \to \mu^-\mu^+\mu^-$ decays in the presence of neutrino oscillations with Z^0-Penguin contribution (right) and box diagrams contribution (left).

1.4.1. Little Higgs model with T-parity

The Higgs boson plays a significant role in the Standard Model by breaking the electroweak symmetry and providing mass to the W and Z^0 gauge bosons. The observed mass of the Higgs boson, approximately 125 GeV, is of the same order as the W mass, which is necessary for weakly-coupled theories. If the Higgs boson's mass was much larger than the W mass, the Higgs self-interactions would be too strong, resulting in the hierarchy problem.

Little Higgs models [64, 65] aim to stabilize the mass of the Higgs boson by using the spontaneous breaking of approximate global symmetries. The Higgs boson is viewed as a pseudo-Goldstone boson, which are massive bosons responsible for symmetry breaking. However, precision electroweak measurements indicate no evidence of new physics up to 7 TeV instead of 1 TeV, which is known as the "little hierarchy problem." This problem can be addressed by introducing a T-parity symmetry for new particles at the TeV scale, where SM particles are T-even, and new particles are T-odd.

Within the littlest Higgs model with T-parity [51, 52, 53], new mirror leptons are introduced $\begin{pmatrix} \ell_H & \nu_H \end{pmatrix}$ as well as heavy gauge bosons W_H, Z^0_H and A_H. These new mirror particles introduce flavour mixing matrices related to the PMNS matrix. So LFV is embedded by the mixing with mirror leptons. The diagrams of flavour–changing interaction through loops of T–odd particles for the $\tau^- \to \mu^-\mu^+\mu^-$ are represented in Figure 1.7.

1.4.2. Supersymmetry models

SuperSymmetry (SUSY) models are based on a hypothetical symmetry that might exist between bosons and fermions. This concept is particularly relevant in attempts to unify gravity with other interactions, such as supergravity and superstring theories. According to SUSY theory, every fermion must have a boson partner, known as the fermion's superpartner, while every boson must have a fermion partner. So SUSY is a symmetry along the spin, superpartner having the same properties except for spin

Table 1.1. – Summary table of the upper limit on the $\tau^- \to \mu^-\mu^+\mu^-$ branching fraction for different NP theoretical frameworks.

New Physics models	Limit BF for $\tau^- \to \mu^-\mu^+\mu^-$
Littlest Higgs with T-parity [51, 52, 53]	10^{-8}
R-parity violating SUSY [54]	10^{-8}
Non-universal Z' [55, 49]	10^{-8}
MSSM + seesaw [56, 57]	10^{-9}
SUSY SO(10) [58]	10^{-10}
SUSY Higgs [59]	10^{-10}
SM + seesaw [60, 61, 62]	10^{-10}
V(1) Leptoquarks [63]	10^{-12}

($1/2 \rightarrow 1$ and $1 \rightarrow 1/2$). If there is an unbroken supersymmetry, then a particle and its superpartners will have identical masses.

Additionally, SUSY provides a more successful extension for gauge coupling unification than the SM and is a candidate to solve the hierarchy problem. Some versions of SUSY also offer natural candidates for cold dark matter.

1.4.2.1. Supersymmetric models with R-parity violation

The SUSY theory introduces additional particles and interactions that have the potential to violate the baryon number and the lepton number. To prevent these violations from occurring within the framework, a discrete symmetry called R-parity is introduced [54]. R-parity varies between +1 for SM particles (R-even) and -1 for SUSY particles (R-odd). It is defined as follows:

$$R = 3B + L + 2S \tag{1.20}$$

Here, B represents the baryon number, L the lepton number, and S the spin. The conservation of R-parity ensures the stability of the lightest supersymmetric particle in the model, but in some cases, the R-parity can violated.

In the absence of R-parity, the MSSM features supplementary Yukawa interactions between charged leptons (ℓ^-) and sneutrinos ($\tilde{\nu}$). These interactions may not be diagonalized when the Higgs-lepton Yukawa interactions are diagonalized, which

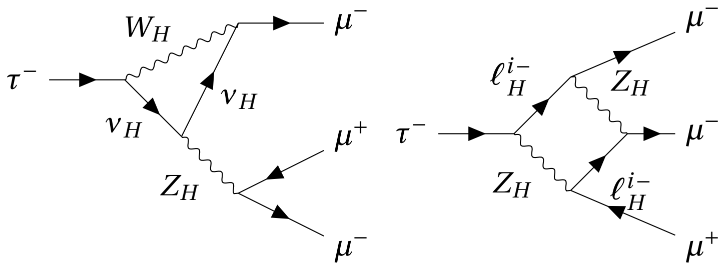


Figure 1.7. – Example Feynman diagrams leading to $\tau^- \rightarrow \mu^-\mu^+\mu^-$ in the Littlest Higgs model with T–parity.

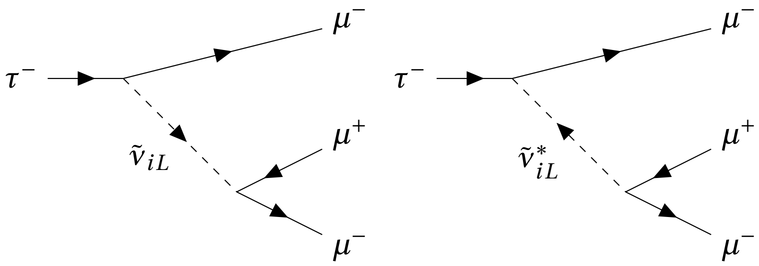


Figure 1.8. – Feynman diagrams of $\tau^- \rightarrow \mu^-\mu^+\mu^-$ in the Minimal Supersymmetric Standard Model (MSSM) without R–parity model.

opens up the potential for tree-level scalar particle-induced cLFV, Figure 1.8. In particular, the $\tau^- \to \mu^-\mu^+\mu^-$ mode follows the effective Lagrangian:

$$\mathscr{L}_{\text{eff}} = \sum_i \left(\frac{1}{m^2_{\tilde{\nu}_{iL}}} \lambda_{i32}\lambda^*_{i22} \left(\bar{\mu}\frac{1+\gamma_5}{2}\mu\right)\left(\bar{\mu}\frac{1-\gamma_5}{2}\tau\right) + \frac{1}{m^2_{\tilde{\nu}_{iL}}} \lambda_{i22}\lambda^*_{i23} \left(\bar{\mu}\frac{1-\gamma_5}{2}\mu\right)\left(\bar{\mu}\frac{1+\gamma_5}{2}\tau\right) \right) \tag{1.21}$$

with the coupling constant λ, $m_{\tilde{\nu}_{iL}}$ the mass of the mediator sneutrino between lepton tau τ and muons μ.

1.4.2.2. Supersymmetric models with see-saw mechanism

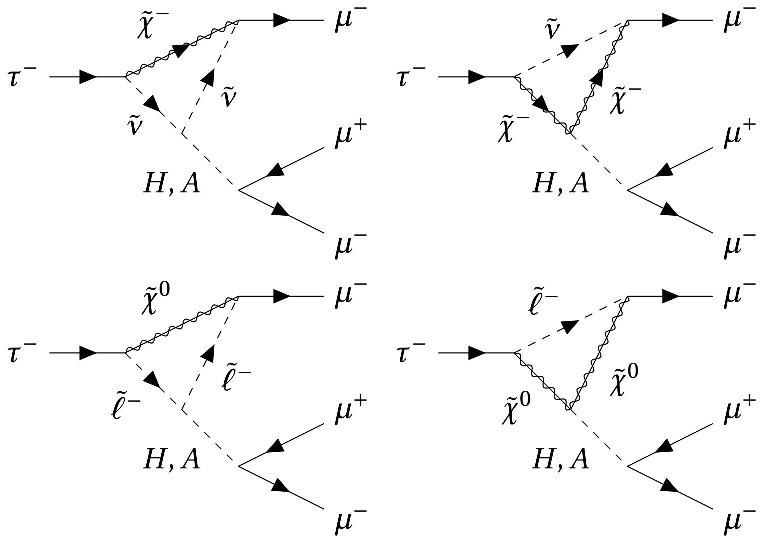


Figure 1.9. – Feynman diagrams of $\tau^- \to \mu^-\mu^+\mu^-$ in the MSSM +seesaw mechanism model.

In the SM, the fermions arise from the Yukawa couplings with the Higgs fields. The Yukawa couplings fail to explain the difference between neutrinos and charged leptons masses. The "seesaw mechanism" offers a response to that problem by introducing right-handed sterile neutrinos $\tilde{\chi}$ in order to couple with left-handed neutrinos through $SU(2) \times U(1)$-violating Dirac mass terms, m_D, while also receiving large, $SU(2) \times U(1)$-invariant Majorana masses, M_R. So the seesaw mechanism defines two types of neutrinos according to their mass: light neutrinos ν primarily left-handed with an extremely small mass m_D^2/M_R and heavy neutrinos primarily right-handed with masses $M_R \sim 1\,\text{TeV}$. As particle physics experiments did not observe such heavy neutrino, it is expected to be sterile, meaning that it doesn't interact through the electroweak interaction.

In Chapter 1, Section 1.3.3, we learned that massive neutrino oscillations lead to LFV. However, in the SM, flavour violation in charged lepton processes, particularly $\tau^- \to \mu^-\mu^+\mu^-$, is suppressed by factors $m^2_{\nu_i}/m^2_W$ simplified as $1/M_R^2$ with the seesaw meachnism because $m_{\nu_i} \approx m_D^2/M_R$ and m$_D$ and m_W are of the same order. In the MSSM with seesaw mechanism [56, 57], LFV can be directly communicated by heavy right-handed neutrinos $\tilde{\chi}$, sleptons $\tilde{\ell}$ and $\tilde{\nu}$ and additional Higgs particles (H and A) introduced in the MSSM, Figure 1.9. In this case, LFV is suppressed by factors of $1/M_{SUSY}^2$ instead of $1/M_R^2$, where $M_{SUSY}^2 \ll M_R^2$. The effective lagrangian for such LFV interaction follows:

$$-\mathscr{L} \simeq \left(2G_F^2\right)^{1/4} \frac{m_\tau \kappa_{32}}{\cos^2\beta} \bar{\tau}_R \quad \mu_L\left(\cos(\beta-\alpha)h^0 - \sin(\beta-\alpha)H - iA\right) + h.c. \tag{1.22}$$

where G_F is the Fermi coupling constant given by $G_F = g^2/(4\sqrt{2}m_W^2)$, κ is a mass mixing parameter, and β is defined as the ratio between the vacuum expectation values of two Higgs particles (v_H/v_A). It gives an expected branching fraction:

$$\mathscr{B}(\tau^- \to \mu^-\mu^+\mu^-) \simeq \left(1\times 10^{-7}\right) \times \left(\frac{\tan\beta}{60}\right)^6 \times \left(\frac{100\text{GeV}}{m_A}\right)^4, \tag{1.23}$$

where m_A represents the mass of a new Higgs field introduced by the MSSM.

1.4.3. Leptoquarks hypothesis

Over the years, one of the most appealing results from experiments such as BaBar, Belle, and LHCb are the Lepton Flavour Universality (LFU) tests. LFU assumes that the interactions between electroweak gauge bosons and leptons do not vary based on their flavour. To test this theory, these experiments have studied semileptonic B meson decays by analyzing the ratio of their branching fractions and comparing them to the predictions of the Standard Model. Two observables $R_{K^{(*)}}$ and $R_{D^{(*)}}$ are defined as:

$$R_{K^{(*)}} = \frac{\mathscr{B}\left(B \to K^{(*)}\mu^+\mu^-\right)}{\mathscr{B}\left(B \to K^{(*)}e^+e^-\right)}, \qquad R_{D^{(*)}} = \frac{\mathscr{B}\left(B \to D^{(*)}\tau\bar{\nu}_\tau\right)}{\mathscr{B}\left(B \to D^{(*)}\ell\bar{\nu}_\ell\right)}, \tag{1.24}$$

where $\ell = e, \mu$.

The two observables' measurements suggest that the LFU test presents anomalies with respect to the SM. One solution to this issue is introducing leptoquarks [49], which are hypothetical particles interacting with quarks and leptons. Leptoquarks are bosons belonging to the colour-triplet ($red, green, blue$) group; they carry both baryon B and lepton numbers L joined together in a fermion number $F = 3B + L$. They introduce unknown interactions between both types of fermions, converting quarks into leptons and vice versa.

Even if the golden channel in τ decays is $\tau^- \to \ell^-\phi$ with potential branching fraction just below the current experimental limit, the vector leptoquarks V_1 can contribute to three-body LFV decays as $\tau^- \to \mu^-\mu^+\mu^-$. Indeed, vector leptoquarks V_1 can induce $\tau^- \to \mu^-\mu^+\mu^-$ at the loop level through photon penguins, Z^0 penguins and box diagrams [63]. The effective Lagrangian relevant for these decays can be expressed as:

$$\begin{aligned}\mathscr{B}(\tau^- \to \mu^-\mu^+\mu^-) =& 2\left(|g_3|^2+|g_4|^2\right)+|g_5|^2+|g_6|^2+\\ &+8e\,\mathrm{Re}\left[C_R^{\mu e}(2g_4^*+g_6^*)+C_L^{\mu e}(2g_3^*+g_5^*)\right]+\\ &+\frac{32e^2}{m_\mu^2}\left\{\ln\frac{m_\mu^2}{m_e^2}-\frac{11}{4}\right\}\left(|C_R^{\mu e}|^2+|C_L^{\mu e}|^2\right);\end{aligned} \tag{1.25}$$

where the photon penguins, Z penguins, and box diagrams contribute to g3, g4, g5 and g6 coefficients. Depending on the mass of the vector leptoquark candidate, the branching fraction is up to 10^{-10} as shown in Figure 1.10.

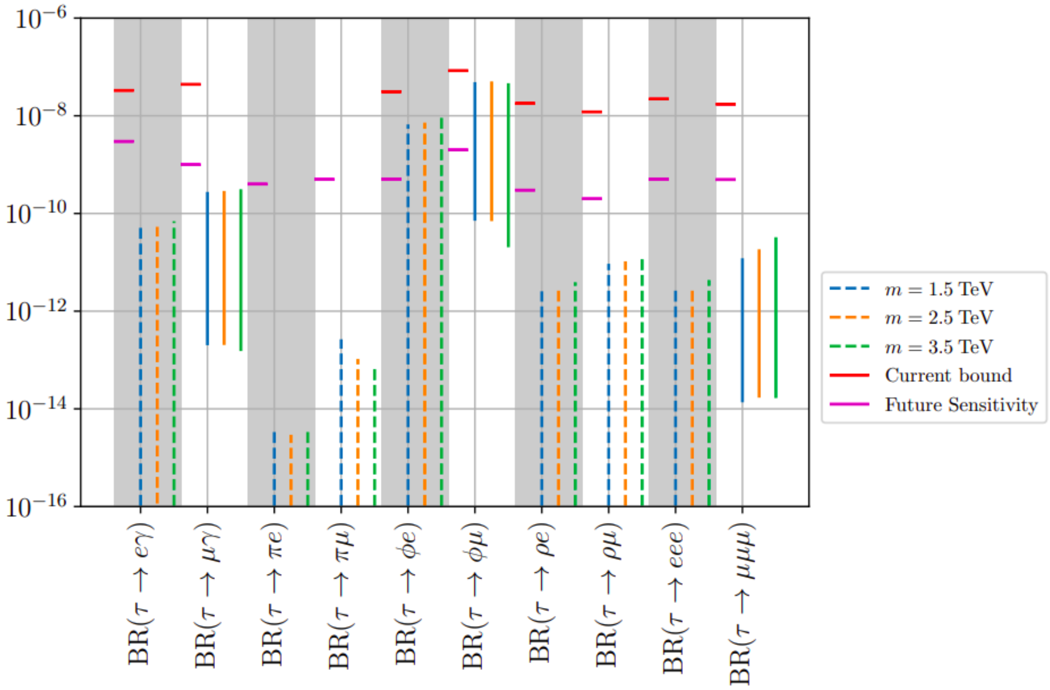


Figure 1.10. – Current and expected upper limits (red and purple horizontal lines) on the branching fractions of lepton flavour violating τ decay modes studied at the Belle II experiment, and predicted ranges at 90% Confidence Level (CL) in the vector leptoquark hypothesis for input masses $m_U = 1.5, 2.5, 3.5\,\mathrm{TeV}/c^2$ (blue, yellow and green solid or dashed lines), Credits [63].

1.4.4. Non-universal Z'

The extension of the electroweak interaction in the SM by extra gauge bosons Z' and W' is introduced in several models. The vector boson Z' model can contribute to $\tau^- \to \mu^-\mu^+\mu^-$ LFV decays [55, 49], as shown in Figure 1.11 by introducing four-leptons operators at the tree level. The branching fraction for $\tau^- \to \mu^-\mu^+\mu^-$ decays in the Z' models can be written as:

$$\mathscr{B}\left(\tau^- \to \mu^-\mu^+\mu^-\right) = X\frac{\left(g_{\ell V}^{33}\right)^4}{16m_V^4}\frac{m_\tau^5\tau_\tau}{192\pi^3}\sin^6\theta_L\cos^2\theta_L, \tag{1.26}$$

where X is a suppression factor due to the non-zero muon mass, τ_τ is the τ^- lifetime, m_τ the τ^- mass, m_V the mass of the vector boson candidate, and θ_L is a parameter arising from the lepton mass basis transformation. Following the current constraints on θ_L parameter, the $\tau^- \to \mu^-\mu^+\mu^-$ LFV decay is predicted at a branching fraction of 10^{-8}. The prediction from vector boson models is already at the experimental sensitivity. The improvements on Upper-Limit (UL) of $\tau^- \to \mu^-\mu^+\mu^-$ will constrain θ_L.

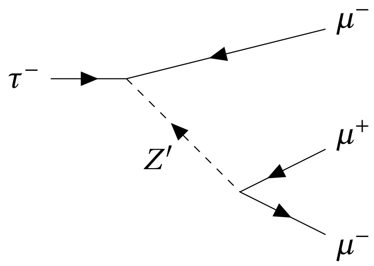


Figure 1.11. – Feynman diagrams leading to $\tau^- \to \mu^-\mu^+\mu^-$ in the models introducing Z' bosons.

1.5. Experiment status in LFV searches

During the last forty years, tau LFV decays searches have been ongoing in multiple particle physics experiments from CLEO, to the B-factories, Belle and BABAR, and the LHC. No evidence of these decays has been found in any of the searches conducted, but the upper limit on the branching fractions has been improved up to $\sim 10^{-8}$ [8] for the current most stringent experiments.

The best upper limit on $\tau^- \to \mu^-\mu^+\mu^-$ was obtained by Belle at the level of 2.1×10^{-8} at 90% CL with $782 fb^{-1}$, while BABAR [7] put a limit at 3.3×10^{-8} with 468 fb^{-1} [8]. This mode was also searched for by LHCb [9] ($3 fb^{-1}$), ATLAS ($20.3 fb^{-1}$) and CMS [11, 66] (33.2 fb^{-1}), which obtained upper limits of 4.6, 38 and 8.0×10^{-8} at 90% CL, respectively [10]. Recently, CMS provided an update adding 2017 and 2018 data for a total of 131 fb^{-1} [66], reaching a limit of 2.9×10^{-8} at 90% CL.

The other $\tau^- \to \ell^-\ell^+\ell^-$ modes are more difficult to study in hadronic environment. Only B-factories have examined final states that involve electrons. An experimental summary for the search of $\tau^- \to \ell^-\ell^+\ell^-$ LFV decays can be found in Table 1.2.

The current experimental bounds at 10^{-8} are close to the theoretical branching fraction given by BSM models (see Section 1.4). Despite the long ongoing work on tau LFV searches, this remains an important topic as we are at the sensitivity edge to probe NP.

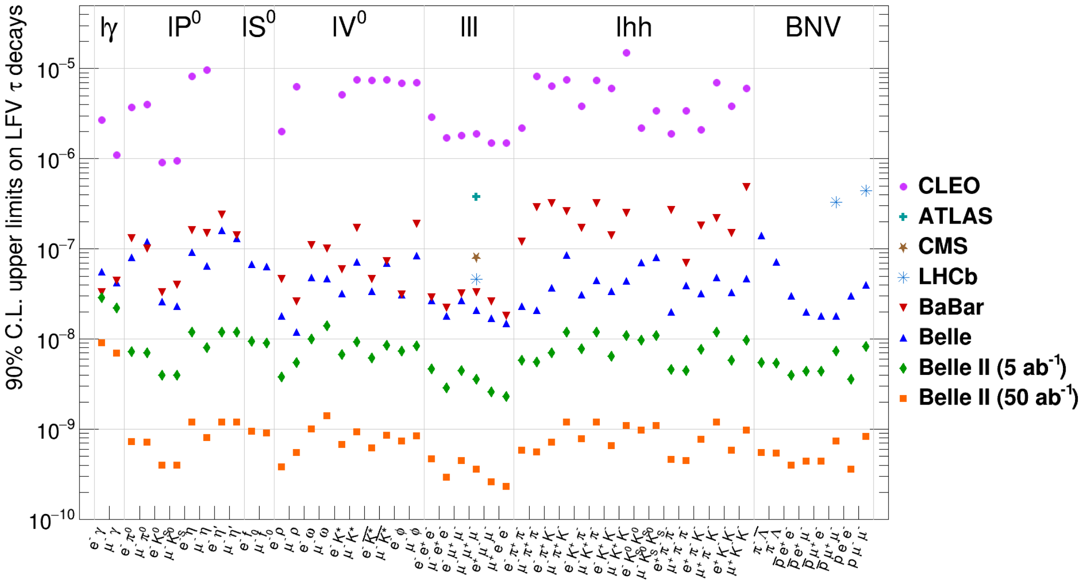


Figure 1.12. – Upper limits on branching fractions at 90% CL for τ^- LFV decays: $\tau \rightarrow \ell\gamma$, $\tau \rightarrow \ell P^0/S^0/V^0$ (neutral pseudoscalar, scalar or vector mesons), $\tau \rightarrow \ell\ell\ell$, $\tau \rightarrow \ell hh$ (two hadrons) with $\ell = e, \mu$. The green and orange marks correspond to expected upper limits at Belle II for integrated luminosities of 5 ab^{-1} and 50 ab^{-1}, the other to current limits from CLEO, ATLAS, CMS, LHCb, BABARand Belle. Credits [67]

At the head of the incoming τ LFV search stand LHC experiments mainly with LHCb and CMS and Belle II. In particular, the Belle II experiment aims to improve the upper limit by one to two orders of magnitude with respectively 5 ab^{-1} and 50 ab^{-1} for a wide range of decays, Figure 1.12. For the $\tau^- \rightarrow \mu^-\mu^+\mu^-$ channel, the future

Table 1.2. – Observed upper limits at 90% C.L. on $\tau \rightarrow \ell\ell\ell$ branching fractions obtained by BABAR [7], Belle [8] and the LHC experiments. Values are given multiplied by 10^8.

Mode	Belle	Babar	LHCb	ATLAS	CMS
$\mu^-\mu^+\mu^-$	2.1	3.3	4.6	3.8	2.9
$e^-e^+e^-$	2.7	2.9	-	-	-
$e^-\mu^+\mu^-$	2.7	3.2	-	-	-
$e^-e^+\mu^-$	1.8	2.2	-	-	-
$e^+\mu^-\mu^-$	1.7	2.6	-	-	-
$\mu^+e^-e^-$	1.5	1.8	-	-	-

Belle II measurements will reach an upper limit on the branching fraction between 10^{-9} and 10^{-10}. In the short future, before the next generation of colliders, τ LFV and especially $\tau^- \to \mu^-\mu^+\mu^-$ experimental searches will be able to challenge the theoretical expectation from the BSM models.

2. The Belle II experiment

Table of contents

Belle II [12, 68, 69] is a particle physics collaboration working on the data collected by the eponymous detector. It has been operating since 2019 at the SuperKEKB electron-positron collider [70] hosted by KEK laboratory in Tsukuba, Japan. The experiment is an upgrade of its predecessor Belle [71], which ran from 1999 to 2010, of which it took over the physics goals.

2.1. The SuperKEKB electron-positron collider

The electron-positron colliders operating at the $\Upsilon(4S)$ energy, called B-factories, were originally designed to search for Charged Parity (CP) violation in B meson decays, meaning that the matter and antimatter do not follow the same natural laws in the B sector. This phenomenon was described, in the 1980s, by Cabibbo-Kobayashi-Maskawa [34, 35] and the eponymous matrix allowing quark flavour mixing within the Standard Model (SM). Their works were confirmed by the BaBar [72, 73] and Belle [74, 71] experiments, which led to a Nobel Prize in 2008 for Kobayashi and Maskawa.

Today, the physics of B factory is complementary to the work performed by the Large Hadron Collider. Indeed e^+e^- collisions lead to a clean environment with lower track multiplicity and detector occupancy, resulting in an excellent laboratory for B, D mesons and τ^+ leptons studies with a high reconstruction efficiency.

2.1.1. The SuperKEKB accelerator

SuperKEKB performs asymmetric collisions between electron and positron beams with an energy of 7 and 4 GeV, respectively. The resulting collision energy at the centre of mass is $\sqrt{s} = 10.58\,\text{GeV}$ and corresponds to the mass of the $\Upsilon(4S)$ resonance, a $b\bar{b}$ bound state. Its branching fraction to decay into B^+B^- or $B^0\bar{B}^0$ is about 96%. The energy asymmetry of the beam leads to a Lorentz gain, β_γ, of the laboratory referential [1] compared to the centre-of-mass system:

$$\beta_\gamma = \frac{P_{e^-} - P_{e^+}}{\sqrt{s}} \simeq \frac{E_{e^-} - E_{e^+}}{\sqrt{4E_{e^-}E_{e^+}}} \simeq 0.28. \tag{2.1}$$

Considering the Belle II boost, B mesons can fly a certain distance, in average $\sim 130\,\mu\text{m}$, before decaying, which can be measured thanks to the excellent resolution of the two vertex detectors. The decay length is important to separate the mixed B^0 mesons as the decay time is a compelling component for CP violation studies. The boost was reduced with respect to the one used in KEKB ($\beta_\gamma = 0.42$) in order to cope with the increase of instantaneous luminosity.

Even though SuperKEKB is the upgrade of KEKB, they use the same infrastructures, including the tunnel with a circumference of 3 km. The electrons are produced by the interaction of a pulsed laser on a cathode in the pre-injector accelerator. The electrons are accelerated to 7 GeV in the linear accelerator (linac) before they enter the High Energy Ring (HER), Figure 2.1. The positrons are created by causing an interaction between a portion of the electrons and a tungsten target located in the first half of the linac. The positrons pass through a damping ring to reduce the beam emittance needed for the desired luminosity before being injected into the Low Energy Ring (LER) after reaching 4 GeV in the second half of the linac. After a final focusing along the z axis [2] of the Belle II barrel by quadrupole magnets, the two beams finally collide

---

[1] In the analysis, we distinguish two referential where the quantities such as momentum and energy are measured. The referential of the laboratory is the referential of an exterior spectator looking at the experiment. The second is the centre-of-mass system corresponding to the referential, where the total momentum of the two colliding electrons/positrons is zero.

[2] In the laboratory system, the origin is defined at the interaction point between the e^+e^- beams. The z axis is in the direction of the electron beam, and the x and y axis are in the transverse plane of the detector.

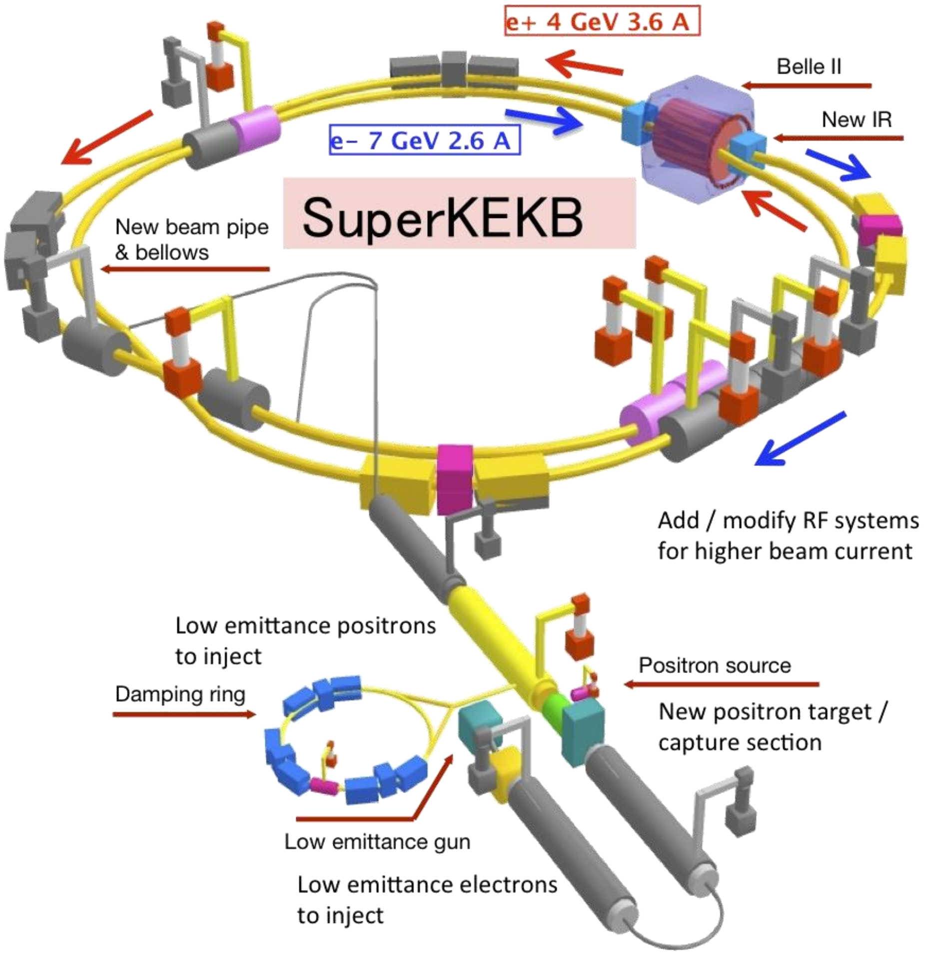


Figure 2.1. – The schematic view of the asymmetric electron-positron collider SuperKEKB systems, from the electron and positron sources to the interaction point under Belle II. Credits [12]

at the Interaction Point (IP).

2.1.2. The nano-beam scheme

Table 2.1. – Machine parameters for KEKB in its final configuration and for SuperKEKB in June 2022 and its final design.

Parameters	Unit	KEKB	SuperKEKB	
		(HER/LER)	2022	Design
Beam Energy	GeV	8.0/3.5	7.0/4.0	7.0/4.0
Beam Current (I)	A	1.19/1.64	1.099/1.321	2.62/3.60
Beam Size at IP (x)	μm	80	16.6/17.9	11.2/10.2
Beam Size at IP (y)	μm	1	0.215	0.0618/0.0483
Beam Size at IP (z)	mm	5	-	5/6
ξ_y		0.090/0.129	0.0279/0.0407	0.088/0.090
β_y^*	mm	5.9/5.9	1.0/1.0	0.41/0.27
Lorentz factor ($\beta\gamma$)		0.43	0.28	0.28
Half-crossing angle	mrad	11	41.5	41.5
Instantaneous Luminosity	$10^{34}\,\text{cm}^{-2}\text{s}^{-1}$	2.1	4.65	80

Belle II was designed to record up to 55.5×10^9 $B\bar{B}$ mesons pairs, also expressed in term of 50 ab^{-1} in integrated Luminosity L[3] through:

$$N = L \times \sigma, \tag{2.2}$$

with the number of produced particles N, and the $B\bar{B}$ production cross-section σ of 1.05 nb for an e^+e^- collision at $\sqrt{s} = m(\Upsilon(4S)) = 10.58\,\text{GeV}$. This amount of collected data represents fifty times the Belle dataset.

The whole Belle II 50 ab^{-1} data collection will be provided within 2035, by reaching a maximum instantaneous luminosity $\mathscr{L}$ of $6 \times 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$ delivered by SuperKEKB:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathscr{L}\sigma. \tag{2.3}$$

The SuperKEKB machine can play on several parameters to deliver this unprecedented instantaneous luminosity described by [75]:

$$\mathscr{L} = \frac{\gamma^{\pm}}{2er_e}\left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)\frac{I^{\pm}\xi_y}{\beta_y}\frac{R_L}{R_{\xi_y}}, \tag{2.4}$$

where e, r_e and $\gamma^{\pm}$ are, respectively, the elementary electric charge, the electron radius, and the Lorentz factor while ± signs distinguish the positrons + and electrons −

---

[3] The luminosity is given in barn unit b which is equivalent to an area $1\,b = 10^{-24}\,\text{cm}^2$.

beam. R_L and R_{ξ_y} stand for the Luminosity and beam-beam reduction factors; those two ratios are approximated by 1. In addition, the term in parenthesis can also be considered equal to one since the ratio between the σ_y^* vertical and σ_x^* horizontal beam sizes is order 10^{-3}. So, the instantaneous luminosity can be reduced to:

$$\mathscr{L} \approx \frac{I^{\pm}\xi_y}{\beta_y}, \tag{2.5}$$

where the remaining parameters to play with for increasing the luminosity are the beam current (I) which is multiplied by a factor of two with respect to KEKB, the vertical beam-beam parameter ξ_y and the beta function β_y at the IP.

Thus the main modification in the nano-beam scheme, proposed by SuperB [76], adopted by SuperKEKB is to minimise the beta function β_y, which can be seen as minimizing the longitudinal size of the beams overlap d, Figure 2.2. This is achieved by introducing a non-zero half-crossing angle ϕ according to:

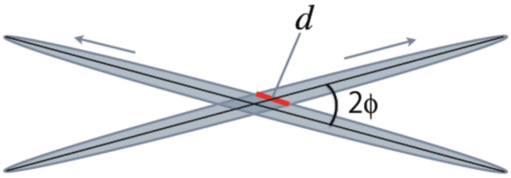


Figure 2.2. – View of the two beams colliding under the nano-beam scheme configuration: the effective beam size d and the half-crossing angle ϕ.

$$d \simeq \frac{\sigma_x^*}{\phi}. \tag{2.6}$$

In Belle II the half-crossing angle ϕ is $\sim 41.5\,\text{mrad}$, which is four times larger than Belle. Nevertheless, the choice of the angle is related to the magnet design, Figure 2.3, and the detector background level. Indeed, passing a certain point, the increase of the half-crossing angle is deteriorating ξ_y. The list of the main upgrades with respect to KEKB machine parameters is given in Table 2.1.

2.1.3. Particle production and beam backgrounds

If the first goal of B-factories is to study B mesons, the $\Upsilon(4S)$ resonance is not the dominant particle produced by e^+e^- collision at $\sqrt{s} = 10.58\,\text{GeV}$, as shown in Table 2.2. The particle production is dominated by Bhabha scattering $e^{\pm} \rightarrow e^+e^-(\gamma)$ and other low-multiplicity processes : $\mu^+\mu^-(\gamma)$, $e^+e^-e^+e^-$, $e^+e^-\mu^+\mu^-$, $\mu^+\mu^-\mu^+\mu^-$... The cross-sections of the other $b\bar{b}$ processes ($u\bar{u}$, $d\bar{d}$, $s\bar{s}$ and $c\bar{c}$) have the same order of magnitude as $\Upsilon(4S)$, which leads to a high background contamination, called continuum background. Belle II is also suitable for τ physics since $\tau^+\tau^-$ process, shown in Figure 2.4, has a cross-section similar to the $B\bar{B}$ one. Trigger systems and skim flags

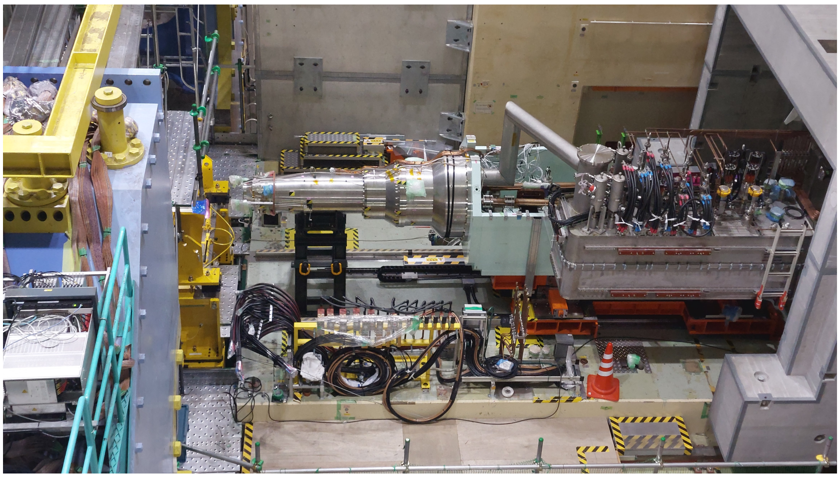


Figure 2.3. – Final focusing magnets removed from the Belle II detector during the Long Shutdown 1 (LS1).

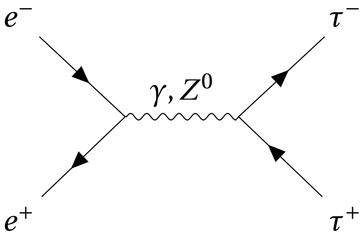


Figure 2.4. – Feynamn diagram of the τ-pair production in the electrons-positrons collider with a cross-section of 0.919 nb.

Table 2.2. – Production cross section for the main physics processes of e^+e^- collisions at $\sqrt{s} = m(\Upsilon(4S)) = 10.58\,\text{GeV}$ [12].

Physics process	Cross section (nb)
$\Upsilon(4S)$	1.110
$u\overline{u}\ (\gamma)$	1.61
$d\overline{d}\ (\gamma)$	0.40
$c\overline{c}\ (\gamma)$	1.30
$s\overline{s}\ (\gamma)$	0.38
$\gamma\gamma\ (\gamma)$	4.99
$e^+e^-\ (\gamma)$	300
$\mu^+\mu^-\ (\gamma)$	1.148
$\tau^+\tau^-\ (\gamma)$	0.919
$e^+e^-\ e^+e^-$	39.7
$e^+e^-\ \mu^+\mu^-$	18.9
$e^+e^-\ \tau^+\tau^-$	0.018
$e^+e^-\ \pi^+\pi^-$	1.895
$e^+e^-\ K^+K^-$	0.079
$e^+e^- p\bar{p}$	0.012
$\mu^+\mu^-\ \mu^+\mu^-$	3.5×10^{-4}
$\mu^+\mu^-\ \tau^+\tau^-$	1.4×10^{-4}
$\tau^+\tau^-\ \tau^+\tau^-$	2.1×10^{-7}
$\pi^-\ \pi^+$ ISR	0.167
$\pi^-\ \pi^+\ \pi^0$ ISR	0.024
K^+K^- ISR	0.016
$K^0\overline{K}^0$ ISR	0.009

are designed to select only interesting physics events and reduce, particularly, the low multiplicity processes.

In addition to these physics processes, the beams can create other particles. There are five different types of background coming from the beams [12]:

- Synchrotron radiation: When charged particles move along a curved path in an electric field, they get accelerated and emit X-rays in the process. The intensity of the radiation is determined by the current of the beam and the magnetic field.
- Beam-gas scattering: When residual gas is in the beam pipe, it can interact and deviate with the electrons of the beam. Deviated electrons can produce secondary particle showers that can leave tracks in the detector. The amount of scattering that occurs is directly proportional to the square of the beam's current and the pressure within the beam pipe.
- Touschek scattering: When the beam is squeezed, Coulomb scattering may occur within it, causing electrons to be deflected and interact with the pipe edges. This type of background is more prevalent than the other two mentioned and is estimated to be 20 times higher than in KEKB at nominal luminosity due to the nano-beam scheme.
- Radiative-Bhabha process and electron-positron pair production: In most cases, collisions do not result in the creation of a $\Upsilon(4S)$ particle. Instead, the two beams interact in two different ways. An electromagnetic interaction occurs between the electrons or positrons, causing them to deviate and produce either a photon or a low-momentum electron-positron pair. The production rate for both is directly proportional to the luminosity, which means it is expected to be 40 times higher than the background level at KEKB under nominal luminosity.

The first three types of beam-induced backgrounds can create unwanted particles far from the detector's centre, making it challenging to simulate accurately. During SuperKEKB commissioning, specific measurement instrumentation (BEAST) was used to qualify the beam background. The collaboration constantly monitors the beam-induced background rates and performs simulations to reproduce the measured level.

2.1.4. SuperKEKB operation

SuperKEKB started its operation in 2016, and to date, it has successfully completed the first three phases:

- Phase 1 (2016): was dedicated to commissioning run to estimate beam-induced background with the BEAST detector, [77]. At this stage, the final focus magnets and the Belle II detectors were not yet installed, and no collisions were performed.
- Phase 2 (2018): was mostly used for commissioning studies with the installed detector except for Vertex Detector. With the $496\,\mathrm{pb}^{-1}$ collected during this phase, the Belle II collaboration has published its first results [78, 79].

- Phase 3 (2019-2022): The first phase dedicated to physics runs with the almost complete Belle II detector. On 15 June 2020, SuperKEKB broke the LHC instantaneous luminosity world record, and in 2022, it was set at $4.7 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Phase 3 ended in summer 2022 with $362\,\mathrm{fb}^{-1}$ at the $\Upsilon(4S)$ resonance [80]. SuperKEKB has also scanned other collision energies, *e.g.* $\Upsilon(4S)$ off-resonance and $\Upsilon(5S)$, for a total dataset of $424\,\mathrm{fb}^{-1}$.

Since autumn 2022, SuperKEKB and Belle II have been in the LS1 phase [81, 82] to prepare the next round of data-taking planned for Winter 2023. The detector upgrades performed during the LS1 are:

- The replacement of the beam-pipe at the interaction point,
- The installation of the second layer of pixel sensors for the PiXel Detector (PXD),
- The replacement of the photomultipliers of the central Particle IDentification (PID) detector (Time-Of-Propagation (TOP)),
- The replacement of the ageing components,
- The upgrade of the data-acquisition system by transitioning to new cards and monitoring.

Regarding the accelerator, work was being done to improve the injection of electrons and positrons in the collider and the final collimators and add shieldings to decrease the backgrounds.

2.2. The Belle II detector

The Belle II detector, Figure 2.5, surrounds the interaction point where electrons and positrons collide in SuperKEKB. Belle II is $7 \times 7.5m^2$ in size and weighs 1400 tons [68]. It is, for a large part, an upgrade of the previous Belle experiment's device, designed as the piling of multiple layers, each component being a sub-detector with its own functions. A simple 3D representation can be found in Figure 2.6 and a more detailed 2D description in Figure 2.7.

The detector's main purpose is to reconstruct both charged and neutral particles that are produced during collisions. It aims to do so by meeting specific requirements, including precise measurements of particle's four-momentum and space-time coordinates. It must also cover a large polar angle, have an efficient PID system, and make accurate trigger decisions at a high frequency.

The upcoming sections will detail the various components of the Belle II detector, starting from the innermost layer and progressing outward.

2.2.1. The Pixel Vertex Detector

Due to the nano-beam design of SuperKEKB, the beampipe radius is tiny ($\sim 10\,\mathrm{mm}$). This architecture has the advantage of allowing the VerteX Detector (VXD) to be close to the IP. In counterpart, the VXD has to handle an extremely high hit rate, mainly

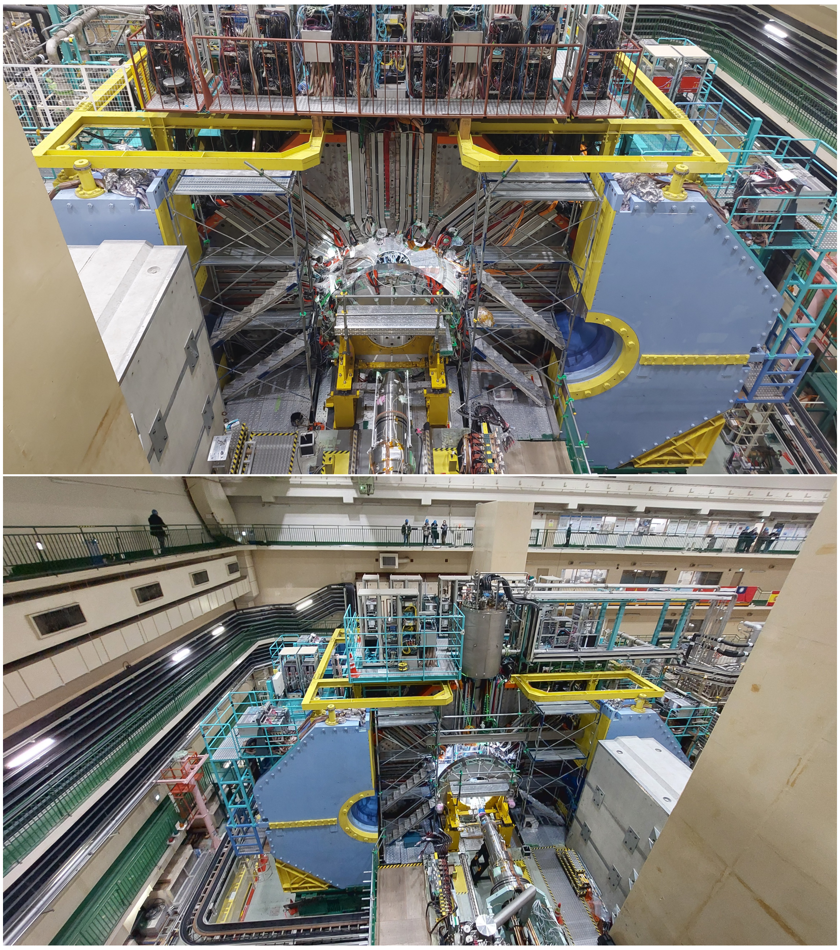


Figure 2.5. – Pictures of Belle II detectors located at KEK, Japan. The detector was opened on the occasion of the maintenance LS1.

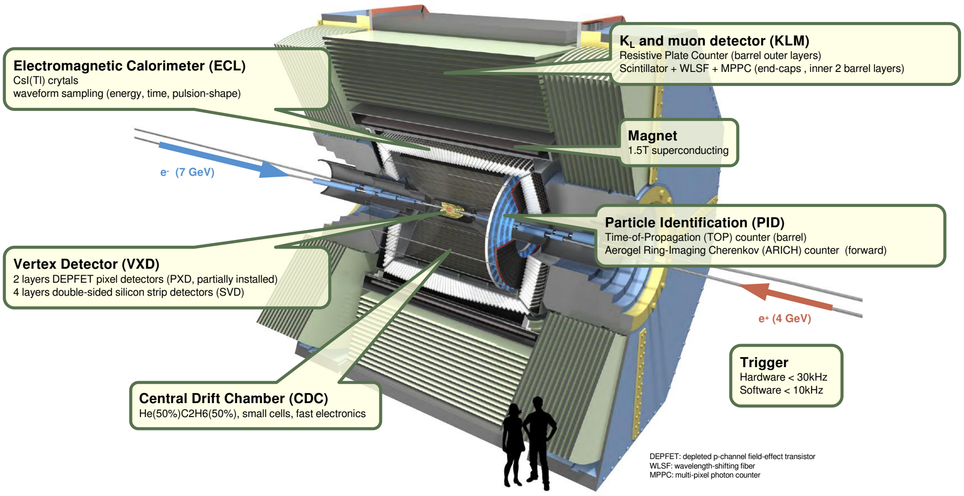


Figure 2.6. – Schematic three dimension view of the Belle II detector. Credits [12]

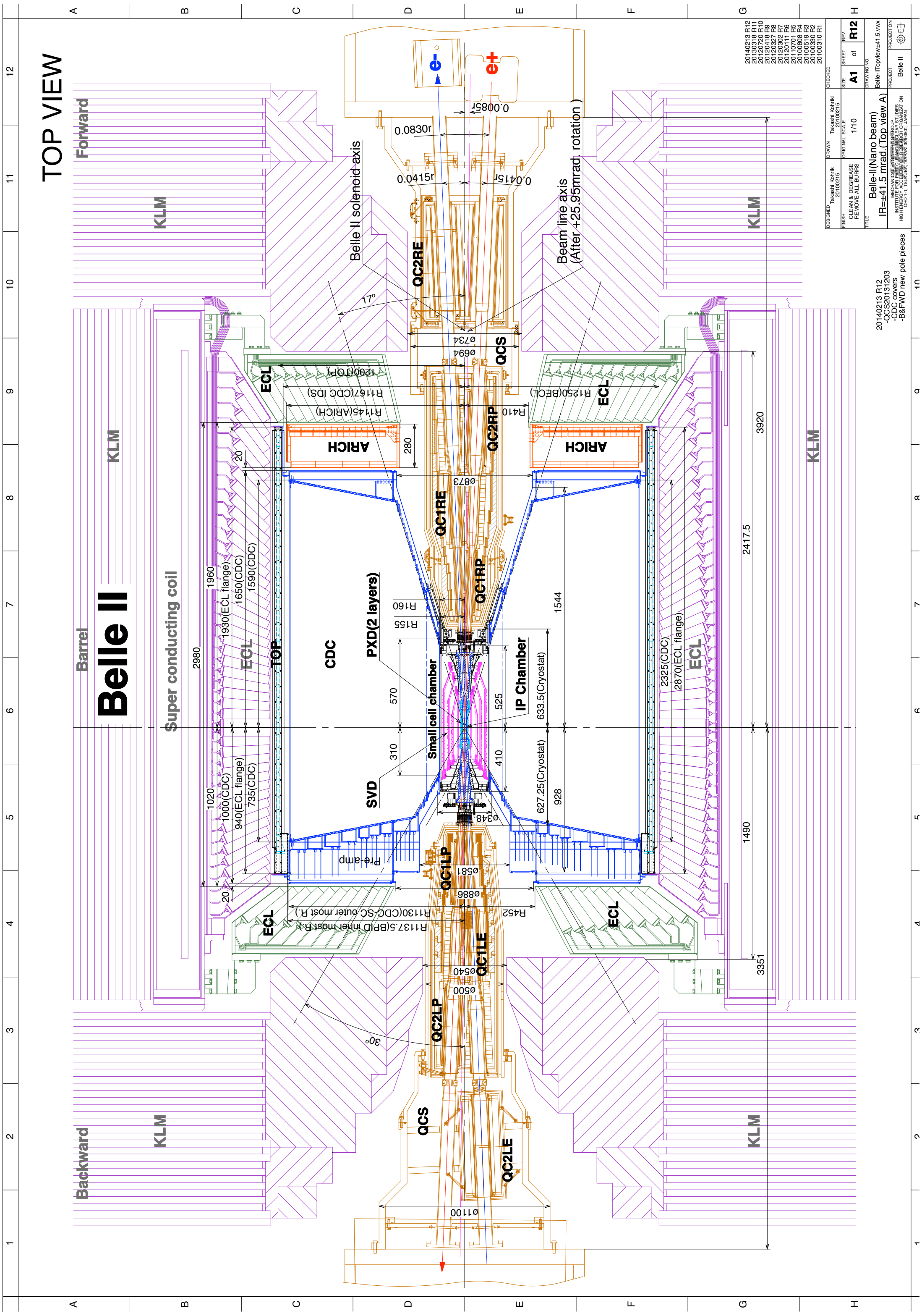


Figure 2.7. – Detailled longitudinal scheme view of the Belle II detector. Credits [12]

from the beam background. Pixel detectors are used for inner VXD layers at small radii, while the outer layers rely on Silicon strips technology [4].

The VXD architecture incorporates two inner layers based on ultrathin DEPleted Field Effect Transistor (DEPFET) pixels (75 μm), the PXD [83]. The readout electronics and their active coolings are exported beyond the acceptance region, reducing the material budget that leads to multiple scattering. The only part of electronics inside the acceptance region is the switcher cooled by skinny nitrogen pipes. The pixel itself is cooled by air, the power dissipation is sufficient thanks to the low consumption of the pixels. The design was tested to experience resistance against radiation.

DEPFET pixels combine detection and amplification in a completely depleted silicon substrate using a p-channel MOSFET structure. An internal gate collects the electrons liberated by the passing charged particles. The inner gate modulates the current at readout time, allowing the detection of the accumulated charge. The sensors of the PXD design are stacked in 8 and 12 planar modules (ladders) to form two layers with 14 mm and 22 mm radii, respectively, shown in Figure 2.8. The sensitive lengths are calculated to match the polar angular acceptance range of 17° to 155°. The pixel sizes $50 \times 55 - 85\,\mu\text{m}^2$ verify the 10 μm hit resolution requirements. The module is read in 20 μs by reading four rows over 1600 in 100 ns simultaneously.

After some issues during the assembly of the PXD layers, only half of the designed PXD was mounted. Early phase 3 started in March 2019 and only used a full first layer and two ladders in the second one. The installation has been completed during the first long shutdown in 2023.

2.2.2. The Silicon Vertex Detector

In the field of physics at the B factories, low-momentum particles play a crucial role. However, they pose a challenge for the track-finding system due to their sensitivity to multiple scattering. For this reason, the Silicon Vertex Detector (SVD) system [84], Figure 2.9, requires a very low material budget and a short shaping time to limit occupancy to just a few per cent. These characteristics are essential for the track-finding algorithms to reject background effectively. To achieve these goals, the sensors are designed in a specific geometrical shape and equipped with appropriate readout electronics. The use of large silicon wafers and trapezoidal slanted sensors in the forward region solves the issue of geometrical shape.

The SVD is made up of 172 double-sided strip sensors, which are arranged in four layers, Figure 2.10, with varying radii. These radii are 39 mm (layer 3), 80 mm (layer 4), 104 mm (layer 5), and 135 mm (layer 6) [84]. The coverage of the azimuthal angle ranges from 17° (in the forward region) to 150° (in the backward region). This asymmetry is due to the Lorentz boost in the laboratory frame, which favours events that are boosted in the forward direction. The strips on the sensors are arranged in perpendicular directions on opposite sides. The u/P side measures the $r\phi$-direction,

---

[4] The hit rate and the beam background are proportional to the transverse distance from the beam. So after a given radius, the use of silicon strips is safe.

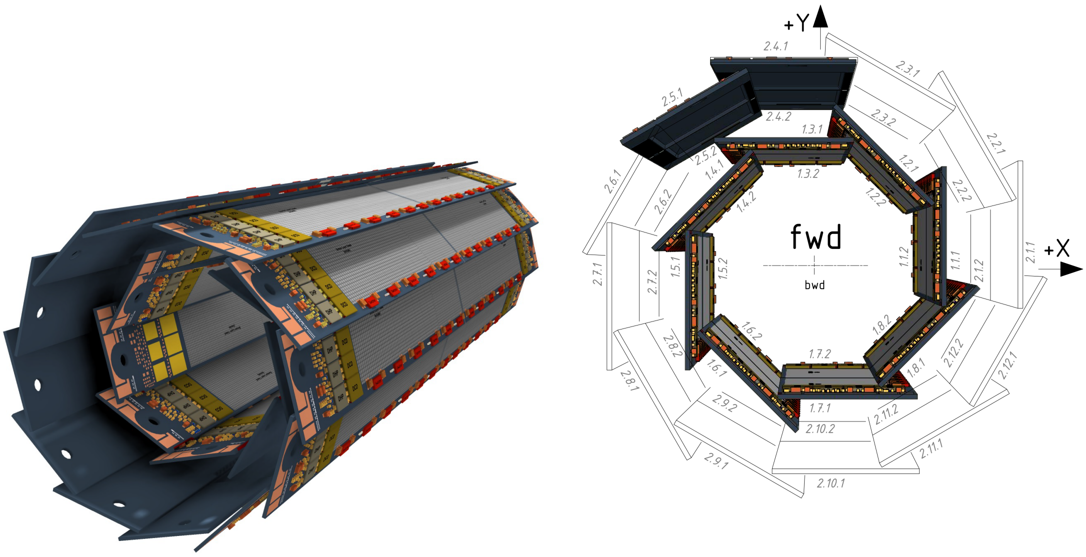


Figure 2.8. – Three-dimensional view of the two PXD layers (left). Transverse scheme of the phase 3 PXD layout with one full layer and two ladders on the second (right). Credits [12]

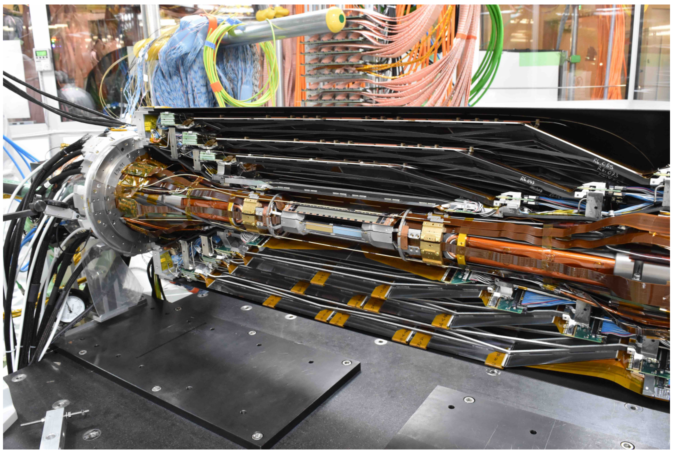


Figure 2.9. – The Belle II SVD already matched with the PXD and ready to be installed at the Interaction Point of the detector. Credits [84].

while the v/N side provides information on the z-coordinate along the beam line. Table 2.3 provides detailed information about SVD sensors.

	Small	Large	Trap.
No. u/P readout strips	768	768	768
No. v/N readout strips	768	512	512
Readout pitch u/P strips (μm)	50	75	50-75
Readout pitch v/N strips (μm)	160	240	240
σ_{dig} v/N strips (μm)	23	35	35
σ_{dig} u/P strips (μm)	7	11	7-11
Sensor thickness (μm)	320	320	300
Active Length (mm)	122.90	122.90	122.76
Active Width (mm)	38.55	57.72	57.59-38.42

Table 2.3. – Geometrical details of the SVD double-sided strips sensors. All sensors have one intermediate floating strip between two readout strips.

The electronic system at the front-end is made up of APV25 chips providing an analogue readout of the collected signal. Each chip has 128 channels and a quick shaping time of 50 ns. It can also tolerate high radiation of up to 100 Mrad of integrated dose. To reconstruct the signal waveform, six consecutive analogue samplings are used. A mixed three/six acquisition mode is also in place to lessen dead time, data size, and occupancy at higher luminosity. For shorter ladders in layer 3, the chips are situated outside the active area, while for longer ladders in layers 4, 5, and 6, the chip-on-sensor, Origami, concept is utilized to reduce the signal propagation distance, capacitance, and noise. With this design, the chips are positioned only on one side of the detector for the middle sensors of the ladders, where a wrapped flex allows for reading out of the sensor side opposite to the chip's position. The chips are thinned to 100 μm to minimize the material budget, and stainless steel pipes for bi-phase CO2 cooling at $-20°C$ are only located on one side.

The size of the pitch p, Table 2.3, already provides a rough indication of the spatial

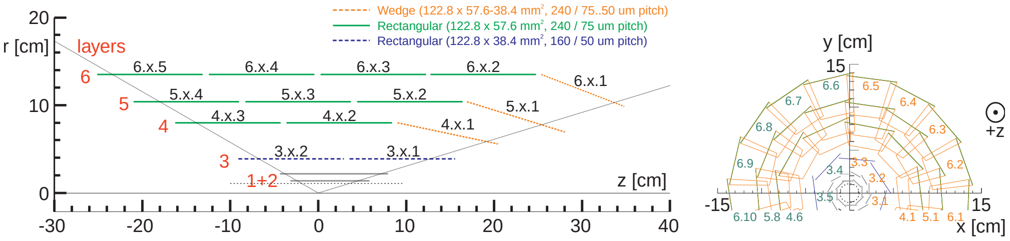


Figure 2.10. – Schematic Layout of the SVD detector. In the left schematic, the yellow sensors are the wedge, green are the large sensors and blue the small ones. Credits [84].

resolution, the digital resolution defined as:

$$\sigma_{\text{dig}} = \frac{p}{2\sqrt{12}} \tag{2.7}$$

where factor 2 at the denominator is needed to take into account the presence of the floating strip.

The detector operations went smoothly and we didn't observe any major hardware issues. All 1748 APVs were working properly for the data we used, and only about 1% of strips were masked, mostly at the edges of the chips. As a result, we anticipate that the detector's resolution won't be impacted by any malfunctions.

2.2.3. The Central Drift Chamber

The Belle II Central Drift Chamber (CDC) [85] detector endorses the same three roles as its predecessor in Belle but aims for better performances. First, it plays a role in charged tracks reconstruction and momentum measurement along with the SVD and PXD. Then, it provides particle identification information by characterizing the energy loss $\mathrm{d}E/\mathrm{d}x$ due to gas ionisation. Finally, it gives efficient and reliable trigger signals.

The CDC cylinder volume, Figure 2.11, extends from a radius of 16 cm to 113 cm to fit with the new VXD and PID detectors, and covers a polar angle from 17° to 150°. It is composed of 14336 tungsten sense wires in an electrical field gradient provided by aluminium wires. All wires are arranged into 56 layers, subdivided into nine super-layers, as shown in Figure 2.12 (left). The super-layers alternate between axial (along the z-direction) and stereo (slanted by an angle of at most 80 mrad), as shown in Figure 2.12 (right), to provide a three-dimensional reconstruction and trigger system. The volume is made of drift cells, from $6 \times 10\,\text{mm}^2$ to $18 \times 18\,\text{mm}^2$, filled by a helium-ethane gas mixture at 50% proportion each, already proven in Belle. The structure is supported by two carbon cylinders and two aluminium end plates where the readout electronics are located.

Figure 2.11. – Picture of the dismantle *Belle* CDC volume exposed at KEK, Japan.

The CDC spatial resolution is about 2 mm in z and 100 μm in r direction while the relative precision on energy loss is 12% for 90° incident track angle.

2.2.4. Particles identification system

The Belle II PID system is divided into two subdetectors, one for the barrel region with the TOP counter and one for the end-cap with the Aerogel Ring-Imaging Cherenkov (ARICH) counter. They both use the Cherenkov effect to assess the charged particle velocity and, thus, the likelihood of the different mass assumptions.

2.2.4.1. Time-Of-Propagation counter

The TOP [87] consists of 16 quartz radiators placed between CDC outer cylinder and the ECL. Two quartz bars are joined together to form a radiator of 2.5 m length, 44 cm with and 2 cm thickness. A mirror is placed on the forward end while at the back end is placed the Micro-Channel Plate PhotoMultiplier Tubes (MCP-PMTs).

A Cherenkov ring is produced by charged particles travelling through the quartz radiator. These photons reflected by the quartz bar either propagate to the front end, where the mirror returns them, or to the back end, where MCP-PMTs collects them, as shown in Figure 2.13. The MCP-PMTs provides the time of propagation and the arriving two-dimensional space information (x, y), combined together to reconstruct the Cherenkov angle θ_C. Consequently, the particle velocity v is inferred as:

$$cos\theta_C = \frac{c}{nv}, \tag{2.8}$$

with c the speed of light in vacuum and the refractive index of the medium n. The likelihood for different mass hypotheses is calculated. The TOP PID's performances

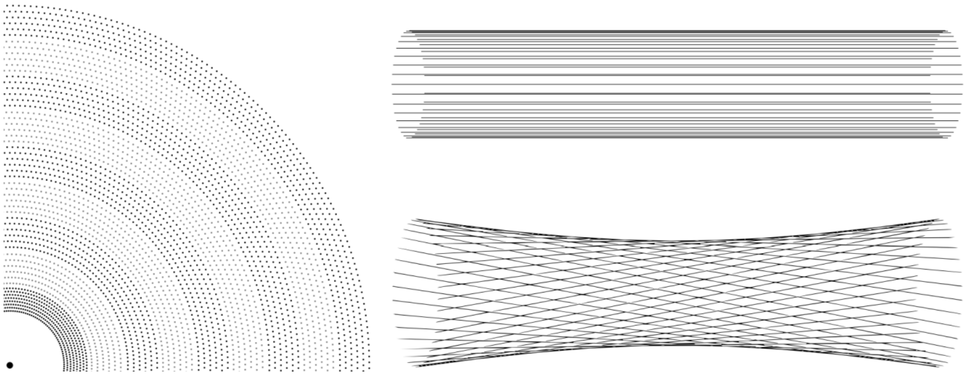


Figure 2.12. – Left: A quadrant of the drift chamber in $r\phi$ projection. The innermost superlayer confines eight layers, and all others contain six. Right: A scheme of stereo wires (bottom) relative to axial wires (top). The skew is exaggerated. Credits [86].

are critically impacted by time resolution broadening due to the photon's chromaticity. This issue is overcome by introducing a focusing system, a concave mirror, which divides the ring images according to the wavelength of Cherenkov photons.

The MCP-PMTs achieve an excellent time resolution for single-photon detection of 50 ps thanks to a transit time spread of ~ 30 ps and a gain of 10^6.

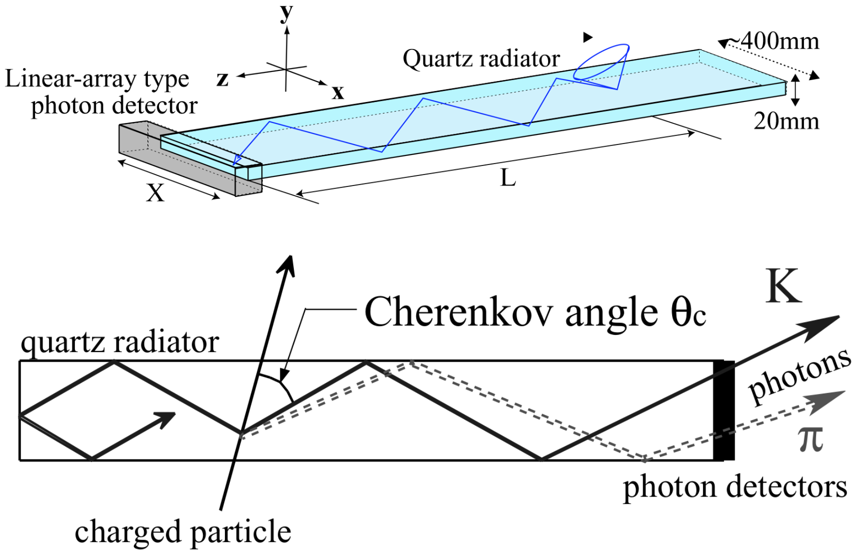


Figure 2.13. – Schematic view of a TOP counter module (top) and representation of the internal reflecting Cherenkov photons emitted by a kaon or a pion (bottom). Credits [68]

2.2.4.2. Aerogel Ring-Imaging Cherenkov Detector

The forward endcap particle identification is handled by the ARICH [89] counter, Figure 2.14, composed of:

- the Aerogel Cherenkov radiator, where Cherenkov photons are produced when a charged particle passes through it,
- an expansion volume, which is a 20 cm gap between the radiator and the photon detector, allowing emitted Cherenkov photons to form well-defined Cherenkov rings,
- an array of position sensitive photons sensor, based on Hybrid Avalanche Photo-Detector (HADP) that detects single photons with high efficiency and good 2D resolution in a high magnetic field,
- the Readout system for the photons detectors.

The key indicator of the ARICH counter performances is the Cherenkov angle resolution per charged tracks $\sigma_{trk} = \sigma_\theta / \sqrt{N_\gamma}$, with the number of detected photons N_γ and the single photon angle resolution σ_θ. The radiator thickness impacts the number

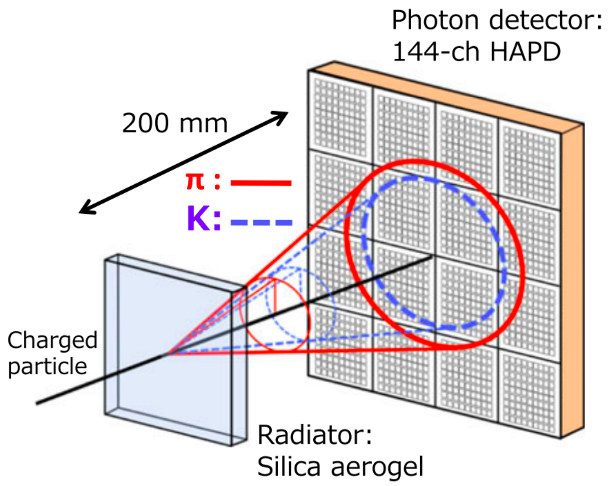
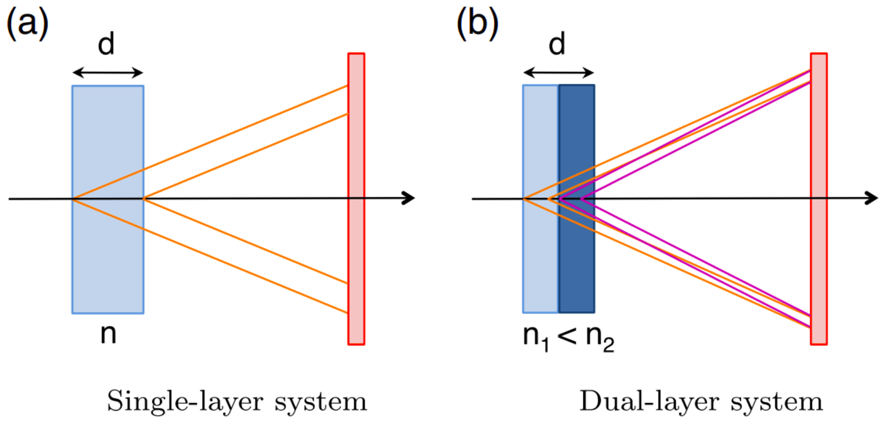


Figure 2.14. – Particle identification principle between π and K using the ARICH counter. The cones shown in solid and dotted lines represent the Cherenkov light emitted by a pion and a kaon, respectively. Credits [88]



Single-layer system

Dual-layer system

Figure 2.15. – The dual-layer focusing scheme: (a) image of a normal Cherenkov counter with a single layer; (b) focusing with dual layer, in which different refractive indices of n_1 and n_2 such as ($n_1 < n_2$) are used. Credits [88]

of Cherenkov photons emitted by charged particles, thus the resolution σ_{trk} by N_γ. However, the emission point uncertainty degrades the single photon resolution σ_θ. The radiator design maximizes both parameters thanks to a focusing arrangement with two layers with refractive index $n_1 < n_2$ for a total thickness of 4 cm. This reduces photon spread in the detector plane and makes the Cherenkov rings overlap, increasing N_γ and reducing the emission point uncertainty, as shown in Figure 2.15.

Thanks to this design, the ARICH counter achieves a resolution per track $\sigma_{trk} = \sigma_\theta / \sqrt{N_\gamma} = 4.5\,\text{mrad}$ assuming 12.7 detected photons in average and a single photon resolution of 14.3 mrad.

2.2.5. The Electromagnetic Calorimeter

In B physics, electromagnetic calorimeters play a critical role by providing an excellent energy resolution since B mesons produce a large amount of π^0 and neutral particles. The Electromagnetic CaLorimeter (ECL) [90] tasks are to detect photons from a wide energy window, from 20 MeV to 4 GeV and to measure their energy and angular coordinates while contributing to particle identification of electrons and K_L^0, together with the KLM. In addition, the ECL takes part in several trigger conditions and the measurement of the instantaneous luminosity.

The choice for the Belle II calorimeter was to reuse the previous Belle device by upgrading the readout electronic system to cope with the high background rates. The ECL is made of a 3 m long and 1.25 m inner radius barrel to cover a polar angle acceptance from 32.20° to 128.70°, forward and backward endcaps for a full range from 12.10° to 155.03° with a gap $\sim 1°$ between barrel and endcaps. The structure contains 8736 CsI(T) crystals shaped in a truncated pyramid with an average size of $6 \times 6 \times 30\,\text{cm}^3$. On the rear of each crystal are glued two $10 \times 20\,\text{mm}^2$ photo-diodes to catch the scintillation light in two independent signals through their preamplifier summed at the later stage. This readout setup allows us to deal with the significant pile-up noise and a high background environment.

The intrinsic energy resolution was measured from tests on prototypes and approximated by:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34\%)^2}, \tag{2.9}$$

where the energy E is measured in GeV.

2.2.6. The K_L^0 and Muon Detector

The outermost detector is called KLM [68]. It's dedicated to provide the PID for high penetration power particles like the muons and K_L^0 mesons. Muons can cross the KLM straight on depending on their momenta and polar angles, while K_L^0 interacts by producing hadronic showers in the ECL and/or KLM.

It consists of an alternating sandwich of 14 (15 in the barrel) 4.7 cm thick iron plate layers and 14 active detector element layers located outside the superconducting

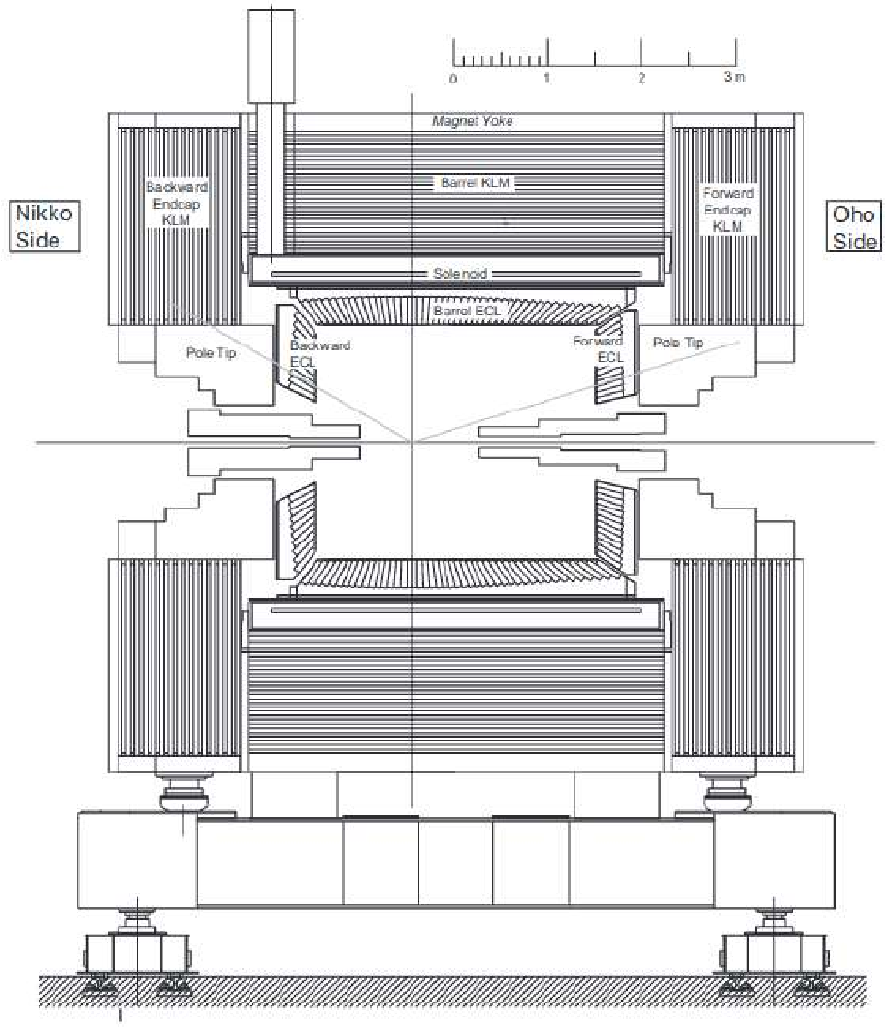


Figure 2.16. – Schematic overview of the outermost Belle II layers, particularly the K_L and Muon detector (KLM) with its barrel, frontcap and endcap components. Credits [68]

solenoid, which provides the 1.5T magnetic fields used to measure the particle's momenta. The iron plates are used as magnetic flux return for the solenoid and to give a 3.9 interaction length in addition to the 0.8 from ECL where the K_L^0 mesons can shower hadronically. The octogonal KLM barrel is divided into 16 sectors to cover a polar angle region from 45° to 125°. This region is enlarged from 20° to 155° by adding the two forward and backward endcaps. The outer layers of the barrel are still equipped with the glass-electrode resistive plate chambers used in Belle. In contrast, the scintillator readout in endcaps and inner layers is provided by silicon photo-multipliers. Resistive plates have a long dead time after a discharge which is unsuitable for the expected high background, the use of silicon photo-multiplier, with its 0.7 ns time resolution, will be generalised to the whole KLM.

2.3. Track reconstruction

The process of charged particle (track) reconstruction in Belle II [91, 86], is called "tracking". Using reconstruction software, the spatial measurements (hits) of a traversing particle are combined into a helical trajectory (tracks) to extract the kinematic properties of the particle. This tracking is divided into the task of partitioning the hits into tracks (track finding) and calculating the track parameters (track fitting).

Due to the very different properties of the three tracking detectors (PXD, SVD and CDC), different algorithms are used for each of them. First, the hits in the CDC are filtered and reconstructed using two separate algorithms: a global track-finding algorithm based on the Legendre algorithm and a cellular automaton algorithm. The outputs of these two algorithms are combined to form CDC tracks. Additionally, a combinatorial Kalman filter is used to enhance the CDC tracks with SVD clusters. The remaining SVD hits are used in a standalone SVD track finder. Finally, the resulting CDC and SVD tracks are extrapolated to the PXD.

The parameters of a track are determined thanks to the track fitting. This includes the position of the point of closest approach to the interaction point, represented in Figure 2.17, defined as:

- d_0, the signed distance of the point of closest approach with respect to the interaction point in the Belle II transverse plane (x,y);
- z_0, z-coordinate of the point-of-closest-approach (beam direction).

The distance (dr, dz) of the point of closest approach between charged particles (tracks) with respect to the interaction point is derived from those two parameters. For particles that are obtained by combining multiple final state particles, like a τ^-, the (dr, dz) variables are taken as the distance between the reconstructed vertex and the interaction point.

Unfortunately, finding tracks is not always flawless, and in multiple cases can lead to a failed reconstruction of tracks. The track finding efficiency quantifies the number of charged particles that are not successfully reconstructed as tracks due to the inefficiency of algorithms or detectors. There are instances where a reconstructed track may be, in fact, a fake track. This occurs when it includes hits from the beam-induced background or combines hits from two different particles. Another misreconstruction possibility is a clone track, where multiple tracks are reconstructed from hits left by the same particle.

In the Belle II experiment, the ability to accurately track particles produced after collisions was studied [92]. Specifically, the focus was on tracking efficiency, which was measured by analyzing $e^+e^- \to \tau^-[\to \pi^-\pi^+\pi^-\nu_\tau]\tau^+[\ell^+\overline{\nu}_\ell\nu_\tau]$ events where ℓ can be either an electron or a muon. By comparing the number of instances where a charged pion was missing or properly reconstructed, it was found that tracking efficiency ranged from 75% at low transverse momenta to 95% around 4 GeV/c. However, the efficiency decreased as the track got closer to the beam axis (at small or large polar angles), but remained constant at around 90% regardless of the azimuthal angle. The calibrated discrepancy between the efficiencies measured in data (ε_{data}) and Monte Carlo simulation (ε_{MC}) is denoted by δ^*, which was found to be below 0.5% overall.

2.4. The trigger system

A significant challenge raised by increasing the instantaneous luminosity is facing the large event rate, considering the beam-induced and luminosity-dependent backgrounds. For Belle II, the estimated event rate for the principal physics processes is 20 kHz for the planned luminosity of $8 \times 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$. The beam-induced and luminosity-dependent background sources are mainly low multiplicity processes: for example, the Touschek effect, the radiative Bhabha ($e^+e^- \to e^+e^-\gamma$), and the two-photon ($e^+e^- \to e^+e^-e^+e^-$) processes can be trivially discriminated from $B\overline{B}$, with high charged track multiplicity, by a trigger based on the number of tracks. Nonetheless, it is tricky to distinguish interesting low multiplicity processes as τ-pair decays or dark photons from low multiplicity backgrounds. The Belle II trigger system architecture is based on two levels, a level-1 hardware trigger [93] followed by the software High Level Trigger (HLT) [94] illustrated in Figure 2.18:

The L1 trigger or low-level is implemented for a 30 kHz maximum output rate and

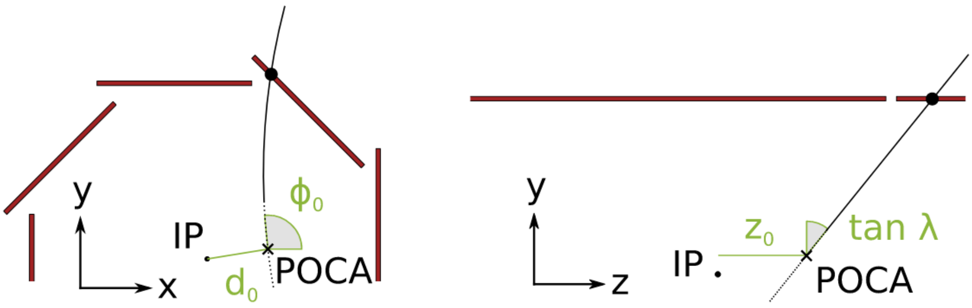


Figure 2.17. – Scheme of the helix trajectory (tracks) in the Belle II detector transverse plane (left) and in the beam direction (right) with the track's parameters: the coordinates of the point of closest approach (POCA) d_0, z_0 and the angles ϕ_0, λ. The first PXD layer is displayed for visual guidance.

a $5\,\mu$s latency. It gathers information from sub-trigger systems, mainly from the CDC and ECL. These are merged through a global reconstruction logic and a global decision logic that gives the trigger output signal. The CDC trigger system is based on 3D tracking, which allows retrieving the z position of the tracks. The z coordinates will enable us to discriminate the background with a displacement along z, such as Touschek intra-bunch scattering. The ECL complete the trigger scheme by adding two more pieces of information: the total energy trigger sensitive to events with high electromagnetic energy deposition and the isolated cluster counter sensitive to multi-hadronic events with low energy cluster or minimum ionizing particles. In addition, the ECL trigger can distinguish the back-to-back topology from the Bhabha scattering.

The second trigger stage is based on software called HLT and aims to reduce the output rate to 10 kHz. Events passing the L1 trigger are fully reconstructed offline using all subsectors information except PXD, where a region of interest is defined. The HLT filter was turned on in 2021.

2.5. Particle Identification

The sub-detectors (SVD, CDC, TOP, ARICH, ECL, and KLM) collect information and analyze it independently to determine the likelihood of each charged particle hypothesis $\mathscr{L}_i^s$. These likelihoods can be combined to create a global likelihood ratio,

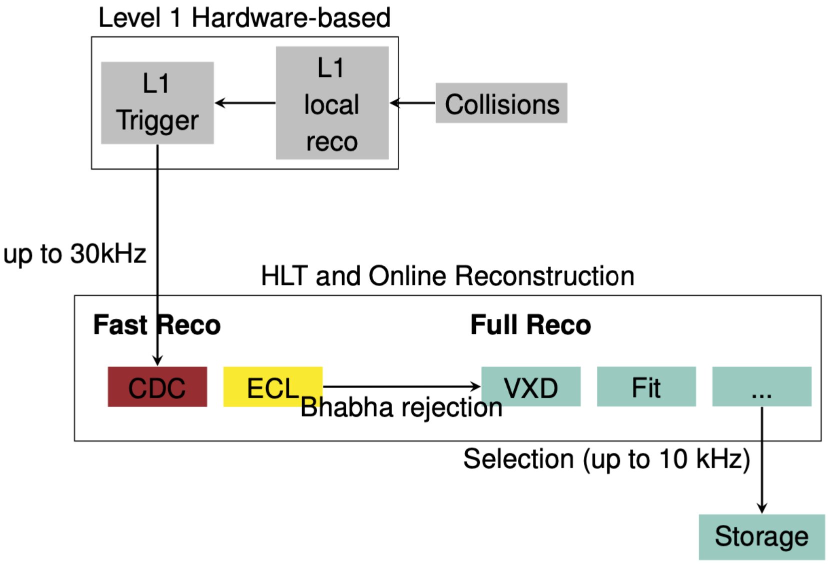


Figure 2.18. – Workflow of the two levels of the trigger system. The first stage is hardware-based and called L1, while the second relies on a fast software reconstruction and is called HLT. Credit [68]

referred to as PID, using the equation:

$$\text{PID} = \frac{\mathscr{L}_i}{\mathscr{L}_e + \mathscr{L}_\mu + \mathscr{L}_\pi + \mathscr{L}_K + \mathscr{L}_p + \mathscr{L}_d} \tag{2.10}$$

Here, $\mathscr{L}_i$ represents the likelihood, calculated as:

$$\mathscr{L}_i = \prod_{s \in S} \mathscr{L}_i^s \tag{2.11}$$

where $S = \{SVD, CDC, TOP, ARICH, ECL, KLM\}$. The variable i can be any of the following particles: electron (e), muon (μ), pion (π), kaon (K), proton (p), or deuteron (d). This method is referred to as "global" as it takes into account all particle types, unlike "binary" calculations that only consider one particle type as a reference. The denominator in the equation represents the sum of likelihoods for i and the reference particle type.

For example, muon identification mainly relies on CDC and KLM detectors. The tracks are extrapolated with the pion mass hypothesis reconstructed by the CDC and check if the predicted KLM outer layer crossed fits with the measured hits. Then the track extrapolation is reran with the muon hypothesis mass from the inner layer. Finally, the muon likelihood is computed by comparing predicted and measured ranges and the goodness of track fit. The muon detection efficiency reaches 90% for momenta larger than 1 GeV/c, while for the K_L^0, the efficiency is about 80% at 3 GeV/c.

Another method of identifying leptons is also available. This method involves combining several ECL measurements, such as shower shapes and pulse shape discrimination variables, with other sub-detector likelihoods in a boosted decision tree. The boosted decision tree is trained in both multi-class modes to separate leptons from other charged particle hypotheses and in binary ℓ versus π mode.

Based on the study of $J/\psi \to \ell^+\ell^-$ events [95] where $\ell = e, \mu$, the electron and muon identification efficiencies were found to be 86% and 88.5%, respectively, at a PID probability threshold of 0.9. Pion misidentification rates were also measured using the $K_S^0 \to \pi^-\pi^+$ decay, with rates of 0.4% for electron candidates and 7.3% for muon candidates.

2.6. Overview of the Belle II Analysis Software

The *basf2* [97, 96] processes the data by executing small processing blocks (*modules*) in a linear specific execution order, depending on the tasks that it intends to do, inside an instance called *path*, as shown in Figure 2.19. At each step of the *path*, the *modules* can read data stored in the *DataStore*. In addition, sets of conditions can be loaded into the modules by a central database, *DBStore*. These conditions consist of a *payload*, a ROOT file [98] attached to the relevant data-taking periods called "interval of validity". *Payloads* are forming collections called *globaltags* with a unique name and valid only for a specific data-taking period range.

In the physics analysis case, the chain of *modules* is defined and executed into a Python script file called *steering file*. In this file, the reconstructed tracks and neutral particles are loaded and combined from final to initial states to recover the studied decay. Sets of requirements could be defined either on the level of the tracks or on the recombined mother particles to select accurately the wanted particles and/or remove backgrounds. Once the path is defined, the steering file loops over all the input events and the outputs requested variables are stored into a ROOT file called "ntuples".

2.7. Dataset production and nomenclature

2.7.1. Skimming

Skims [96] refer to a selection applied to data and Monte-Carlo (MC) simulated datasets tailored for specific analyses. Their purpose is to reduce the size of the original samples by applying specific criteria, allowing analysts to focus on interesting events. Skimmed samples are usually around 90% smaller than the original data and MC samples. At Belle II, where 50 ab^{-1} of data will be collected, skimming is essential to manage CPU load and the number of input files during the production of ntuples.

Skims are run on the mDST format of the produced MC samples and processed data at Belle II to reconstruct a specific list of particles passing different requirements. The reconstructed particle information is then written in a new file format, microDST. The

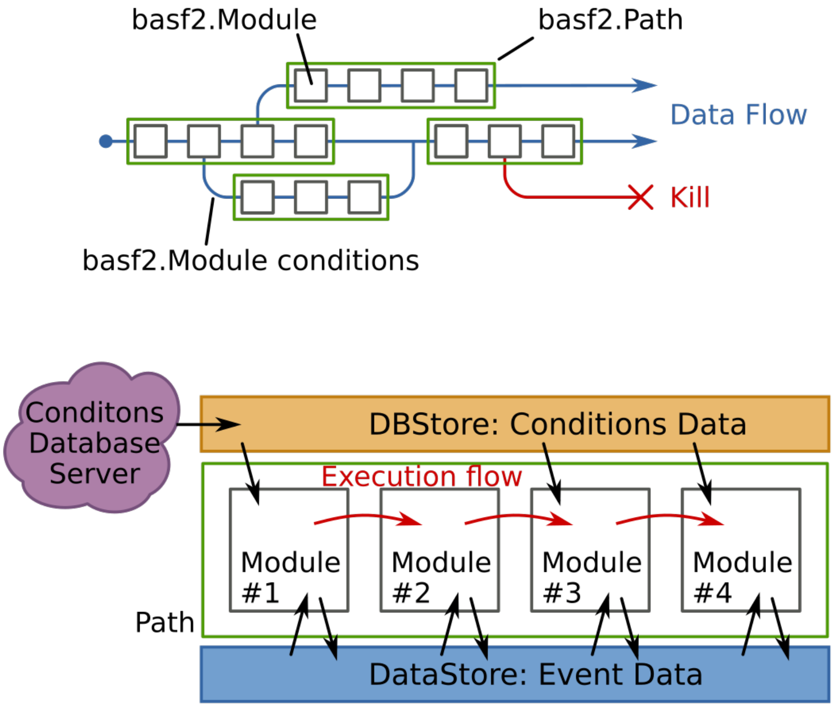


Figure 2.19. – Worflow for the event processing chain under *Belle II* Analysis Software Framework (*basf2*). Credits [96]

skim filter removes events that do not meet the skim's criteria.

For example, the Lepton Flavour Violation (LFV) searches in the decays of the τ lepton have their own dedicated skim criteria. If the final state of the signal τ^- decays contains only one charged particle, the events must satisfy:

- Number of good tracks < 5,
- $1. < M_\tau < 2.\,\text{GeV}/c^2$,
- $-1.5 < \Delta E_\tau < 0.5\,\text{GeV}$.

If the final state of the signal τ^- decays contains more than one charged particle:

- Number of good tracks < 7,
- $1.4 < M_\tau < 2.\,\text{GeV}/c^2$,
- $-1 < \Delta E_\tau < 0.5\,\text{GeV}$.

Where ΔE_τ is the difference of energy between the τ^- and the beam.

The τ LFV skim is used for all data and MC simulation samples, except for simulated $e^+e^- \to \tau^+\tau^-$ events, which have an excessively high retention rate.

2.7.2. Experimental data

Experimental datasets, also called data, are characterised according to different criteria. The first one is the data-taking period, made up of two numbers: the experiment number refers to the long period (a few months for a total of three per year) where the experimental condition is relatively stable, and the second is the run number which denotes the short period where the detector is taking data between a start and a stop caused by the 8 hours safety stop or an issue in the data taking.

The data analysed in Chapter 4 have been collected from March 2019 to Autumn 2022 and is also called LS1 dataset. More details are given in Table 2.4. The analysed data gathers the samples collected at different energies of collision, $\Upsilon(4S)$, $\Upsilon(4S)$ off-resonance and $\Upsilon(5S)$, for a total luminosity of 424 fb^{-1}.

Table 2.4. – Experimental datasets analysed in the following studies and their statistical uncertainty on integrated luminosities with their statistical and systematic uncertainties [80].

Experiment	Beam Energy	Data Taking Period	Offline luminosity fb^{-}1
Total	$\Upsilon(4S)$		$361.654 \pm 0.021 \pm 2.170$
	$\Upsilon(4S)_{offres}$		$42.279 \pm 0.007 \pm 0.254$
	$\Upsilon(5S)scan$		$19.661 \pm 0.004 \pm 0.118$

2.7.3. Monte-Carlo simulated data

Physics analysis requires some data samples produced by simulation in addition to experimental data. They allow us to get a lot of unknown information in data, like the true identity of the particles used in the reconstructed decay chain. To get these simulated samples, events are generated following a given physics process thanks to a dedicated generator based on Monte-Carlo simulation packages, listed in Table 2.5. To mimic the data-taking conditions simulations, simulated beam backgrounds and static detector conditions are stored in the *basf2* release used to run the simulations.

Table 2.5. – Generators for each simulated physics process [100].

Physics process	Generator
B/D meson decays (exclusive final state)	*EvtGen* [101]
B/D meson decays (inclusive final state) Light $q\overline{q}$ continuum	*PYTHIA* [102]
$e^+e^- \to \tau^+\tau^-$	*KKMC* [103]
τ decays	*TAUOLA* [104]
$e^+e^- \to e^+e^-(\gamma)/\gamma\gamma(\gamma)$	*BABAYAGA.NLO* [105, 106]
$e^+e^- \to e^+e^-\ell^+\ell^-$ $(\ell = e, \mu)$	*AAFH* [107]

On top of the generated events is applied the interaction with the detector simulated

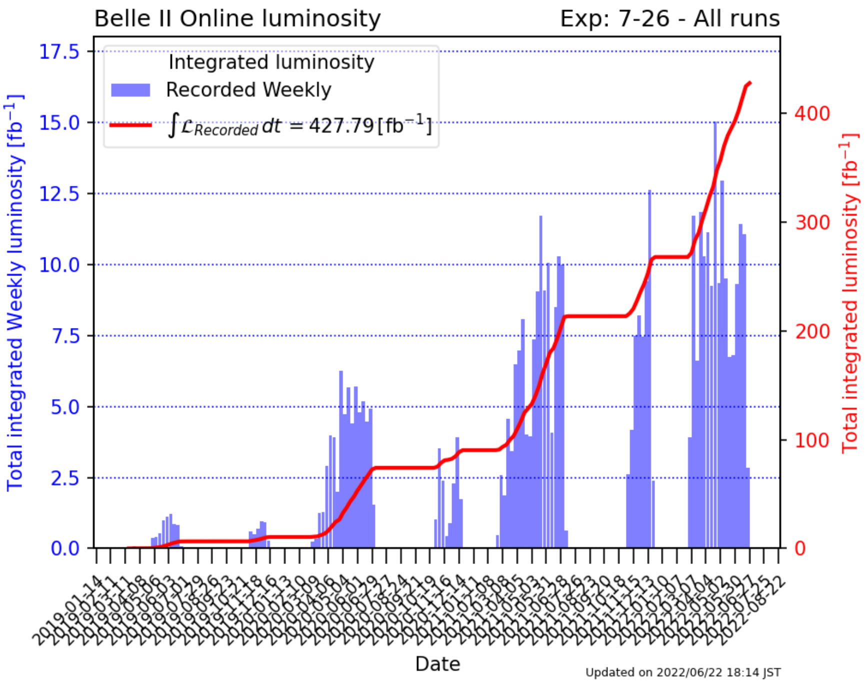


Figure 2.20. – Recorded integrated luminosity per week and total through the phase 3 data taking from 2019 to 2022. Credits [99]

thanks to *Geant4* [108] while *Opera3D/TOSCA* [109] software gives a magnetic field. In the end, the trigger is simulated by Trigger SIMulation (TSIM) [96].

In the search for $\tau^- \to \mu^-\mu^+\mu^-$ decays presented in Chapter 4, the used simulated MC dataset was produced during the 15th official Monte Carlo (MC15) production campaign. The simulated samples used are composed of the following:

- signal samples with one over the ten million of $e^+e^- \to [\tau^- \to \mu^-\mu^+\mu^-]\tau^+$ signal events generated. The simulation is done with `Tauola` by imposing a Lorentz-invariant phase space without any new physics processes,
- background samples: 8 ab^{-1} of $\tau^-\tau^+$ SM and continuum $q\overline{q}$ processes with τ^- LFV skim applied at the generation level and the low-multiplicity background samples.

More details on background simulated MC samples and their luminosity are given in Table 2.6.

Table 2.6. – List of the background simulation samples with corresponding integrated luminosities.

Background types	Background decay modes	Sample name	Integrated luminosity
τ^--pairs background	$\tau^+\tau^-$	`taupair`	8000 fb^{-1}
Continuum $q\overline{q}$ background	$u\bar{u}$	`uubar`	8000 fb^{-1}
	$d\bar{d}$	`ddbar`	8000 fb^{-1}
	$s\bar{s}$	`ssbar`	8000 fb^{-1}
	$c\bar{c}$	`ccbar`	8000 fb^{-1}
B-pairs background	$B^0\overline{B}^0$	`mixed`	1000 fb^{-1}
	B^+B^-	`charged`	1000 fb^{-1}
Low multiplicity background	$\gamma\gamma$	`gg`	500 fb^{-1}
	$e^+e^-\gamma$	`ee`	100 fb^{-1}
	$\mu^+\mu^-$	`mumu`	1000 fb^{-1}
	$e^+e^-e^+e^-$	`eeee`	200 fb^{-1}
	$e^+e^-\mu^+\mu^-$	`eemumu`	200 fb^{-1}
	$e^+e^-\tau^+\tau^-$	`eetautau`	2000 fb^{-1}
	$e^+e^-\pi^+\pi^-$	`eepipi`	1000 fb^{-1}
	$e^+e^-K^+K^-$	`eeKK`	1000 fb^{-1}
	$e^+e^-p^+p^-$	`eepp`	1000 fb^{-1}
	$\mu^+\mu^-\mu^+\mu^-$	`mumumumu`	2000 fb^{-1}
	$\mu^+\mu^-\tau^+\tau^-$	`mumutautau`	2000 fb^{-1}
	$\tau^+\tau^-\tau^+\tau^-$	`tautautautau`	10000 fb^{-1}
Low multiplicity initial state radiation background	$\pi^-\pi^+$	`pipiISR`	2000 fb^{-1}
	$\pi^-\pi^+\pi^0$	`pipipi0ISR`	2000 fb^{-1}
	K^-K^+	`KKISR`	2000 fb^{-1}
	$K^0\overline{K}^0$	`K0K0barISR`	2000 fb^{-1}

3. Measurement of the *Belle II* SVD cluster position resolution

Table of contents

In Section 2.2.2, we observed that the Silicon Vertex Detector (SVD) plays a vital role in reconstructing the decay vertex and low-momentum particles. It offers stand-alone tracking capabilities and contributes to charged particle identification through ionisation energy-loss information. Additionally, it helps extrapolate the tracks towards the PiXel Detector (PXD) and define the area of interest to minimise the PXD data size. A precise cluster position resolution is essential for SVD reconstruction, as it provides crucial input for improving the quality of reconstructed tracks and vertices and correctly propagating the uncertainty on the track extrapolation position. In this chapter, we will explore a method that uses overlapping sensors to measure the resolution of cluster positions.

3.1. Definition of Cluster Position Resolution

3.1.1. The cluster and its position

A charged particle passing through the silicon bulk knocks electrons from the atoms, creating electron-hole pairs. Electrons and holes drift to the opposite sensor sides and are collected by the sensing strips. The arrangement of sensing strips in the sensor is shown in Figure 3.1. To identify the stips activated by a single charged particle, a clustering algorithm gathers groups of adjacent activated strips into an object called cluster [12, 68, 84]. To build a cluster, strips must pass the online zero suppression: a minimum signal of three times the strip noise (N_i) and consecutive strips have to get at least one strip with a minimum signal of five times N_i. The available information for each strip is the six sampled amplitudes, the noise from local calibrations, and the value of the gain from local calibrations that allows reconstruction of the electrical charge.

The cluster position (m) is computed using the charge (S_i) and position of the centre of the readout implant (X_i) of its strips, with the Center-of-Gravity (CoG) algorithm:

$$m = \frac{\sum_{i=1}^{N} X_i S_i}{\sum_{i=1}^{N} S_i}, \tag{3.1}$$

where N is the number of strips in the cluster, also called cluster size later. The associated cluster position error is parameterized as a function of $x = \sqrt{N \cdot \mathrm{SNR}}$, SNR is the signal over noise ratio, as:

$$\sigma_m = \mathrm{pitch} \cdot \sqrt{\left(\frac{a_1}{a_2 + x}\right)^2 + (bx)^2 + c} \tag{3.2}$$

were the parameters a_1, a_2, b and c have been extracted from the Monte-Carlo (MC) for the different sensor types and sides to match the MC resolution. Scale factors for the errors are applied to data to correct the resolution disagreement between data and MC.

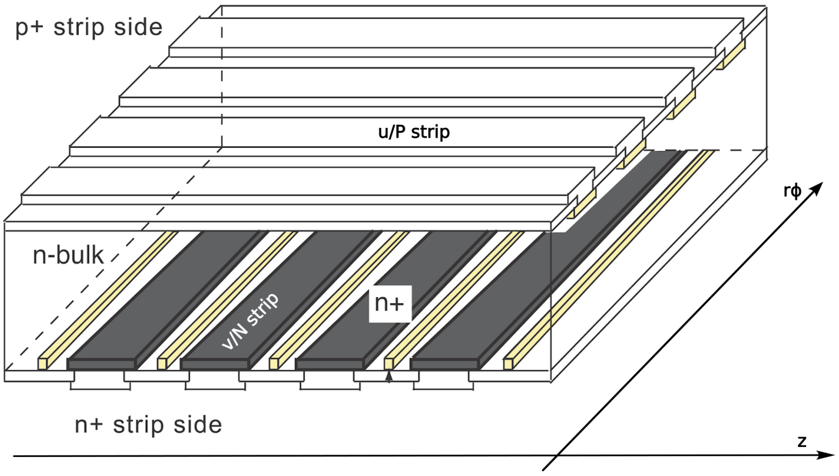


Figure 3.1. – Scheme of the SVD sensors. The sensing strips are implanted in a Silicon bulk on the two sensor sides. Following the two sides, the strips are arranged orthogonally. So respectively, u/P (v/N) sides strips measure the position along the $r\phi$ (z) direction.

Each cluster is assigned its position m, the position of the unbiased extrapolation of the track on the sensor t and its error σ_t, and the true position, which is only known in simulation samples, as shown in Figure 3.2. The track extrapolated position is unbiased because the track is extrapolated on the sensor plane, not considering the cluster on that sensor used for the resolution measurement. The distances, called residuals, between these three positions are computed such as:

$$\varepsilon_m = m - x, \tag{3.3}$$

$$\varepsilon_t = t - x, \tag{3.4}$$

$$R = m - t, \tag{3.5}$$

where ε_m is the cluster position true residual, ε_t the track position true residual and R the measured residual. Example distributions of the measured residuals, true cluster and track residual, and track extrapolation error are shown in Figure 3.3 for layer 4 u/P clusters.

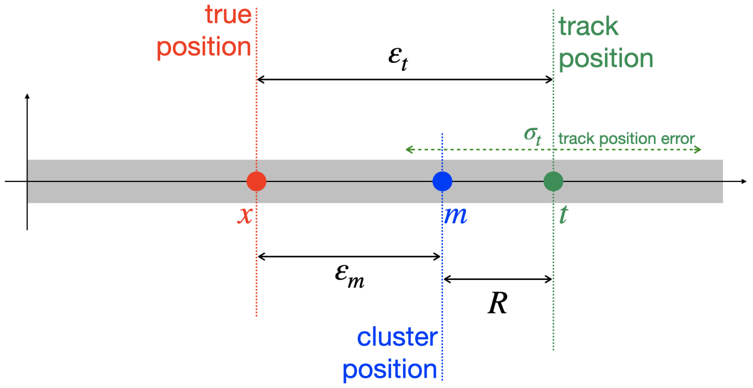
3.1.2. The cluster position resolution

By definition, the cluster position resolution σ_{cl} is the deviation of the cluster position m with respect to the true position t [110]. The σ_{cl} can be computed as the square root of the variance of the true cluster residual $\varepsilon_m = m - x$, such as:

$$\sigma_{cl}^2 = E[(m-x)^2] - E[(m-x)]^2, \tag{3.6}$$

where $E[y]$ is expectation value of y. Assuming that the reconstructed cluster position is unbiased we have $E[(m-x)] = 0$, as shown in Figure 3.3. Hence:

$$\sigma_{cl}^2 = E[(m-x)^2] = E[\varepsilon_m^2]. \tag{3.7}$$



[th]

Figure 3.2. – Simplified representation of the true position x, reconstructed cluster position m, extrapolated track position t, true cluster residual ε_m, true track residual ε_t, and measured residual R. Credits [110]

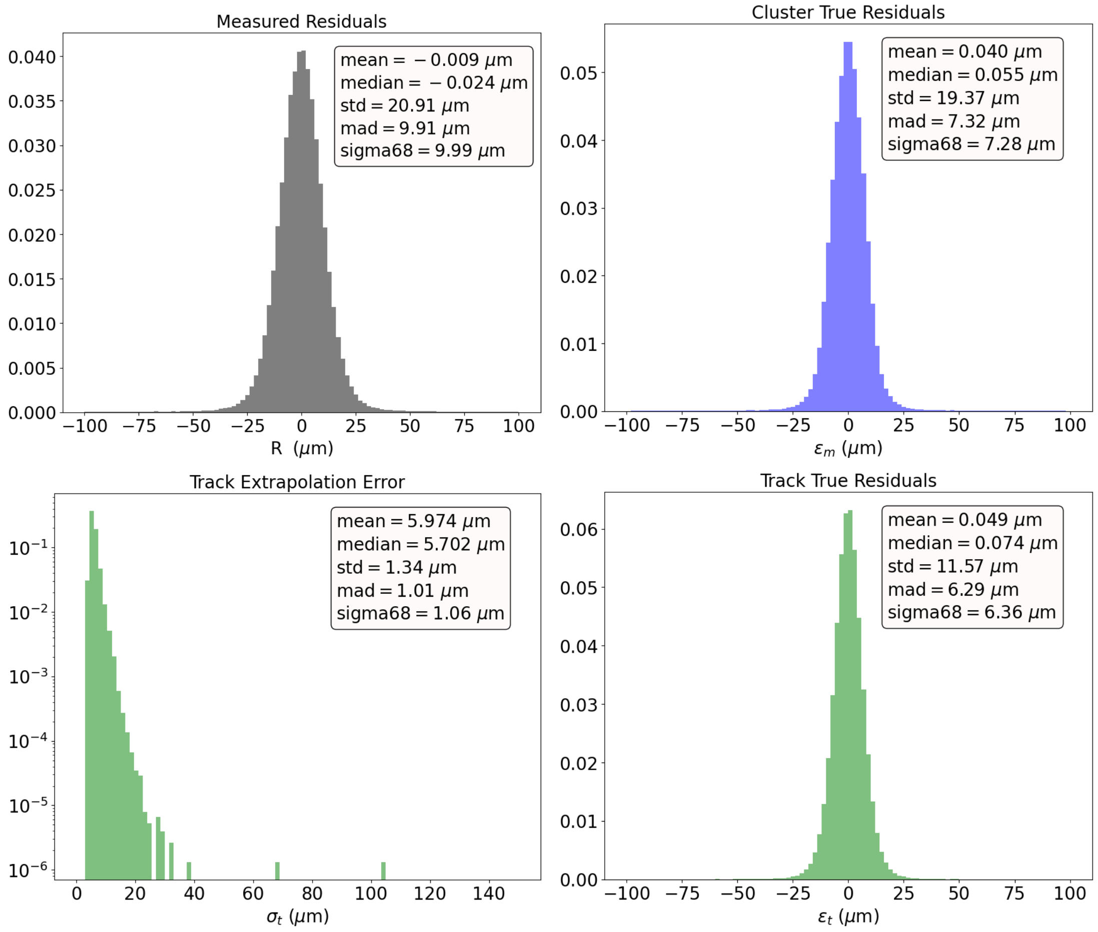


Figure 3.3. – Distributions of the measured residuals R (top left), true cluster residual ε_m (top right), true track residual ε_t (bottom right) and track extrapolation error σ_t (bottom left), for the Layer 4 u/P side clusters from simulated di-muon events. Credits [110]

According to this definition, the resolution σ_{cl} cannot be measured in data samples where the true position x is unknown. But a true resolution σ_{true} can be derived in simulation to be used as a standard to evaluate the correctness of a σ_{cl} estimator.

So the cluster position resolution σ_{cl} has to be estimated using the available quantities: m, t, σ_t and $R = m - t$. Several dependencies must be taken into account *e.g.* the track extrapolation error depends on the lever arm of the extrapolation, on the track incident angle, and the error on the nearest measurement. Nonetheless, the method detailed below is limited by the statistics. The only dependencies taken into account are the SVD layer and the strip side u/P or v/N.

The resolution can be computed as the variance of the measured residuals $R = m - t$, given by:

$$Var(R) = E[(m-t)^2] \tag{3.8}$$

thanks to the unbiased measured residuals $E[(m-t)] = 0$, which is usually true thanks to the alignment calibration. The term $E[(m-t)^2]$ can be expanded by summing and subtracting x:

$$\begin{aligned}
Var(R) &= E[(m-x+x-t)^2 && (3.9)\\
&= E[(m-x)^2+(x-t)^2+2(m-x)(x-t)] && (3.10)\\
&= E[(m-x)^2]+E[(x-t)^2]+2E[(m-x)(x-t)] && (3.11)\\
&= \sigma_{\text{cl}}^2+E[\varepsilon_t^2]+2E[(m-x)(x-t)]. && (3.12)
\end{aligned}$$

It is assumed that there is no correlation between the true cluster residuals and track residuals, meaning that $E[(m-x)(x-t)] = 0$. Moreover, the track extrapolation is computed without considering the cluster position. The validity of such approximation is checked in Figure 3.4, where no hints for any correlation between ε_t and ε_m is shown and the median of $\varepsilon_m \cdot \varepsilon_t$ is compatible with null measurement, being smaller than $0.1\,\mu\text{m}^2$. Under this assumption, Equation 3.9 can be rewritten as:

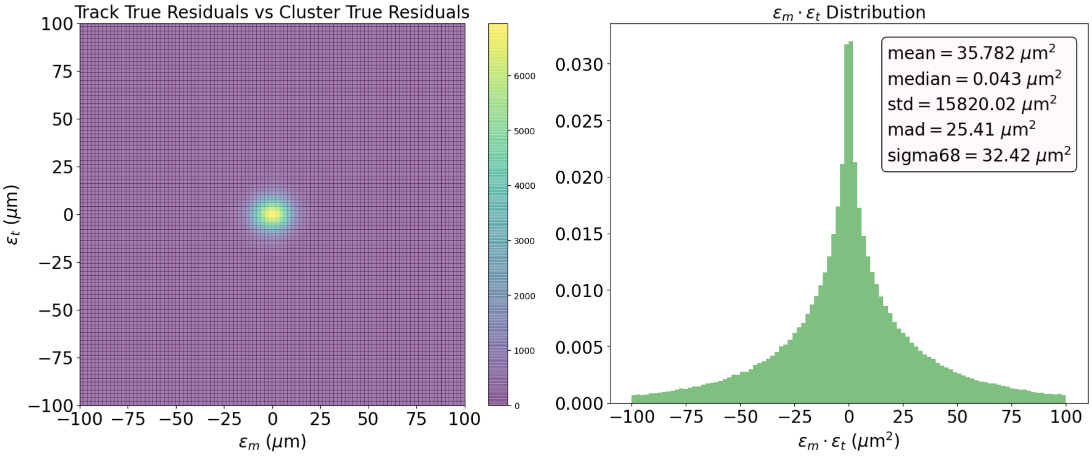


Figure 3.4. – Layer 4 u/P side true cluster residual ε_m vs true track residuals ε_t (left) and $\varepsilon_m \cdot \varepsilon_t$ distribution (right). Credits [110]

$$\sigma_{cl}^2 = Var(R) - E[\varepsilon_t^2]. \tag{3.13}$$

The track extrapolation error can be used as an estimate of the standard deviation of the track's true residuals, and we can write $E[\varepsilon_t^2] = E[\sigma_t^2] = E[\sigma_t]^2 + Var(\sigma_t)$. The assumption is detailed in Appendix A.1. Consequently, the Equation 3.13 becomes:

$$\sigma_{cl}^2 = Var(R) - E[\sigma_t]^2 - Var(\sigma_t). \tag{3.14}$$

3.1.2.1. The true cluster resolution

For optimization purposes, we also consider the true cluster position resolution from simulation, which is defined as:

$$\sigma_{true}^2 \equiv E[\varepsilon_m^2] = E[\varepsilon_m]^2 + Var(\varepsilon_m) = Var(\varepsilon_m). \tag{3.15}$$

Its distribution, as shown by the plots in Figure 3.3 and 31, has two components and large tails that point to a non-Gaussian behaviour.

A dedicated study for the optimal estimator is described in Appendix A.2, and the true resolution can finally be extracted from a fit to the true residuals on the simulation by using the 68% coverage, calculated by taking half the difference between the 84th and 16th quantiles, which results in:

$$\sigma_{true}^2 = E[\varepsilon_m^2] \simeq \texttt{sigma-68}^2(\varepsilon_m) = 7{,}3\,\mu\text{m}, \tag{3.16}$$

The calculation was performed using the entire range of ε_m. It's worth noting that this value is actually better than the estimated value based on the pitch (11 µm) listed in Table 2.3, which was expected.

3.2. Datasets

In this section, we briefly describe the cluster reconstruction algorithms, the data and the simulation MC samples used to perform the measurements, including the selection criteria.

We have used the basf2 release-06 for the reconstruction and simulation and the prompt calibration of the data.

3.2.1. Data Samples and Calibrations

The resolutions have been measured on events of e^+e^- annihilation into a pair of muon (di-muon events), collected during the summer and autumn runs of 2020.

Only clusters associated with selected tracks are then considered. Tracks are required to satisfy the same criteria in data and simulation[1], reported in Table 3.1.

---

[1] though they are slightly different for the different methods of extraction of the resolution.

Simulated events are required to pass the same selection criteria applied to the data (see table 3.1), and additionally, the true position x is required to be less than 20 cm for both directions (u/P and v/N), to ensure well-defined cluster position m.

3.3. Measurement strategy with overlapping sensors

The track extrapolation error plays a crucial role in the computation of the cluster position resolution. From simulated data, it is known that this error can be inaccurately estimated for some tracks, and there is no available information on the accuracy of the track error estimation in real data. An alternative method relying on overlapping sensors, Figure 3.5, is proposed in this discussion to mitigate the impact of this uncertainty on the final results.

The overlap or pairs method is used to determine the cluster position resolution using overlapping sensors. This method was previously implemented by the CMS experiment [15]. Only tracks that cross two consecutive ladders (internal, i and external e) of the same layer are considered. The resolution is obtained from the distribution of a combination of the residual of the internal ladder R_i and the residual of the external ladder R_e, known as the double residual ΔR:

$$\Delta R = \frac{R_i - R_e}{\sqrt{2}}. \tag{3.17}$$

The predicted hit is here computed as the intercept on a given layer of the track fitted by excluding the clusters on the layer under study when applying the track reconstruction algorithm and extrapolating the results to the overlaps region, *i.e.* an unbiased extrapolation.

This strategy can cancel errors in extrapolated track positions through the double residual difference. As a result, it is possible to separate the contribution of tracking precision from the actual cluster position resolution. Additionally, the measurement is only marginally affected by Coulomb scattering due to the small radial distance between the two overlapped sensors.

In the overlap method, the cluster position resolution is defined, as described in

Variable	Selection criteria
transverse momentum	$p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$
number of PXD hits	> 0
number of CDC hits	> 20

Table 3.1. – Summary of the track selection criteria used for the different resolution extraction methods.

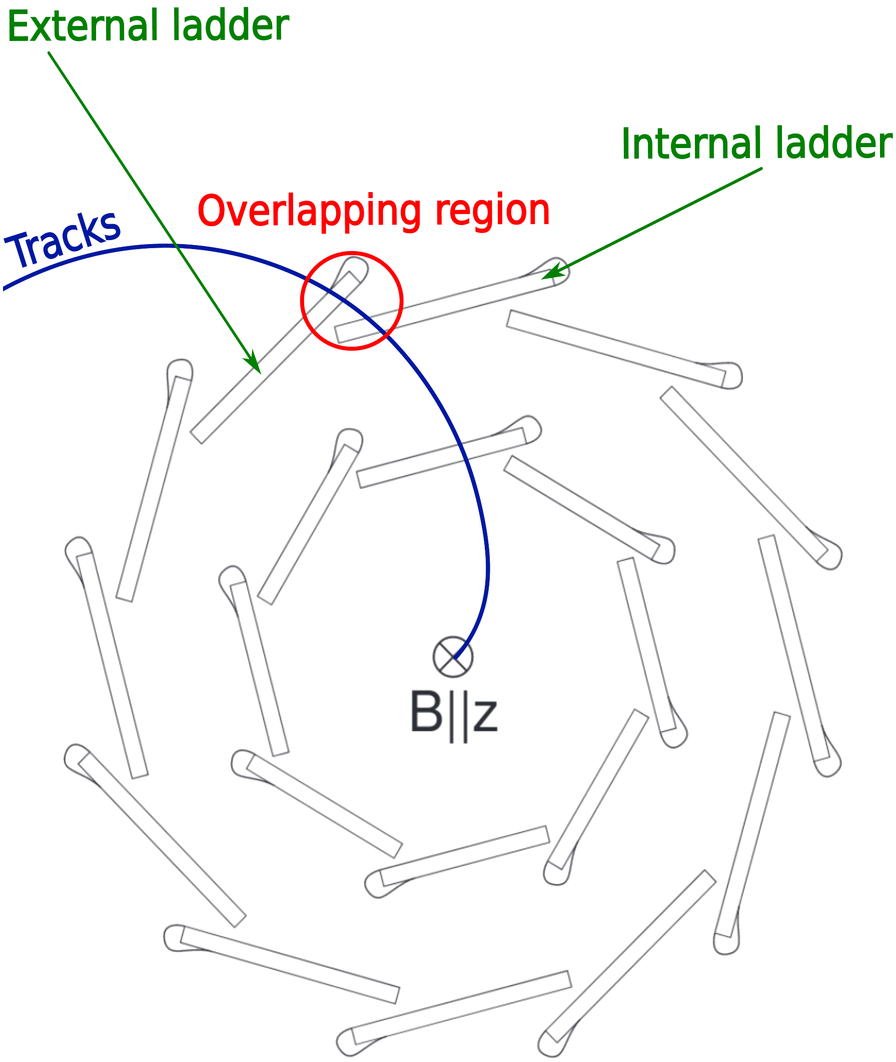


Figure 3.5. – Schematic view of the SVD volume in the $r\phi$-direction with tacks passing through two consecutive sensors (internal and external) of the same layer called overlapping region.

Appendix A.3, as the double residual variance:

$$\sigma_{cl}^2 = Var(\Delta R). \tag{3.18}$$

The overlap method assumes that both residuals are measured in the same plane, which is not the case for the SVD because of its windmill architecture (refer to Figure 3.5). Therefore, in order to properly compare residuals measured on the same plane, the projection of the external residual onto the internal ladder that is parallel to the track is computed for the u/P side sensors by applying a correction factor C and re-normalize event by event:

$$\Delta R = \frac{R_i - CR_e}{\sqrt{1+C^2}}, \tag{3.19}$$

with:

$$C = \frac{cos(a_e)}{cos(a_i)}. \tag{3.20}$$

Further details on how this geometrical correction is determined are reported in Appendix A.4.

Three important preliminary steps are needed for the overlaps measurement: first, we study whether we have enough candidates to develop the method; second, the definition of the track selection process. And finally, the way to extract the variance of the double residual distribution.

3.3.1. Evaluation of the feasibility

One major challenge when using this method is ensuring a sizeable overlapping region and enough track candidates. A small overlapping region can create two issues. Firstly, if the available area is too small, the number of tracks traversing the overlapping region on two consecutive layers will be proportionally reduced. Secondly, on the u/P side, the number of available strips is limited because the strips are aligned along the length of the overlapping region in the u-direction. The strips considered for the study in u/P are illustrated in Figure 3.6 where only the strip positions that meet the requirement of overlapping with the sensors are shown.

The estimated number of strips N_{strips} per layer is determined from the total range in strip positions divided by the pitch [111]:

$$N_{\text{strips}} = \frac{\texttt{std}(x_{\text{strips}})}{p}, \tag{3.21}$$

where std refers to the usual standard deviation, x_{strips} is the SVD strip position range in the overlaps area and p the pitch readout depending on the sensor type:

- Layer 3 (u/P): $p = 50\,\mu$m,
- Layers 4, 5 & 6 (u/P): $p = 75\,\mu$m.

The results are provided in Table 3.2. It should be noted that whereas the number of strips in the outer layers is reasonable, the case of layer 3 is more problematic,

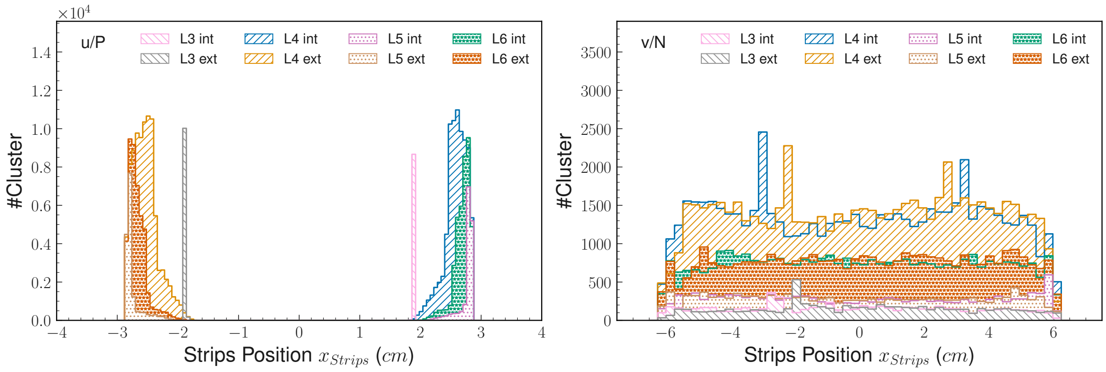


Figure 3.6. – Position of activated strips by a hit of the u/P (left) and v/N (right) for each layer and the relative ladder's position of the overlap.

Table 3.2. – Estimated number of strips N_{strips} of the u/P side in the overlapping region for data.

	Ladders	
Layers	Internal	External
3	5.76	6.15
4	49.49	56.12
5	36.04	40.78
6	38.44	43.09

with a maximum number of strips in the overlaps range equal to 6. Additionally, the requirement of hits in two consecutive ladders on the same layer further reduces the sample size available for the study.

3.3.2. Computation of the cluster position resolution

In the overlapping method, the cluster position resolution σ_{cl} is extracted from the double residual ΔR distribution, defined in Equation 3.17, as a function of the sensor sides (u/P and v/N) and layers (from 3 to 6), as in the Figure 3.7. The ΔR histogram is fitted with a student's t-distribution:

$$T(X,\nu,\mu,\sigma) = \frac{e^{\Gamma(\frac{\nu+1}{2})-\Gamma(\frac{\nu}{2})}}{\sigma\sqrt{\pi\nu}}\left(1+\frac{(X-\mu)^2}{\sigma^2\nu}\right)^{-\frac{\nu+1}{2}}, \tag{3.22}$$

with the parameters:

- the number of degrees of freedom ν,
- the mean of the distribution μ,
- the variance $\sigma^2\frac{\nu}{\nu-2}$.

The student's t-distribution is found to describe non-Gaussian tails better. The smaller values of the parameter ν account for heavier tails, while for larger values of ν, the distribution tends to be a Gaussian. Moreover, the student's t-fit results in the best χ^2 agreement and the smallest fit parameters uncertainties among the tried functions.

The resolution σ_{cl} is then assessed using the `sigma-68` estimator, which involves finding the half distance between the 84^{th} and 16^{th} quantiles of the student's t-fitted distribution [112].

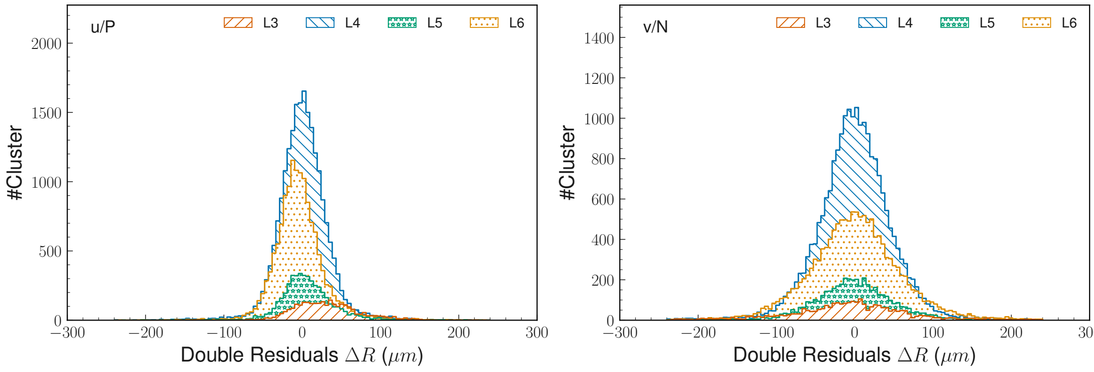


Figure 3.7. – Distribution of the Double residuals (ΔR) for each side and layer using di-muon data.

3.3.3. Estimation of uncertainties in the resolution

While the `sigma-68` estimator is a robust measure of the distribution's width, the resolution uncertainty could no longer be determined by propagating the error from the fit parameters. So, the statistical uncertainty is estimated by varying the fit's parameters within their maximal errors. The procedure to estimate the uncertainties follows:

1. Vary all the fit parameters $p = \left(N, \mu, \nu, \sigma\right)$ in their maximal error ranges $p^{Test} \in \left[p - error(p), p + error(p)\right]$ determined during the fit,
2. Compute the the student's t-distribution T^{Test} for each sets of parameter's values p^{Test} and get the `sigma-68` resolution,
3. Statistical error is taken as half the maximal variation of the resolution: $error(\texttt{sigma-68}) = \frac{max(\texttt{sigma-68}) - min(\texttt{sigma-68})}{2}$.

3.3.4. Fiducial Area Selection

When extrapolating the track position t, it is important to select a fiducial area to exclude the edge sensor region, where the effect of masked strips has not been simulated and could introduce a bias in the track position evaluation. This selection is particularly significant for the overlap method, as it impacts the already low number of strips selected, and a simple geometrical reduction of such area cannot be deployed to not limit too much the sample size.

Figure 3.8 depicts the distribution of the double residual ΔR, with bumps in the tail due to the masked strips. The fiducial area selection is optimized as a function of the SVD layer and side separately. The optimal cut value can be determined by comparing the resolutions obtained from data, and MC simulations at different t cut values, as shown in Figure 3.9. The resolution increases with the number of selected clusters until it reaches a plateau where the overlapping region falls into the fiducial area definition. A threshold is set just before the plateau to ensure enough statistics while maintaining good resolutions. The optimized fiducial area selection is outlined in detail in Table 3.3.

Table 3.3. – Fiducial area selection relying on the extrapolated track postion t (in cm) projected in the u/P and v/N sides to remove tracks passing in masked strips in the sensor edges.

Layer 3	$\lvert t^{u/P}\rvert < 1.908\,\text{cm}$ & $\lvert t^{v/N}\rvert < 5.9\,\text{cm}$
Layer 4	$\lvert t^{u/P}\rvert < 2.82\,\text{cm}$ & $\lvert t^{v/N}\rvert < 5.9\,\text{cm}$
Layer 5	$\lvert t^{u/P}\rvert < 2.84\,\text{cm}$ & $\lvert t^{v/N}\rvert < 5.9\,\text{cm}$
Layer 6	$\lvert t^{u/P}\rvert < 2.82\,\text{cm}$ & $\lvert t^{v/N}\rvert < 5.9\,\text{cm}$

Figure 3.8 demonstrates the impact of the fiducial area selection, particularly in layer 3, where it is more significant. Considering the v/N side case, the position's tracks passing the overlap sensor requirement cover a broader range, making it feasible to keep the geometrical-based selection. As a general rule, this selection leads to a more symmetrical, bell-shaped distribution of the residuals with a smaller width. The fiducial area selection improves the measured resolutions σ_{cl} on data and simulation.

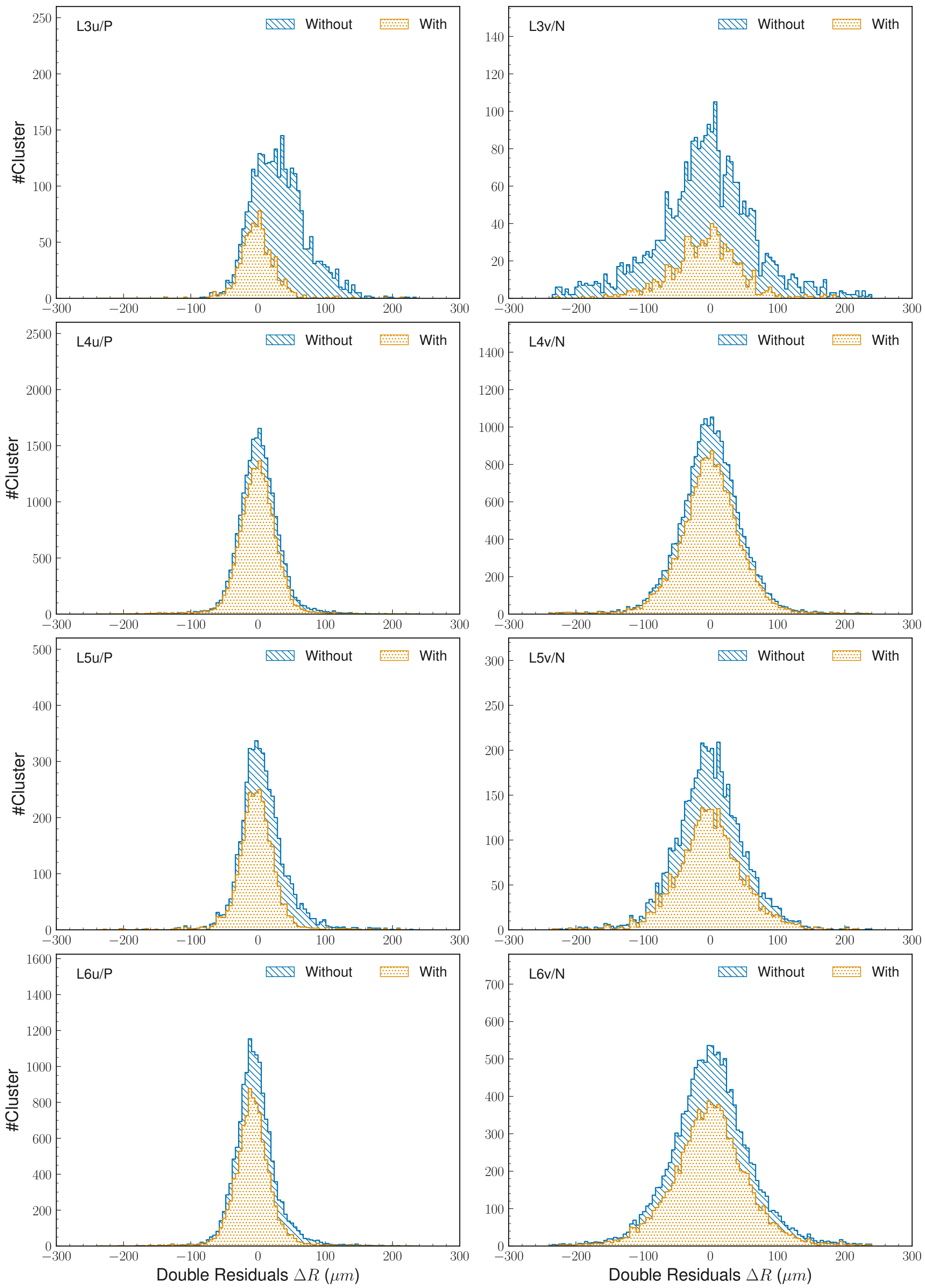


Figure 3.8. – Comparison of the double residual distributions before (blue) and after (orange) applying the fiducial area selection in experiment 14 sample.

After the selections presented in 3.2.1 and the fiducial area selection, the remaining numbers of clusters by the side and layers are summarized in Table 3.4.

Table 3.4. – Number of clusters passing all selections.

		Number of Clusters		
		After Selections	Before Selections	Relative efficiency
Side u/P	Layer 3	860	3539	24.30%
	Layer 4	18776	36565	51.35%
	Layer 5	3168	7465	42.44%
	Layer 6	10484	20222	51.84%

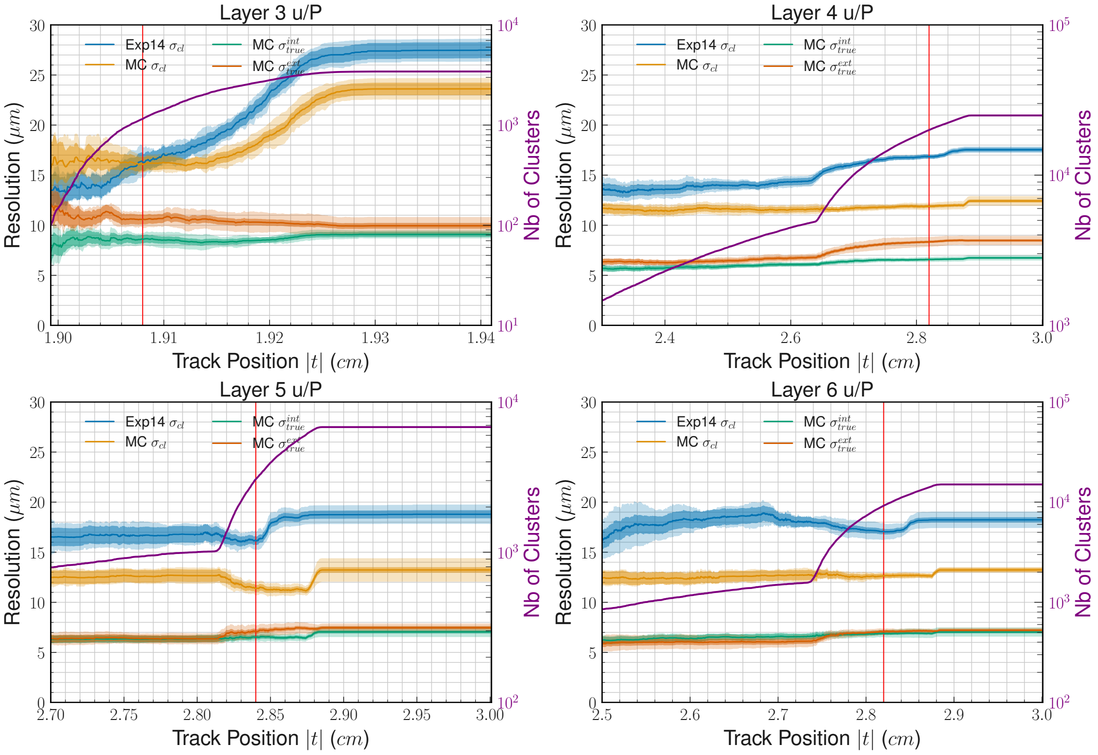


Figure 3.9. – Fiducial area selection optimization for each layer to reduce the Data/MC discrepancy.

3.4. Results on the spatial resolution

Using the method of the overlapping sensors, the spatial resolutions, true σ_{true} (Monte-Carlo simulations) and measured σ_{cl} (Monte-Carlo simulations and data) are given in Figure 3.10, in functions of the SVD layer and side. The detailed values are reported in Table 3.5.

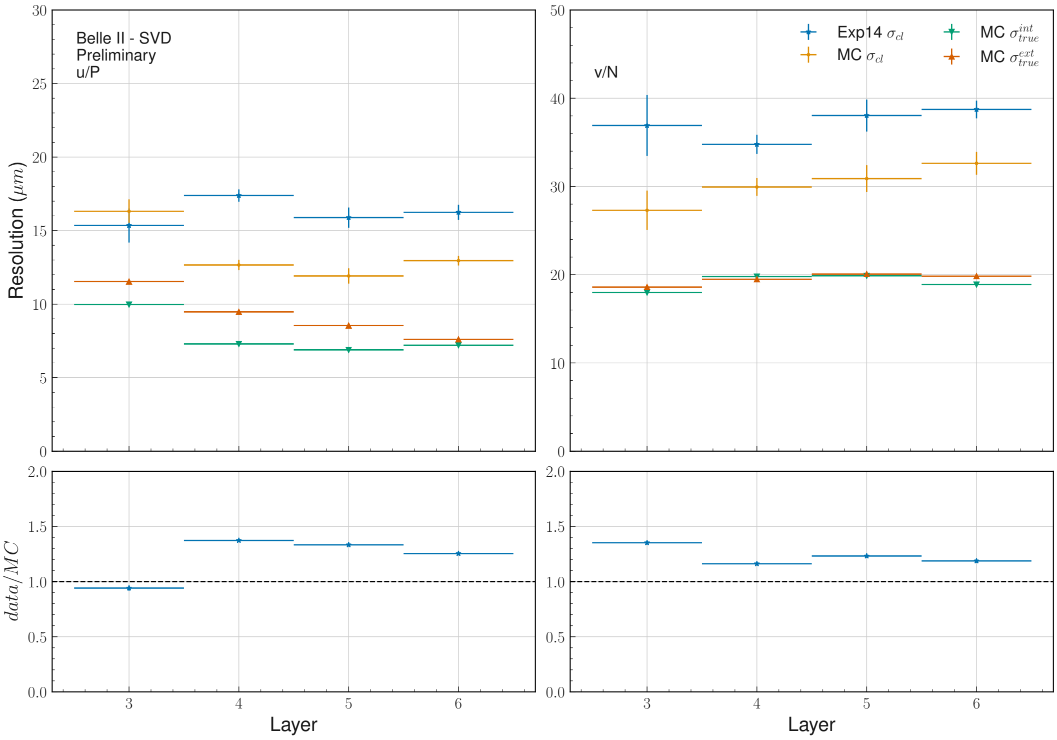


Figure 3.10. – Comparisons of the measured spatial resolution between data-exp14 in blue, Monte-Carlo simulations in orange and the true resolutions in green (internal sensor) and dark orange (external sensor) in top plots. The different resolutions are given in functions of SVD sides; u/P in left plots and v/N side in right plots; and layers. The bottom plots show the resolution ratio between data and simulations.

Table 3.5. – Spatial resolutions (μm) results for data and simulations in the sensor layer and side functions. Only statistical uncertainties are given.

	Data		Simulation	
Layer	u/P side	v/N side	u/P side	v/N side
3	15.3 ± 1.1	36.9 ± 3.4	16.31 ± 0.81	27.3 ± 2.2
4	17.39 ± 0.42	34.7 ± 1.0	12.66 ± 0.36	29.94 ± 1.0
5	15.88 ± 0.69	$38. \pm 1.0$	11.92 ± 0.52	30.8 ± 1.5
6	16.24 ± 0.52	$38. \pm 1.0$	12.95 ± 0.33	32.6 ± 1.2

The measured resolutions on data are around 16 μm along the $r\phi$ direction (u/P side) and 35 μm along the z-direction (v/N side). Regarding Figure 3.11, the spatial resolution results on the v/N side are similar to the resolution obtained from the strip pitch size. However, the overlapping method yielded a slightly higher resolution on the u/P side than the strip pitch. This is expected due to the intrinsic limitation of the overlapping method in accessing a wider range of incident angles on the u/P side.

Furthermore, Figure 3.10 shows a significant discrepancy between the spatial resolution (σ_{cl}) and the true resolution σ_{true} in simulations. This disagreement in simulation is discussed in Appendix A.5 to infirm potential non-cancellation of the tracking uncertainties.

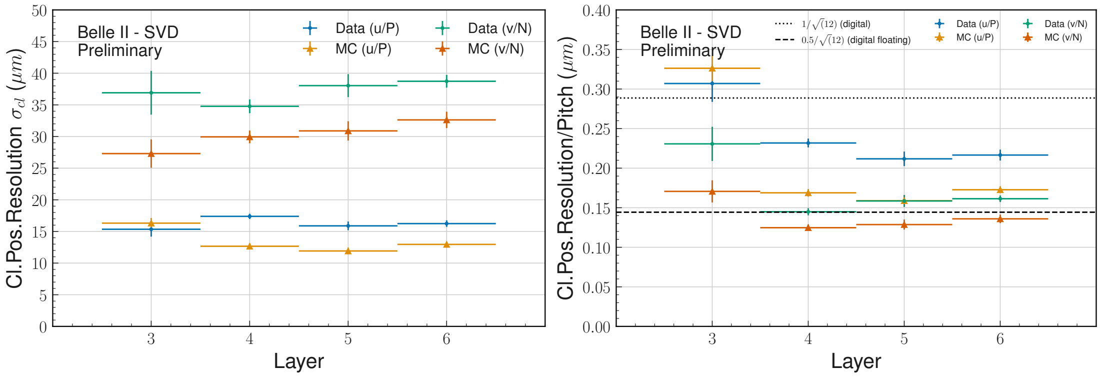


Figure 3.11. – Measured spatial resolutions (right) obtained for data (blue) and simulation (orange) following sides and layers. Measured resolutions divided by the sensor pitch (left) obtained for data (blue) and simulation (orange) following sides and layers.

3.5. Conclusions on the spatial resolution

This chapter presents the first results of the Belle II's SVD spatial resolution obtained from a method exploiting the overlapping sensor of the same layer described by CMS [15]. After adaptation to the SVD geometry, the method gives respectively 16 μm (35 μm) for the $r\phi$ (z) direction with a good agreement with the expectations given by the strip pitch.

This method was accomplished within a wider study to measure the spatial resolution of the SVD with different methods presented at the 30th Vertex Workshop and published [110]. The overlapping method can be compared, in Table3.6, to the Event-by-event and Global methods [110] where the track uncertainty is explicitly subtracted at the events or distributions levels, in the same angular range accessible to the overlaps method. Unlike the other methods, the overlapping method is robust against multiple Coulomb scattering and tracking uncertainty thanks to the small distance between the two sensors used. Nevertheless, the limited incident angle range may have caused worse performances for the u/P side compared to the two others, while the overlapping method gives similar results for the v/N side.

Table 3.6. – Summary of the digital (strip pitch based) and measured resolution on data taken at the normal incidence for the Event-by-event and Global methods [110, 16]. For the overlapping sensor method, the average on the whole accessible angular range is shown.

	Digital	Event-by-event	Global	Overlap
Layer 3 u/P (μm)	7	7	9	15
Layer 456 u/P (μm)	11	10	11	16-17
Layer 3 v/N (μm)	23	24	23	33
Layer 456 v/N (μm)	35	32	35	29-36

To improve the method and the study, a better knowledge of the track uncertainty cancellation in the $r\phi$ direction is necessary. To perform such studies, an idea will be to access the tracking information obtained by doing a fit considering only one of the two sensors in the overlapping regions. A fudge factor could also be designed in order to correct the simulations in order to give a more accurate resolution close to the one obtained in the data. Finally, we can design a mode to perform the spatial resolution measurement in data acquisition live to monitor the SVD performances.

4. Search for $\tau^- \to \mu^-\mu^+\mu^-$ lepton flavour violating decays

Table of contents

According to the Standard Model, lepton flavour conservation happens accidentally. However, the discovery of neutrino oscillations shows that there is a direct violation of neutrino flavour and also a charged-lepton flavour violation through the charged currents of weak interaction. Decays that involve Lepton Flavour Violation (LFV) are estimated to occur at exceedingly low rates of 10^{-50} [2], making them impossible to detect experimentally. If LFV decays were observed, it would provide solid proof of non-SM physics. In this chapter, we will detail the method to search for $\tau^- \to \mu^-\mu^+\mu^-$ LFV decays at Belle II using the 424 fb^{-1} of collected data [80].

4.1. Untagged analysis strategy

The commonly adopted method to search for $\tau^- \to \ell^-\ell^+\ell^-$ [1] decays in B-factories rely on the reconstruction of both τ leptons produced by the e^+e^- collisions [8, 7, 30]. The signal τ_{sig} is reconstructed as the searched decay $\tau^- \to \ell^-\ell^+\ell^-$, which is a 3-prong decay [2], while the other tau, called tag, is reconstructed from a Standard Model (SM) decay. The most used tag decays are:

- $\tau^- \to e^-\overline{\nu}_e\nu_\tau$,
- $\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$,
- $\tau^- \to \pi^-\nu_\tau$,
- $\tau^- \to \pi^-\pi^0\nu_\tau$, where the π^0 is usually not explicitely reconstructed.

These four 1-prong decays cover about 85% of all τ decays as shown in Figure 4.1. The so-called 3x1 reconstruction is usually accompanied by a requirement for the

---

[1] ℓ could be either an electron e^- or a muon μ^-.

[2] Prong refers to the number of charged tracks in the final state.

total number of tracks in the event to be exactly four. The 3 × 1-prong topology reconstruction has an efficiency limited by the tag side branching fractions and the requirement of having 4 tracks only in the event.

The idea described in this Section is to perform an inclusive (or *untagged*) reconstruction, as was done already for the search for the $\tau \to \ell\phi$ at Belle II [17]. The signal τ is reconstructed into three muons, and all the other tracks and clusters are used to form a Rest-of-Event (ROE), as illustrated in Figure 4.2 and Figure 4.3. An additional selection (or *mask*) is applied to clean the ROE from background tracks and clusters. With the inclusive method, the reconstruction doesn't only target the 1-prong decays of the τ^- tag, but also the 3-prong ones, which represent about 15% of the total τ^- decays. This method also selects events with an odd number of tracks, which can happen if the tag side is not fully reconstructed or in case of additional fake (beam background) tracks, as illustrated in Figure 4.4 where the good tracks are defined as tracks with impact parameters $|dz| < 3\,\text{cm}$ and $|dr| < 1 cm$. The downside of the inclusive method is to enlarge the number of reconstructed background events.

Like the other τ-LFV studies performed by B-factories, the signal and background yields are evaluated in the 2D plane formed by the reconstructed mass of the signal τ, and the energy difference between the three leptons and the beam energy in the centre of mass system:

$$M_{3\mu} = \sqrt{E_{3\mu}^2 - P_{3\mu}^2}, \tag{4.1}$$

$$\Delta E_{3\mu} = E_{3\mu}^{CM} - E_{beam}^{CM} = E_{3\mu}^{CM} - \frac{\sqrt{s}}{2}, \tag{4.2}$$

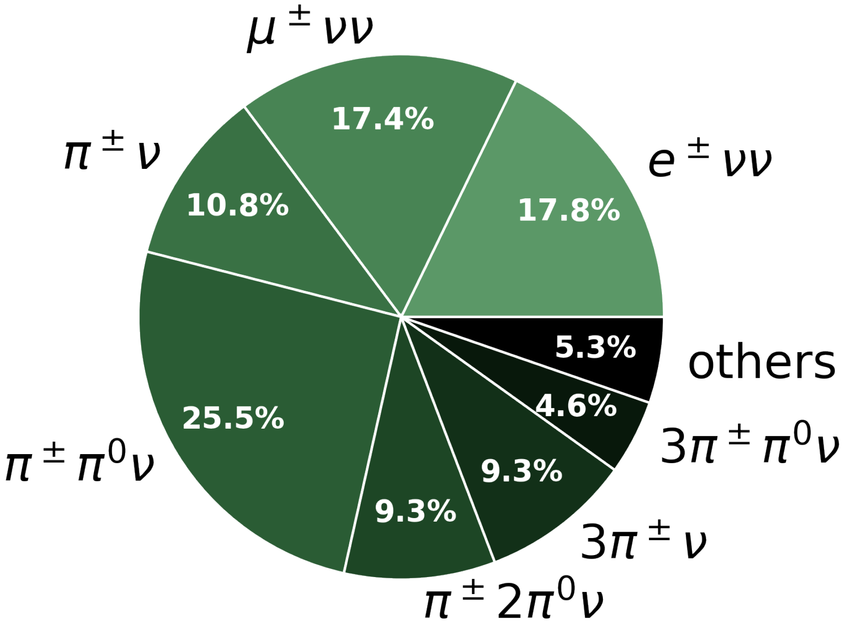


Figure 4.1. – Diagrams of the τ^- lepton branching fractions into the main decays final state [13].

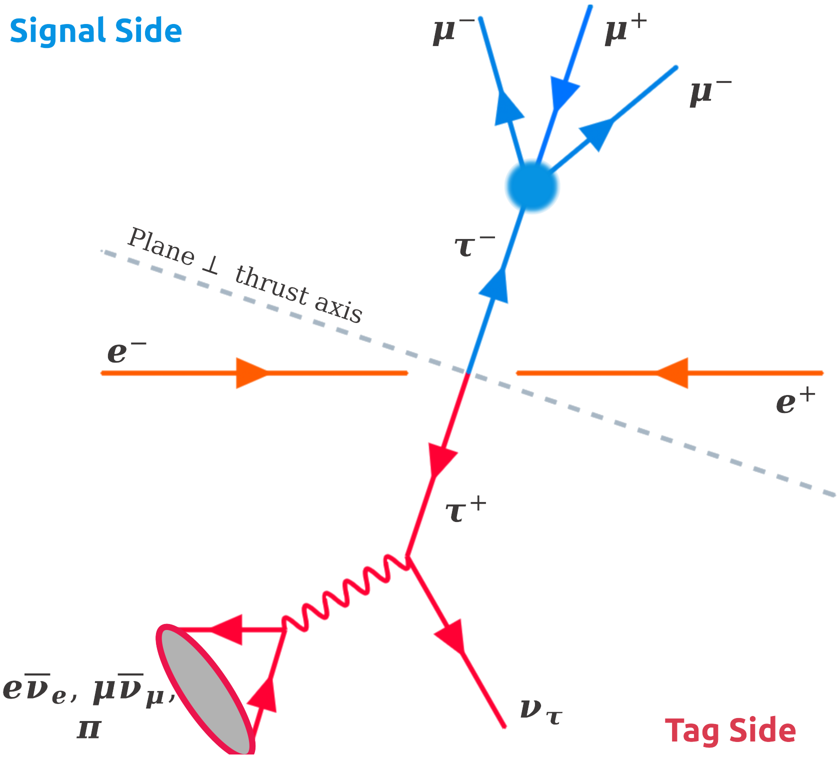


Figure 4.2. – Diagrams of the $e^+e^- \to \tau^+\tau^-$ events reconstruction for the $\tau^- \to \mu^-\mu^+\mu^-$ LFV decays in 3 × 3-topology reconstruction strategy.

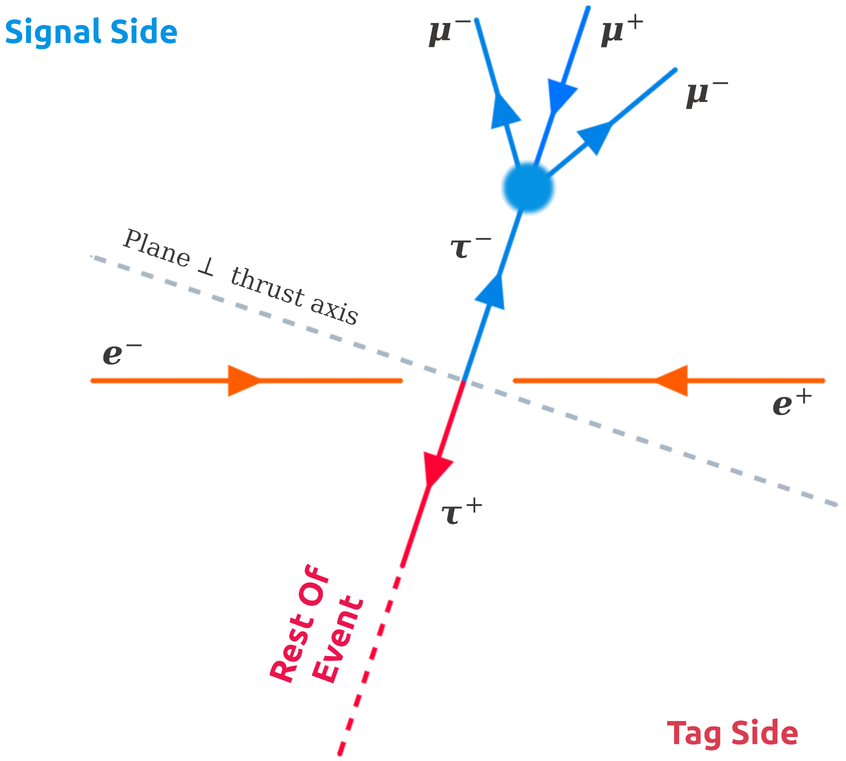


Figure 4.3. – Diagrams of the $e^+e^- \to \tau^+\tau^-$ events reconstruction for the $\tau^- \to \mu^-\mu^+\mu^-$ LFV decays in inclusive reconstruction strategy.

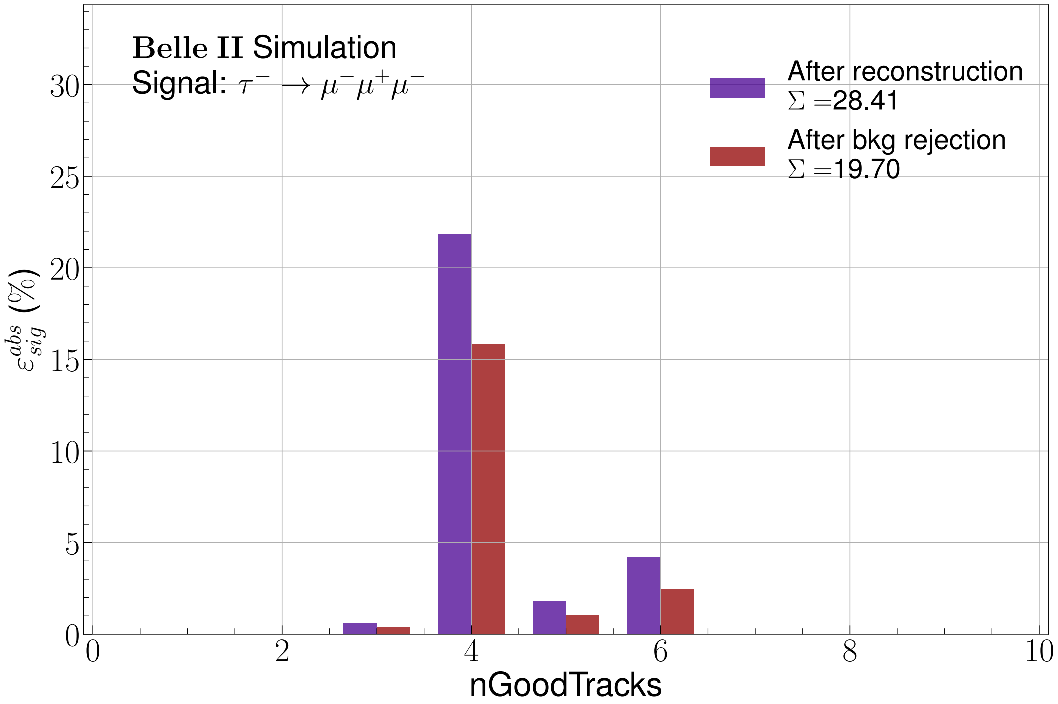


Figure 4.4. – Distribution of the number of good tracks (are defined as tracks with impact parameters $|dz| < 3\,\mathrm{cm}$ and $|dr| < 1 cm$) per event after the inclusive reconstruction (purple) and background rejection (red). The total absolute signal efficiency is 28.41% after the reconstruction and 19.70% after the background rejection.

where $E_{3\mu}$ and $P_{3\mu}$ are the sum of the energies and the summed momenta's magnitude of the three leptons. For signal, it is expected that these quantities peaks at $M_{3\mu} \sim m_\tau = 1776.86 \pm 0.12\,\mathrm{MeV}$[13] and $\Delta E_{3\mu} \sim 0\,\mathrm{GeV}$ due to the final state without neutrinos. On the contrary, background events are smoothly distributed without peaking structures. Different regions of the $(M_{3\mu}, \Delta E_{3\mu})$ plane are defined according to the resolution of these variables. The regions are either used to optimize the background rejection, evaluate the number of expected background events, or search for the signal. The signal region is kept hidden until the analysis is finalized in order to avoid potential experimenters' bias.

The procedure to suppress the background is divided into two steps, as detailed in Section 4.3, optimised to give the highest Punzi figure of merit [19]:

$$FOM = \frac{\varepsilon_{sig}}{a/2 + \sqrt{b}}, \tag{4.3}$$

where ε_{sig} is the absolute signal efficiency, b the number of background events, and $a = 1.28$ is the desired 90% confidence level. Firstly, a set of cut-based preselections is designed to limit contamination by the low-multiplicity background, followed by a k-folded Boosted Decision Tree (BDT), which targets $\tau^+\tau^-$ and $q\overline{q}$ backgrounds.

Finally, sources of systematic uncertainties are identified and evaluated in Section 4.4, before computing the branching fraction's upper limit at 90% Confidence Level (CL):

$$\mathscr{B}^{90}_{UL} = \frac{s^{90}}{2 \times L \times \sigma_{\tau^+\tau^-} \times \varepsilon_{\tau^-\to\mu^-\mu^+\mu^-}}, \tag{4.4}$$

where s^{90} is the upper limit on the signal yield, $L = \int \mathscr{L} dt = 424\,\mathrm{fb}^{-1}$ the integrated luminosity of the LS1 dataset, $\sigma_{\tau^+\tau^-}$ the $\tau^+\tau^-$ production cross-section measured as $0.919 \pm 0.003\,\mathrm{nb}$ and $\epsilon_{\tau^-\to\mu^-\mu^+\mu^-}$ the $\tau^- \to \mu^-\mu^+\mu^-$ signal efficiency. The computation of the upper limit is based on the CL_s method [113, 23] described in Section 4.5.

The data and Monte-Carlo (MC) simulated datasets used in this analysis are described in Chapter 2. Since they correspond to the $\Upsilon(4S)$ energy, the analysis is optimised for $L = 362\,\mathrm{fb}^{-1}$, and the $\Upsilon(5S)$ and off-resonance data are added at the end of the study.

4.2. Event reconstruction

4.2.1. Particle lists

In this study, the muon candidates are first combined to form a $\tau^- \to \mu^-\mu^+\mu^-$ decay candidate. The event shape and the kinematic calculators use the list of photons and charged particles with an assumed pion mass. Finally, the rest of the event is built from all the photons and charged particles which are not used in the reconstruction of the $\tau^- \to \mu^-\mu^+\mu^-$ decay chain. Pions are also used to reconstruct the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample. The charged and neutral particle definitions are given in the following.

4.2.1.1. Charged particle lists requirements

The muons and pions lists are reconstructed as charged tracks coming from the Interaction Point (IP), with the transverse (dr) and longitudinal (dz) projections of their point of closest approach smaller than 1 cm and 3 cm, respectively [114]. Muons are identified by exploiting the likelihood-based muon identification variables. The particle IDentification (ID) requirements are loose and tuned at a later stage. Lepton IDentification (LID) corrections provided by the performance group are applied to the simulation. No particle ID requirement is applied to the pion list in order to use all particles coming from the IP to build the event shape and kinematics, as described later. All selections on charged-particle candidates are summarized in Table 4.1.

Table 4.1. – Selection criteria for charged particle lists.

Variable	Requirement μ^-	π^-	π^- control sample
$\|dz\|$ (cm)	<3	<3	<3
$\|dr\|$ (cm)	<1	<1	<1
$muonID$	> 0.5	-	< 0.5
$electronID^{BDT}$	-	-	< 0.5
$pionID$	-	-	-

4.2.1.2. Photon and neutral particle lists requirements

The photon lists are filled with the Electromagnetic CaLorimeter (ECL) clusters within the Central Drift Chamber (CDC) acceptance, $cos\theta \in [-0.8660, 0.9563]$ and not associated with any charged track, with requirements on the distance between clusters and tracks, energy and cluster timing. The sum of the weights of all crystals in an ECL cluster should be larger than 1.5. According to the applied energy selection and their usage in the event reconstruction, two types of photons are defined:

1. Photons for π^0 candidate reconstruction. Candidates π^0 are made by combining two photons with an energy greater than 100 MeV passing the invariant mass requirement $0.115 < M_{\gamma\gamma} < 0.152$ GeV/c^2.
2. Photons not belonging to the previous category and having an energy of at least 200 MeV. Those photons are used in the calculators of the event shape and kinematics.

4.2.1.3. Event shape and kinematics

The charged pions and photons defined in the lists above are used to build the event shape and kinematics variables, such as the visible and missing energy of the event, the missing momentum and mass. The visible energy is equal to the sum of the energy of every input particle. The missing momentum (energy) is calculated by subtracting the sum of the momentum (energy) of all particles in the input from the sum of the beam momenta (energy). These quantities are important for background suppression, being sensitive to the emission of undetected neutrinos or not.

An additional event shape variable relevant for this analysis is the thrust axis $\mathbf{n}_T$ [96] which is the vector maximizing the quantity:

$$T = \max_{\mathbf{n}_T}\left(\frac{\sum_i |\mathbf{p}_i \cdot \mathbf{n}_T|}{\sum_i |\mathbf{p}_i|}\right), \tag{4.5}$$

where $\mathbf{p}_i$ is the momentum vector of the i-th particle in the final state, either charged or neutrals.

4.2.2. Signal reconstruction

In the $e^+e^- \to \tau^+\tau^-$ process, τ^- leptons are produced back-to-back in the centre of mass frame, so their decay products are well geometrically separated. Thus, we combine three muons, requiring that they are in the same hemisphere defined by the plane orthogonal to the thrust axis. In addition, a constrain is applied in the 2D plane formed by the 3μ invariant mass and $\Delta E_{3\mu}$:

- $1.4 < M_{3\mu} < 2.0$ GeV/c^2,
- $-1.0 < \Delta E_{3\mu} < 0.5$ GeV.

By doing this, it is possible to eliminate background events occurring outside the expected peak area. Finally, the signal τ vertex is fitted by the *treeFitter* tool [115], which updates the momenta and position of the final state particles.

4.2.3. Rest-of-Event building

Only the τ^- lepton decaying into the searched LFV channel is reconstructed. As described in Section 4.1, our method does not constrain the opposite τ^- decay. Instead, the rest of all remaining tracks and ECL clusters are gathered in the ROE. The mass of the tracks is assigned following the largest particle ID likelihood. The ROE is cleaned from the background by masking tracks and clusters not passing some requirements: tracks are required to lie in the CDC acceptance, have a transverse momentum greater than 75 MeV/c, $|dz| < 3$ cm, and $|dr| < 1$ cm [114]. In the photon's case, it is required to be in the CDC acceptance and to have an energy of at least 200 MeV.

The ROE properties, such as the mass and the difference of energy with respect to beam one, are stored in variables which are used for background suppression.

4.2.4. Additional requirements

4.2.4.1. Signal region

For the signal LFV decay, the $M_{3\mu}$ and $\Delta E_{3\mu}$ variables are peaking thanks to the absence of neutrinos in the final state. The resolution of these variables called δ, is extracted by an unbinned fit, with `RooFit` library [116], using an asymmetric Gaussian probability density function, as seen in Figure 4.5. The fit is performed in a limited range to only take into account the signal peak. From the fit are extracted the mean and the up and down resolutions $\delta^{up(down)}$, which are reported in Table 4.2.

Table 4.2. – Fitted resolution for $M_{3\mu}$ and $\Delta E_{3\mu}$ used as units to define signal regions.

Mode	Variable	Mean $\bar{\mu}$	Down Resolution δ^{down}	Up Resolution δ^{up}
$\tau^- \to \mu^-\mu^+\mu^-$	$M_{3\mu}$ (MeV/c^2)	1777.35±0.07	4.80±0.07	4.44±0.06
	$\Delta E_{3\mu}$ (MeV)	0.7±0.3	14.9±0.3	10.0±0.5

The $(M_{3\mu}, \Delta E_{3\mu})$ two-dimensional plane is subdivided into different regions with sizes and shapes based on the parameters extracted from the one-dimensional distributions. Two shapes can be used. The first one is the asymmetric rectangular boxes centred around the fitted means $\bar{\mu}$ and with semi-axis based on multiples n of the resolutions $\delta_{up(down)}$ such as:

$$\bar{\mu}_M - n \times \delta_M^{down} < M_{3\mu} < \bar{\mu}_M + n \times \delta_M^{up}, \tag{4.6}$$

$$\bar{\mu}_{\Delta E} - n \times \delta_{\Delta E}^{down} < \Delta E_{3\mu} < \bar{\mu}_{\Delta E} + n \times \delta_{\Delta E}^{up}. \tag{4.7}$$

The second type of region is the asymmetric ellipse, such as:

$$\begin{cases} \dfrac{x^2}{(n \times \delta_M^{down})^2} + \dfrac{y^2}{(n \times \delta_{\Delta E}^{down})^2} \leq 1 & \text{if } M \leq \bar{\mu}_M \text{ and } \Delta E \leq \bar{\mu}_{\Delta E} \\ \dfrac{x^2}{(n \times \delta_M^{up})^2} + \dfrac{y^2}{(n \times \delta_{\Delta E}^{up})^2} \leq 1 & \text{if } M > \bar{\mu}_M \text{ and } \Delta E > \bar{\mu}_{\Delta E} \\ \dfrac{x^2}{(n \times \delta_M^{up})^2} + \dfrac{y^2}{(n \times \delta_{\Delta E}^{down})^2} \leq 1 & \text{if } M \leq \bar{\mu}_M \text{ and } \Delta E > \bar{\mu}_{\Delta E} \\ \dfrac{x^2}{(n \times \delta_M^{down})^2} + \dfrac{y^2}{(n \times \delta_{\Delta E}^{up})^2} \leq 1 & \text{if } M > \bar{\mu}_M \text{ and } \Delta E \leq \bar{\mu}_{\Delta E} \end{cases} \tag{4.8}$$

with:

$$\begin{aligned} x &= (\Delta E_{3\mu} - \bar{\mu}_{\Delta E}) \times cos(\theta) - (M_{3\mu} - \bar{\mu}_M) \times sin(\theta), \\ y &= (\Delta E_{3\mu} - \bar{\mu}_{\Delta E}) \times sin(\theta) + (M_{3\mu} - \bar{\mu}_M) \times cos(\theta), \end{aligned} \tag{4.9}$$

where θ is the correlation angle between $M_{3\mu}$ and $\Delta E_{3\mu}$ extracted by a linear fit of $M_{3\mu}$

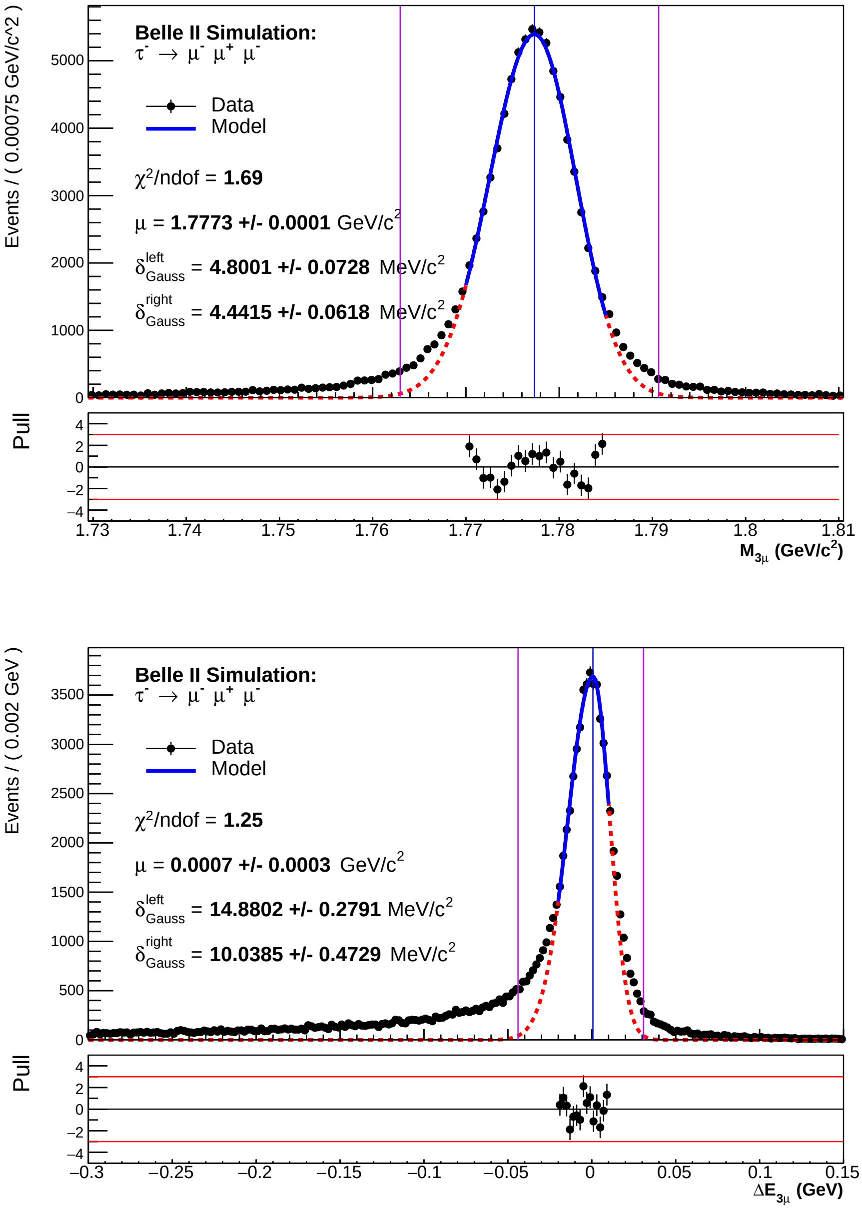


Figure 4.5. – Fits of the $M_{3\mu}$ (top) and $\Delta E_{3\mu}$ (bottom) distributions using signal simulation. The fit curve is represented with a solid line. The vertical magenta lines represent the 3δ region.

profile plot as function of $\Delta E_{3\mu}$, see Figure 4.6.

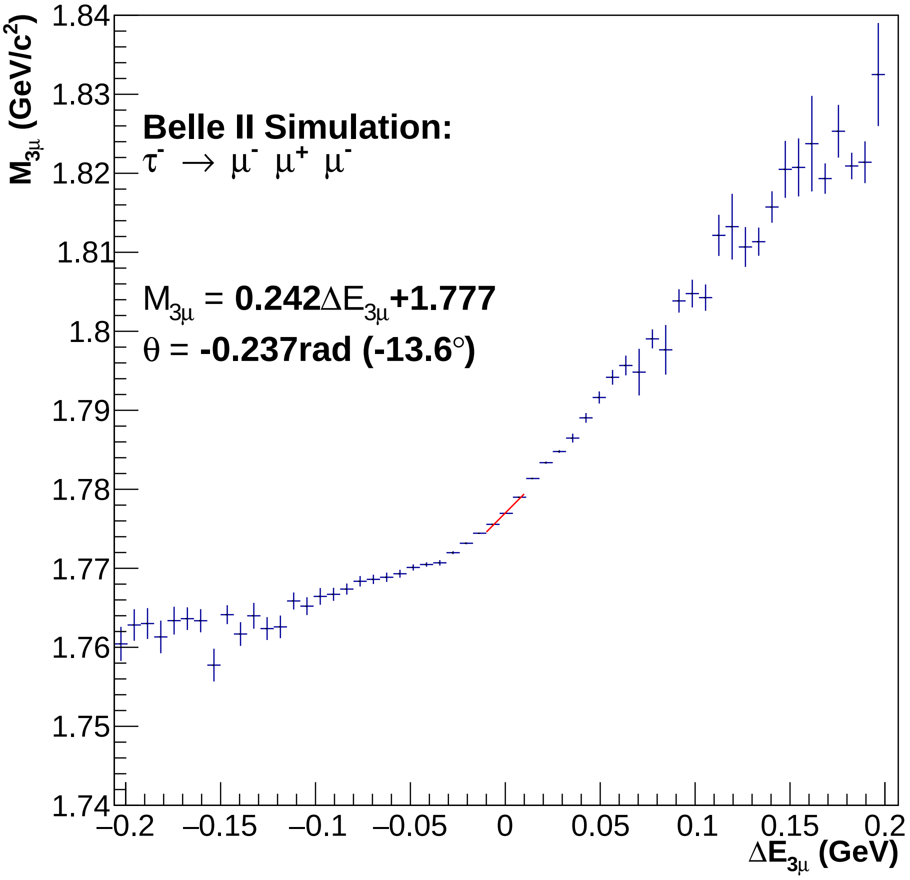


Figure 4.6. – Profile plots of the $M_{3\mu}$ vs $\Delta E_{3\mu}$ for the $\tau^- \to \mu^-\mu^+\mu^-$ final state for signal MC events. The correlation angle between signal region variables is extracted with a linear fit around $\Delta E = 0\,\text{GeV}$ where the signal is peaking.

When optimising background suppression, a $\pm 20\delta$ wide rectangular area is used, while the events outside this box are rejected. This keeps enough background events to perform the BDT's training. However, to optimize the requirement on the BDT output, the region is restricted to a $\pm 5\delta$ wide asymmetric box. The final yield extraction is performed on the best-performing region between $\pm 3\delta$ asymmetric rectangular and elliptic Signal Region (SR), as explained in Section 4.3.3. The 5δ rectangular region is kept hidden in the data, and the selection validation is performed in the Side Bands (SB) defined as the area covering $\pm 20\delta_M$, $\pm 10\delta_{\Delta E}$ while hiding the $\pm 5\delta$ wide box. All of these regions are represented in Figure 4.7.

4.2.4.2. Trigger lines and skimming

The trigger lines considered for this study are the low multiplicity (lml) lines, the hie line and the lines based on the CDC (fff, ffo, fyo and ffy). The signal efficiency, on top

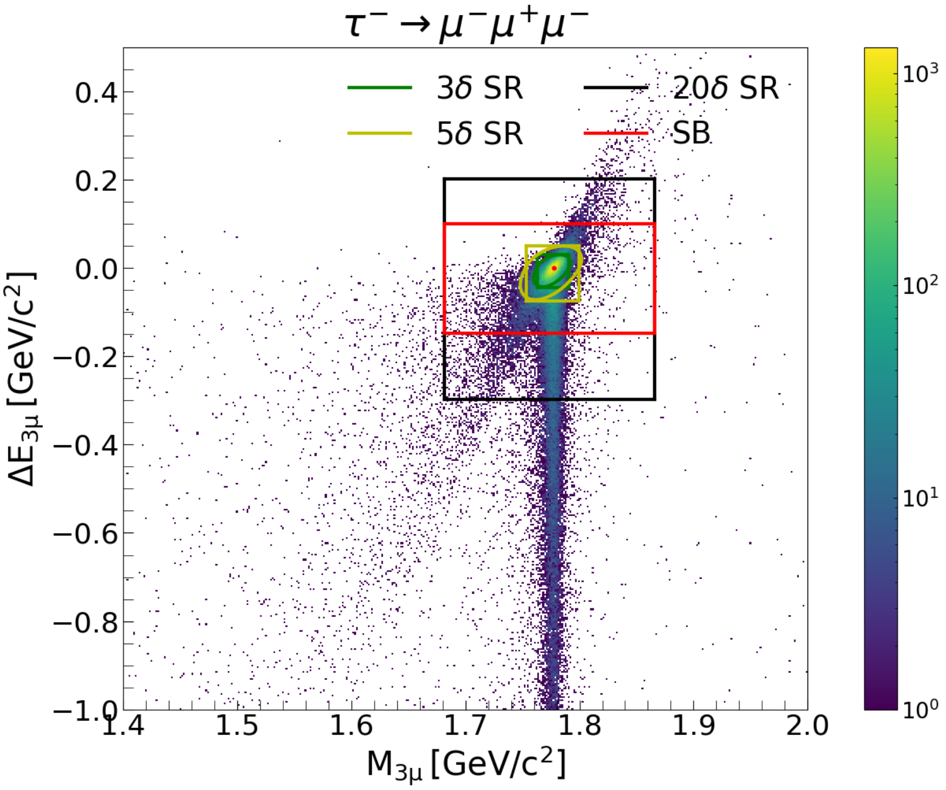


Figure 4.7. – Distributions of $\tau \to \mu\mu\mu$ simulated signal events in the $(M_{3\mu}, \Delta E_{3\mu})$ plane. The edges of the ±3, ±5 and ±20δ regions defined from the previous fits are marked as green, yellow and black rectangles, respectively. The red rectangle represents the sidebands region, ±20δ in $M_{3\mu}$ and ±10δ in $\Delta E_{3\mu}$. The corresponding elliptical regions are also shown.

of reconstruction, for the different lines based on the simulation Trigger SIMulation (TSIM) is shown in Figure 4.8. The most important ones for the $\tau^- \to \mu^-\mu^+\mu^-$ are the ffy, fyo, hie, lml10 and lml12, but we decided to use in addition lml 6,7,8,9 to maximize the efficiency. The definition of those trigger lines is described in Appendix B. The others are discarded since they have been prescaled at some point during the data taking, and their simulated efficiency is difficult to determine. The overall trigger efficiency for $\tau^- \to \mu^-\mu^+\mu^-$ is about 95% on reconstructed signal candidates.

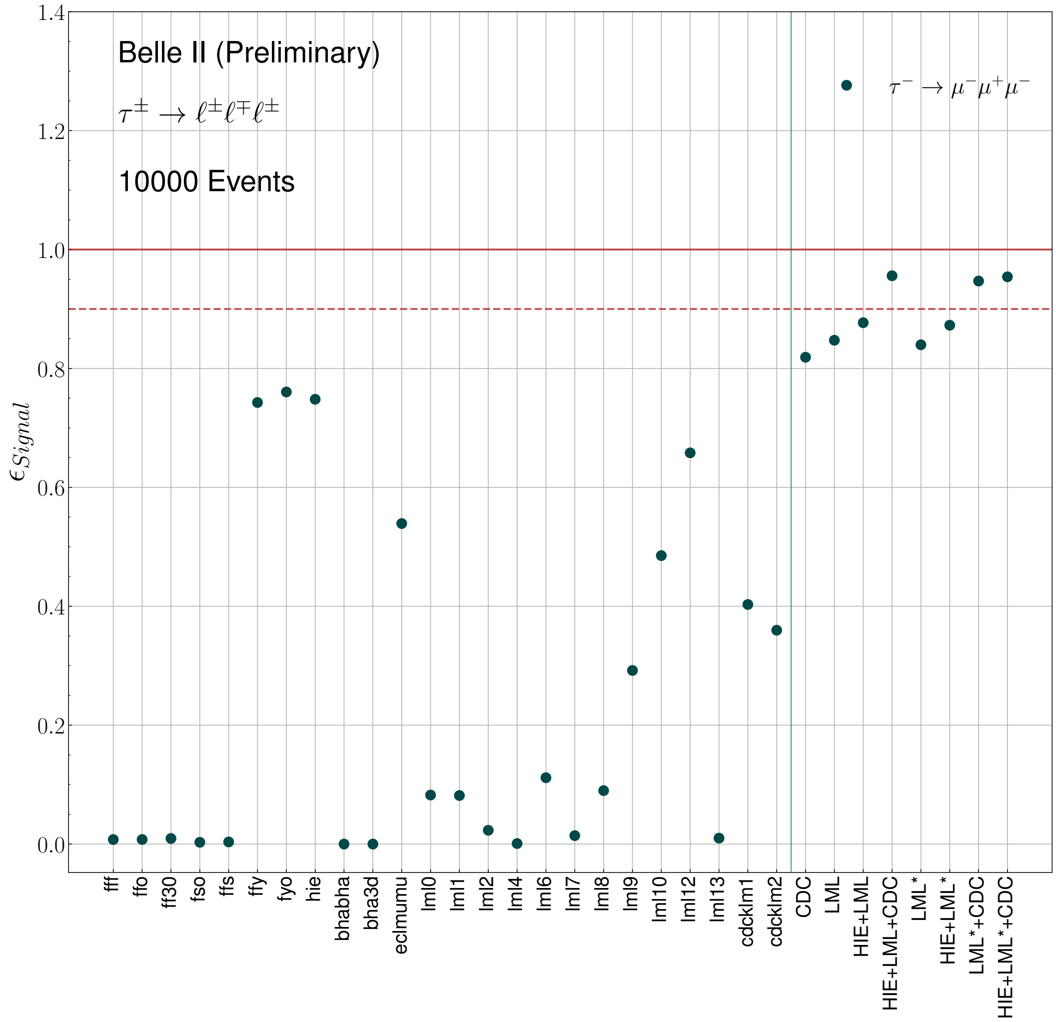


Figure 4.8. – Trigger efficiency obtained from simulation for the $\tau^- \to \mu^-\mu^+\mu^-$ decays. The combination of trigger lines CDC, LML (LML*) and HIE are respectively `fyo/ffy`, `lml0/1/2/4/6/7/8/9/10/12/13` (`lml6/7/8/9/10/12`) and `hie`.

The events are required to pass the τ-LFV skim defined in Section 2.7.1.

4.2.4.3. Signal truth matching

In order to ensure the correctness of the $\tau^- \to \mu^-\mu^+\mu^-$ reconstruction for the signal sample, a set of requirements based on MC information is applied ashing:

- at least one of the two τ decay in a 3 prong final state
- the three muons have a τ in their ancestors,
- the three muons are correctly identified.

This requirement on simulated signal sample is important for background suppression optimization, where signal and background candidates must be well identified.

4.2.5. Signal efficiency and background composition

The signal efficiency at the different reconstruction steps defined previously, as well as the expected background yields, are given in Table 4.3.

Table 4.3. – Relative and absolute signal efficiencies and number of background events retained after the reconstruction and after each offline requirements step for $\tau^- \to \mu^-\mu^+\mu^-$ in the two simulated samples of 4 ab^{-1} each, called train and test. The yields are normalized to 362 fb^{-1}. N_{bkg} is the sum of the individual background components.

		ε^{rel}_{sig}	ε^{abs}_{sig}	N_{bkg}	$N_{\tau\tau}$	$N_{q\bar{q}}$	$N_{B\bar{B}}$	N_{lowm}
Reconstruction	train	46.39%	46.39%	30157.59	1439.31	20569.29	422.09	7726.89
	test	46.48%	46.48%	30483.66	1452.89	20864.05	440.92	7725.80
TruthMatch	train	99.79%	46.30%	30157.59	1439.31	20569.29	422.09	7726.89
	test	99.78%	46.38%	30483.66	1452.89	20864.05	440.92	7725.80
21δ SR box	train	88.47%	40.96%	2220.60	14.03	953.69	0.72	1252.16
	test	88.49%	41.04%	2225.49	13.39	965.73	1.45	1244.92
Trigger	train	94.57%	38.74%	1994.35	9.86	922.01	0.72	1061.75
	test	94.57%	38.81%	2013.99	8.51	936.49	1.45	1067.54
TauLFV Skim	train	98.56%	38.18%	1994.35	9.86	922.01	0.72	1061.75
	test	98.55%	38.25%	2013.99	8.51	936.49	1.45	1067.54
LID correction for 0.5 cut	train	89.85%	34.30%	1803.70	7.98	830.10	0.70	964.92
	test	89.81%	34.35%	1819.67	6.15	842.73	1.23	969.56

4.3. Background suppression

The background suppression is a two-level strategy, with the first level aiming to remove low multiplicity backgrounds using a cut-based selection, followed by a second level aiming to remove primary background components $q\overline{q}$ with a BDT classifier.

4.3.1. Cut-based preselection

The loose lepton identification requirement applied at the reconstruction stage leads to a high background contamination in the $\pm 20\delta$ SR. The first step of background suppression is thus to refine the LID selection to maintain high signal efficiency and reduce background. The performance group only provides simulation corrections for a few threshold values (0.5, 0.9, 0.95, 0.99), hence we are restricted to those thresholds. Cutting at a value of 0.95 for the three muons would result in a too-low signal efficiency (about 68%) while still retaining 24% of the background.

The low muon identification efficiency is mainly caused by slow muons that cannot reach the K_L and Muon detector (KLM). The idea to improve the muonID selection is first to rank the LID of the three muons. The resulting distributions are shown in Figure 4.9. The main efficiency loss comes from the muon with the lowest LID value. By making a selection on the first two muons only:

$$\mu ID_{lead} > 0.95, \tag{4.10}$$

$$\mu ID_{sub} > 0.95, \tag{4.11}$$

$$\mu ID_{third} > 0.5, \tag{4.12}$$

we obtain a relative (absolute) signal efficiency of 97% (32%) with a background efficiency of 61%. These efficiencies are obtained by applying LID weights to adjust the simulation distribution and improve the agreement between data and simulation. Each muon in the $\tau^- \to \mu^-\mu^+\mu^-$ final state is assigned a weight $w_{LID,\ell i}$ based on whether it is real or fake (misidentified pion or kaon). The event weight w_{LID} is given as:

$$w_{LID} = \prod_{i=1}^{3} w_{LID,\ell i}. \tag{4.13}$$

Despite the significant degradation of background suppression caused by the loose selection on the third LID, the signal retention gain remains commendable. The BDT-based background suppression copes with this background increase.

Although the $\tau^-\tau^+$ and $q\overline{q}$ backgrounds are accurately simulated, there exist some discrepancies with low multiplicity backgrounds with respect to the data. In particular, certain di-photon backgrounds with energy along the beam direction are not properly simulated, resulting in peaking data at the edges of the polar angle of the missing momentum edges θ^{CM}_{mis} and at high thrust axis magnitude (as depicted in Figure 4.11). To eliminate contamination from these backgrounds, we have implemented a selection criterion of $0.3 < \theta^{CM}_{mis} < 2.7$.

The BDT classifier is designed to reject $\tau^-\tau^+$ and $q\overline{q}$ components. Hence, it is crucial to reduce the low-multiplicity backgrounds before training. To reject those low-multiplicity backgrounds, some cut-based preliminary selections are defined. We searched for discriminant variables with peaking low-multiplicity structures as illustrated in Figure 4.10. The selections are determined to eliminate any peaks present in the background distributions.

The discriminating variables are grouped to form several preselection combinations.

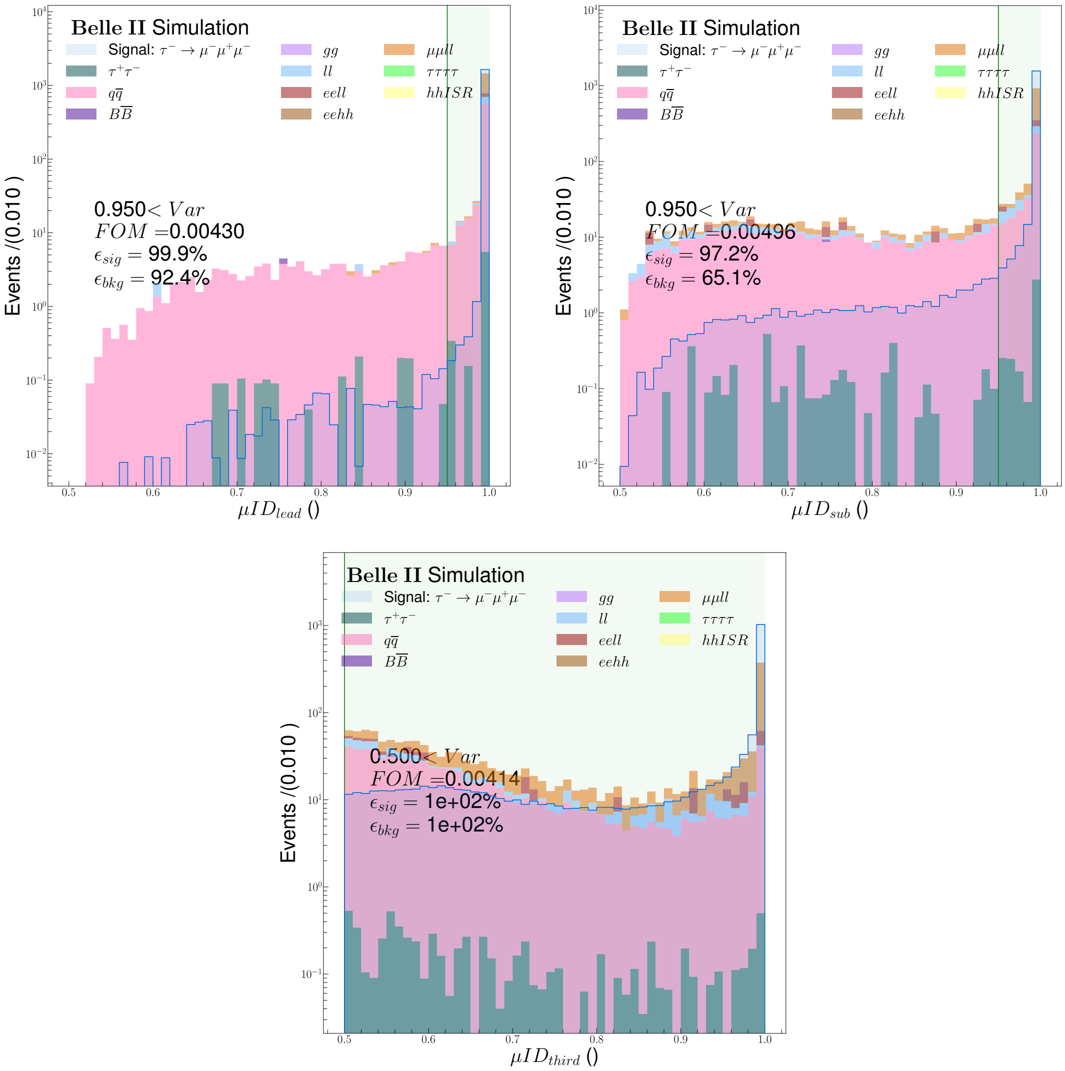


Figure 4.9. – Distributions of the three ranked leptonID in the signal side of the $\tau^- \to \mu^-\mu^+\mu^-$ MC signal and background sources (left: leading LID, right: subleading LID, bottom: third LID). The green area represents a cut at 0.95.

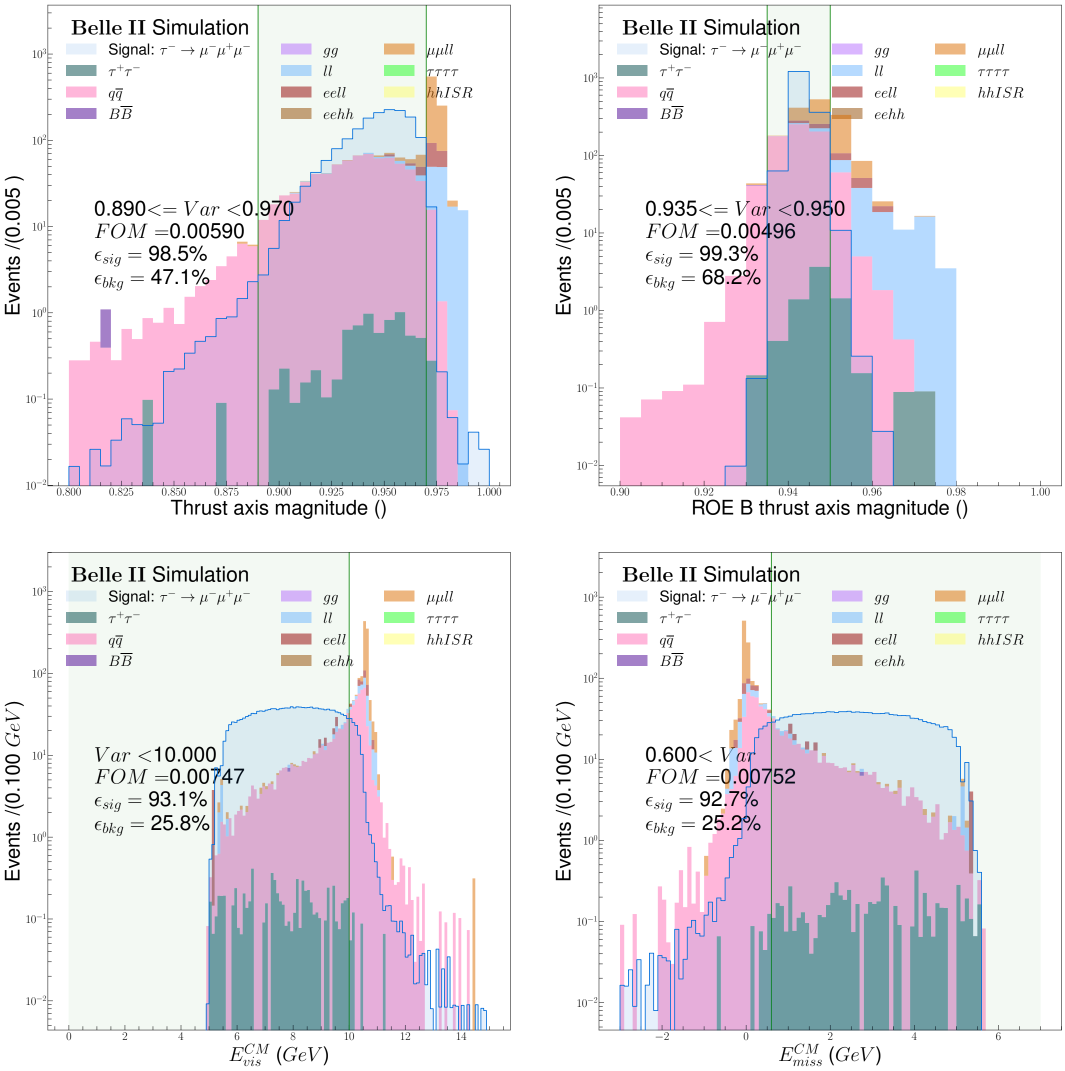


Figure 4.10. – Distribution of the sets of discriminated variables used to define several preliminary selections aiming to remove the low-multiplicity backgrounds. The green area represents the kept region after selection.

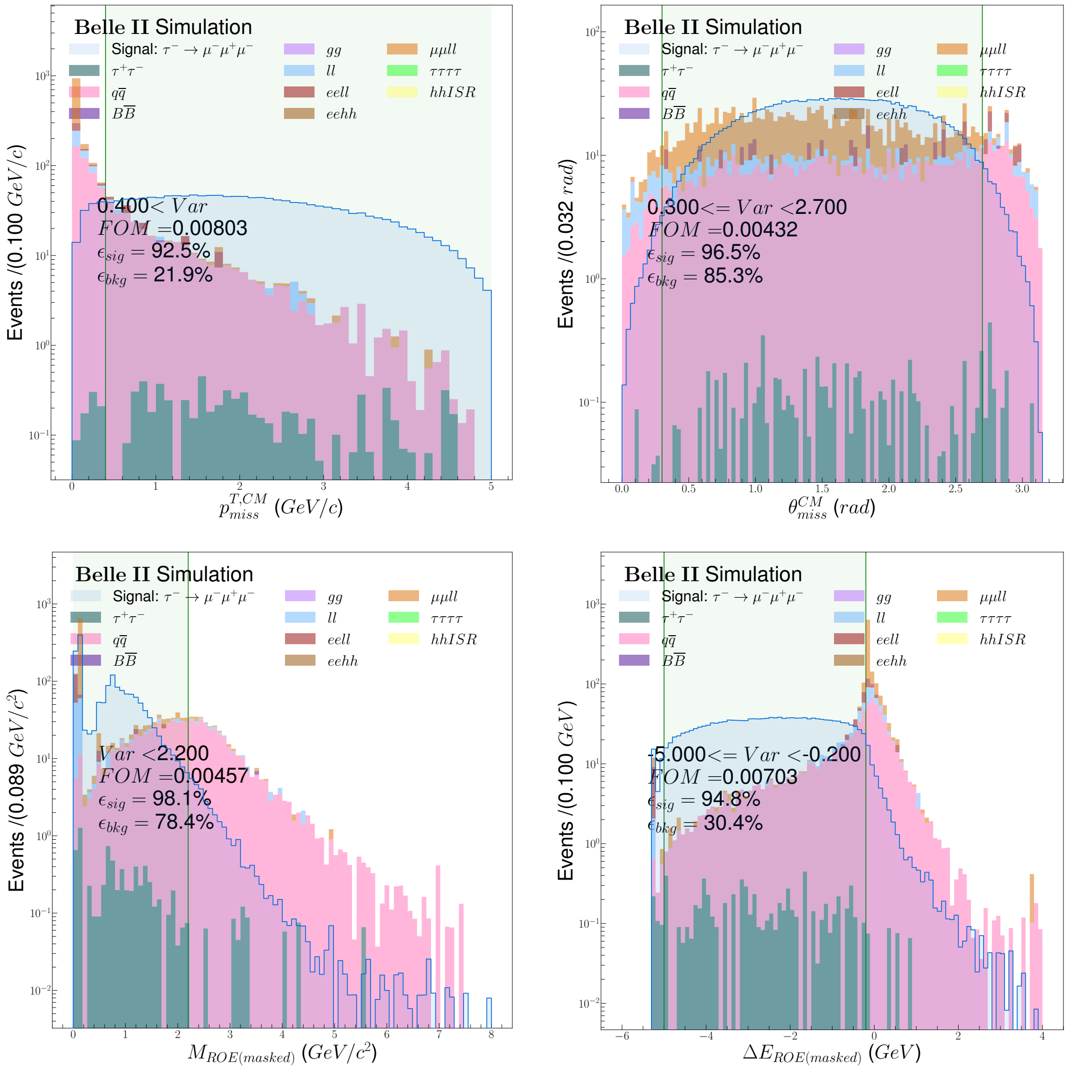
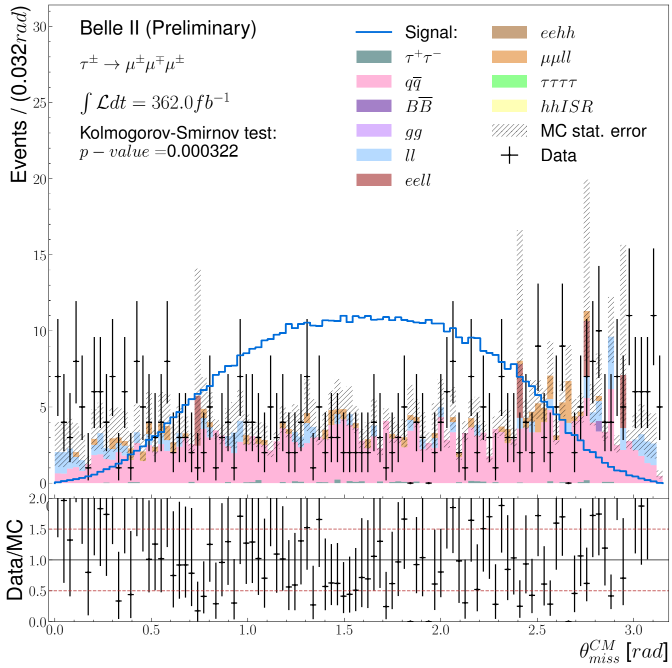


Figure 4.10. – Distribution of the sets of discriminated variables used to define several preliminary selections aiming to remove the low-multiplicity backgrounds. The green area represents the kept region after selection.



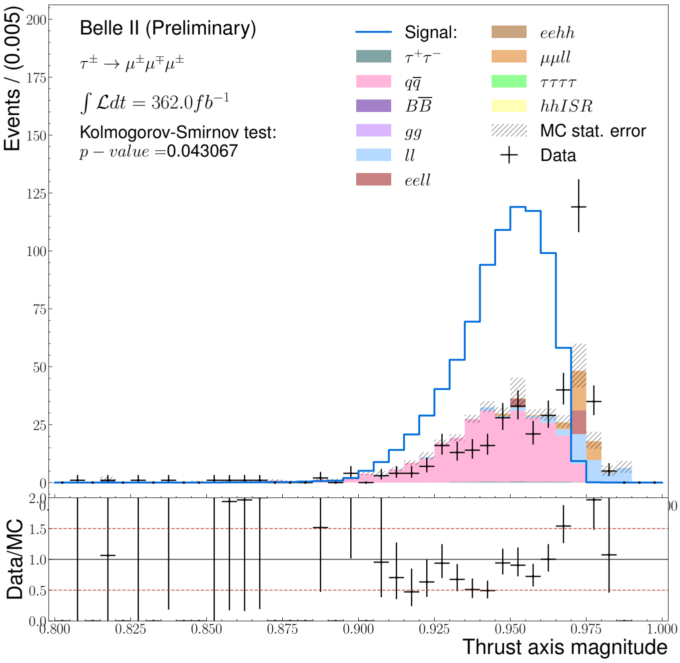


Figure 4.11. – Comparison between the data and simulation for the polar angle of the missing momentum (top) and thrust axis magnitude (bottom) for events in the sideband region.

The signal efficiency and background rejection for each combination are shown in Table 4.4, and detailed background composition is listed in Appendix C. The final set of preselection is chosen to maximise the Punzi figure of merit computed after applying the BDT classifier, with the hyper-parameters optimised and trained for a given set of preselection, as discussed in Section 4.3.3.

Table 4.4. – Efficiencies and background yields for several sets of preselection. The number of background events is scaled to 362 fb^{-1} retained in the 20δ region after applying the `muonID` requirement.

Name	Preselection	ε^{rel}_{sig} (%)	ε^{abs}_{sig} (%)	ε^{rel}_{bkg} (%)	N_{bkg}	$N_{\tau-pair}$	$N_{q\bar{q}}$	$N_{B\bar{B}}$	N_{lowm}
Reference	$0.3 < \theta^{CM}_{miss} < 2.7$	96.88	31.11	89.99	938.82	3.08	287.52	0	648.22
Set 1	$0.3 < \theta^{CM}_{miss} < 2.7$ $0.89 < Thrust < 0.97$	95.48	30.67	30.83	321.64	2.96	270.87	0	47.82
Set 2	$0.3 < \theta^{CM}_{miss} < 2.7$ $0.935 < Thrust_{\tau ROE}$ $0.95 > Thrust_{\tau ROE}$	96.35	30.94	61.78	644.50	2.58	244.94	0	396.97
Set 3	$0.3 < \theta^{CM}_{miss} < 2.7$ $E^{CM}_{vis} < 10.$	90.54	29.08	14.89	155.30	2.98	127.81	0	24.52
Set 4	$0.3 < \theta^{CM}_{miss} < 2.7$ $E^{CM}_{miss} > 0.6$	90.22	28.98	14.69	153.29	2.91	125.85	0	24.52
Set 5	$0.3 < \theta^{CM}_{miss} < 2.7$ $p^{T,CM}_{miss} > 0.4$	91.12	29.26	15.89	165.74	2.77	135.08	0	27.90
Set 6	$0.3 < \theta^{CM}_{miss} < 2.7$ $M_{ROE} < 2.2$ $-5. < \Delta E_{ROE} < -0.2$	90.76	29.15	16.49	172.08	2.62	106.08	0	63.39

In addition, in the last preselection combination, we attempted to enhance the learning of our BDT model by applying a preliminary selection on the most BDT effective variables, such as the mass of the ROE M_{ROE} and the difference in energy between the ROE and beam ΔE_{ROE}. The aim of such pre-selection is to allow the BDT algorithm to learn more about the less-used input variables.

4.3.2. Boosted Decision Tree classifier

The BDT selection aims at removing the remaining background events. It is trained on simulated signal sample, with 176,376 events, and background samples made of $\tau^-\tau^+$ and $q\bar{q}$ with respectively 45 and 3,399 events. The BDT used in this analysis is based on the `XGBoost` library [18, 117].

4.3.2.1. Boosted Decision Tree algorithm

BDT classifiers [118] are a type of machine learning algorithm that utilizes multiple decision trees to create a highly accurate predictive model. This particular algorithm is a form of gradient-boosting, where weaker learners (decision trees) are sequentially trained to enhance the overall performance of the model. Boosted decision tree classifiers are especially useful for handling complex datasets and generating precise predictions. They are commonly applied in a variety of machine-learning tasks, such as regression and classification.

The decision trees are an advanced version of the cut-based approach. They use a multivariate technique to analyze events meeting specific criteria for either signal or background. This is done through a binary tree structure, as shown in Figure 4.12, where nodes are split following the sequence:

1. The node receives all input variables and events.
2. The optimal separation between signal and backgrounds for every variable is searched.
3. The variable with the highest separation score is retained.
4. Events are split into two classes: events fulfilling the selection of the best separation variable on one side and events failling it on the other.
5. The two classes are sent to two different daughter nodes.

The binary tree is created through a recursive process of splitting nodes using the same method. Stopping parameters are introduced to stop the node splitting when one of them is satisfied. Nodes at the ends of the tree are called leaves where a weight, such as purity $s/(s+b)$, is computed to measure the similarity of the events in the leaf to the signal.

The boosting consists of combining multiple small decision trees, resulting in a stable classifier. This is done by recursively creating decision trees on the residual of the previous iteration. For a given event, the prediction for an event i is expressed as the sum of the K decision tree leaf weights.

For a dataset $\mathscr{D} = \{(x_i, y_i)\}$, where x_i is the input variables vector and y_i is the target [3] the decision tree prediction $\hat{y}_i$ is given by:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \tag{4.14}$$

where each f_k takes into count the structure of the k^{th} tree and the leaf weights w. The training of the boosted decision tree is done by minimizing the objective function:

$$\mathscr{L} = \sum_i l\big(\hat{y}_i, y_i\big) + \sum_k \Omega(f_k), \tag{4.15}$$

where $l\big(\hat{y}_i, y_i\big)$ is a loss function that measures the difference between the prediction and the target. In our case, we use a logarithmic loss function. $\Omega(f_k)$ is a regularisation function that penalises the complexity of the model. In the used XGBoost library, the

---

[3] In our case the target function is $y_i = 0$ if the event is simulated background and 1 if it is simulated signal

minimisation of the objective function is done using an extreme gradient boosting algorithm.

4.3.2.2. k-folding algorithm

Training a BDT for the $\tau^- \to \mu^-\mu^+\mu^-$ channel is a challenge due to the low background level of this channel. To still be able to use a BDT classifier in the background rejection strategy despite the low background statistic, we work on two different aspects. The first was to improve the provided simulated background statistic by requiring additional samples with the `TauLFV` skim applied at the generator level. This increases the statistic from 1 ab^{-1} to 8 ab^{-1} for the dominant background sources.

In addition to the statistical increase, the BDT algorithm was completed by a method to save the statistics during the sample splitting. Indeed, usually, with machine learning methods, a dataset $\mathscr{L}$ is split into three mutually exclusive subsamples. The first subsample $\mathscr{L}_{\texttt{Train}}$ called `Train`, is used for the BDT classifier training. The second $\mathscr{L}_{\texttt{Validation}}$, called `Validation`, is used to evaluate the performances of the BDT and take care of potential overtraining or overfitting. The final sample $\mathscr{L}_{\texttt{Test}}$ `Test` is reserved throughout the optimisation procedure to provide a clean subsample to extract the final yields and efficiencies of the signal/background classification. A sufficient statistic is thus needed to have three large enough samples.

In this study, we increase the statistics in the train and test samples by employing the K-Folding algorithm for cross-validation [20]. The k-fold algorithm allows us to use the

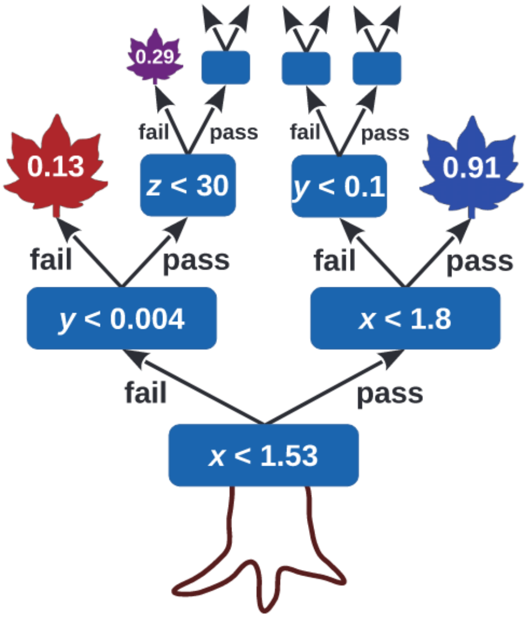


Figure 4.12. – Diagram of a binary decision tree. The blue rectangles represent the internal nodes and their associated splitting selection. The leaves are the terminal nodes and indicate their purity.

same sample in both the training and validation phases. In our strategy illustrated in Figure 4.13, the full simulation dataset (8 ab^{-1}) is split in two: a `Train` $\mathscr{L}_{\texttt{Train}}$ sample of 4 ab^{-1} for the optimisation, the training, and checks the BDT performances, and a `Test` $\mathscr{L}_{\texttt{Test}}$ sample of 4 ab^{-1} to extract the final yields. During the optimisation of the BDT hyperparameters and the training of the BDT, the `Train` dataset is divided into K equally sized folds $\mathscr{L}_{\text{k}}$: so $\mathscr{L}_{\texttt{Train}} = \bigcup_{k=1,K} \mathscr{L}_{\text{k}}$. The first BDT classifier T_1 is trained on $\bigcup_{k=2,K} \mathscr{L}_{\text{k}}$ while $\mathscr{L}_1$ is used to validate the classifier T_k. After training one classifier, the roles of the folds are changed, and a new BDT $T_{k'}$ is trained on another permutation of $k-1$ folds and validated on the remaining $\mathscr{L}_{k'}$ sample. This procedure is repeated until all K fold combinations are used. It results in K BDT classifiers. Each one applied on its independent validation sample in order to have an independent dataset. When applying the BDTs to the test and data samples, the K BDT outputs are averaged since, by definition, those samples were not used in the training. The cross-validation allows having a larger training sample since each BDT is trained on $k-1$ folds, which minimizes possible overtraining issues. In addition, this also allows us to keep a bigger independent test sample. This technique has been used in HEP for years, see e.g. [119], and is supported in some machine learning software, such as TMVA. That is not the case in the XGBoost library. We thus had to implement it.

In this study, the MC dataset is split into two `Train` and `Test` samples with the equal statistic of 4 ab^{-1} for the main background contribution. The training sample is split into $K = 10$ folds to perform the cross-validation and $K = 4$ for the BDT

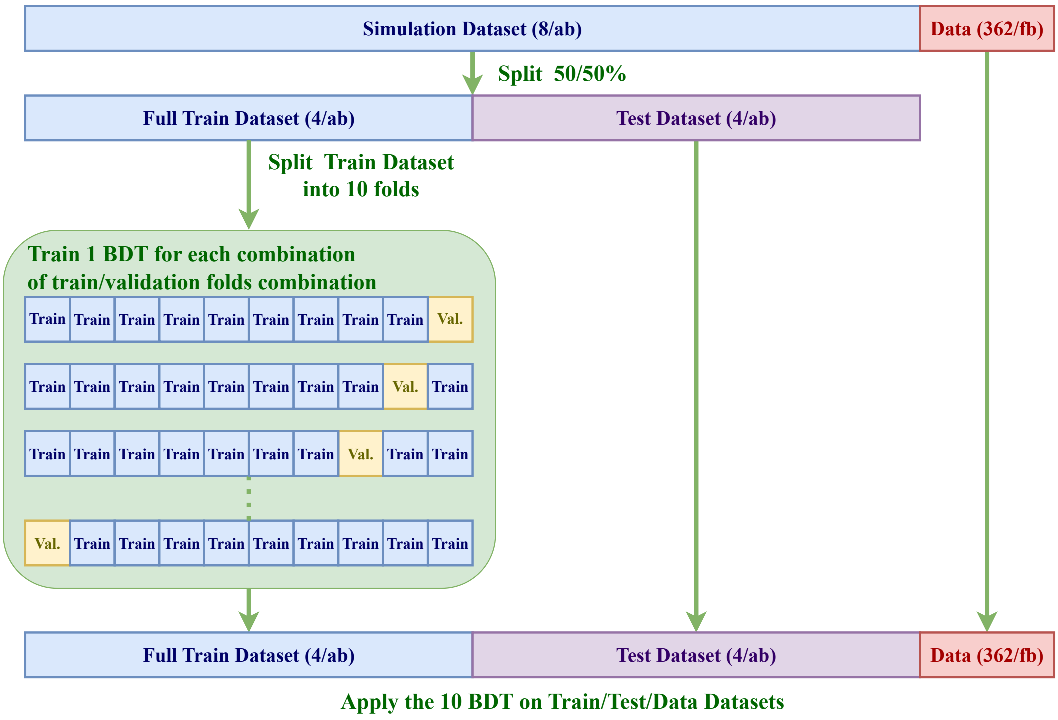


Figure 4.13. – Scheme of the `Train` sample split into K folds during the training procedure.

hyperparameters optimisation.

4.3.2.3. Discriminating variables for the BDT classifier

The set of variables and their definitions used as input features for the BDT classifier are listed in Table 4.5, the BDT importance of each input variable is shown in Appendix D. A comparison of the signal and background simulated distributions is given in Figure 4.3.2.3.

Table 4.5. – List of all the signal/background discriminating variables used as input for BDT classifier algorithm.

Related to	Name	Definition	Unit
Events	$thrust$	Thrust axis magnitude	
	-	Number of good tracks	
	-	Number of good photons	
	E_{γ}^{CM}	Sum of all photons energy in the laboratory frame	GeV
	$q^{\tau} \times q^{ROE}$	Product of the total charge of the signal tau side and the ROE side	
	$p_{miss}^{T,CM}$	Transverse momentum in the center-of-mass frame	GeV/c
	θ_{miss}^{CM}	Polar angle in the center-of-mass frame	rad
	$cos\theta_{miss-\ell i}$	Cosine of the angle between missing momentum and the muon i	
	M_{miss}^2	Invariant mass squared of the missing momentum	GeV^2/c^4
Signal side τ	$t^{\tau-flight}/t_{err}^{\tau-flight}$	τ flight time normalized by its error	
	$\theta_{\tau-closest}$	Angle between τ and closest track	
	$cos\theta_{p-vertex}$	Cosine of the angle between the τ momentum and the direction obtained from its production and decay vertices	
	-	χ^2 probability of τ vertex fit result	

4.3.2.4. BDT parameters optimisation

Several hyperparameters are optimised to provide the best separation and avoid overfitting from the BDT classifier. The python library `Optuna` [21] is used to scan and find the best configuration in the hyperparameter space. Two inputs are given to the library: an objective function representing the indicator to minimize or maximize and a phase space for the hyperparameters, listed in Table 4.6. The sampling of the

Table 4.5. – List of all the signal/background discriminating variables used as input for BDT classifier algorithm.

Related to	Name	Definition	Unit
Lepton	$\theta_{\ell i}^{CM}$	Polar angle in the center-of-mass frame	rad
	$E_{\ell i}^{CM}$	Energy in the center-of-mass frame	GeV
	$p_{lead}^{T,CM}$	Transverse momentum of the leading muons in the centre-of-mass frame	GeV/c
	$p_{sub}^{T,CM}$	Transverse momentum of the second muons in the centre-of-mass frame	GeV/c
	$p_{third}^{T,CM}$	Transverse momentum of the third muons in the centre-of-mass frame	GeV/c
Rest Of Event	M_{ROE}	Mass of the ROE	GeV/c^2
	ΔE_{ROE}	Difference between the energy of the ROE and half the centre of mass energy	GeV
	-	Number of electrons in ROE	
	-	Number of muons in ROE	
	-	Number of pions in ROE	
	-	ROE thrust axis magnitude	
	$cos\theta_{ROE\tau-ROE}^{thrust}$	Cosine of angle between thrust axis of the signal τ and thrust axis of ROE	
	$cos\theta_{ROE\tau-z}^{thrust}$	Cosine of angle between thrust axis of the signal τ and z axis	

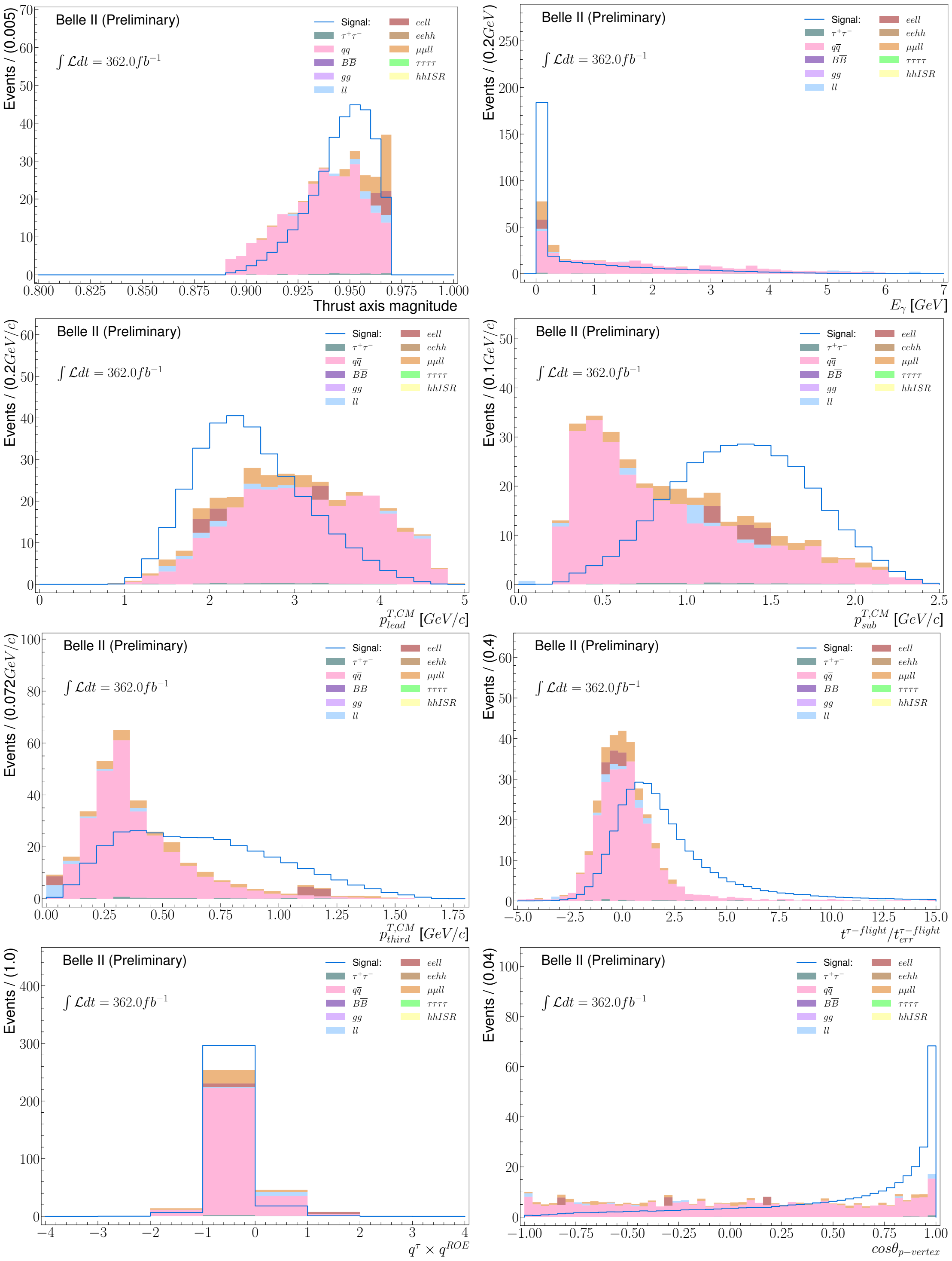


Figure 4.14. – Signal and background comparison in simulation in the 20δ SR of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for discriminant variables taken as inputs to the BDT, after the preselection.

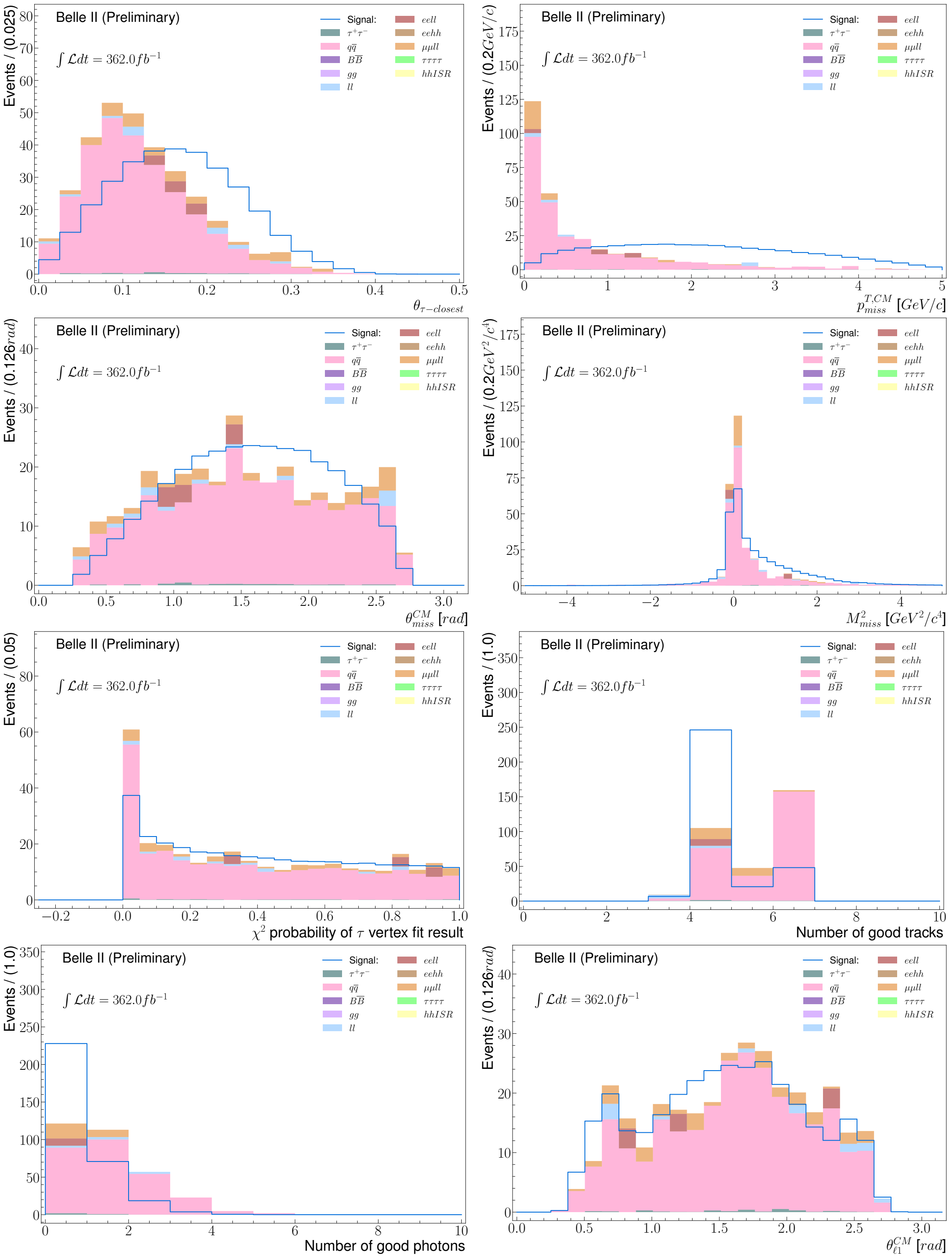


Figure 4.14. – Signal and background comparison in simulation in the 20δ SR of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for discriminant variables taken as inputs to the BDT, after the preselection.

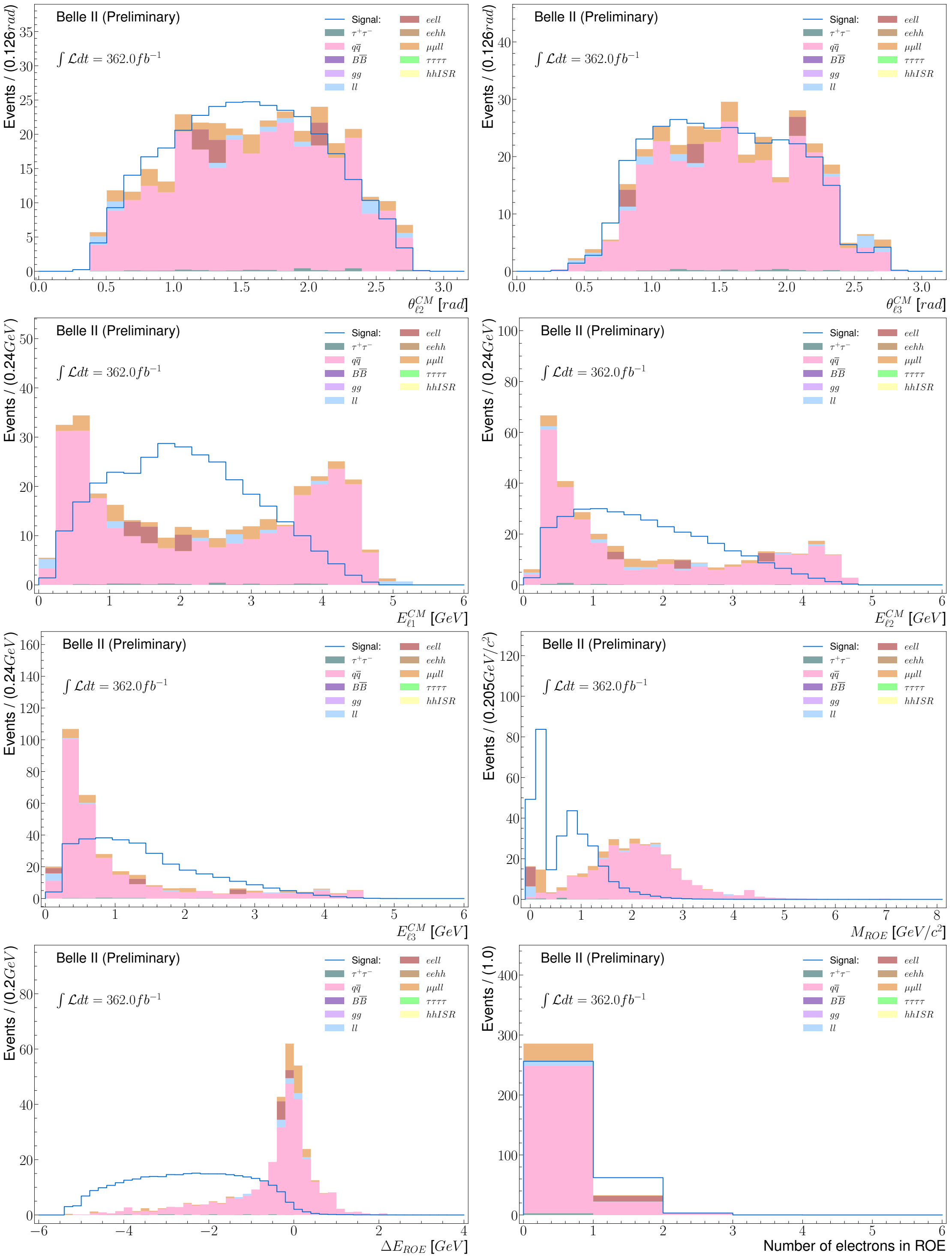


Figure 4.14. – Signal and background comparison in simulation in the 20δ SR of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for discriminant variables taken as inputs to the BDT, after the preselection.

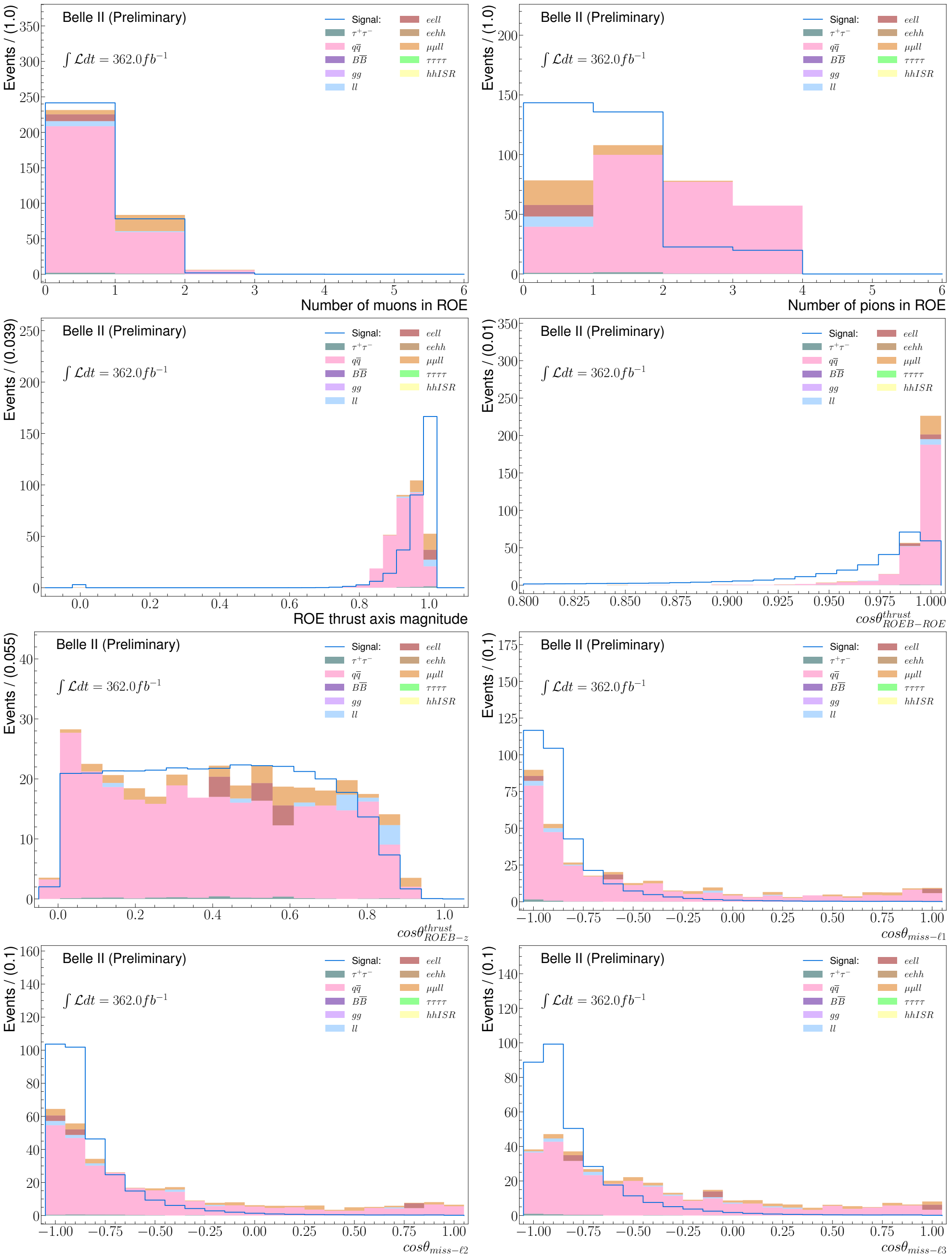


Figure 4.14. – Signal and background comparison in simulation in the 20δ SR of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for discriminant variables taken as inputs to the BDT, after the preselection.

hyperparameters is done with a Tree-structured Parzen Estimator [120] implemented in the library, and it provides an efficient independent sampling.

Table 4.6. – Definition and range value of the `XGBoost` BDT classifier hyperparameters optimised to extract the best performances and avoid overfitting. The range value for each parameter is chosen to be around the default value.

Name	Definition	Scanned Values	Step	Default
`n_estimators`	Number of boosting rounds.	$[100,500]$	50	-
`max_depths`	Maximum depth of a tree, number of intermediate nodes.	$[1,9]$	1	6
`min_child_weight`	Minimum sum of instance weight needed in a child node.	$[0,10]$	1	1
`learning_rate`	Step size shrinkage used after each boosting update.	$[0.01,0.5]$	0.01	0.3
`gamma`	Minimum loss reduction required to make a further partition on a leaf node of the tree	$[0,10]$	1	0

To avoid overfitting, the definition of the objective function is crucial. Here, the choice was made to minimize the last iteration of the logarithmic loss function defined as:

$$L_{log}(y,p^{BDT}) = -(y * log(p^{BDT}) + (1-y) * (log(1-p^{BDT}))), \tag{4.16}$$

where y is the input target (0 for background and 1 for signal), and p^{BDT} is the output probability of being signal. Doing this minimization in validation folds will automatically provide a performant separation and avoid overfitting as the logarithmic loss function evaluates the goodness of the predicted probability with the target one. The scan of all the parameters for the $\tau^- \to \mu^-\mu^+\mu^-$ channel is given in Figure 4.15, where the best parameters are linked to a high learning rate with a low number of epochs.

4.3.3. Background rejection results

4.3.3.1. Optimisation results

BDT classifiers are optimised and trained for each set of preselection presented in Section 4.3.1. In each configuration, the rectangular and ellipsoidal signal regions are tried. The absolute signal efficiency, retained number of background events and Punzi FOM for each combination of preselection and BDT are shown in Figure 4.16. The best background rejection configuration is chosen according to the Punzi figure

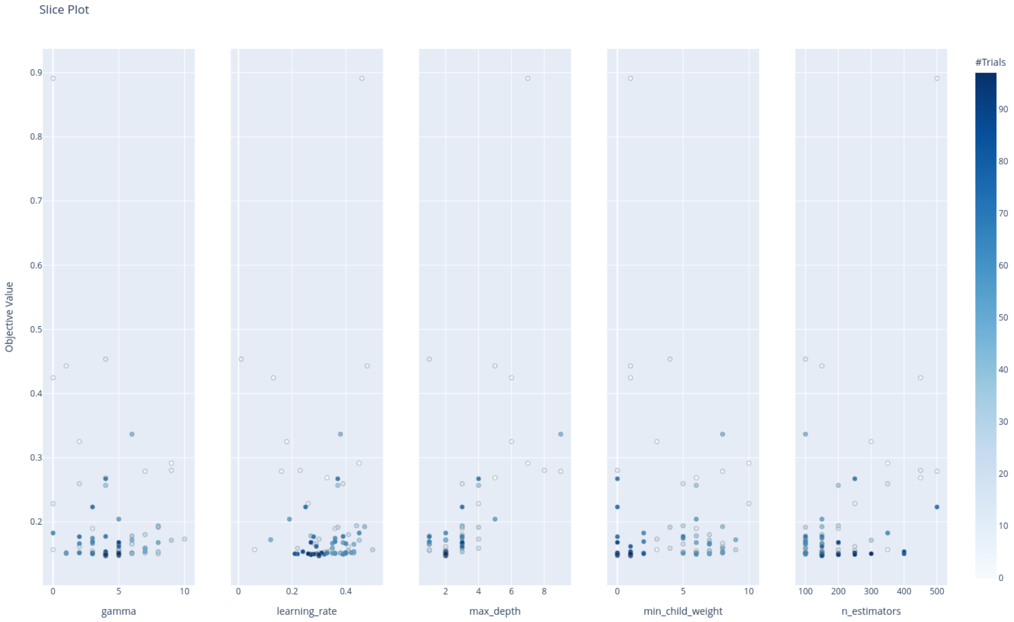


Figure 4.15. – Two-dimensional plots of the `Optuna` objective value, here logarithmic loss function of the last validation epoch, versus the values of the hyperparameters during the BDT optimisation for $\tau^- \to \mu^-\mu^+\mu^-$. At each `Optuna` trial, the algorithm chooses a set of parameters in the available ranges and trains a BDT. The logarithmic loss function is extracted from each model. A single point in each plot represents each trial. The best hyperparameter set is the lower point.

of merit evaluated on the training sample[4]. The reference and set 1 performance are very close, while the other preselection sets give worse results. Performances of rectangular and elliptical regions are very similar. The elliptical region is preferred as it has a smaller area and will thus be less polluted by background events. Table 4.7 lists the final selection cuts.

Table 4.7. – Final background rejection selection after optimizing the Punzi FOM on the training sample.

$\tau^- \to \mu^-\mu^+\mu^-$
$muID_{lead} > 0.95$
$muID_{sub} > 0.95$
$muID_{third} > 0.5$
$0.3 < \theta^{CM}_{miss} < 2.7$
$0.89 < thrust < 0.97$
$p^{BDT} > 0.89$
3δ ellipse region

The signal and background probability distributions of the BDT output are compared for the train and validation (test) samples in the left (right) plot of Figure 4.17. As stated earlier, the requirement on the BDT output is optimized on the validation folds by maximizing the Punzi figure of merit in the 5δ SR, as illustrated by the lower plot in Figure 4.17 (left). Optimizing in the validation folds allows for keeping an independent test sample to evaluate the performances. The corresponding signal efficiency and expected background events before and after the BDT selection for different regions of the 2D plane are given in Table 4.8.

The final absolute signal efficiency is $19.70^{+0.06}_{-0.06} stat\%$ for $0.08^{+0.21}_{-0.07} stat$ remaining background events. The numbers of background events are obtained from the 4 ab^{-1} test sample scaled down to 362 fb^{-1}.

4.3.3.2. Background composition after preselection and BDT

The composition of the remaining background is analyzed using the 4 ab^{-1} of simulated MC test sample. As demonstrated by Table 4.9, in the signal region after applying preselections, the main remaining background is coming from $q\overline{q}$ samples where the pions and kaons present in the final state are misidentified as muons. Those pions and kaons backgrounds are coming directly from the IP. The other surviving background events fall outside the signal region, as shown in Figure 4.18. We can easily distinguish the distribution of $\tau^-\tau^+$ background events at low $M_{3\mu}$ and low $\Delta E_{3\mu}$ due to the invisible energy carried by neutrinos. The remaining low multiplicity backgrounds are well located outside the 20δ SR used for the background rejection optimisation. When the selection on the BDT classifier output is added, the only surviving event in the test

---

[4] as explained in Sec. 4.3.2, the training sample is split into sub-train and sub-validation samples. In contrast, the final test sample remains independent until the end of the optimization process.

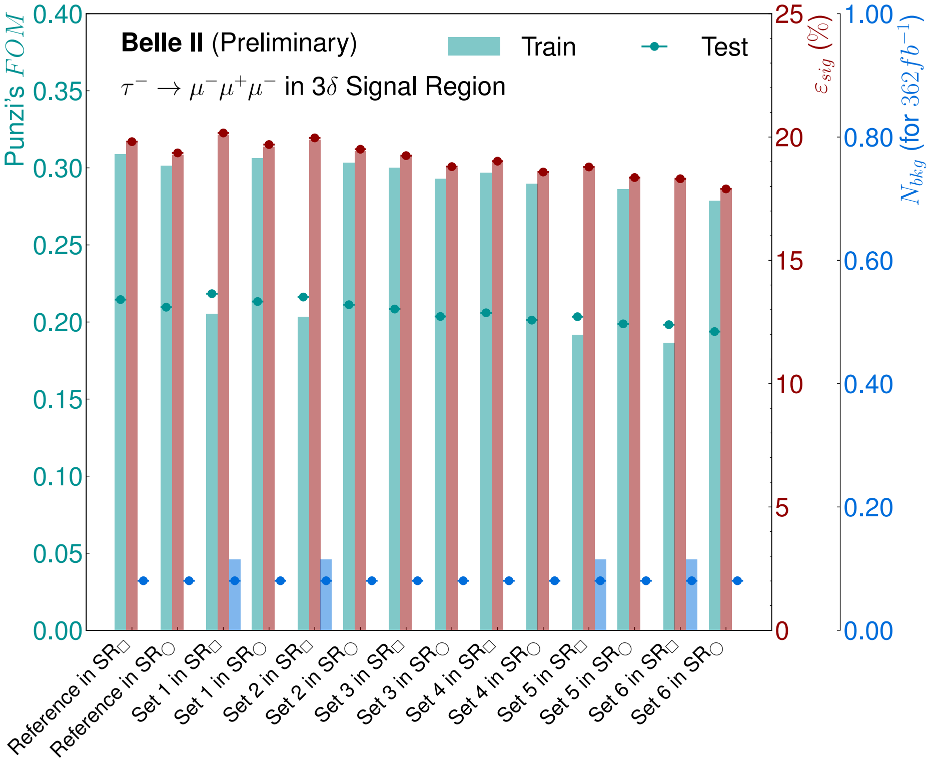


Figure 4.16. – Final background rejection performances, number of surviving backgrounds (blue), absolute signal efficiency (red) and the Punzi's figure of merit (cyan) for the `Train` and `Test` sample for the different preselection and signal region (3δ rectangle or ellipse) setups. ∘ label is for the configuration evaluated in the elliptical signal region, while □ is for the rectangular box one.

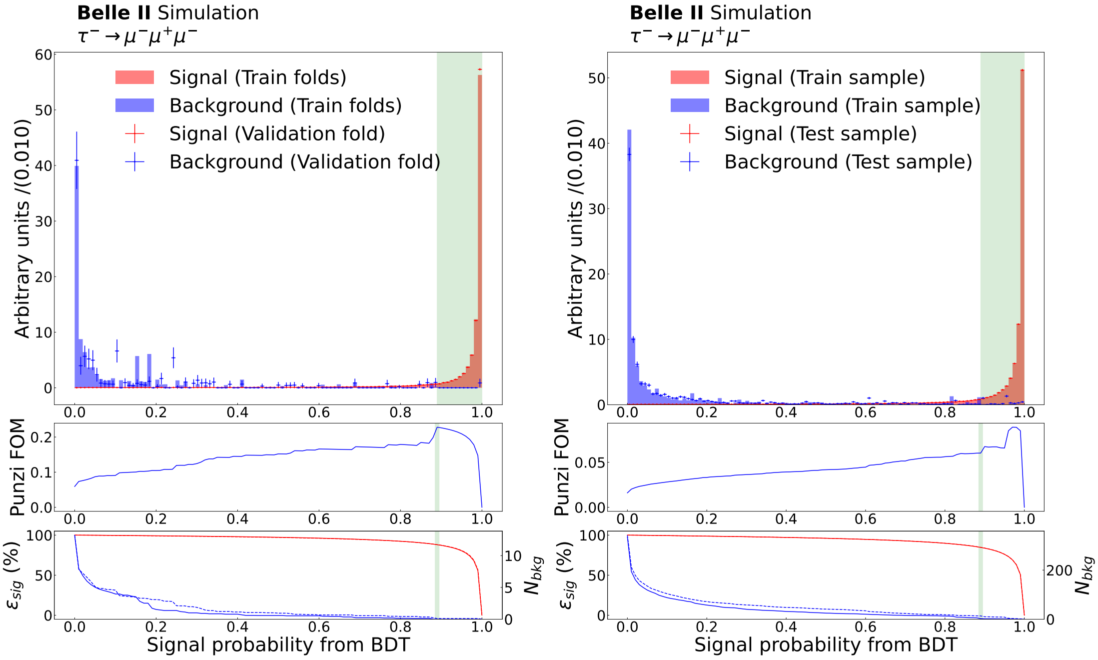


Figure 4.17. – Averaged BDT classifier probability output of being signal for `Train` and `Validation` (left) or `Test` (right) samples. In the middle plot, the row represents the Punzi figure of merit, evaluated on the `Validation`/`Test` sample as a function of a selection on the BDT output. The bottom plot represents the absolute signal efficiency (red) and remaining background (blue) scaled to 362 fb^{-1} for `Train` (solid line) and `Validation`/`Test` (dashed lines). The green area in both plots corresponds to values selected as a result of the FOM optimisation in 5δ Signal Region.

sample is a $q\overline{q}$ event with misidentified pions and kaons in the final state, presented in Table 4.10.

4.3.4. Background rejection validation

4.3.4.1. BDT overfitting monitiring

The BDT is prevented from overfitting by monitoring the logarithmic loss function behaviour for train and validation samples during the boosting in Figure 4.19. The tendency of the validation curve shows if the BDT is overfitting or not. Indeed, the curve decreasing to a floor without increasing indicates the absence of overfitting. Indeed, the overfitting behaviour is traduced by an increasing logarithmic loss function in the validation sample while the train is flat [118]. The divergence from the test curve is not worrisome as long as the two curves have the same general behaviour.

4.3.4.2. BDT peaking structure monitoring

In Figure 4.20, we show the profile plots of the BDT versus the $M_{3\mu}$ and $\Delta E_{3\mu}$ variables for simulated background events. No peaking structure is visible.

4.3.4.3. Data/Simulation comparison in sidebands

The consistency between data and simulation for the variables entering the BDT classifier is checked in the sidebands region, defined by the rectangle of $\pm 20\delta$ along $M_{3\mu}$ and $\pm 10\delta$ in $\Delta E_{3\mu}$, as shown in Figure 4.21. The distributions for $M_{3\mu}$, $\Delta E_{3\mu}$ and

Table 4.8. – $\tau^- \to \mu^-\mu^+\mu^-$ signal efficiencies, background yields weighted for a luminosity of 362 fb^{-1} and Punzi FOM for the train and test samples before and after applying the BDT selection.

Sample	Signal Region	Pre Selection on BDT		Post Selection on BDT		
		$\varepsilon^{abs}_{signal}$ (%)	N_{bkg}	$\varepsilon^{abs}_{signal}$ (%)	N_{bkg}	Punzi FOM
Train	20δ	30.67	333.96	25.96	2.43	0.11799
Test	20δ	30.69	341.49	26.03	13.55	0.06024
Train	5δ	25.32	13.19	22.29	0.11	0.22775
Test	5δ	25.40	13.23	22.38	0.17	0.21281
Train	Elliptical 5δ	24.82	10.48	21.88	0.11	0.22361
Test	Elliptical 5δ	24.88	10.35	21.96	0.16	0.21216
Train	3δ	22.64	4.35	20.09	0.11	0.20527
Test	3δ	22.72	4.30	20.16	0.08	0.21836
Train	Elliptical 3δ	22.08	3.80	19.61	0.00	0.30639
Test	Elliptical 3δ	22.16	3.77	19.70	0.08	0.21328

Table 4.9. – Simulation truth identification of final state particles for candidates passing the preliminary selection in the 3δ SR.

$PDG^{MC}_{\ell 1}$	$PDG^{MC}_{\ell 2}$	$PDG^{MC}_{\ell 3}$	type	sample	occurrence
π	π	π	$q\overline{q}$	$u\overline{u}$	19
π	π	π	$q\overline{q}$	ddbar	7
K	π	π	$q\overline{q}$	$s\overline{s}$	3
K	π	π	$q\overline{q}$	$u\overline{u}$	2
π	π	π	$q\overline{q}$	$s\overline{s}$	2
π	K	π	$q\overline{q}$	$c\overline{c}$	2
μ	π	π	$q\overline{q}$	$u\overline{u}$	1
π	K	K	$q\overline{q}$	$u\overline{u}$	1
K	π	π	$q\overline{q}$	$d\overline{d}$	1
π	NaN	π	$q\overline{q}$	$d\overline{d}$	1
π	π	NaN	$q\overline{q}$	$u\overline{u}$	1
π	K	π	$q\overline{q}$	$s\overline{s}$	1
μ	NaN	π	$q\overline{q}$	$c\overline{c}$	1
π	π	K	$q\overline{q}$	$u\overline{u}$	1
π	π	μ	$q\overline{q}$	$u\overline{u}$	1
π	μ	π	$q\overline{q}$	$d\overline{d}$	1
K	K	K	$q\overline{q}$	$s\overline{s}$	1

Table 4.10. – Simulation truth identification of final state particles for candidates passing the BDT selection in the 3δ SR.

$PDG^{MC}_{\ell 1}$	$PDG^{MC}_{\ell 2}$	$PDG^{MC}_{\ell 3}$	type	sample	occurrence
K	π	π	$q\overline{q}$	$s\overline{s}$	1

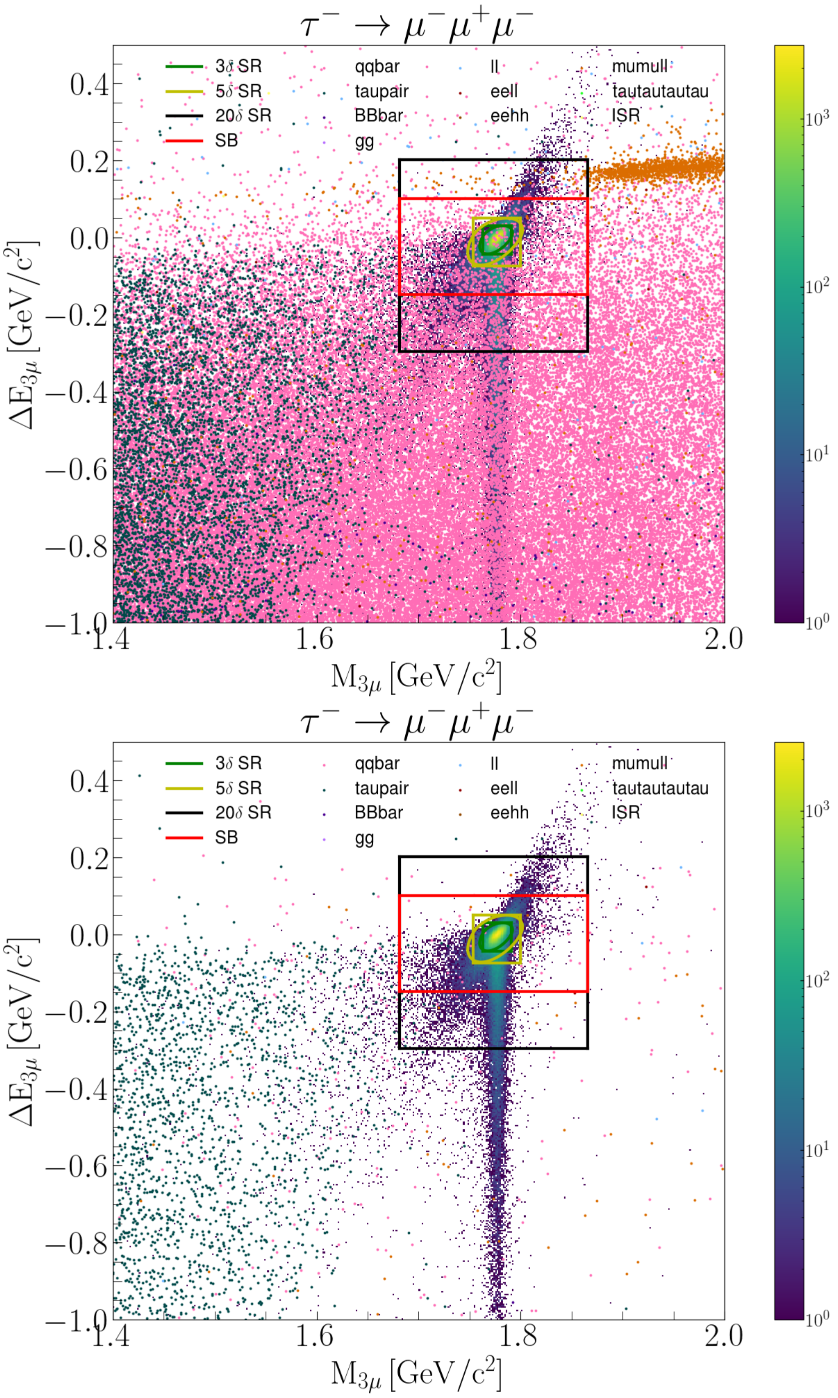


Figure 4.18. – Scatter plot of the different simulated background samples in coloured dots and the signal in density map in the plan ($M_{3\mu}$, $\Delta E_{3\mu}$). The two-dimensional distribution of the backgrounds is given at different steps of the background rejection: after applying the nominal pre-selection (top) and after applying the BDT classifier (bottom). The different definitions of the signal region are given as reference: 20δ SR (black box), SB (red box) and the 5δ and 3δ ellipse SR.

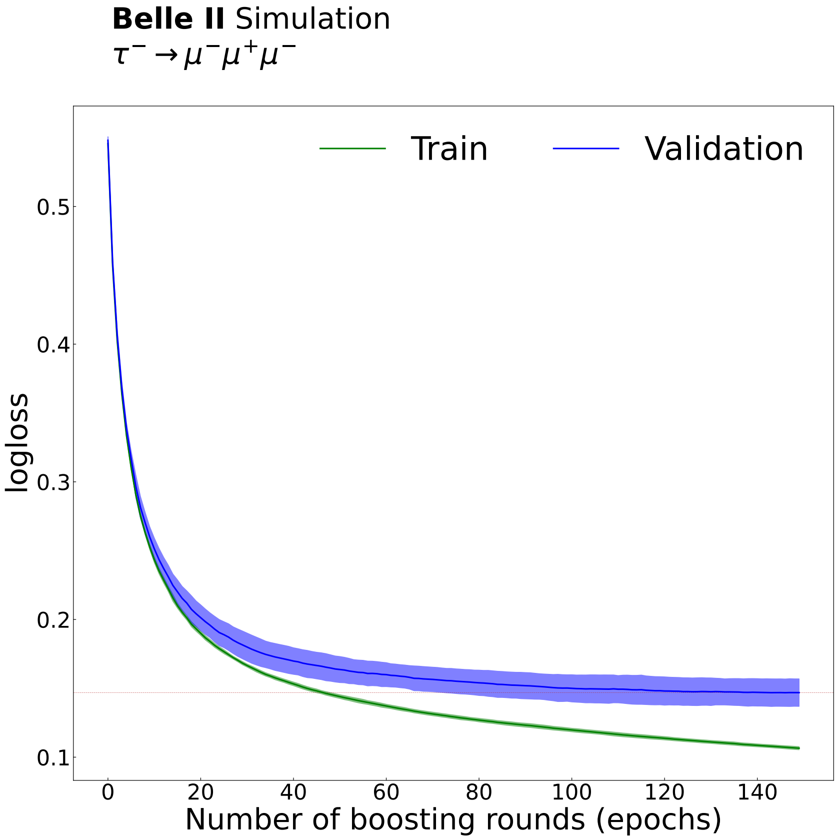


Figure 4.19. – Average (solid line) and fluctuations (filled area) of the logarithmic loss function as a function of the boosting rounds of the K BDT classifiers for the $\tau^- \to \mu^-\mu^+\mu^-$ channel. The quantity is estimated in both `Train` (green) and `Validation` (blue) samples to visualise potential overfitting effects.

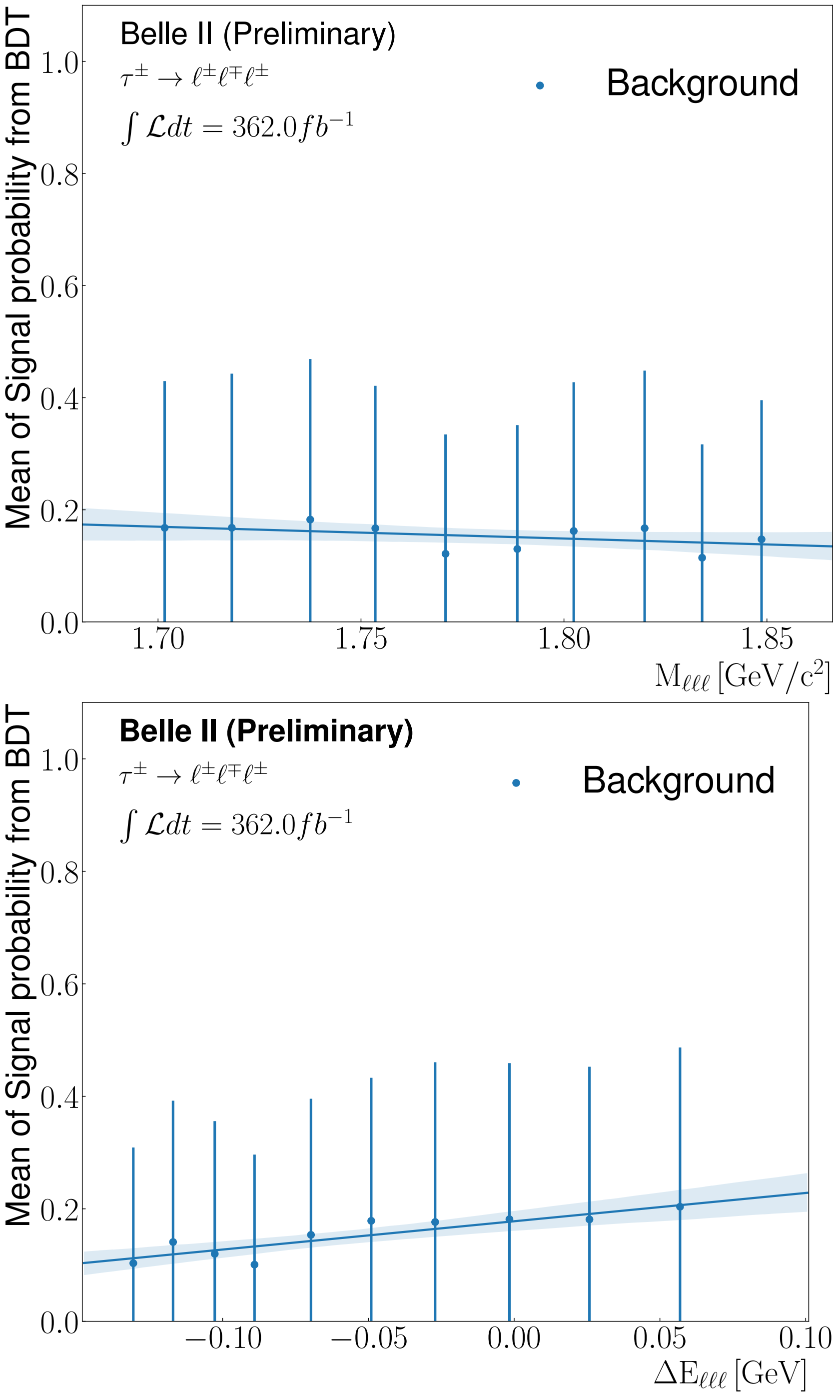


Figure 4.20. – Profile plot of the averaged signal probability BDT output function of the M_τ (left) and ΔE_τ (right).

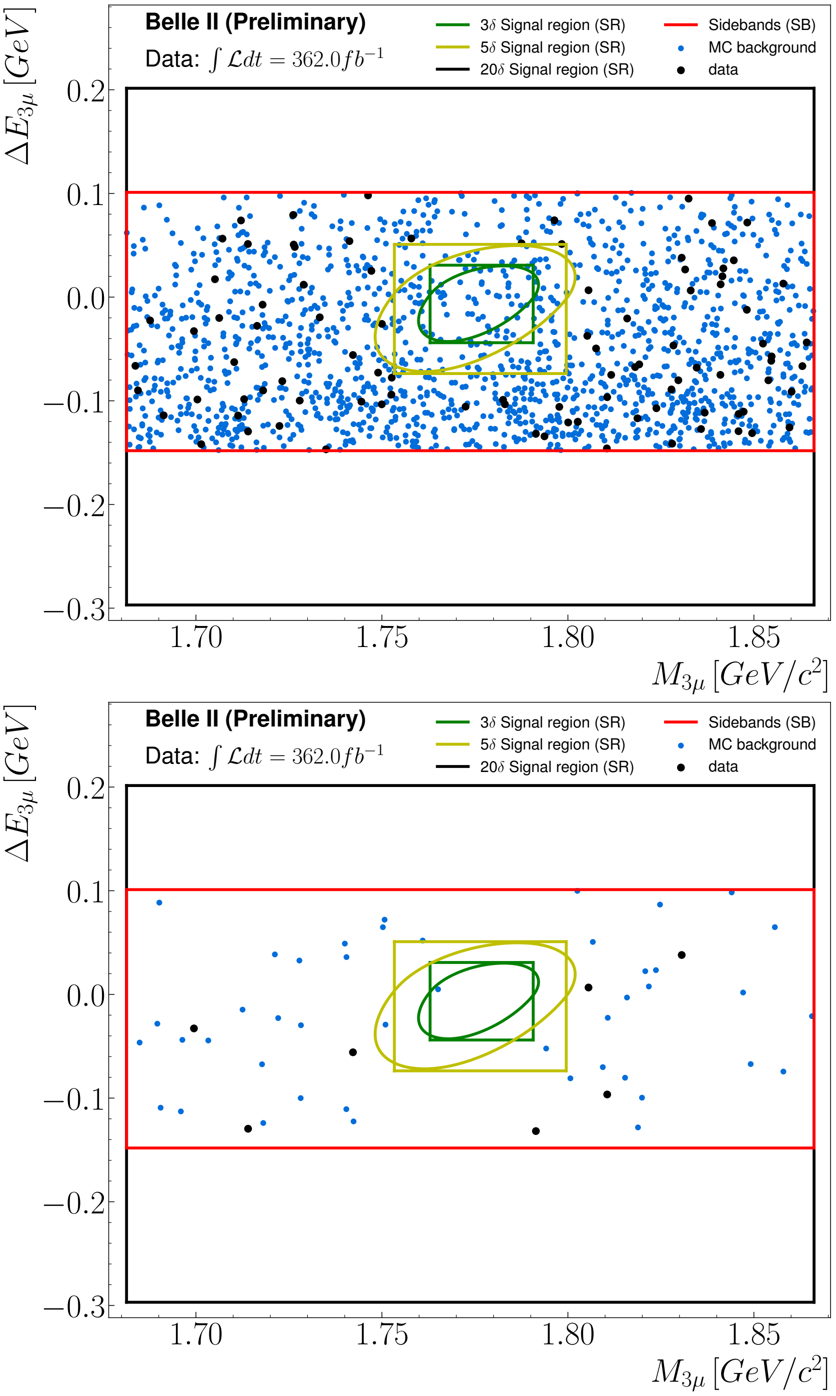


Figure 4.21. – Scatter plots (unweighted) of surviving events after preselection (top) and selection on BDT signal probability output (bottom), in the $(M_{3\mu}, \Delta E_{3\mu})$ sidebands region (red). The $\pm 5\delta$ box (yellow) is hidden for data to prevent unblinding. The luminosity for background simulated samples is half of the one listed in Table 2.6.

the BDT output p_{BDT} after the optimal preselection are shown in Figure 4.22 (see Appendix F, Figure 41 for other variables) and after the full background rejection selection in Figure 4.23 (see Appendix F, Figure 42 for other variables). After the preselection, the number of events in the simulation is $118.1^{+9.8}_{-5.1} stat$ while it is $94.0^{+10.7}_{-9.7} stat$ in data. It is known that the $q\overline{q}$ simulated events are overrepresented with respect to the actual data cross-section. A correction factor must be applied to ensure agreement in blinded sidebands. The correction factor is known to be selection-dependent, and we compute it after applying the preselections by:

$$w_{q\overline{q}} = \left| \frac{N^{data} - N^{MC}_{notq\overline{q}}}{N^{MC}_{q\overline{q}}} \right|, \tag{4.17}$$

where $N^{data} = 94$ is the number of event in data sample, $N^{MC}_{notq\overline{q}} = 13.91$ the number of simulated events which are not $q\overline{q}$ and $N^{MC}_{q\overline{q}} = 104.24$ the number of $q\overline{q}$ simulated events. The correction factor is equal to 0.77, and is applied for the data/simulation comparison plots after the preselection. After the BDT cut, there are $3.3^{+1.2}_{-0.7} stat$ events in simulation for $7.0^{+3.8}_{-2.6} stat$ events in data, the statistics is not significant enough to compute a $q\bar{q}$ correction factor. In addition, since the expected background yield is obtained without relying on the simulation, as explained in the next section, the correction factor has no impact on the analysis.

4.3.5. Expected background yield

The simplest strategy for estimating the number of expected background events in the signal region is usually to extrapolate this number from the data sidebands (sidebands method), as was done in $\tau \to \ell\phi$ analysis [17]. This assumes that one knows the yield ratio between the sideband and the signal regions from MC. The difficulty in applying this method in low background level decays such as $\tau^- \to \mu^-\mu^+\mu^-$ comes from the low remaining number of MC backgrounds in the signal region. Indeed, this will lead to a ratio between sidebands and the signal region very close to 0. For such reason, we will also discuss two other alternative methods, the relaxed LID sidebands method and the ABCD method.

4.3.5.1. Sidebands method

To estimate the number of expected background events $N^{Data}_{expected,SR}$ inside the 3δ elliptical SR, an extrapolation is performed from the data events N^{Data}_{SB} left in the $M_{3\mu}$ sidebands within $\pm 10\delta\Delta E_{3\mu}$ band, referred as SB, as illustrated in Figure 4.21 (right). The extrapolation is done as follows:

$$N^{Data}_{expected,SR} = N^{Data}_{SB} \times r^{MC}, \tag{4.18}$$

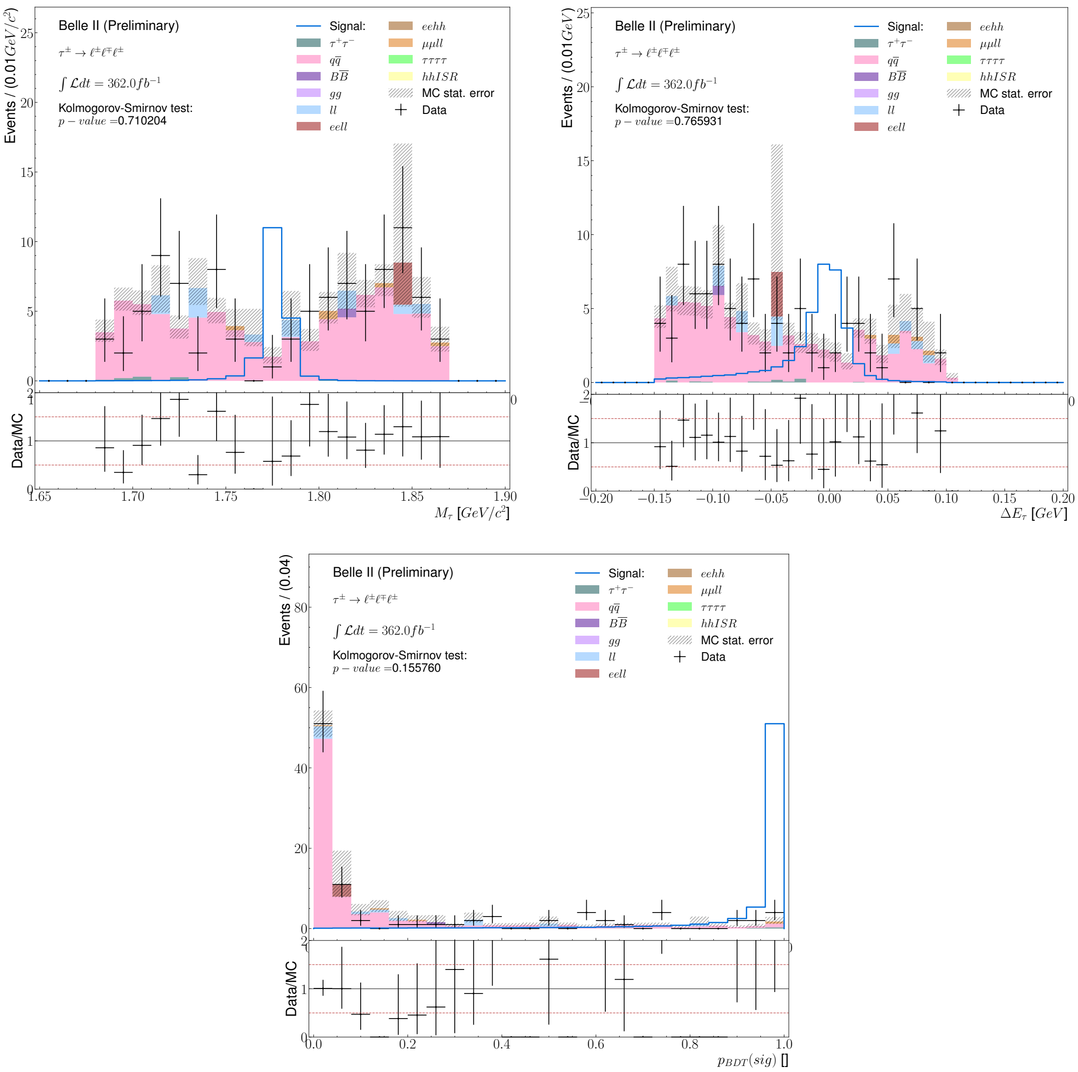


Figure 4.22. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for M_τ, ΔE_τ and the BDT output probability p_{BDT} after the preselection. The $q\bar{q}$ correction factor is applied in those plots.

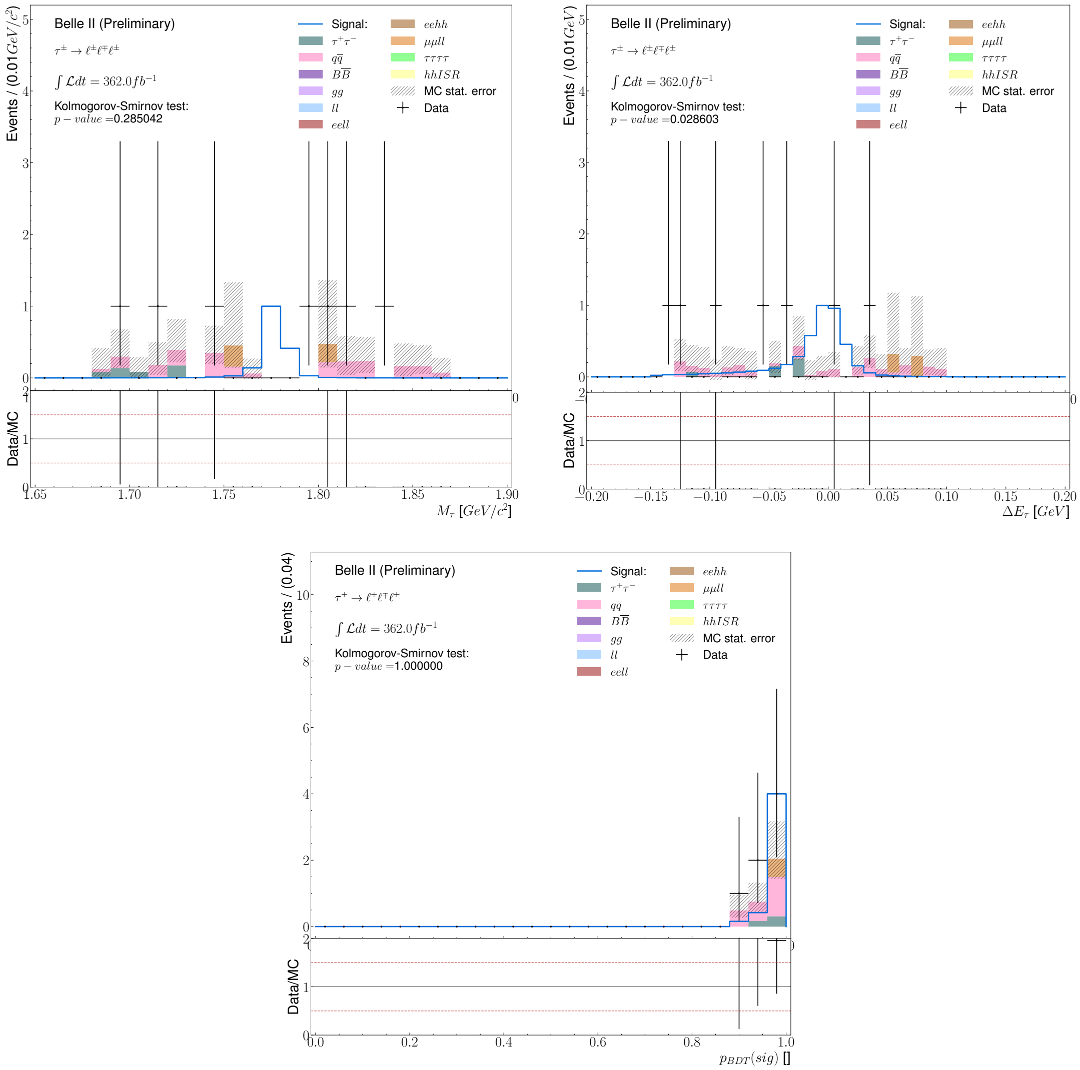


Figure 4.23. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for M_τ, ΔE_τ and the BDT output probability p_{BDT} after the preselection and the BDT selection applied.

where r^{MC} is the ratio measured on retained simulated background events between the ones inside the SR and the ones in SB, $r^{MC} = N_{SR}^{MC}/N_{SB}^{MC}$. In order to avoid the unblinding risk in data sidebands, we are considering blinding the events in the 5δ signal region. Following this method, the computed values and their asymmetrical statistic errors are given in Table 4.11.

Table 4.11. – Number of background events and their asymmetrical statistic errors left after the nominal background rejection selection in the different regions of the $(M_{3\mu}, \Delta E_{3\mu})$ plane for $\tau^- \to \mu^-\mu^+\mu^-$ with the sidebands method.

N_{SB}^{MC}	$3.29^{+1.24}_{-0.73} stat$
N_{SR}^{MC}	$0.08^{+0.21}_{-0.07} stat$
r^{MC}	$0.02^{+0.06}_{-0.02} stat$
N_{SB}^{Data}	$7.00^{+3.77}_{-2.58} stat$
$N_{expected,SR}^{Data}$	$0.17^{+0.46}_{-0.17} stat$

4.3.5.2. Relaxed LID sidebands method

A similar approach could be done relaxing the LID requirements from $\mu ID_{lead} > 0.95$ *and* $\mu ID_{sub} > 0.95$ to $\mu ID_{lead} > 0.5$ *and* $\mu ID_{sub} > 0.5$. This increases the statistic, shown in Figure 4.24, in the SR. Still, the efficiency on LID requirements ε_{LID}^{rel} has to be estimated in simulation assuming it is flat in the $(M_{3\mu}, \Delta E_{3\mu})$ SR. Thus the equation 4.18 became:

$$N_{expected,SR}^{Data} = N_{SB}^{Data} \times \varepsilon_{LID}^{MC} \times r^{MC}, \tag{4.19}$$

The computed values and their asymmetrical statistic errors are given in Table 4.12.

Table 4.12. – Number of background events and their asymmetrical statistic errors left after the nominal background rejection selection in the different regions of the $(M_{3\mu}, \Delta E_{3\mu})$ plane for $\tau^- \to \mu^-\mu^+\mu^-$ with the relaxed LID sidebands method.

N_{SB}^{MC}	$7.15^{+1.36}_{-0.89} stat$
N_{SR}^{MC}	$0.09^{+0.21}_{-0.07} stat$
r^{MC}	$0.01^{+0.03}_{-0.01} stat$
ε^{rel}	$43.88\%^{+17.65}_{-10.72} stat$
N_{SB}^{Data}	$9.00^{+4.11}_{-2.94} stat$
$N_{expected,SR}^{Data}$	$0.05^{+0.12}_{-0.05} stat$

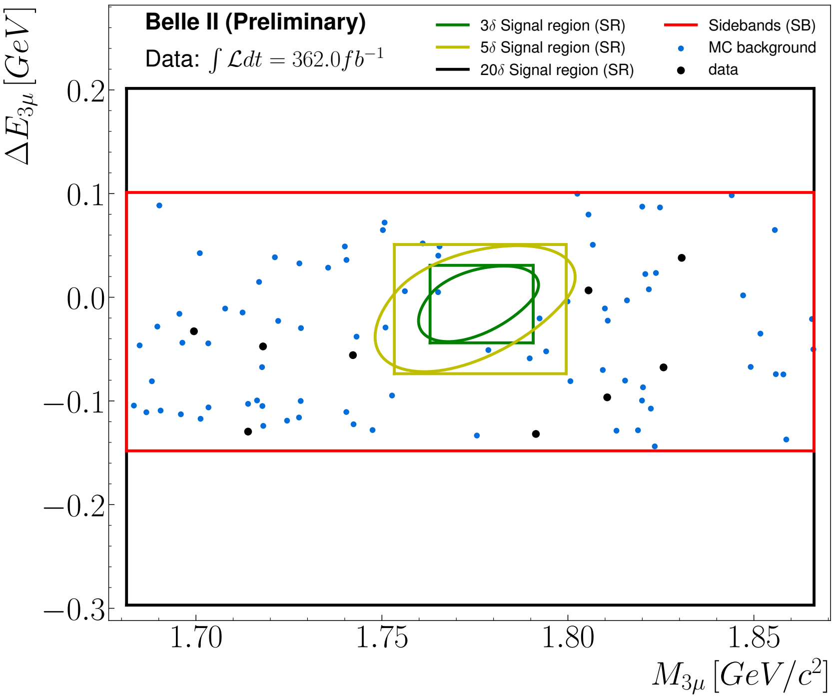


Figure 4.24. – Scatter plots (unweighted) of surviving events after applying the relaxed LID at 0.5, preselection and selection on BDT signal probability output, in the ($M_{3\mu}$, $\Delta E_{3\mu}$) sidebands region (red) for $\tau^- \to \mu^-\mu^+\mu^-$. Data are hidden in the ±5δ box (yellow) to prevent unblinding. The luminosity for background simulated samples is half of the one listed in Table 2.6.

4.3.5.3. Data driven ABCD method

The data-driven methods to extract the expected number of backgrounds are often more reliable as they don't rely on a perfect agreement between data and simulations. The "ABCD" method, used here, was originally devised by LHC experiments [22, 121] to deal with mis-modelled multijet QCD backgrounds. The two-dimensional plane chosen here is defined by the elliptical distance in the $(M_{3\mu}, \Delta E_{3\mu})$ plane (see Equation 4.8) which corresponds to the 2 cases *in* or *outside* the SR, and the BDT signal probability output $p^{BDT}(sig)$. This two-dimensional plane is divided into four regions, as shown in Figure 4.25:

- region A: Outside the elliptical $\pm 5\delta$ SR and $0.2 < p^{BDT}(sig) < 0.5$,
- region B: Inside the elliptical $\pm 3\delta$ SR and $0.2 < p^{BDT}(sig) < 0.5$,
- region C: Outside the elliptical $\pm 5\delta$ SR and $p^{BDT}(sig) > 0.89$,
- region D: Inside the elliptical $\pm 3\delta$ SR and $p^{BDT}(sig) > 0.89$.

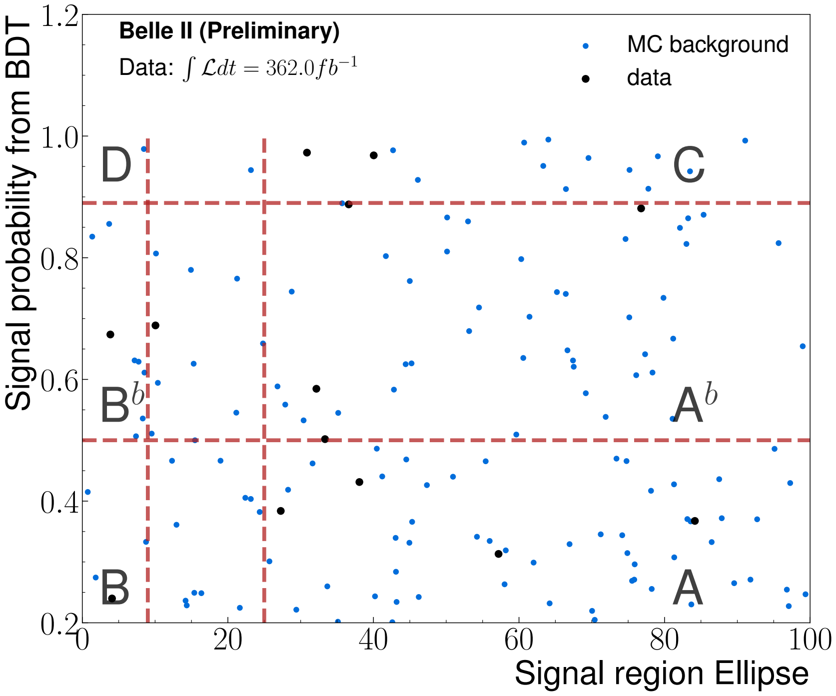


Figure 4.25. – Scatter plots (unweighted) of surviving events after applying the LID, preselection, in the plane between the elliptical distance from the $(M_{3\mu}, \Delta E_{3\mu})$ centre and the BDT signal probability output for $\tau^- \to \mu^-\mu^+\mu^-$. Red dashed lines define the limit of the four ABCD regions. Data are hidden in D regions to prevent unblinding. The luminosity for background simulated samples is half of the one listed in Table 2.6.

The ABC regions are used as control regions, while D is a region where the background is estimated, in our case, where the signal is expected to peak. To avoid potential signal contamination in B and C, the events falling between 3δ and 5δ, or $0.5 < p^{BDT}(sig) <$

0.89 are not used. In addition, an upper bound at $\pm 10\delta$ ellipse is set for the regions A and C to be strictly included in the SB. The number of events in D is extrapolated from the distribution in C as:

$$N_D^{expected} = N_C \times R_{B/A}, \tag{4.20}$$

where $R_{B/A} = N_B/N_A$ is a transfer factor between the region A and B.

In the ABCD method, it is assumed that the transfer factor between A and B is the same as the one between C and D. The regions A^b and B^b between AB and CD are defined to avoid possible data unblinding by taking a certain distance to high BDT output. Thanks to the upper bound on the BDT output at 0.5, the signal efficiency is about 0.4%, which is reasonable to avoid unblinding. Region C is the second area that is at risk. It has a signal efficiency of 2.2%, which results in 0.2 signal events with the Belle's branching fraction (2.1×10^{-8}) for 0.9 background events. However, the high signal efficiency in region C is under control thanks to the blinding of the 5δ SR ellipse. The signal efficiency in the other region is given in Table 4.15. We also used these regions A^b and B^b to control the transfer factor uniformity along the BDT output, comparing $R_{B/A}$ and $R_{Bb/Ab}$. An additional check is done by looking at the distribution for the two variables used in the method, Figure 4.26. The data and simulated backgrounds are uniformly distributed along the two variables, which ensures a coherent transfer factor between the different regions. The "ABCD" method efficiently evaluates the background from only one background type. From this figure, one can also see that the $q\bar{q}$ background dominates. The computed final number of events and their asymmetrical statistic errors are reported in Table 4.13.

Table 4.13. – Number of backgrounds and their asymmetrical statistic errors after the nominal background rejection selection for $\tau^- \to \mu^-\mu^+\mu^-$ with the ABCD method in data.

N_A	$4.00^{+3.16}_{-1.91} stat$
N_B	$1.00^{+2.30}_{-0.83} stat$
N_{Ab}	$4.00^{+3.16}_{-1.91} stat$
N_{Bb}	$1.00^{+2.30}_{-0.83} stat$
$R_{B/A}$	$0.25^{+0.61}_{-0.24} stat$
$R_{Bb/Ab}$	$0.25^{+0.61}_{-0.24} stat$
N_C	$2.00^{+2.64}_{-1.29} stat$
$N_D^{expected}$	$0.50^{+1.38}_{-0.50} stat$
$N_{Db}^{expected}$	$0.50^{+1.38}_{-0.50} stat$

As a final check, we also apply the ABCD method on simulated events, and results are shown in Table 4.14. The expected yields computed from regions ABC and $AbBbC$ are compatible with the observed yield in D.

Table 4.14. – Number of backgrounds and their asymmetrical statistic errors after the nominal background rejection selection for $\tau^- \to \mu^-\mu^+\mu^-$ with the ABCD method in simulation.

N_A	$5.43^{+1.87}_{-0.89} stat$
N_B	$0.46^{+0.32}_{-0.14} stat$
N_{Ab}	$3.47^{+1.84}_{-0.82} stat$
N_{Bb}	$0.55^{+0.43}_{-0.22} stat$
$R_{B/A}$	$0.08^{+0.07}_{-0.03} stat$
$R_{Bb/Ab}$	$0.16^{+0.15}_{-0.07} stat$
$R_{D/C}$	$0.09^{+0.25}_{-0.09} stat$
N_C	$0.92^{+0.99}_{-0.41} stat$
$N_D^{expected}$	$0.08^{+0.10}_{-0.04} stat$
$N_{Db}^{expected}$	$0.14^{+0.21}_{-0.09} stat$
N_D	$0.08^{+0.21}_{-0.07} stat$

Table 4.15. – Checks to control the unblinding risk in the different zone used for the ABCD method, using the simulation. An expected number of signals in each zone is obtained with the signal efficiency and assuming the branching fraction corresponding to the Belle limit. This number of signals is compared to the number of simulated backgrounds.

	ε_{sig}^{abs} (%)	$N_{sgn}^{expBelleBR}$	N_{bkg}
Zone A	$0.11^{+0.00}_{-0.00} stat$	$0.02^{+0.00}_{-0.00} stat$	$5.43^{+1.87}_{-0.89} stat$
Zone B	$0.40^{+0.01}_{-0.01} stat$	$0.06^{+0.00}_{-0.00} stat$	$0.46^{+0.32}_{-0.14} stat$
Zone Ab	$0.45^{+0.01}_{-0.00} stat$	$0.06^{+0.00}_{-0.00} stat$	$3.47^{+1.84}_{-0.82} stat$
Zone Bb	$1.89^{+0.02}_{-0.02} stat$	$0.26^{+0.00}_{-0.00} stat$	$0.55^{+0.43}_{-0.22} stat$
Zone C	$2.23^{+0.02}_{-0.02} stat$	$0.31^{+0.00}_{-0.00} stat$	$0.92^{+0.99}_{-0.41} stat$
Zone D	$19.70^{+0.06}_{-0.06} stat$	$2.75^{+0.01}_{-0.01} stat$	$0.08^{+0.21}_{-0.07} stat$

4.3.5.4. Final background yield

The three methods estimate the background yield in the SR to be $0.17^{+0.46}_{-0.17}$ (sideband), $0.05^{+0.05}_{-0.12}$ (relaxed LID), and $0.50^{+1.38}_{-0.50}$ (ABCD). Given the large statistical uncertainties, the three methods agree, but the ABCD method has the advantage of not relying on simulation. It is thus the most robust one. One can note that it is also a conservative choice since it gives the highest expected yield.

4.4. Study of the systematics uncertainties

In this section, we will analyze the primary sources of systematic uncertainty that can affect various quantities involved in the measurement of branching fraction on $\tau^- \to \mu^-\mu^+\mu^-$, given in Eq. 4.29. It is necessary to consider these uncertainties when calculating the upper limit accurately. The systematic uncertainties come either from detector effects, as resolutions and efficiencies that have to be measured on real data, or from the differences between the simulation used for analysis optimization and for the signal efficiency estimation. Both types of contributions need to be estimated with dedicated performance and data validation studies: some are specific to this analysis and derived from comparisons between data and simulation, while others are provided as uncertainty estimations common to all analyses provided by the Belle II performance group. A summary of these sources, along with their relation to the relevant quantities, is provided in Table 4.19.

4.4.1. Uncertainty on signal efficiency

The calculation of signal efficiency is highly sensitive to simulation mismodelling as it is entirely based on simulations. The main factors contributing to this uncertainty are particle identification, trigger, and track reconstruction efficiencies. We examine the differences between the data and the simulation to estimate the errors.

4.4.1.1. Particle identification efficiency and misidentification probability

It is acknowledged that the Lepton identification probabilities are not accurately represented in simulations, causing disparities in comparison to real data.

From the lepton identification performance studies [122, 95], weights are provided in order to correct the simulation and take into account the statistical and systematic uncertainties on LID efficiencies and fake rates. The given corrections are track dependant weights $w_{LID,\ell i}$, a global correction $w_{LID}^{\overset{sys}{stat},\uparrow\downarrow}$ is defined as:

$$w_{LID}^{\overset{sys}{stat},\uparrow\downarrow} = \prod_{i=1}^{3} w_{LID,\ell i} \times w_{LID,\ell i}^{\overset{sys}{stat},\uparrow\downarrow}. \tag{4.21}$$

Then, the signal efficiency is recomputed using the statistical and systematics LID

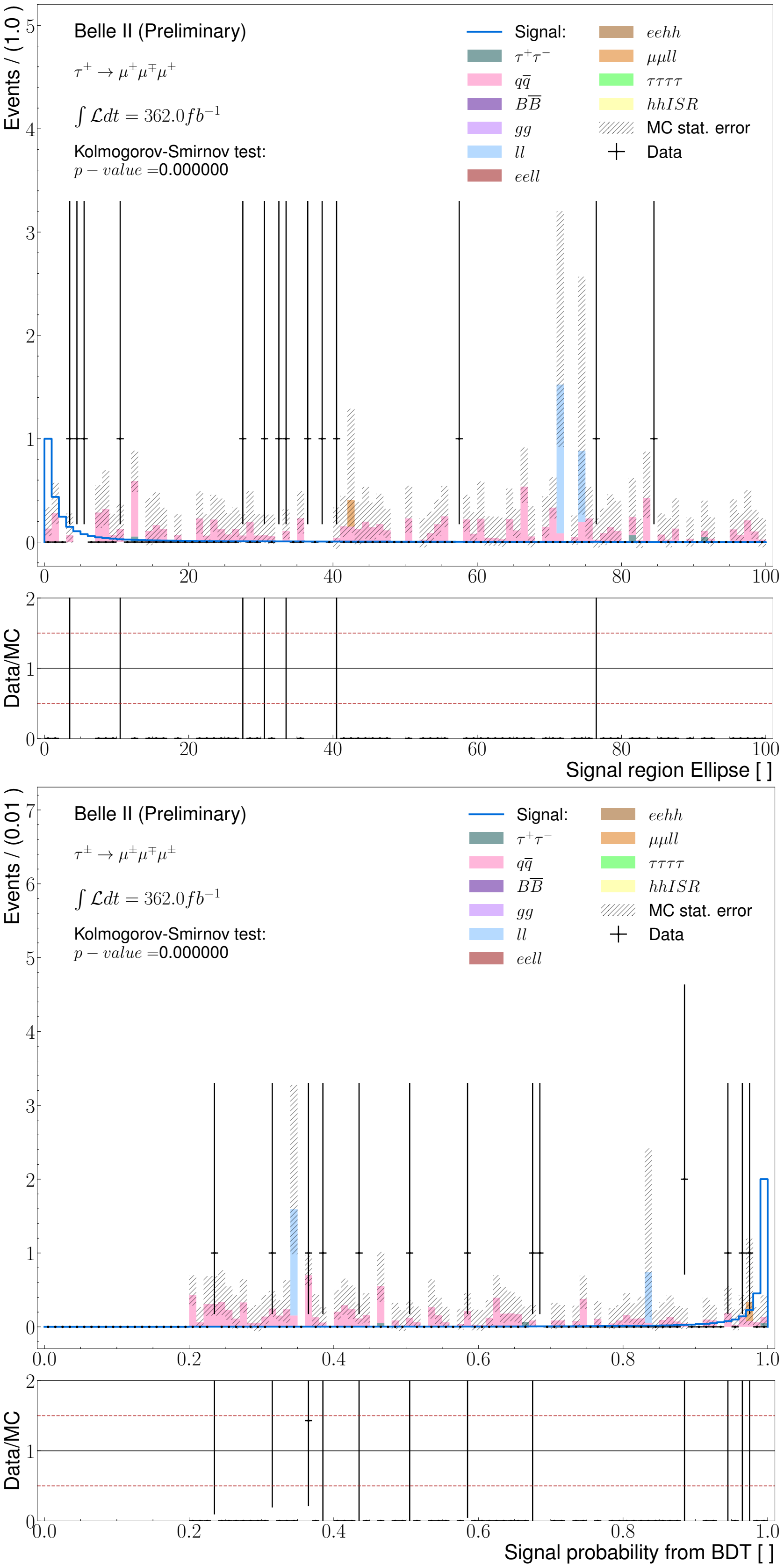


Figure 4.26. – Comparison between data and simulation for the variables entering in the "ABCD" method: elliptical distance from the $(M_{3\mu}, \Delta E_{3\mu})$ centre, Eq. 4.8 (top) and the BDT signal probability output (bottom).

variation, and a relative error is computed as:

$$\sigma_{LID}^{\overset{sys}{stat},\uparrow\downarrow} = \sqrt{\left(\frac{n_{LID} - n_{LID}^{\overset{sys}{stat},\uparrow\downarrow}}{n_{produced}}\right)^2} \times \frac{1}{\varepsilon_{sig}^{abs}}, \tag{4.22}$$

where n_{LID} ($n_{LID}^{\overset{sys}{stat},\uparrow\downarrow}$) is the number of signal events computed with the nominal LID weight w_{LID} (the systematics or statistical LID weights $w_{LID}^{\overset{sys}{stat},\uparrow\downarrow}$) and $n_{produced}$ is the number of simulated signal events produced.

The detailed errors are in Table 4.16. Adding the maximal error for the statistics and systematic contribution in quadrature gives an overall relative systematic uncertainty of 2.39%.

Table 4.16. – Detailled signal efficiencies and deviation with the different LID correction variations. After applying the whole background rejection selections, the numbers are obtained in the Signal Region.

LID correction variation	N_{sgn}	ε_{sn}^{abs} (%)	Deviation	σ_{LID}^{abs}	σ_{LID}^{rel}
Nominal	98478.74	19.70	-	-	-
Statistical ↓	96411.04	19.28	2067.70	0.004135	2.10
Statistical ↑	100769.97	20.15	2291.23	0.004582	2.33
Systematics ↓	98129.62	19.63	349.12	0.000698	0.35
Systematics ↑	99008.57	19.80	529.83	0.001060	0.54

4.4.1.2. Trigger efficiency

The trigger is a crucial element in data collection, but it is difficult to simulate accurately. As explained in Section 4.2.4.2, the selected events are required to fire the ECL-based or the CDC-based trigger lines. In order to assess the overall difference in trigger efficiency between data and simulation, we analyze the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample [123]. The systematic uncertainty is estimated as the total variation observed in the ratio between the data and simulation efficiencies.0

The systematics uncertainty is derived from the agreement between data and simulation for the trigger efficiency. The trigger efficiency can't be absolutely computed in data since the number of generated events, *i.e.* before requiring the L1 trigger, is unknown. To handle this issue, the method is to evaluate the efficiency using orthogonal trigger selections: the efficiency of ECL-based trigger lines (hie and lmlX, X=6,7,8,9,10,12) are evaluated on events triggered by the CDC, while the efficiency of CDC-based trigger lines (ffy and fyo) are evaluated on events passing ECL trigger requirements, as defined by :

$$\varepsilon_{ECL} = \frac{(lmlX \text{ OR } hie) \text{ AND } (ffy \text{ OR } fyo)}{ffy \text{ OR } fyo} \tag{4.23}$$

$$\varepsilon_{CDC} = \frac{(lmlX \text{ OR } hie) \text{ AND } (ffy \text{ OR } fyo)}{lmlX \text{ OR } hie} \tag{4.24}$$

For each trigger group, the difference between data ε^{Data} and simulation ε^{MC} efficiencies relative to unity is computed as:

$$\delta_{ECL} = \left|1 - \frac{\varepsilon^{Data}_{ECL}}{\varepsilon^{MC}_{ECL}}\right|, \qquad \delta_{CDC} = \left|1 - \frac{\varepsilon^{Data}_{CDC}}{\varepsilon^{MC}_{CDC}}\right|. \tag{4.25}$$

To prevent any risks of signal unblinding, the study is based on the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample using 362 fb^{-1} of data and 1 ab^{-1} of simulation samples, including generic and low-multiplicity processes. The Control Sample (CS) is reconstructed similarly to $\tau^- \to \mu^-\mu^+\mu^-$, as described in Section 4.2. Three charged pions from a τ^- replace the combination of three muons. To ensure orthogonal reconstruction and avoid accidental unblinding, a veto on LID is applied since pions can be identified as muons and vice versa. Additionally, some requirements must be met, such as passing the considered trigger lines and the τ^- LFV skim requirement.

The trigger efficiencies are computed as a function of the transverse momentum of the leading and third muon, as seen in Figure 4.27. A linear fit is used to obtain the efficiencies reported in Table 4.17. Since the ECL lines trigger a majority of the events, the systematic uncertainty is computed using the weighted average:

$$\sigma_{TRG} = \varepsilon^{MC}_{ECL}\delta_{ECL} + (\varepsilon^{MC}_{tot} - \varepsilon^{MC}_{ECL})\delta_{CDC} = 1.0\%, \tag{4.26}$$

where the ε^{MC} are the ones from Figure 4.8 and the δ are the averaged values obtained from the two variables.

Table 4.17. – Trigger efficiency discrepancy between data and simulations measured on the $\tau \to \pi\pi\pi\nu$ control sample.

Variable	δ_{CDC}	δ_{ECL}
p^T_{lead}	3.63%	0.5%
p^T_{third}	5.04%	0.4%

4.4.1.3. Tracking efficiency

Tracking efficiency was measured using tag-and-probe techniques in $e^+e^- \to \tau^+\tau^-$ events, targeting 1-prong ($\tau^- \to e^-\overline{\nu}_e\nu_\tau$) and 3-prong ($\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$) decays [92]. Three high-quality tracks were utilised to tag τ-pair events. These tracks have a total charge of ±1. The existence of a fourth track can be inferred from charge conservation

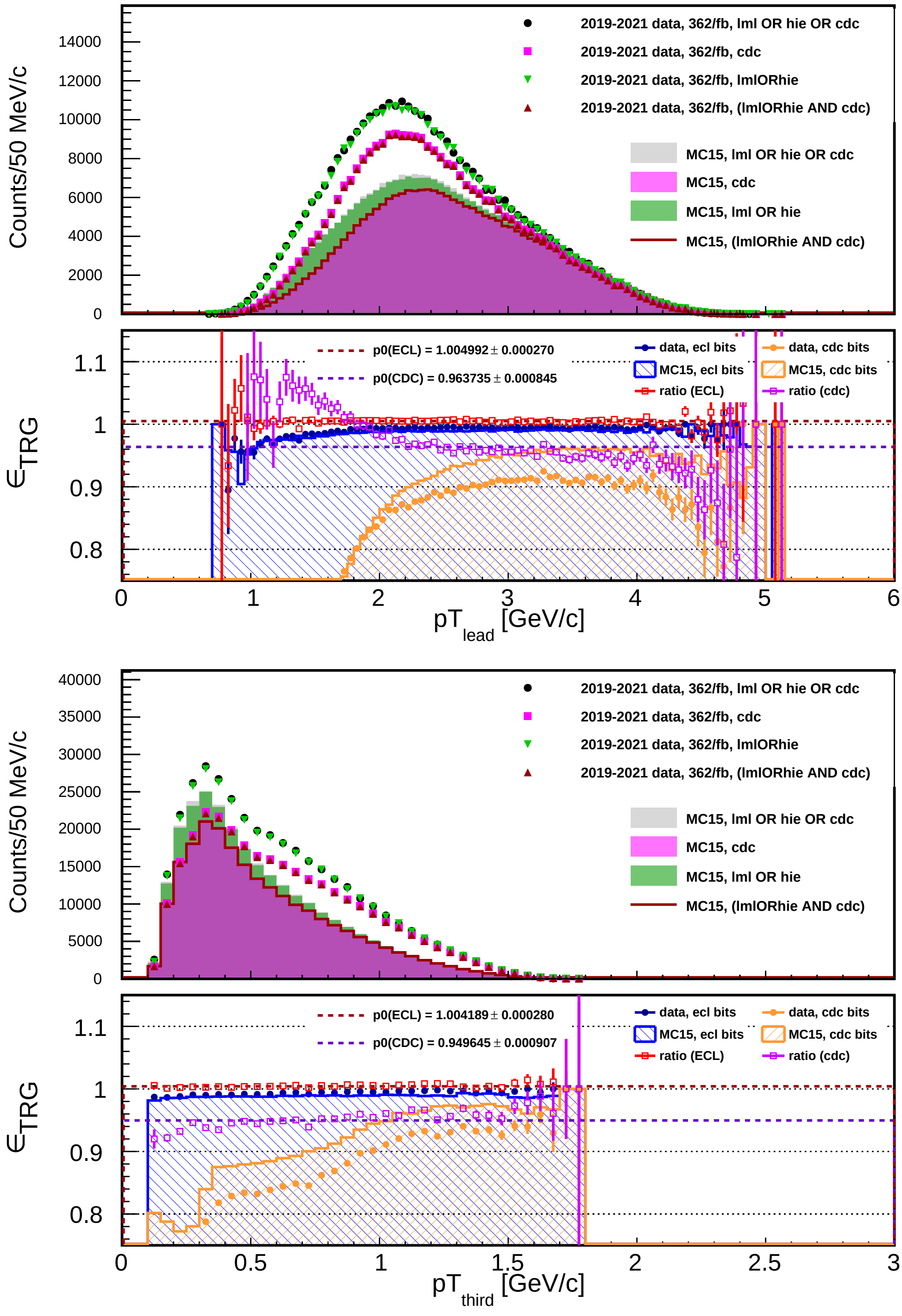


Figure 4.27. – Distribution of the leading momentum pion (top) and third momentum pion (bottom) in data and simulation with the CDC or ECL triggers fired. Trigger efficiency for the ECL and CDC and the discrepancy between data and simulation. The trigger systematics is derived as the constant fit of the discrepancy between data and simulation.

since the total charge of the event must be equal to zero. The per-track reconstruction efficiency $\varepsilon_{tracking}$ was computed:

$$\varepsilon_{track} = \frac{N_4}{N_3 + N_4}, \tag{4.27}$$

where N_4 is the number of events where all four tracks are found, while N_3 is the number of events where the fourth track is not found. Finally, the systematic uncertainty is extracted from the mismodeling of the efficiency in MC simulation computed by $\delta = 1 - \frac{\varepsilon^{data}_{tracking}}{\varepsilon^{MC}_{tracking}}$.

The study led by the performance group established an uncertainty of 0.24% per track applied. Summing this contribution for the three reconstructed tracks, the associated systematic uncertainty is 0.72%.

4.4.2. Uncertainty on the expected background yield

4.4.2.1. Track momentum scale

Errors in the magnetic field map used for data reconstruction can cause a bias in track momentum compared to the true value and the MC simulated samples. To address this calibration issue, correction factors for momentum scaling are calculated based on the observed D^0 mass shift in $D^{*+} \to [D^0 \to K^-\pi^+]\pi^+$ events [124]. These scale factors are then directly applied to the data according to the Performance group's recommendations. The value of the scale factor is set at $0.99987(+3.8/-5.7) \times 10^{-4}$, with two additional data samples created by applying variations in both directions. The systematics uncertainties are calculated from the number of data in the sidebands region N^{data} varying the scale factor as:

$$\sigma^{rel}_{MomentumScale} = \left| \frac{N^{data}_{True\ Scale} - N^{data}_{Low,High\ Scale}}{N^{data}_{True\ Scale}} \right|. \tag{4.28}$$

The detailed numbers are given in Table 4.18. Taking the highest deviation between low and high variation, the final relative systematics uncertainties is 5.3%.

Table 4.18. – Detailled number of data events in the blind sidebands after applying the preselections for each variation of the momentum scale applied. Systematics uncertainties are derived from the difference in number of events.

	N^{data}	$\sigma^{\downarrow}_{stat}$	$\sigma^{\uparrow}_{stat}$	σ^{abs}_{sys}	σ^{rel}_{sys} (%)
True scale	94.0	9.68	10.73	0.0	-
Low scale	92.0	9.57	10.63	2.0	2.13
High scale	99.0	9.93	10.98	5.0	5.32

4.4.2.2. Extraction method

Since the ABCD method doesn't depend on simulation, and the checks done show no bias, we don't assign any systematic uncertainty to the method used to extract the background yield.

4.4.3. Other sources

- Luminosity: the systematic uncertainty on the integrated luminosity has been measured in [78], using Phase 2 data, on two samples corresponding to either Bhabha or di-photons events. The combination of both measurements gives a relative uncertainty of 0.6%.
- Tau pair cross-section: the uncertainty on this parameter was determined in [125] as 0.003 nb.

4.4.4. Systematics uncertainties summary

Table 4.19 displays the systematics uncertainties taken into consideration for computing the upper limit of exclusion at a 90% CL.

Table 4.19. – Relative systematic uncertainties entering the upper limit computation as a function of the decay mode.

Affected quantity	Source	value
$\varepsilon_{\tau^-\to\mu^-\mu^+\mu^-}$	Particle identification	2.39%
	Tracking efficiency	0.72%
	Trigger efficiency	1.0%
N_{exp}	Momentum scale	5.0%
L	Luminosity	0.6%
$\sigma_{\tau\tau}$	Tau-pair cross section	0.3%

4.5. Branching fraction upper-limit estimation

Once the signal efficiency and the numbers of expected background events have been derived, an upper limit on the branching fraction for the $\tau^- \to \mu^-\mu^+\mu^-$ LFV decays:

$$\mathscr{B}^{90}_{UL} = \frac{s^{90}}{2 \times L \times \sigma_{\tau^+\tau^-} \times \varepsilon_{\tau^-\to\mu^-\mu^+\mu^-}}, \tag{4.29}$$

4.5.1. CL_s method

The upper limit is computed using the `RooStat` package and the CL_s method [113, 23]. The CL_s method is parametrized by looking at the expected number of events n

which follows a Poisson distribution with an expectation value:

$$E[n] = \mu s + b, \tag{4.30}$$

where respectively μs and b are the signal and background yields. Following Eq. 4.29, the signal yield can be expressed as:

$$\mu s = L \times 2\sigma_{\tau^+\tau^-} \times \epsilon_{\tau^- \to \mu^-\mu^+\mu^-} \times \mathscr{B}(\tau^- \to \mu^- \mu^+ \mu^-). \tag{4.31}$$

Thanks to this parameterization, the parameter of interest, denoted as μ, can be defined as the $\tau^- \to \mu^-\mu^+\mu^-$ branching fraction $\mu = \mathscr{B}(\tau^- \to \mu^-\mu^+\mu^-)$. And so s is set as the product of the luminosity L, the τ-pair cross section $\sigma_{\tau^+\tau^-}$, and the signal efficiency of the selection criteria for the $\tau^- \to \mu^-\mu^+\mu^-$ decay mode $\epsilon_{\tau^- \to \mu^-\mu^+\mu^-}$: $s = L \times 2\sigma_{\tau^+\tau^-} \times \epsilon_{\tau^- \to \mu^-\mu^+\mu^-}$.

The branching fraction (parameter of interest) is estimated using the toy-based calculator of `RooStat` [126] for 25 points evenly distributed between 0 and 5×10^{-8} using 10k toys at each point. The upper-limit at 90% C.L. on $\mathscr{B}(\tau^- \to \ell^-\ell^+\ell^-)$ can be computed by setting $CL_s = 10\%$, where:

$$CL_s = \frac{CL_{s+b}}{CL_b}, \tag{4.32}$$

with CL_{s+b} and CL_b are respectively the "signal+background" and "background" only hypothesis p-values.

4.5.2. Statistical uncertainites

The statistical uncertainties are assigned to the data and MC yields by a frequentist approach. The Poisson likelihood is integrated iteratively in order to find the values λ_1, λ_2 such as:

$$P(n \leq N_{bin}|\lambda_1) \leq 0.16, \tag{4.33}$$

$$P(n \geq N_{bin}|\lambda_2) \leq 0.16, \tag{4.34}$$

with N_{bin} the measured unweighted yield in each bin of the given distributions. Giving the values λ, the error bars are defined as:

$$\sigma^{stat}_{low} = N_{bin} - \lambda_2, \tag{4.35}$$

$$\sigma^{stat}_{high} = \lambda_1 - N_{bin}, \tag{4.36}$$

$$\tag{4.37}$$

where σ^{stat}_{low} and σ^{stat}_{high} are respectively the lower and upper statistical uncertainties.

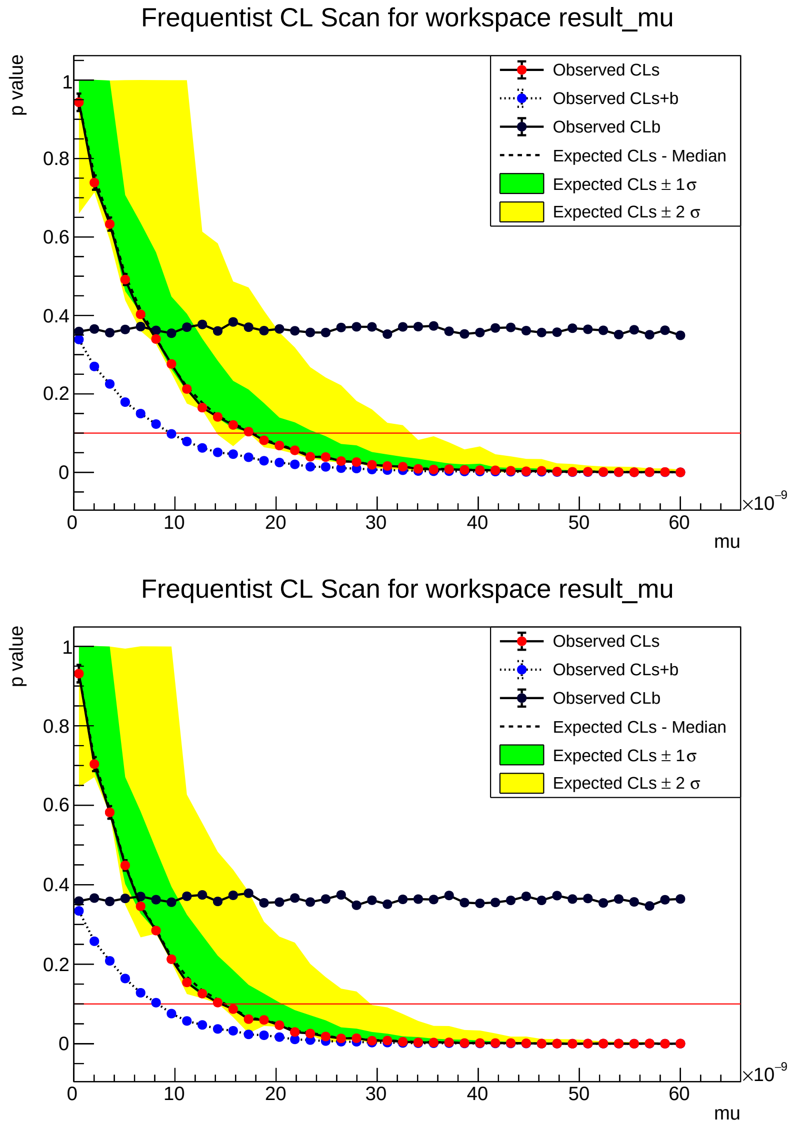


Figure 4.28. – Expected CL_s as a function of the upper limit on the branching fraction of $\tau^- \to \mu^-\mu^+\mu^-$ for statistics of 362 fb^{-1} (top) and 424 fb^{-1} (bottom). The red line corresponds to the 90% confidence level.

4.5.3. Upper limit results

The upper limit of the branching fraction has been estimated for the data collected at an energy of $\Upsilon(4S)$ with a luminosity of 362 fb^{-1}. Additionally, an estimation has been made for the extended dataset with the $\Upsilon(4S)$, $\Upsilon(5S)$ and off-resonance energy, which has a total luminosity of 424 fb^{-1}. By using the signal efficiency after the background rejection, reported in Table 4.8, the expected number of background events, reported in Table 4.13, and the systematic uncertainties reported in Table 4.19, we obtain the expected CL_s, in Figure 4.28 and a corresponding expected 90% C.L.upper-limit on $\mathscr{B}(\tau \to \mu\mu\mu)$ for 362 (424) fb^{-1}:

$$\mathscr{B}^{90}_{UL,exp}(\tau \to \mu\mu\mu) = 1.77(1.51) \times 10^{-8}. \tag{4.38}$$

For the limit evaluation on the 424 fb^{-1} dataset, the tau pair cross section is computed as a weighted average corresponding to the different dataset energies, which is actually equal to 0.919 nb.

This calculation is cross-checked using the Feldman Cousin interval [127], which is 2.44, assuming 0 observed events. Without taking the uncertainties into account, the upper limit is then

$$\mathscr{B}^{90}_{UL,exp}(\tau \to \mu\mu\mu) = \frac{s^{90}}{2 \times L \times \sigma_{\tau^+\tau^-} \times \varepsilon_{\tau\to\mu\mu\mu}} = 1.56 \times 10^{-8}, \tag{4.39}$$

for 424 fb^{-1}.

Conclusion

In this thesis, I presented the search for lepton flavour violating $\tau^- \to \mu^-\mu^+\mu^-$ decays using $e^+e^- \to \tau^+\tau^-$ events in the $424\,\mathrm{fb}^{-1}$ of data collected by the Belle II experiment between 2019 to 2022. This analysis is based on a novel-untagged reconstruction, where only the signal τ^- produced in $e^+e^- \to \tau^+\tau^-$ events is reconstructed, the other tau being left unconstrained. With this approach, the goal is to increase the efficiency of the signal by allowing all the decay in one or three charged particles. Such inclusive reconstruction also allows events with additional or missing tracks. With this strategy, the amount of background is increased. To remove it, we use a three-step optimized selection process to reject it. This involves a selection based on the lepton identification variable, a preliminary cut-based selection, and finally, a boosted-decision-tree classifier to eliminate any remaining $q\overline{q}$ continuum backgrounds. The rejection was optimised using Monte Carlo simulation samples to maximise the background rejection and signal efficiency inside the signal region. The signal region corresponds to an ellipse with a semi-axis equal to three times the resolution of the expected signal peak in the $(M_{3\mu},\Delta E_{3\mu})$ plane. The estimated upper limit at a 90% confidence level on the branching fractions of the $\tau^- \to \mu^-\mu^+\mu^-$ channel is computed using the CL_s method. Input ingredients are the signal efficiency on MC samples, the estimated number of data computed with the "ABCD" method and the estimated uncertainties. The expected upper limits before full unblinding are estimated for $362\,\mathrm{fb}^{-1}$ and $424\,\mathrm{fb}^{-1}$.

Following the method described previously, the efficiency of the signal is increased by a factor of 2.5 with respect to the Belle analysis[8], and 1.4 to a Belle II study relying on a "classical" reconstruction with only one charged particle in the tag side and a cut-based background rejection. At the same time, the number of expected background events is compatible with 0 for the three studies. With these performances on signal efficiency and background rejection, the expected upper limit on $\tau^- \to \mu^-\mu^+\mu^-$ branching fraction is 1.77×10^{-8} using $324\,\mathrm{fb}^{-1}$ and 1.51×10^{-8} using $424\,\mathrm{fb}^{-1}$. The strategy of untagged reconstruction plus BDT background rejection, described in this thesis, leads to an improvement of about 20% to the current most stringent limit measured by Belle at 2.1×10^{-8} using $782\,\mathrm{fb}^{-1}$.

The Belle II experiment has completed its first data-taking phase in 2022 and is now undergoing a long shutdown to install upgrades. The collisions are planned to resume in early 2024 for the second data-taking phase, which plans to collect at least $5\,\mathrm{ab}^{-1}$ by 2027, as shown in Figure 4.29. The experiment will end with a dataset ranging from $25\,\mathrm{ab}^{-1}$ to $50\,\mathrm{ab}^{-1}$ due to various scenarios about the instantaneous luminosity increase. The upper limits at the target integrated luminosities (5, 25 and $50\,\mathrm{ab}^{-1}$) are estimated by the CL_s method with as input: the signal efficiency 19.7% and the expected number of data renormalized to the target luminosity. Respectively the

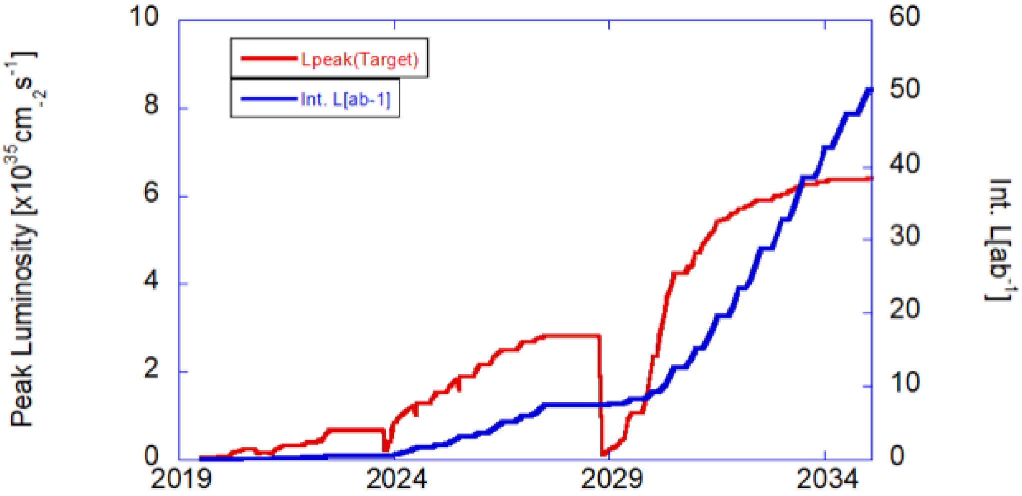


Figure 4.29. – The future projection for the instantaneous (red) and integrated (blue) luminosity until reaching the 50 ab^{-1} scenario in 2035. Credits [99]

estimated upper limits are 1.3×10^{-9}, 2.7×10^{-10} and 1.3×10^{-10} at 5, 25 and 50 ab^{-1}. We can notice that the estimated upper limit is a linear function of the luminosity, and it is better than estimation [128, 67] at 50 ab^{-1}.

In the upcoming year, Belle II will be a leader in tau LFV searches while considering the rise in luminosity and possible improvements,*e.g.* lepton identification based on machine learning alternatives, triggers and simulation description. In addition, the reconstruction and background rejection strategy has room for improvement by making the classifier more robust by training on a bigger sample or using more complex classifiers. A final room for improvement might be using a fitting procedure to count the number of expected data.

On a more long-term perspective, the future of tau LFV searches is covered by the e^+e^- colliders for Super Charm and Tau factories, developed in Novosibirsk [129], Russia and in Hefei, China [130] with high $\tau^-\tau^+$ production. We also notice the proposal for a fixed target experiment to study τ LFV decays [131]. The goal of those experiments is to reach upper limits at the level of 10^{-10}. Considering the estimations of future experimental sensitivities, we can expect to observe LFV decays if the physics beyond the SM is described by models such as Little Higg T-Parity, Supersymmetry or Z' bosons.

In addition, this thesis also presents the first measurement of the vertex detector spatial resolution with a method exploiting overlapping sensors. The $e^+e^- \to \mu^+\mu^-$ data events are selected to keep only the case where a particle has left two hits in the same detector layer. The method estimates the spatial resolution from the difference of the residuals measured in the two layers, allowing it to cancel out tracking error and Coulomb scattering effects. The spatial resolution is measured between 16 μm and 35 μm, depending on the layer and the sensor side. The resolution is consistent with other methods implemented at Belle II.


τ
Lipton
μ
Lipton
μ
Lipton
μ
Lipton

APPENDICES

A. SVD cluster position resolution

A.1. Assumption on the track's true residuals

Examining the simulation, the assumption that "track extrapolation error can be used as an estimate of the standard deviation of the track's true residuals" is not entirely accurate. The track extrapolation pulls[5] do not have a unit width. As depicted in Figure 30, the tracking error is underestimated by approximately 10% on average. This suggests that certain entries in the tails considerably widen the pulls.

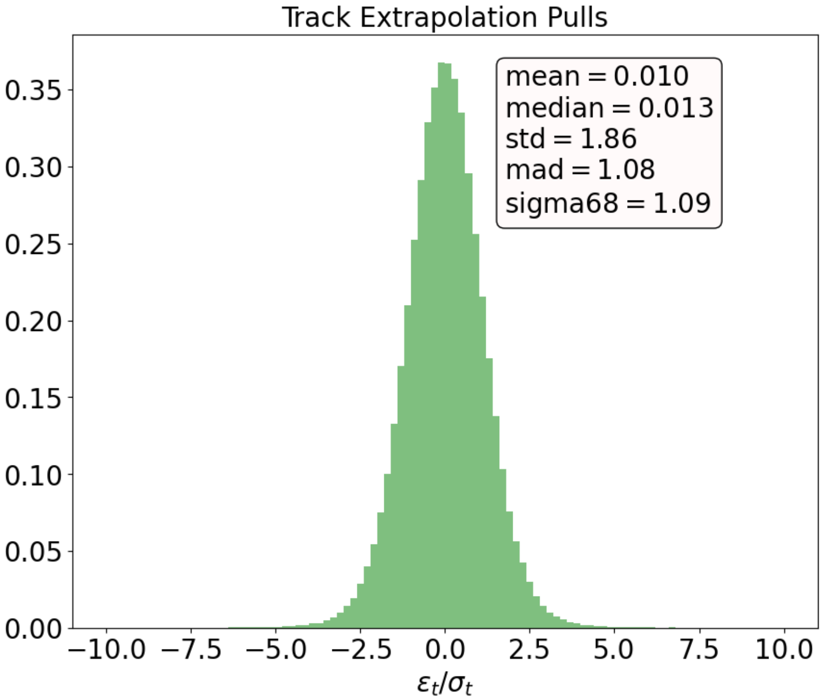


Figure 30. – Layer 4 u/P side track position extrapolation pulls ε_t/σ_t. Credits [110]

As there is no clear method to rectify the estimated track extrapolation error to align with the track extrapolation resolution, and since the number of tracks with considerably incorrect errors is low, we will proceed with the assumption that the estimated track error (σ_t) is a reliable estimator of the true track error's sigma (ε_t).

---

[5] Pulls is the ratio between true residuals and estimated error.

Consequently, we have $E[\epsilon_t^2] = E[\sigma_t^2] = E[\sigma_t]^2 + Var(\sigma_t)$. It has to be noticed that σ_t and ϵ_t have distinct distributions. While ϵ_t is centred at zero, σ_t is not. However, if we draw ϵ_t from a Gaussian distribution with a sigma equal to the σ_t of the corresponding event, it would work for each event.

A.2. The true cluster resolution

The true resolution is defined form the eq. 3.7 by assuming that $E[\epsilon_m] = 0$, as:

$$\sigma_{true}^2 \equiv E[\epsilon_m^2] = E[\epsilon_m]^2 + Var(\epsilon_m) = Var(\epsilon_m). \tag{.40}$$

As shown in Figure 3.3 and 31, it is evident that the distribution of the true residual ϵ_m is not Gaussian. Instead, it has two components: a narrow component representing accurately measured clusters, easily described by a Gaussian distribution, and a wider component, characterized by large tails, describing clusters with poorer resolution. The choice of the measure of the ϵ_m variance (`std`, `sigma-68`) must be carefully considered to adequately address the two regimes presented by the ϵ_m distribution.

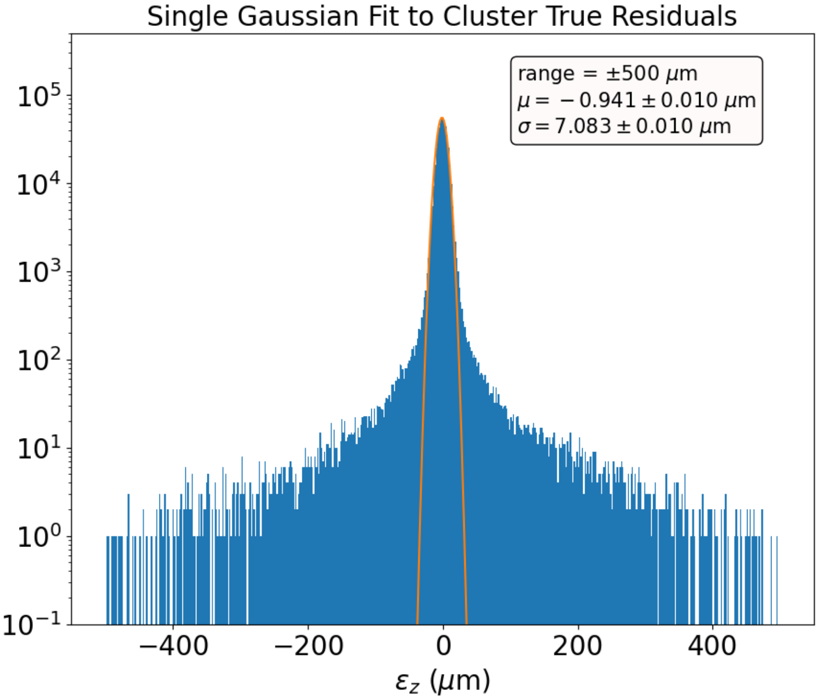


Figure 31. – Layer 4 u/P side true cluster residual normalized distributions, fitted with a single-gaussian for different ranges. Credits [110]

In the Gaussian case, the standard deviation (`std` $= \sqrt{\frac{\sum_{i=1}^{N}(y - mean(y))^2}{N-1}}$) is a reliable measure of how spread out the data is. However, it can be influenced by extreme values and outliers. An alternative measure that is more robust in large tails is the 68%

coverage sigma (`sigma-68`), calculated by taking half the difference between the 84th and 16th quantiles. See Figure 32 for a comparison of the estimators in the function of the range of the distributions.

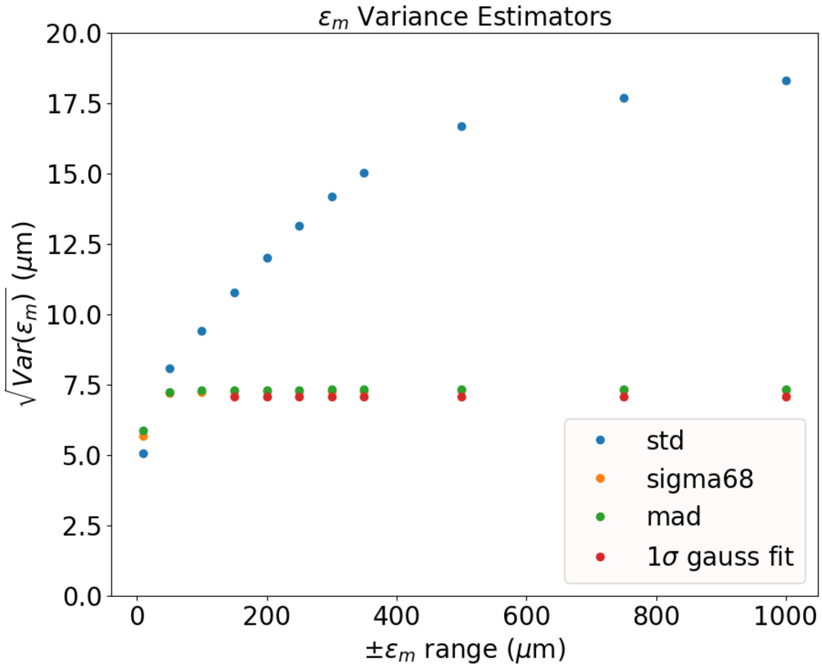


Figure 32. – Estimators of the square root of the variance of the true position residual for Layer 4 u/P side clusters as a function of the ε_m ranges, from ±30 µm to ±1 mm. Credits [110]

We chose the 68% coverage sigma estimator, denoted by `sigma-68`, because it is more robust against large tails than the standard deviation. As a result, the definition of the true resolution has been established.

$$\sigma_{true}^2 = E[\varepsilon_m^2] \simeq \texttt{sigma-68}^2(\varepsilon_m) = 7{,}3\,\mu\text{m}, \quad (.41)$$

The calculation was performed using the entire range of ε_m. It's worth noting that this value is actually better than the estimated value based on the pitch (11 µm) listed in Table 2.3, which was expected.

A.3. Definition of the cluster position resolution with overlap method

To find the relation between the double residual ΔR and the cluster position resolution σ_{cl}, we have to develop the variance of ΔR similarly to work done in Section 3.1.2, as:

$$\begin{aligned} Var(\Delta R) &= Var\left(\frac{R_i - R_e}{\sqrt{2}}\right) && (.42)\\ &= \frac{1}{2}E\left[((m_i - x_i) - (m_e - x_e) - (t_i - x_i) + (t_e - x_e))^2\right] && (.43)\\ &= \frac{1}{2}E\left[(\varepsilon_{mi} - \varepsilon_{me} - \varepsilon_{ti} + \varepsilon_{te})^2\right] && (.44)\\ &, && (.45) \end{aligned}$$

By developing the square sum:

$$\begin{aligned} (\varepsilon_{mi} - \varepsilon_{me} - \varepsilon_{ti} + \varepsilon_{te})^2 \;&=\; \varepsilon_{m_i}^2 + \varepsilon_{me}^2 \\ &+\; \varepsilon_{t_i}^2 + \varepsilon_{te}^2 - 2\varepsilon_{ti}\varepsilon_{te} && (.46)\\ &-\; 2\varepsilon_{mi}\varepsilon_{me} - 2\varepsilon_{mi}\varepsilon_{ti} + 2\varepsilon_{mi}\varepsilon_{te} + 2\varepsilon_{me}\varepsilon_{ti} - 2\varepsilon_{me}\varepsilon_{te}, \\ &=\; \varepsilon_{m_i}^2 + \varepsilon_{me}^2 \\ &+\; \left(\varepsilon_{ti} - \varepsilon_{te}\right)^2 && (.47)\\ &-\; 2\varepsilon_{mi}\varepsilon_{me} - 2\varepsilon_{mi}\varepsilon_{ti} + 2\varepsilon_{mi}\varepsilon_{te} + 2\varepsilon_{me}\varepsilon_{ti} - 2\varepsilon_{me}\varepsilon_{te}. \end{aligned}$$

As seen in Section 3.1.2, we assume that each mixing term $E[\varepsilon_m \varepsilon_t] = 0$ and also $E[\varepsilon_{mi}\varepsilon_{me}] = 0$. The variance of the double residual becomes:

$$\begin{aligned} 2Var(\Delta R) &= E\left[\left(\varepsilon_{mi}\right)^2\right] + E\left[\left(\varepsilon_{me}\right)^2\right] + E\left[\left(\varepsilon_{ti} - \varepsilon_{te}\right)^2\right], && (.48)\\ &= \sigma_{cl}^2 + \sigma_{cl}^2 + E\left[\left(\sigma_{ti} - \sigma_{te}\right)^2\right]. && (.49) \end{aligned}$$

By using overlapping sensors, we assume that any tracking error between the two ladders is eliminated, resulting in a value of zero for the expected square of the difference between the internal and external ladder errors, $E\left[\left(\sigma_{ti} - \sigma_{te}\right)^2\right] = 0$. Finally, the overlap method determines the cluster position resolution as the double residual variance:

$$\sigma_{cl}^2 = Var(\Delta R). \tag{.50}$$

A.4. Geometrical corrections of residuals

The overlap method assumes that both residuals are measured in the same plane, which is not the case for the Silicon Vertex Detector (SVD) because of its windmill architecture (refer to Figure 3.5). The SVD layout has a large angle (i) between two consecutive ladders, which affects the track incident angles (a_i and a_e) for rectangular-barrel sensors. To properly compare residuals measured on the same plane, the projection of the external residual onto the internal ladder is computed. This study does not consider forward-slanted sensors, which add more complexity because of their non-zero polar angle.

On the u/P Side, the pairs of residual could be represented in $r\phi$ projection, as shown in Figure 33. By drawing a plane parallel to the internal ladder passing by the external cluster position, a triangle is formed between the parallel internal plane, external ladder and the tracks.

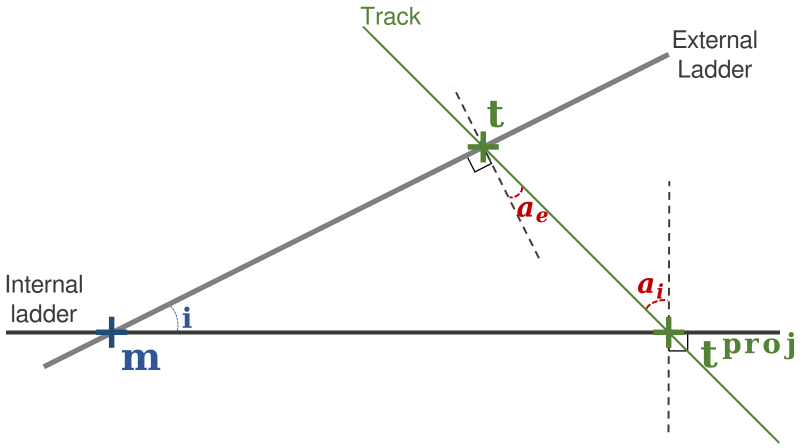


Figure 33. – Schematic view in the $r\phi$-direction of the external residual onto a parallel plane to the internal ladder parallel to the tracks.

Using the sine law in the above triangle allows us to calculate the projection of the external residual:

$$\frac{|t - t^{proj}|}{sin(i)} = \frac{|m - t|}{sin(\pi/2 - a_i)} = \frac{|m - t^{proj}|}{sin(\pi/2 + a_e)}, \tag{.51}$$

$$\frac{|t - t^{proj}|}{sin(i)} = \frac{R_e}{cos(a_i)} = \frac{R_e^{proj}}{cos(a_e)}. \tag{.52}$$

By defining a correction factor C and re-normalizing event by event, the double residual ΔR could be defined in the u/P side by:

$$\Delta R = \frac{R_i - CR_e}{\sqrt{1 + C^2}}, \tag{.53}$$

with:

$$C = \frac{cos(a_e)}{cos(a_i)}. \tag{.54}$$

On the v/N Side, the geometrical correction does take into account the 3D geometry of the SVD because the two ladders can't be represented in the same plane following the v/N side.

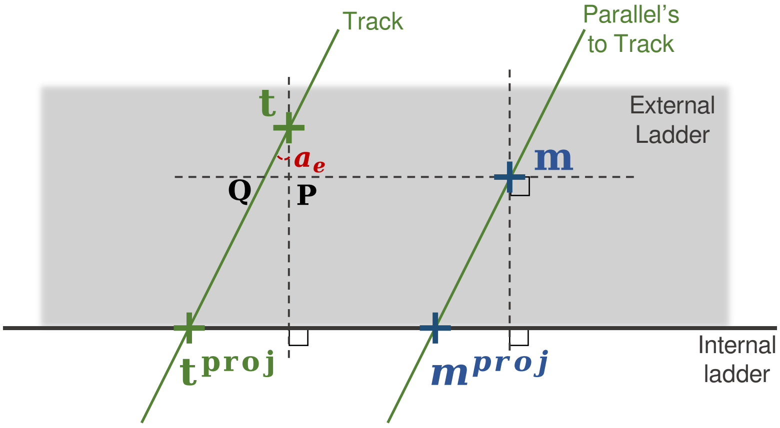


Figure 34. – Schematic view in the z-direction, orthogonal to the internal ladder, of the external residual onto a parallel plane to the internal ladder parallel to the tracks.

In Figure 34, the situation can be effectively represented by a plane that is perpendicular to the internal ladder. This enables us to define the projection of the external ladder residual as $R_e^{proj} = m^{proj} - t^{proj}$, while the external residual is defined by $R_e = m - P$, instead of $m - t$, to account for the removal of the u/P contribution. As a result, the projection can be divided into two contributions R_e and the QP length, as:

$$R_e^{proj} = R_e + |Q - P| \tag{.55}$$

$$= R_e + |t - P| tan(a_e) \tag{.56}$$

The distance between t and P, is the projection of the u/P external residual on our viewing plane, Figure 35, and is $sin(i) R_e^{u/P}$. So:

$$R_e^{proj} = R_e + R_e^{u/P} sin(i) tan(a_e), \tag{.57}$$

and the double residual ΔR in the v/N side becomes:

$$\Delta R = \frac{R_i - (R_e + R_e^{u/P} sin(i) tan(a_e))}{\sqrt{2}}. \tag{.58}$$

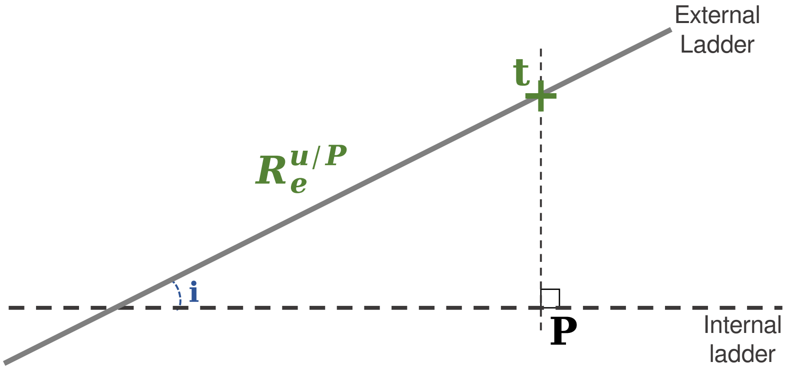


Figure 35. – Projection of the u/P external residual contribution in the v/N projection view 34.

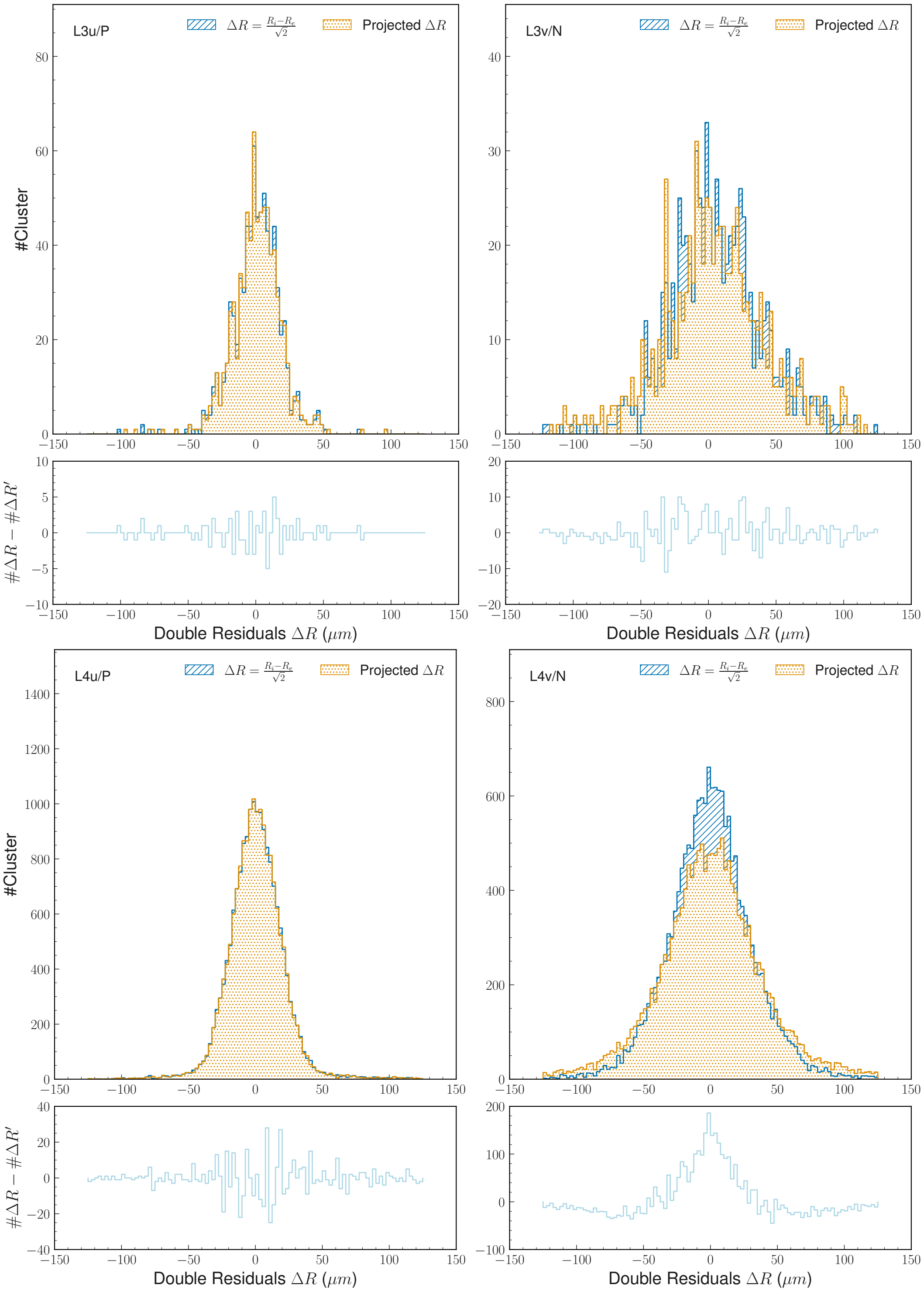


Figure 36. – Double residuals distribution without (red) and with (blue) the geometrical correction at the top, and the height bin difference between both of the distributions at the bottom. The correction does not affect the distribution so much because the correction factors are close to one.

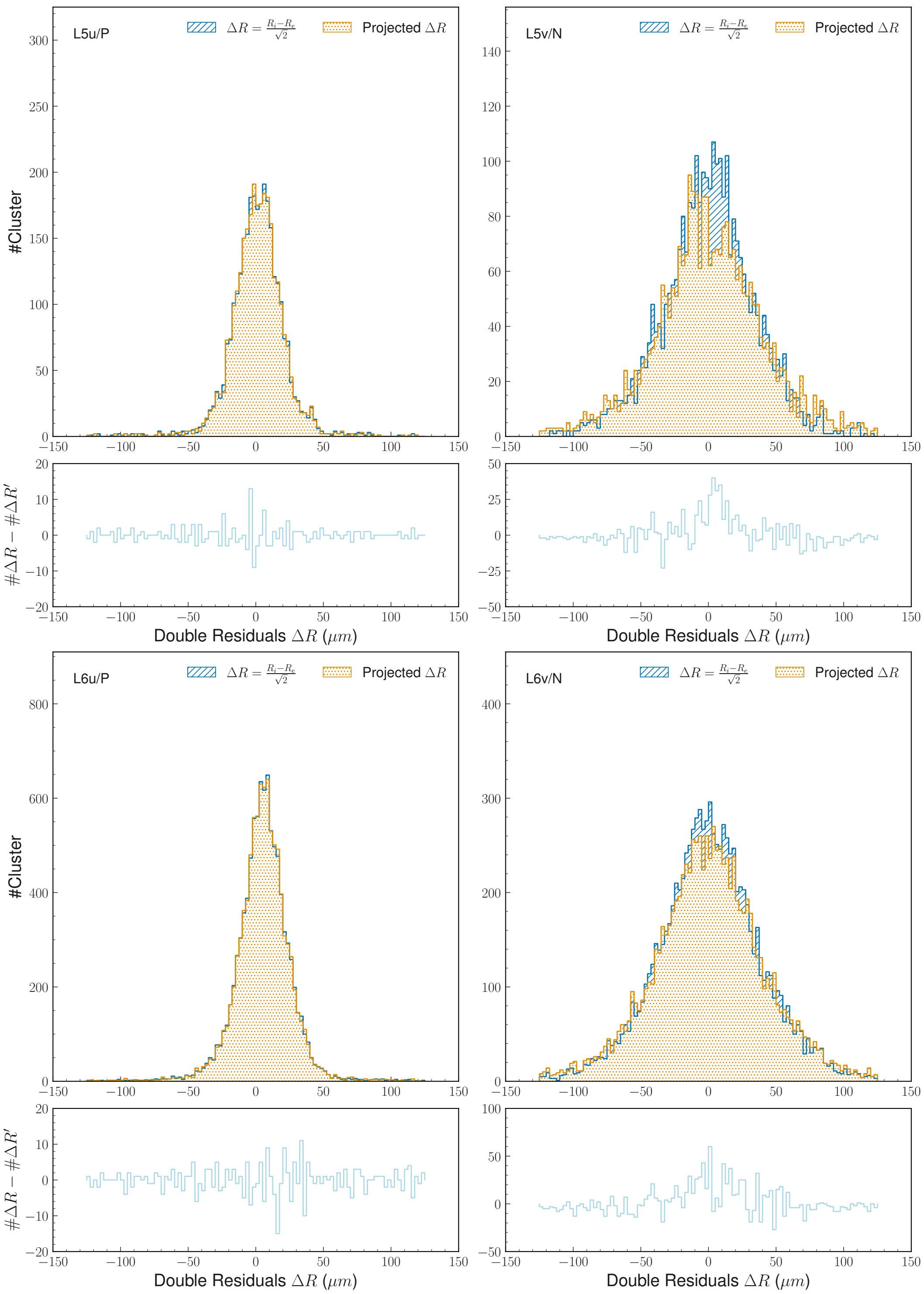


Figure 36. – Double residuals distribution without (red) and with (blue) the geometrical correction at the top, and the height bin difference between both of the distributions at the bottom. The correction does not affect the distribution so much because the correction factors are close to one.

A.5. Overlapping method discrepancy checks

The assumption motivating the overlap method is that the errors on the track extrapolation cancel out in the double residuals, but this might not be the case due to the dependence on the track incident angle and the non-trivial projection of the external residuals onto the internal ladder.

A.5.1. Extrapolated track error cancellation

Our first check aims to verify this possibility. We show in Figure 37 the difference between the errors on the intercepts $\Delta\sigma_t$ for the internal and external residuals on the simulation samples: the plotted distributions have a non-null mean and respectively a width of $\sim 0.5\,\mu$m for u/P and $\sim 5\,\mu$m for v/N, which points to a non-complete cancellation of the track errors in pair method. However, we observed that on simulation, this effect is of the order of $\sim \sqrt{2} \times 0.5(5) \sim 1(8)\,\mu$m for u/P (n/V) side and cannot fully explain the discrepancy with respect to other methods. From this study, we conclude the non-cancellation of the track errors cannot be the only responsible for the observed differences.

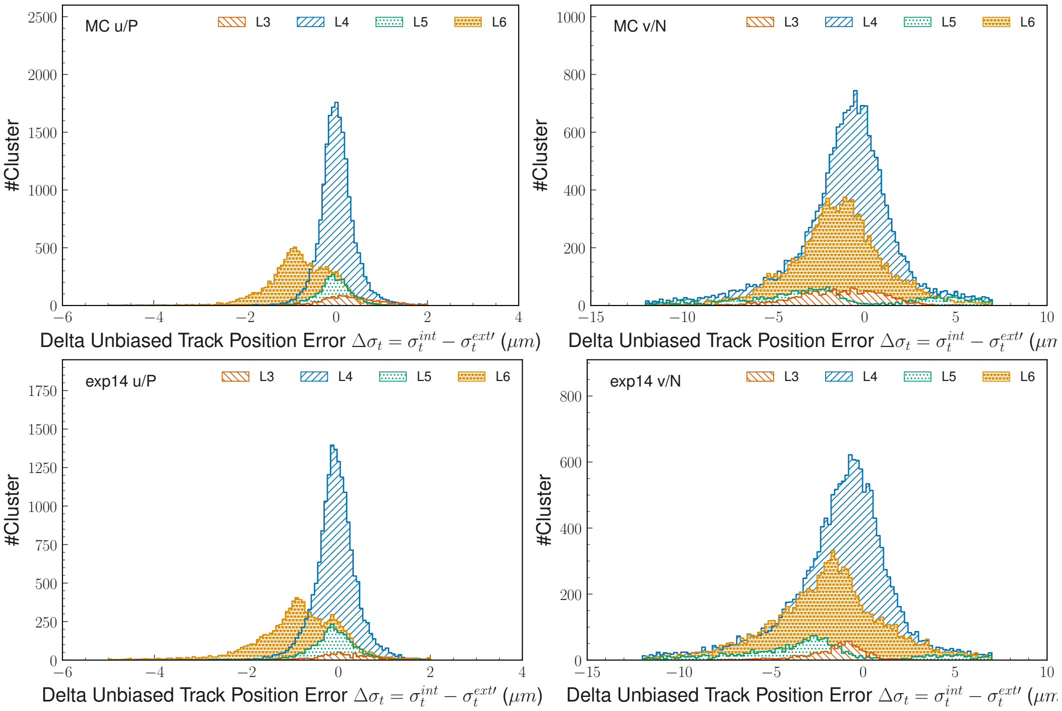


Figure 37. – Difference between internal and external unbiased track position error for each side and layer for simulated samples, top for MC samples and data in the bottom plots. Geometrical correction is applied when subtracting the external track error.

A.5.2. Double residual decomposition

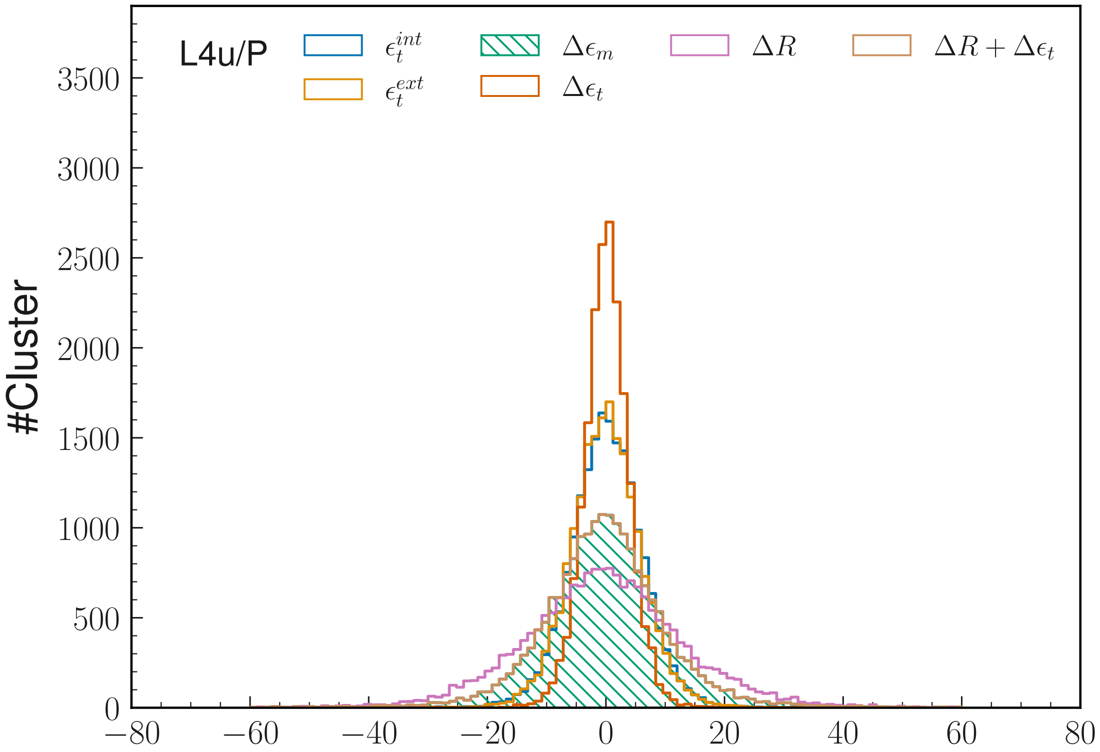


Figure 38. – Comparison between: double true residual $\Delta\varepsilon_m$, double track residual $\Delta\varepsilon_t$, double residual ΔR, and the sum of double track residual and double residual.

Secondly, we investigate the decomposition of the double residuals as:

$$\Delta R = \Delta\varepsilon_m - \Delta\varepsilon_t \tag{.59}$$

and we want to check whether the following hypothesis is correct:

$$Var(\Delta R) = Var(\Delta\varepsilon_m) + Var(\Delta\varepsilon_t) \tag{.60}$$

The above equation, in fact, would imply:

$$Var(\Delta\varepsilon_m) \overset{?}{=} Var(\Delta R) - Var(\Delta\varepsilon_t) \tag{.61}$$

and therefore, we could estimate the true resolution, here corresponding to the width of the double true residuals $\Delta\varepsilon_m$, as the difference between the width of the double residual $\sigma(\Delta R)$ and the width of the double true track residuals $\Delta\varepsilon_t$, which needs to be computed from the simulation.

This study is performed on simulation since it's the only place where we can access the true residuals $\Delta\varepsilon_t$. We select layer 4 on the u/P side, which is the one with a larger acceptance region and higher statistics available for the test. In Figure 38, it is shown that the decomposition of equation .59 holds, but the relation among the widths of the single distributions instead (equation .60) is not satisfied, as reported in the table 20. Therefore, the left side of equation .61 computed as the width of $\Delta\varepsilon_m$ doesn't equal the difference between the widths of ΔR and $\Delta\varepsilon_t$, and the correct expression of the previous relation needs to account for the correlation between $\Delta\varepsilon_m$ and $\Delta\varepsilon_t$ with the additional covariance term:

$$Var(\Delta R) = Var(\Delta\varepsilon_m) + Var(\Delta\varepsilon_t) - 2Cov(\Delta\varepsilon_m, \Delta\varepsilon_t) \qquad (.62)$$

This time we check the relation:

$$Var(\Delta\varepsilon_m) \overset{?}{=} Var(\Delta R) - Var(\Delta\varepsilon_t) + 2Cov(\Delta\varepsilon_m, \Delta\varepsilon_t) \qquad (.63)$$

$$81.18 \overset{?}{=} 144.36 + 2 \times (-28.41) \qquad (.64)$$

$$81.18 \simeq 87.54 \qquad (.65)$$

which from the numerical results reported in the table20 holds within a $\sim 6\ (\mu\text{m})^2$ difference. The observed correlation between $\Delta\varepsilon_m$ and $\Delta\varepsilon_t$ hasn't been observed with the other methods and could be explained by the fact that the track fit is not sufficiently unbiased. The correlation might be due to the fact that when the track intercept is computed for the external layer, the hit from the internal one is used, causing a correlation between ε_m^{int} and ε_t^{ext}, and similarly for the other couple ε_m^{ext} and ε_t^{int}. A viable way to solve this would be to have a fully unbiased track fit that does not use any of the two hits, but it would require rerunning the track fit in the reconstruction of the data used, implying a complete reprocessing of the ntuple and not affordable at present as further offline check on the method.

Estimating the true resolution as $\sqrt{87.54} = 9.4\ \mu$m gives a result closer to what is measured in the other methods for the same layer 4, u/P side, but this correction is based on the MCTruth information from simulation and wouldn't be accessible on data. This would require further validation studies, and we do not think it is already applicable.

Table 20. – Position and width of the different double residuals, taken with the `median` and `sigma-68` respectively in the MC sample for the Layer 4 u/P.

	`median` (μm)	`sigma-68` (μm)
Internal track residual ε_t^{int}	$0.14 \pm 0.00 \pm 0.02$	$5.74 \pm 0.06 \pm 0.05$
External track residual ε_t^{ext}	$-0.01 \pm 0.00 \pm 0.01$	$5.58 \pm 0.06 \pm 0.03$
Delta true residual $\Delta\varepsilon_m$	$-0.32 \pm 0.00 \pm 0.00$	$9.01 \pm 0.09 \pm 0.29$
Delta track residual $\Delta\varepsilon_t$	$0.14 \pm 0.00 \pm 0.01$	$3.52 \pm 0.04 \pm 0.08$
Double residual ΔR	$-0.44 \pm 0.00 \pm 0.05$	$12.52 \pm 0.13 \pm 0.40$

B. L1 trigger lines definition

In Chapter4, we consider the ECL low-multiplicity trigger lines lmlX and some CDC lines, defined in Tables 21 and Table 22, respectively. Among all the lmlX and CDC trigger lines, we are only using the ones that were not prescaled during the data taking. Performance of the used triggers has been extensively studied within the Tau Working Group of Belle II in [123].

Table 21. – Definitions of L1 low-multiplicity trigger lines (lmlX) used in Chapter 4. The centre of the mass frame is referred as "CM".

Trigger line	ECL cluster(s)	Energy of cluster(s) E	Polar angle of cluster(s)	Opening angle in CM $\Delta\phi^{\mathrm{CM}}$	Other requirements
$lml6$	1	$E^{\mathrm{CM}} > 1\,\mathrm{GeV}$	$[32.2°; 128.7°]$	-	no other cluster with $E > 300\,\mathrm{MeV}$
$lml7$	1	$E^{\mathrm{CM}} > 1\,\mathrm{GeV}$	$[18.5°; 31.9°]$ or $[128.7°; 139.3°]$	-	no other cluster with $E > 300\,\mathrm{MeV}$
$lml8$	2	$E > 250\,\mathrm{MeV}$	-	$[170°; 190°]$	no cluster with $E^{CM} > 2\,\mathrm{GeV}$
$lml9$	2	$E > 250\,\mathrm{MeV}$ $E < 250\,\mathrm{MeV}$	-	$[170°; 190°]$	no cluster with $E^{CM} > 2\,\mathrm{GeV}$
$lml10$	2	-	$\sum \theta^{\mathrm{CM}} \in [160°; 200°]$	$[160°; 200°]$	no cluster with $E^{CM} > 2\,\mathrm{GeV}$
$lml12$	≥ 3	≥ 1 with $E > 500\,\mathrm{MeV}$	$[18.5°; 139.3°]$	-	not an ECL Bhabha
hie	-	$E_{tot} > 1\,\mathrm{GeV}$	-	-	not an ECL Bhabha

Table 22. – Definitions of L1 CDC trigger lines used in Chapter 4. The centre of the mass frame is referred as "CM".

Trigger line	Full track(s)	Short track(s)	Neuro 3D track(s) with $\lvert z\rvert < 20\,\mathrm{cm}$	Opening angle $\Delta\phi$	Other requirements
ffy	≥ 3	-	≥ 1	-	-
fyo	≥ 2	-	≥ 1	$> 90°$	not an ECL Bhabha

C. Detailed background composition after cut-based preselctions

Table 23. – Detailled background composition normalized to 362 fb^{-1} after applying each set of preselection and LID requirements.

Name	Preselection	ε_{sig}^{rel} (%)	ε_{sig}^{abs} (%)	ε_{bkg}^{rel} (%)	N_{bkg}	$N_{\tau-pair}$	$N_{q\overline{q}}$	$N_{B\overline{B}}$	N_{lowm}	$N_{\gamma\gamma}$	$N_{\ell\ell}$	$N_{ee\ell\ell}$	N_{eehh}	$N_{\mu\mu\ell\ell}$	$N_{4\tau}$	N_{hhISR}
Reference	$0.3 < \theta_{miss}^{CM} < 2.7$	96.88	31.11	89.99	938.82	3.08	287.52	0.00	648.22	0.00	49.29	44.32	0.00	554.61	0.00	0.00
Set 1	$0.3 < \theta_{miss}^{CM} < 2.7$ $0.89 < Thrust < 0.97$	95.48	30.67	30.83	321.64	2.96	270.87	0.00	47.82	0.00	8.56	9.63	0.00	29.63	0.00	0.00
Set 2	$0.3 < \theta_{miss}^{CM} < 2.7$ $0.935 < Thrust_{ROE\tau} < 0.95$	96.35	30.94	61.78	644.50	2.58	244.94	0.00	396.97	0.00	13.25	29.66	0.00	354.07	0.00	0.00
Set 3	$0.3 < \theta_{miss}^{CM} < 2.7$ $E_{vis}^{CM} < 10.$	90.54	29.08	14.89	155.30	2.98	127.81	0.00	24.52	0.00	6.91	12.62	0.00	4.98	0.00	0.00
Set 4	$0.3 < \theta_{miss}^{CM} < 2.7$ $E_{miss}^{CM} > 0.6$	90.22	28.98	14.69	153.29	2.91	125.85	0.00	24.52	0.00	6.91	12.62	0.00	4.98	0.00	0.00
Set 5	$0.3 < \theta_{miss}^{CM} < 2.7$ $p_{miss}^{T,CM} > 0.4$	91.12	29.26	15.89	165.74	2.77	135.08	0.00	27.90	0.00	7.11	15.50	0.00	5.28	0.00	0.00
Set 6	$0.3 < \theta_{miss}^{CM} < 2.7$ $M_{ROE} < 2.2$ $-5. < \Delta E_{ROE} < -0.2$	90.76	29.15	16.49	172.08	2.62	106.08	0.00	63.39	0.00	14.34	22.93	0.00	26.11	0.00	0.00

D. Input BDT variables importance

The BDT input variables ranked by importance are given in Figure 39.

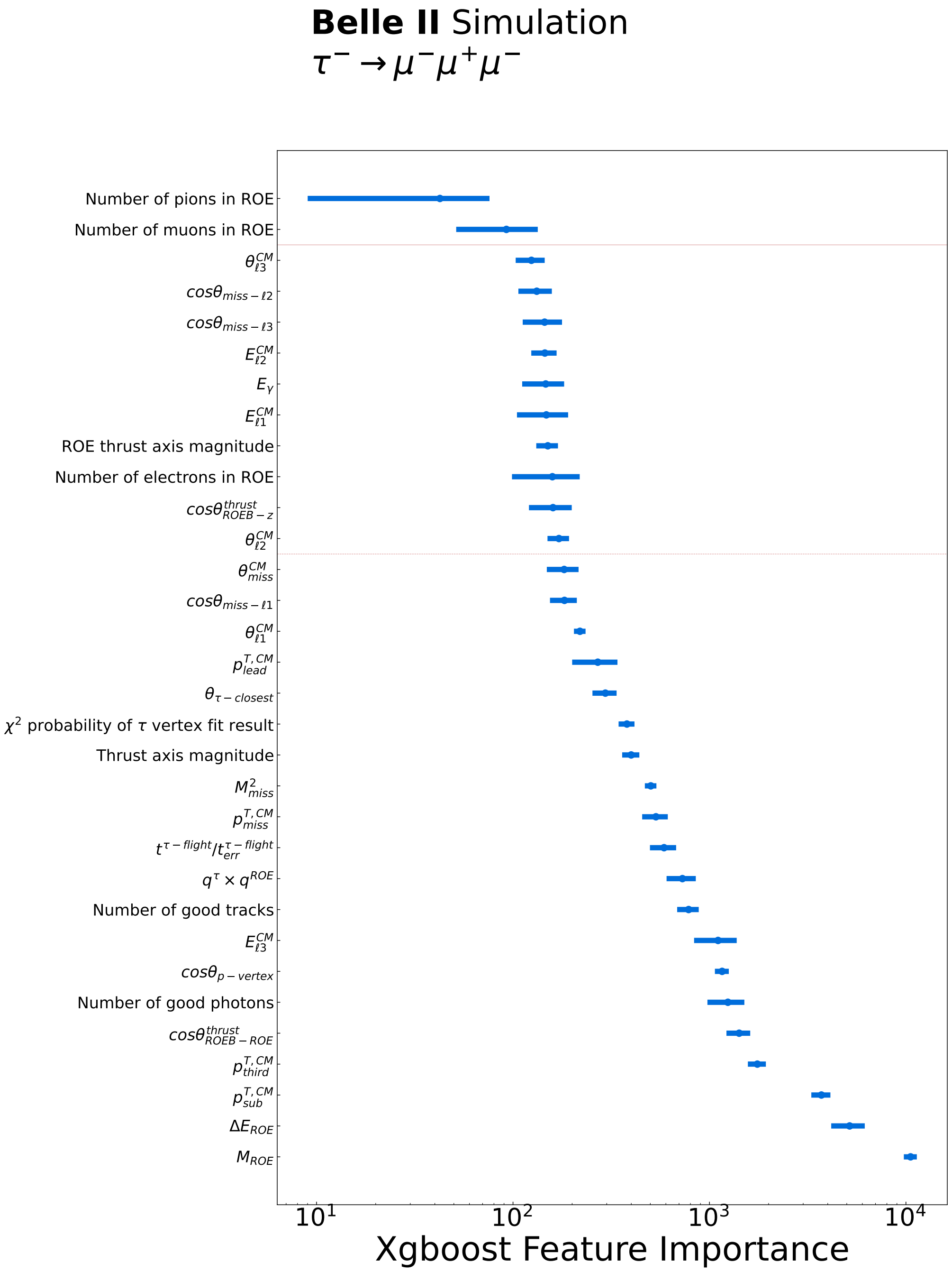


Figure 39. – Variables used as input to the BDT based background rejection for the $\tau^- \to \mu^-\mu^+\mu^-$ mode, ranked by their importance (best variable at the bottom) computed as "the average gain of splits which use the feature"[117].

E. Dalitz plots

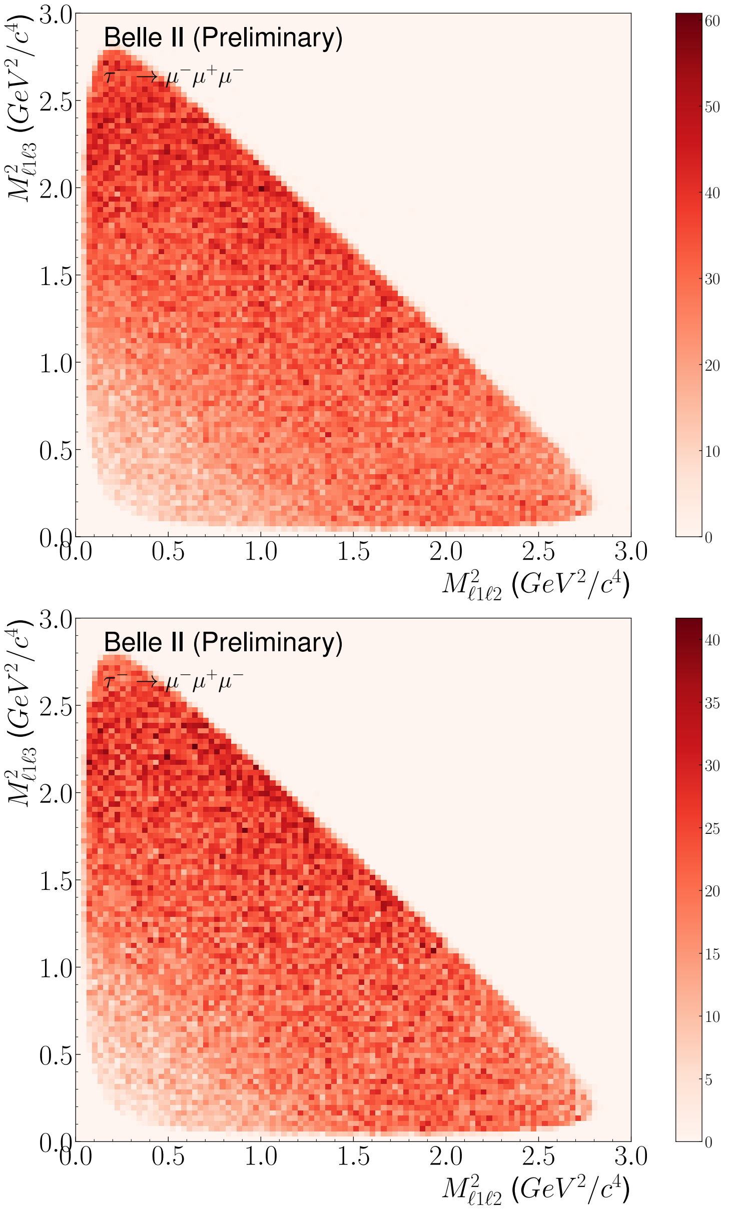


Figure 40. – Signal distribution in the 2D plane defined by the mass squared of the opposite charge muons. They are given after the reconstruction (top) and after the background rejection (bottom).

To prevent the exclusion of a phase space region that may affect various New Physics (NP) model candidates for $\tau^- \to \mu^-\mu^+\mu^-$ LFV decays, we examined the Dalitz plots shown in Figure 40. The region at low masses is unfavoured nevertheless, this feature is present before and after the background rejection.

F. Data-MC Comparison in sidebands

F.1. After preselections

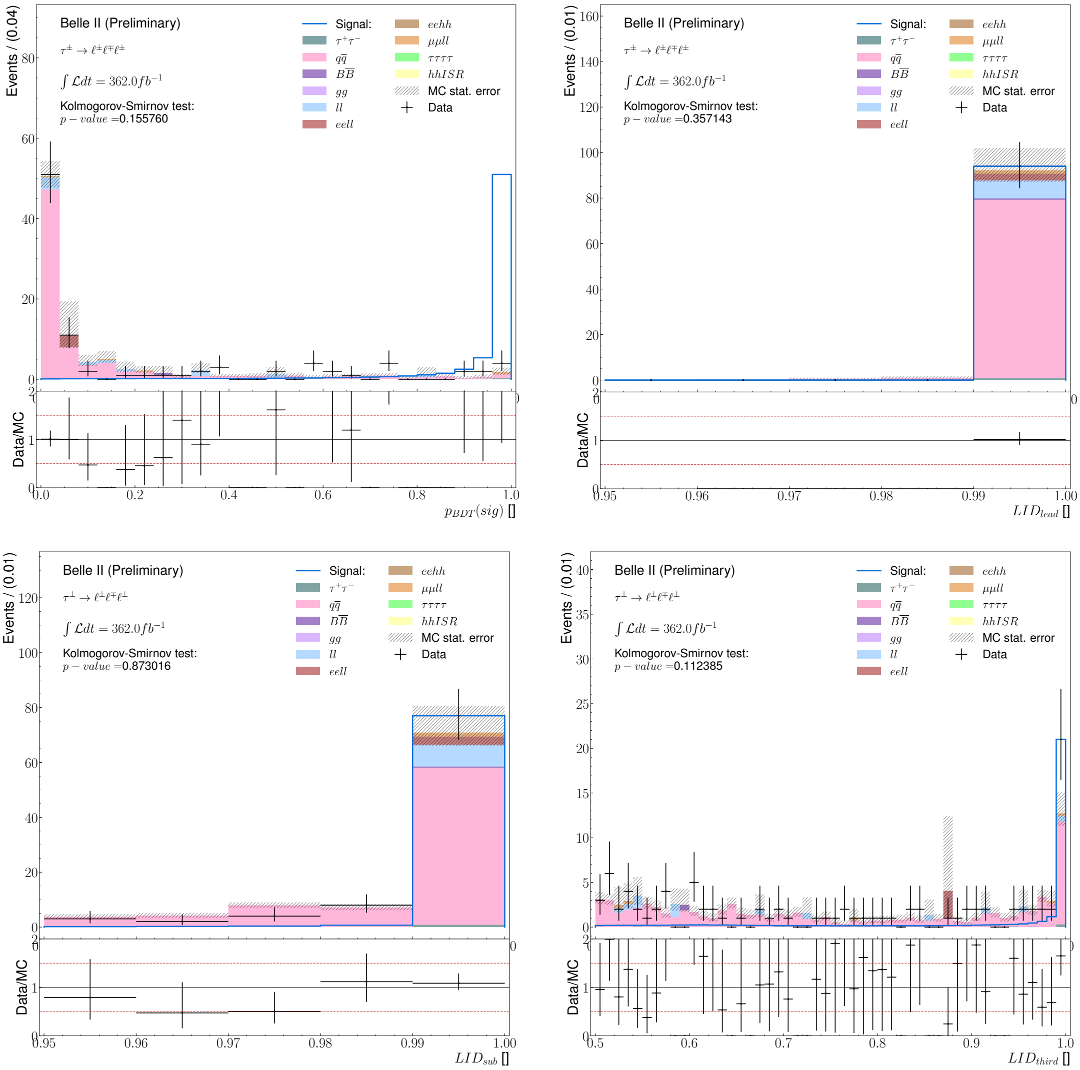


Figure 41. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

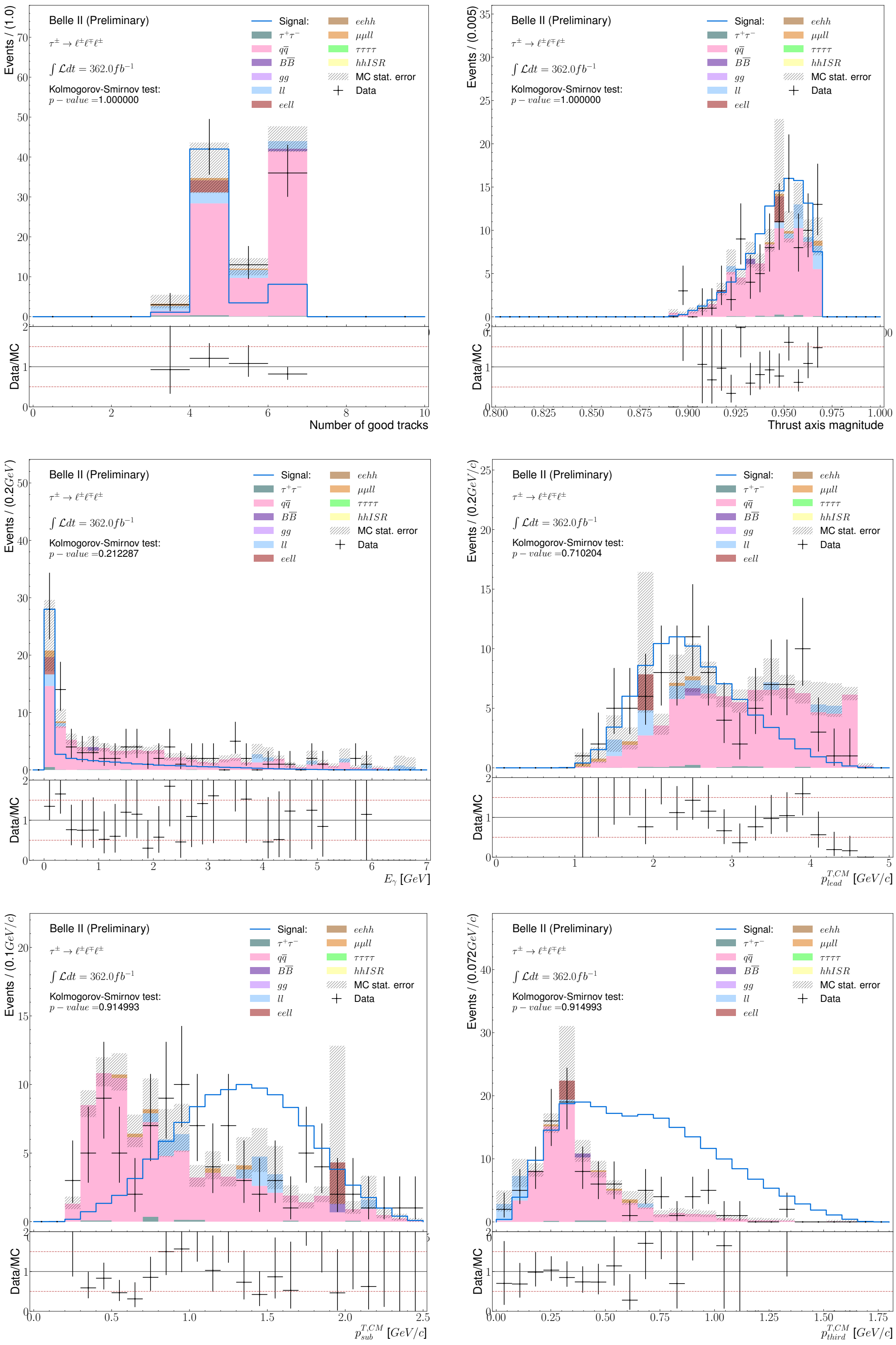


Figure 41. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \rightarrow \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables are taken as inputs to the BDT, after the preselection.

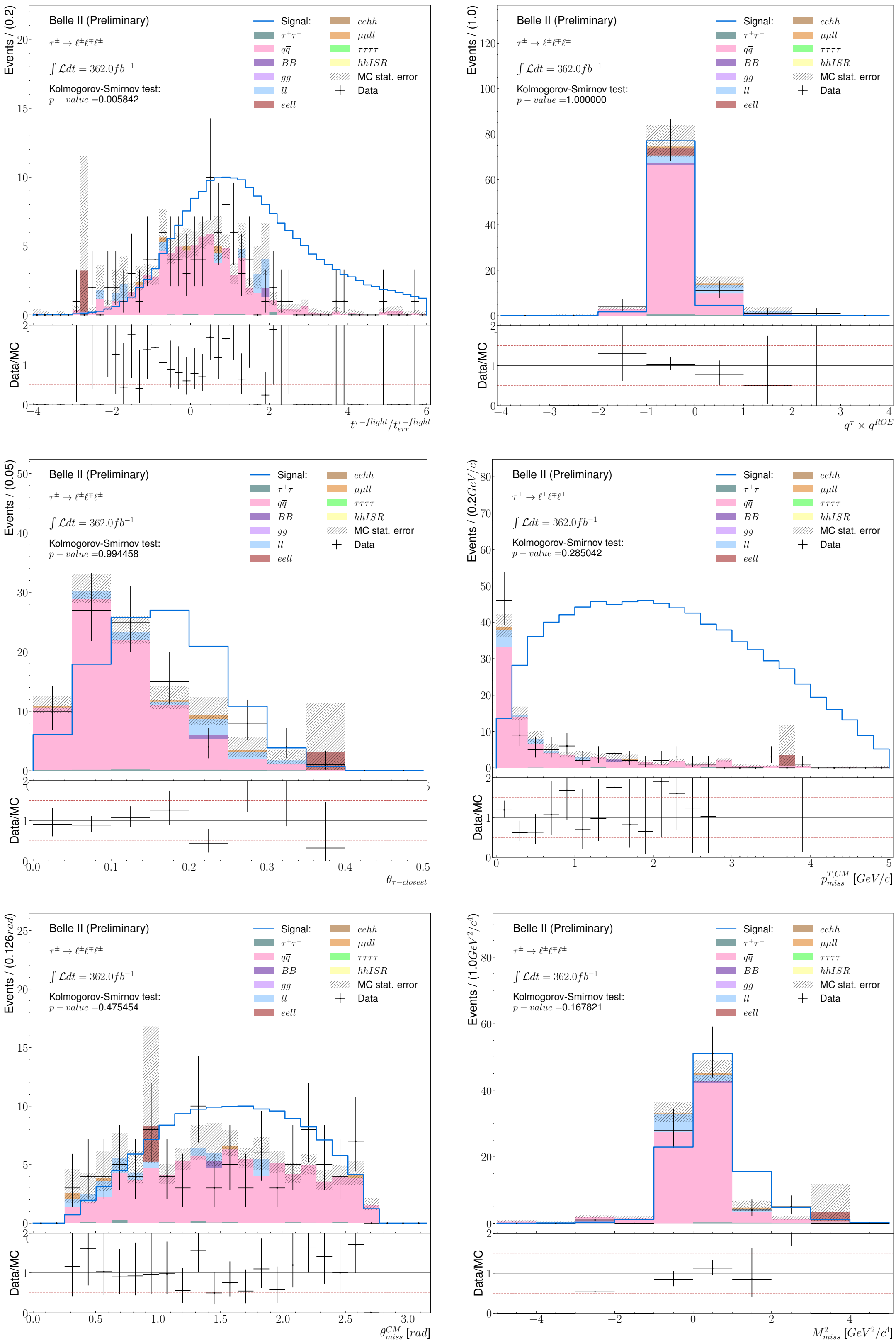


Figure 41. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables are taken as inputs to the BDT, after the preselection.

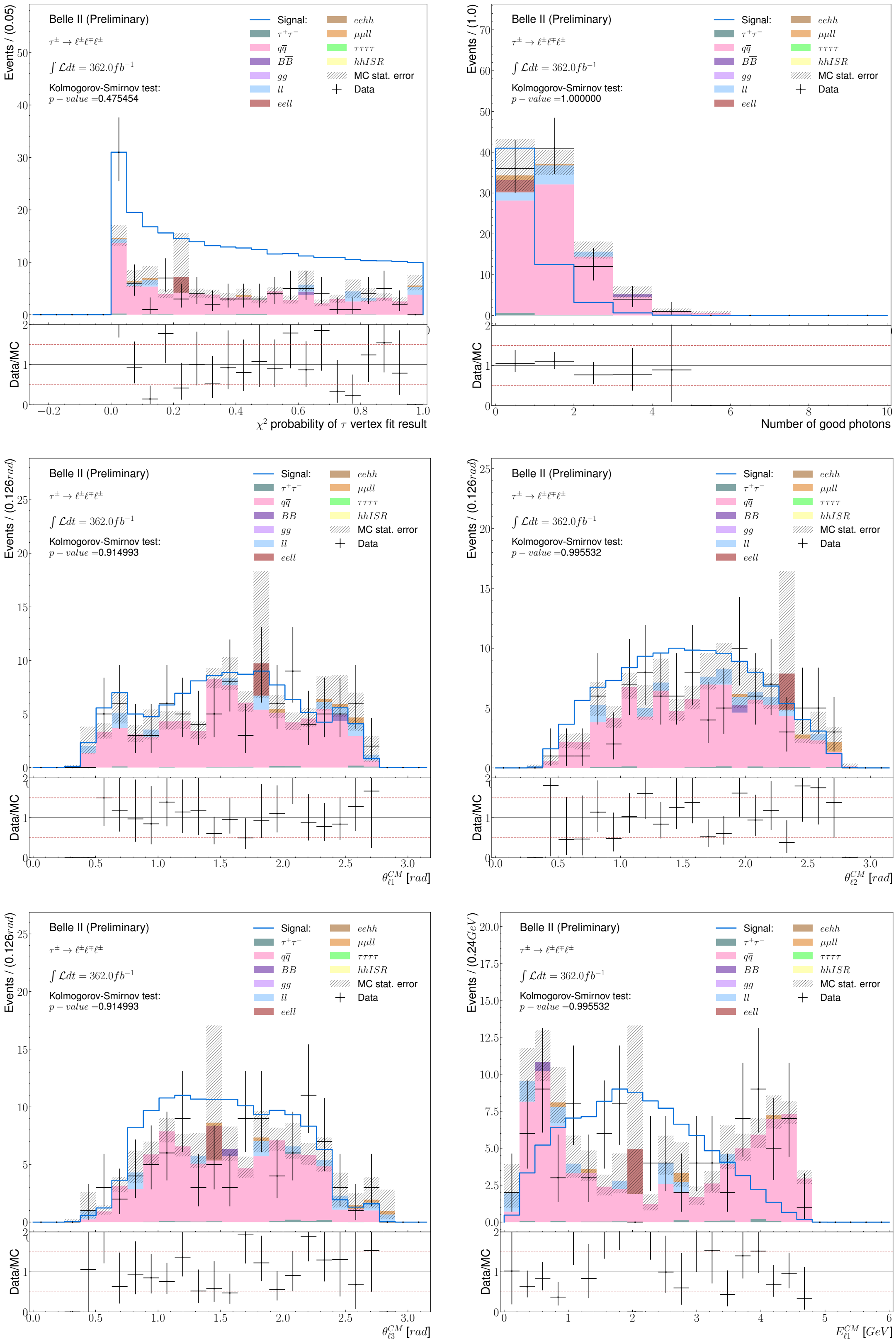


Figure 41. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables are taken as inputs to the BDT, after the preselection.

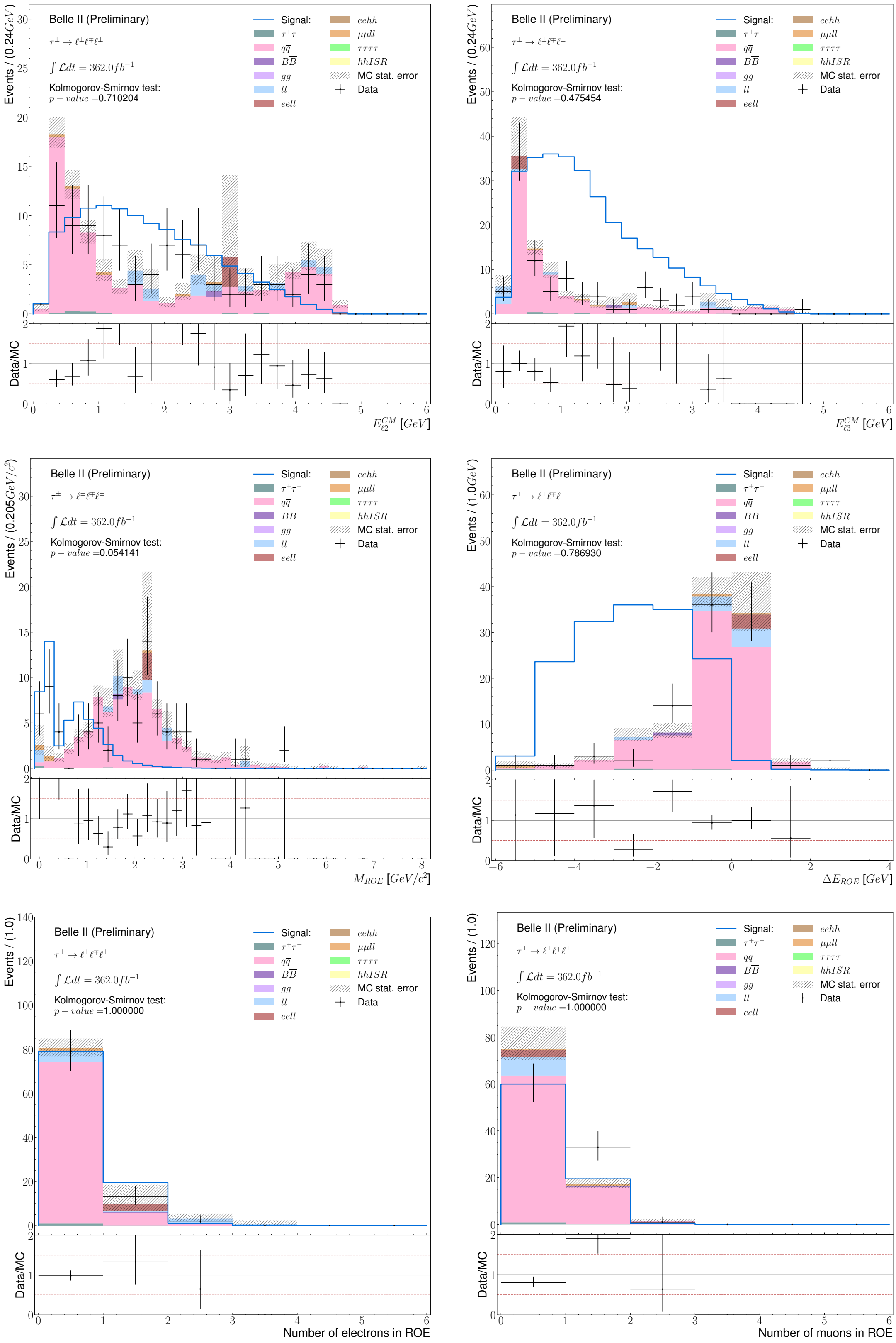


Figure 41. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables are taken as inputs to the BDT, after the preselection.

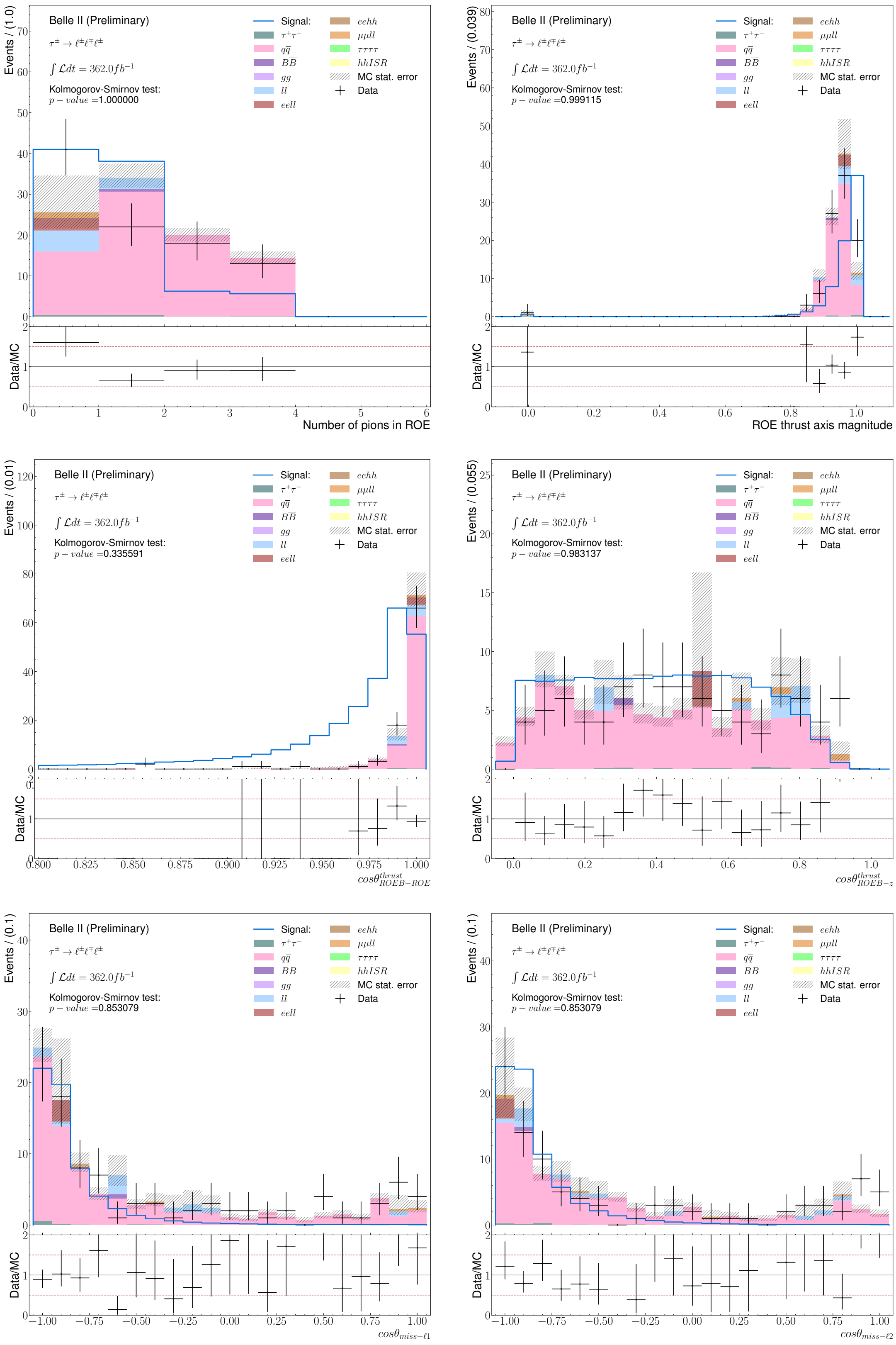


Figure 41. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables are taken as inputs to the BDT, after the preselection.

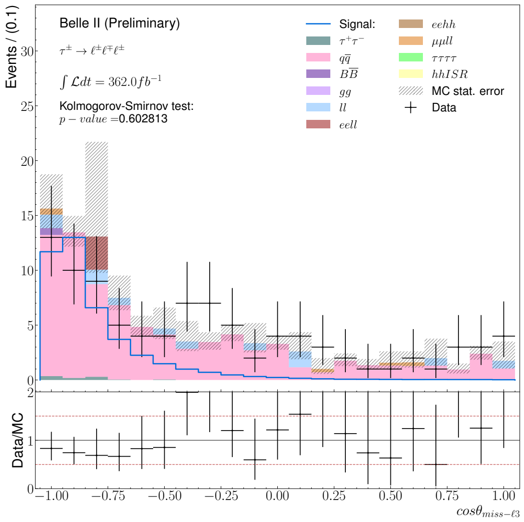


Figure 41. – Data-MC comparison in the $5 - 20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

F.2. After applying BDT output selection

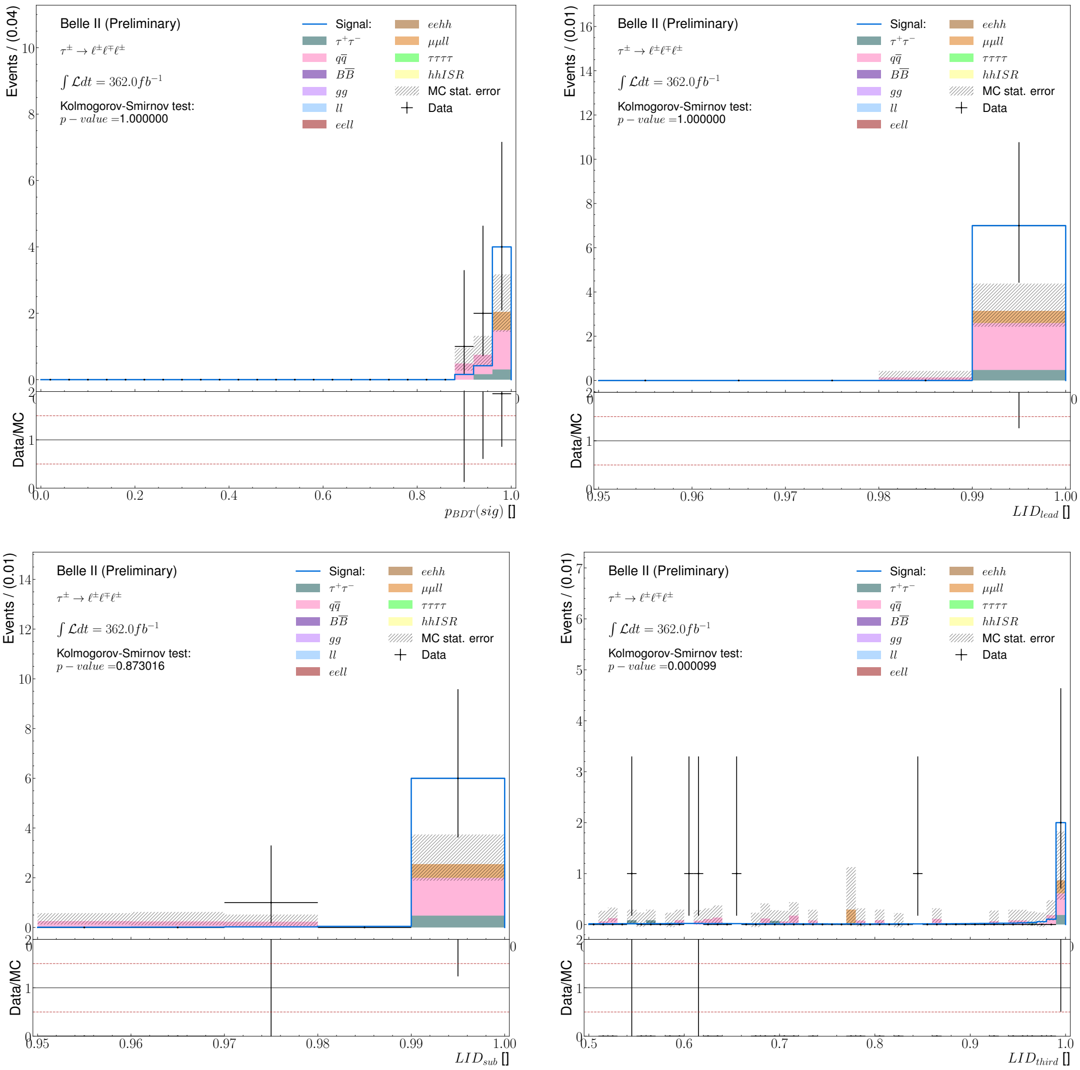


Figure 42. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

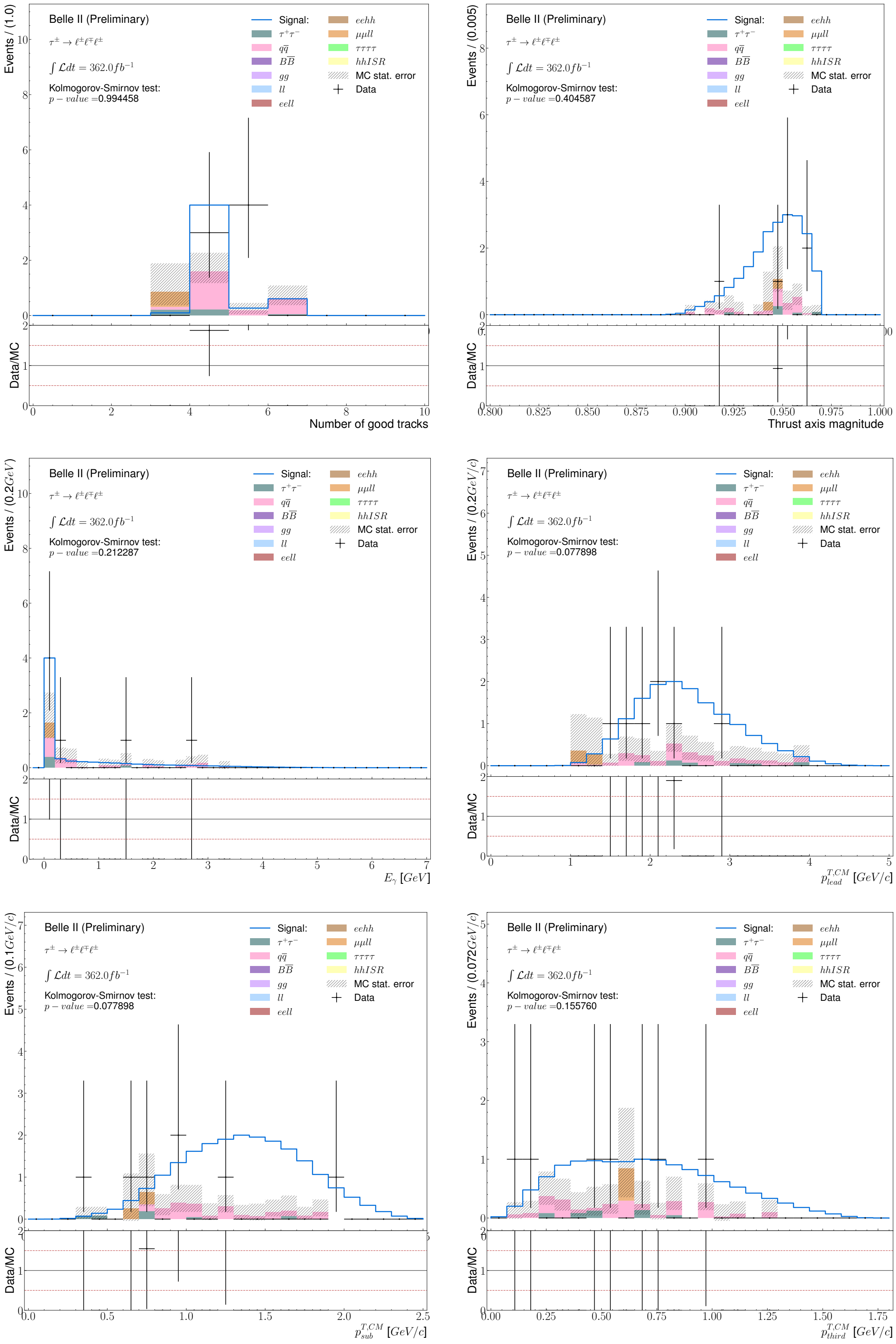


Figure 42. – Data-MC comparison in the 5 – 20(10)δ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

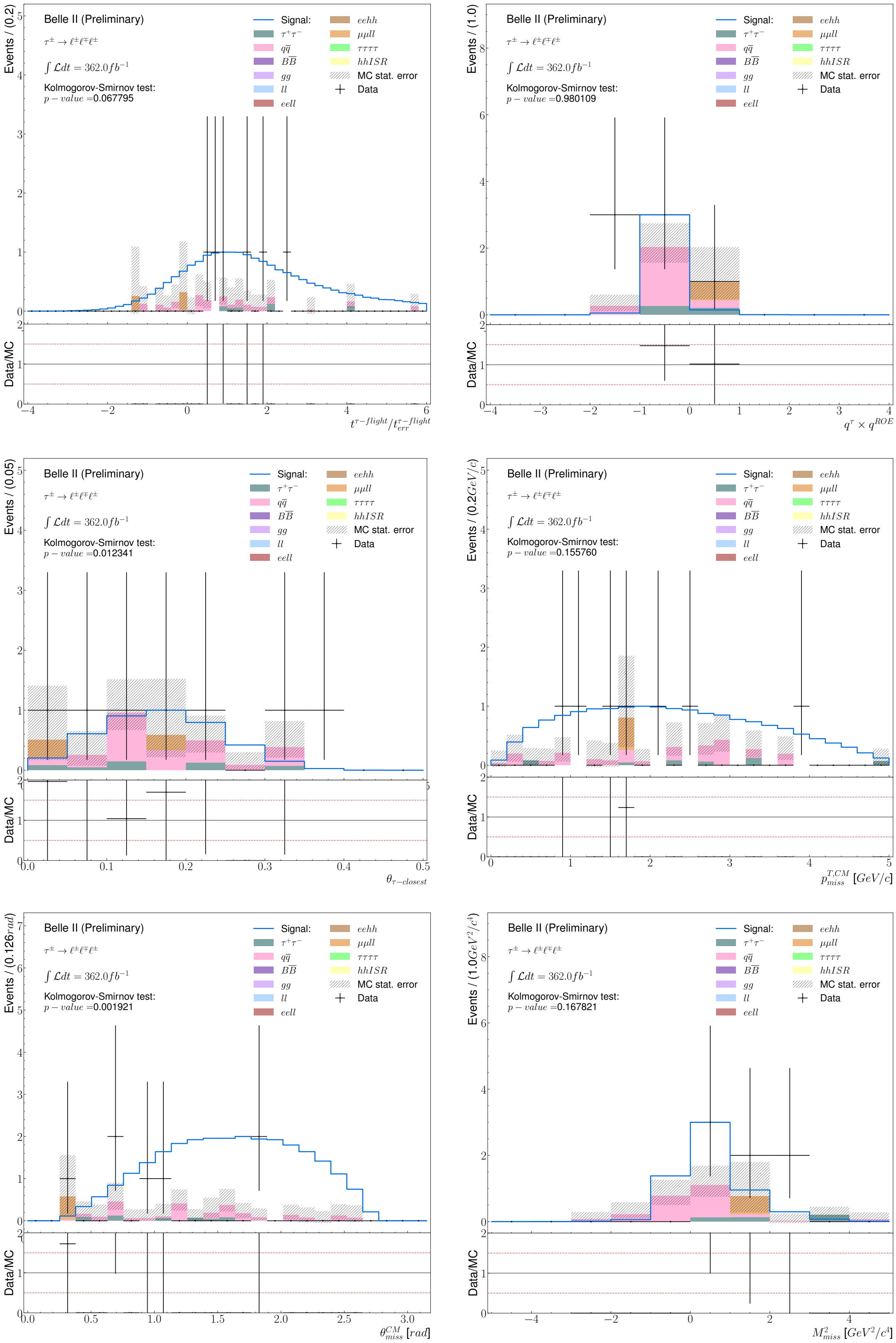


Figure 42. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

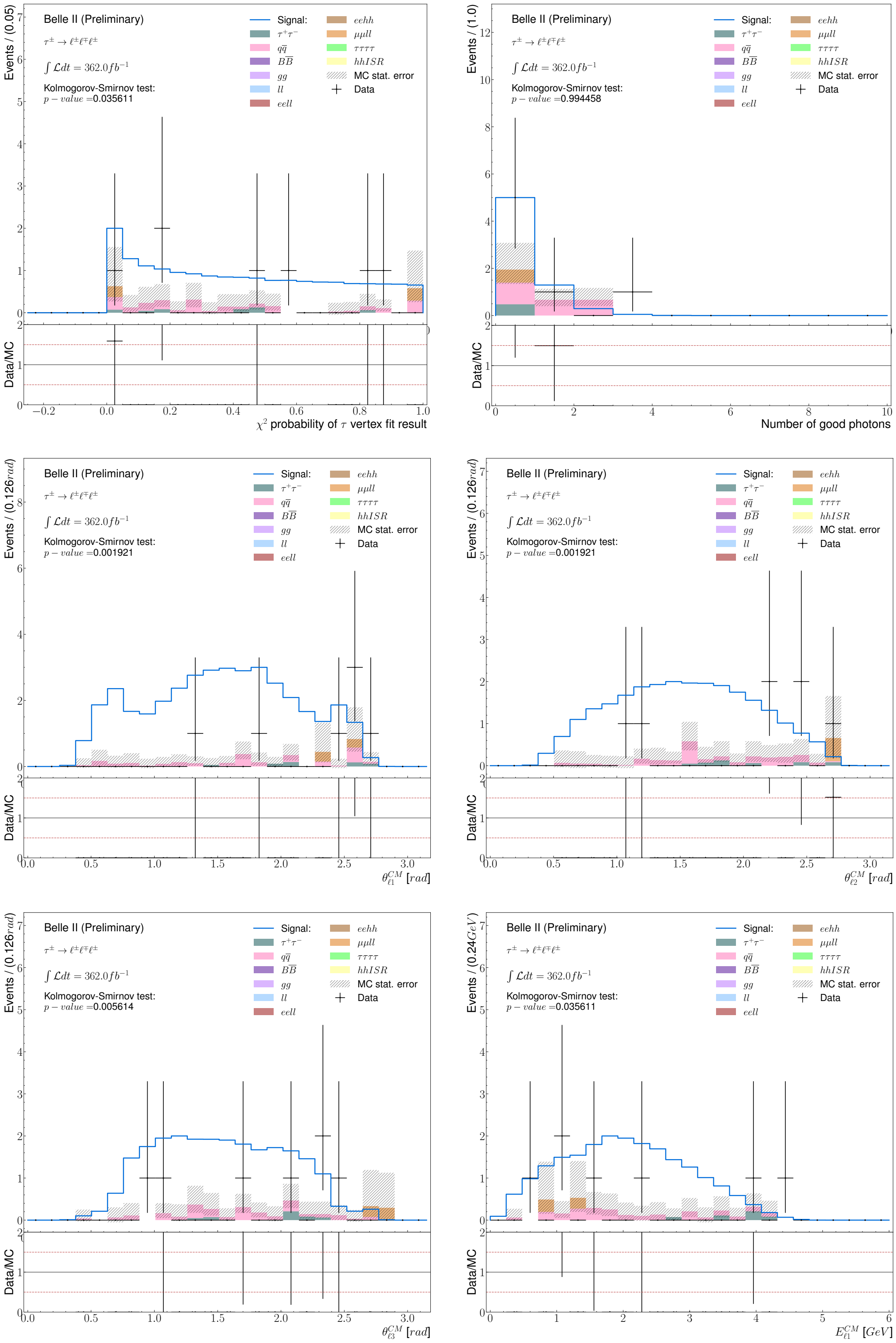


Figure 42. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

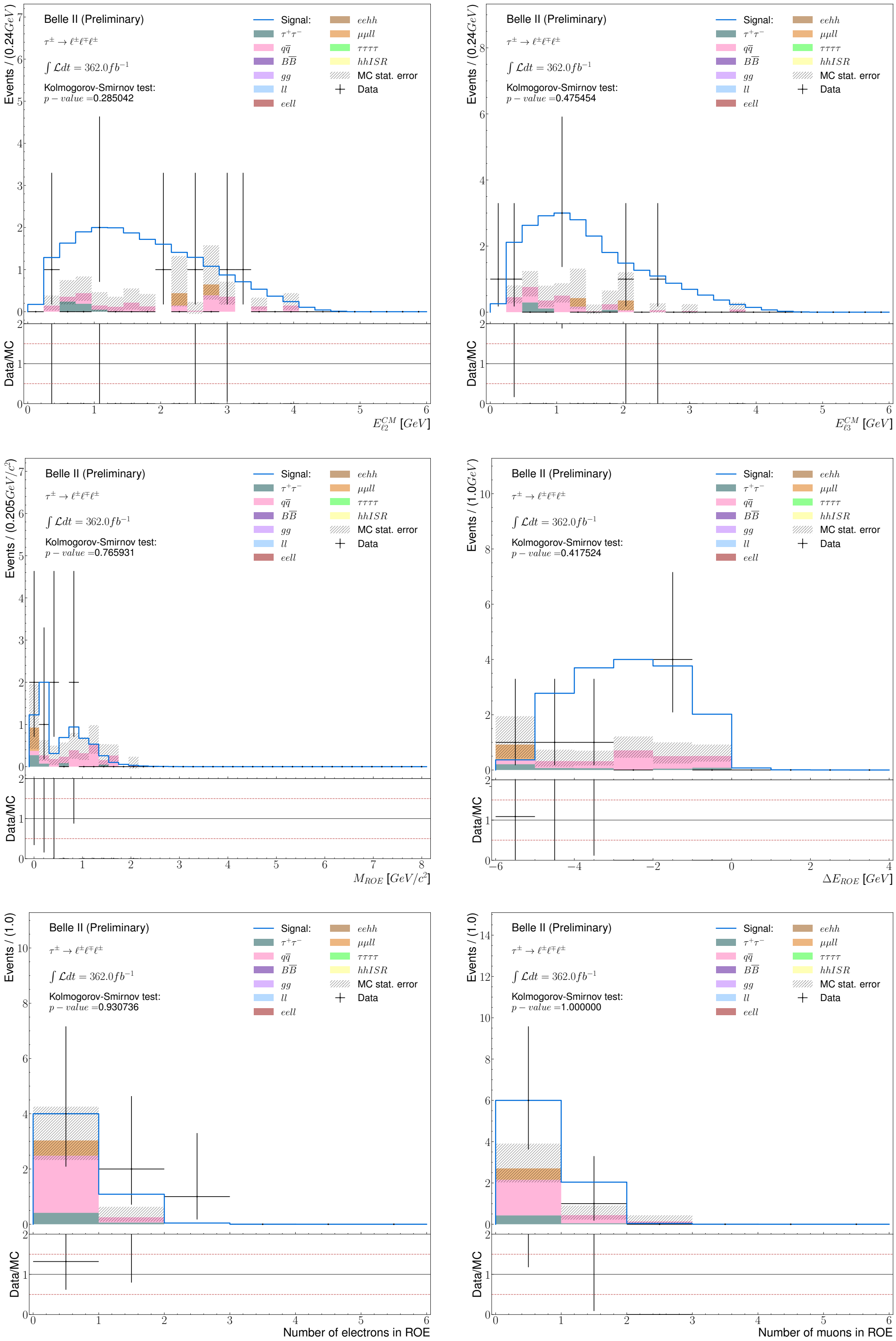


Figure 42. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

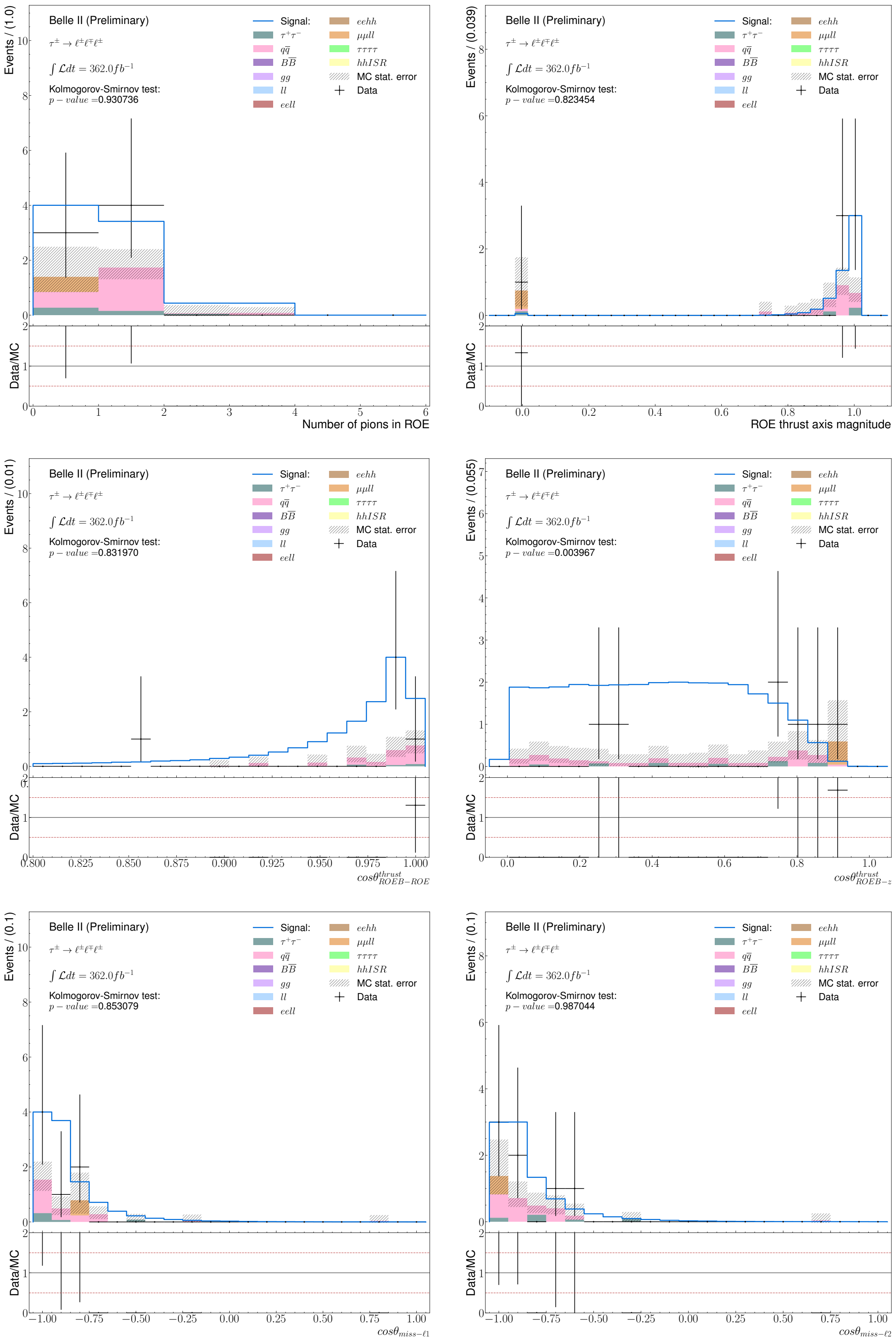


Figure 42. – Data-MC comparison in the $5-20(10)\delta$ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

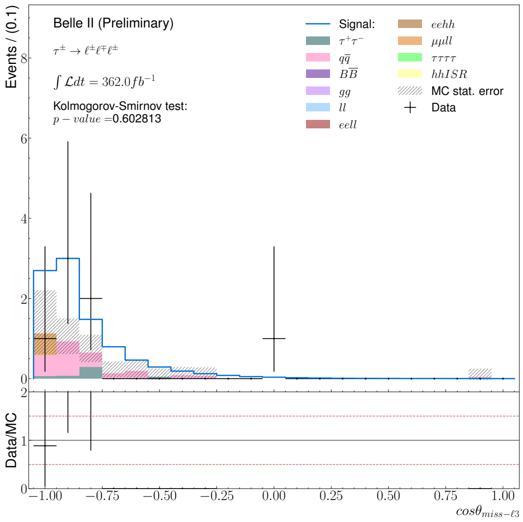


Figure 42. – Data-MC comparison in the 5 – 20(10)δ sidebands of reconstructed $\tau^- \to \mu^-\mu^+\mu^-$ events for BDT signal probability output, LID and variables taken as inputs to the BDT, after the preselection and BDT applied.

List of Figures

List of Tables

List of acronyms

HADP

Hybrid Avalanche Photo-Detector. 67

HER

High Energy Ring. 12, 52

HLT

High Level Trigger. 72, 73

ID

IDentification. 104, 105

IP

Interaction Point. 54, 55, 59, 104, 129

KLM

K_L and Muon detector. 13, 69–71, 73, 74, 112, 191

LER

Low Energy Ring. 12, 52

LFU

Lepton Flavour Universality. 46

LFV

Lepton Flavour Violation. 16, 28, 38, 39, 41–43, 45–50, 76, 78, 98, 100, 101, 105, 106, 110, 149, 152, 157, 174, 190, 192

LID

Lepton IDentification. 104, 112, 113, 138, 141–143, 148, 149, 193, 195, 199

LS1

Long Shutdown 1. 56, 59, 60, 76, 191

MC

Monte-Carlo. 13, 19, 75, 78, 81, 85, 91, 103, 108, 111, 113, 120, 129, 138, 151, 153, 156, 193

MCP-PMTs

Micro-Channel Plate PhotoMultiplier Tubes. 66, 67

MSSM

Minimal Supersymmetric Standard Model. 44–46, 190

NP

New Physics. 36, 42, 43, 49, 174, 197

PID

Particle IDentification. 59, 65, 66, 69, 74

PMNS

Pontecorvo-Maki-Nakagawa-Sakata. 39–41, 43

PXD

PiXel Detector. 12, 59, 62, 63, 65, 71–73, 80, 191

QCD

Quantum ChromoDynamics. 32, 34, 35

QFT

Quantum Field Theory. 31, 33, 36

ROE

Rest-of-Event. 17, 18, 20, 99, 105, 117, 197

SB

Side Bands. 108, 134, 138, 141, 194

SM

Standard Model. 28, 30–33, 35–47, 52, 78, 98, 157

SR

Signal Region. 108, 112, 123–126, 129, 133, 134, 138, 141, 144, 194, 199

SUSY

SuperSymmetry. 43, 44

SVD

Silicon Vertex Detector. 12–14, 62, 64, 65, 71, 73, 80, 81, 84, 87, 88, 93–96, 162, 189, 192, 198

TOP

Time-Of-Propagation. 13, 59, 66, 67, 73, 191

TSIM

Trigger SIMulation. 78, 110

UL

Upper-Limit. 48

VXD

VerteX Detector. 12, 59, 62, 65

Bibliography

[1] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. 2nd ed. Cambridge ; New York: Cambridge University Press, 2007. 272 pp. ISBN: 978-0-521-85249-4.

[2] G. Hernández-Tomé, G. López Castro, and P. Roig. "Flavor Violating Leptonic Decays of τ and μ Leptons in the Standard Model with Massive Neutrinos". In: *Eur. Phys. J. C* 79.1 (Jan. 2019), p. 84. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-019-6563-4.

[3] G. Hernández-Tomé, G. López Castro, and P. Roig. "Erratum to: Flavor Violating Leptonic Decays of τ and μ Leptons in the Standard Model with Massive Neutrinos". In: *Eur. Phys. J. C* 80.5 (May 2020), p. 438. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-020-7935-5.

[4] Ana M. Teixeira. "Theoretical Aspects of Charged Lepton Flavour Violation". In: *J. Phys.: Conf. Ser.* 888 (Sept. 2017), p. 012029. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/888/1/012029.

[5] M. Giffels, J. Kallarackal, M. Krämer, et al. "Lepton-Flavor-Violating Decay $\tau \to \mu\mu\bar{\mu}$ at the LHC". In: *Phys. Rev. D* 77.7 (Apr. 28, 2008), p. 073010. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.77.073010.

[6] Swagato Banerjee, Vincenzo Cirigliano, Mogens Dam, et al. "Snowmass 2021 White Paper: Charged Lepton Flavor Violation in the Tau Sector". Version 2. In: (2022). DOI: 10.48550/ARXIV.2203.14919.

[7] J. P. Lees, V. Poireau, E. Prencipe, et al. "Limits on τ Lepton-Flavor Violating Decays into Three Charged Leptons". In: *Phys. Rev. D* 81.11 (June 10, 2010), p. 111101. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.81.111101.

[8] K. Hayasaka, K. Inami, Y. Miyazaki, et al. "Search for Lepton-Flavor-Violating τ Decays into Three Leptons with 719 Million Produced $\tau^+\tau^-$ Pairs". In: *Physics Letters B* 687.2-3 (Apr. 2010), pp. 139–143. ISSN: 03702693. DOI: 10.1016/j.physletb.2010.03.037.

[9] The LHCb collaboration, R. Aaij, B. Adeva, et al. "Search for the Lepton Flavour Violating Decay $\tau^- \to \mu^-\mu^+\mu^-$". In: *J. High Energ. Phys.* 2015.2 (Feb. 2015), p. 121. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2015)121.

[10] Atlas Collaboration, G. Aad, B. Abbott, et al. "Probing Lepton Flavour Violation via Neutrinoless $\tau \to 3\mu$ Decays with the ATLAS Detector". In: *Eur. Phys. J. C* 76.5 (May 2016), p. 232. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-016-4041-9.

[11] The CMS collaboration, A. M. Sirunyan, A. Tumasyan, et al. "Search for the Lepton Flavor Violating Decay $\tau \to 3\mu$ in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV". In: *J. High Energ. Phys.* 2021.1 (Jan. 2021), p. 163. ISSN: 1029-8479. DOI: 10.1007/JHEP01(2021)163.

[12] E. Kou, P. Urquijo, W. Altmannshofer, et al. "The Belle II Physics Book". In: *Prog. Theor. Exp. Phys.* 2019.12 (Dec. 1, 2019), p. 1808.10567. ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex, physics:hep-lat, physics:hep-ph].

[13] Particle Data Group, P A Zyla, R M Barnett, et al. "Review of Particle Physics". In: *Prog. Theor. Exp. Phys.* 2020.8 (Aug. 14, 2020), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104.

[14] K. Adamczyk, L. Aggarwal, H. Aihara, et al. "The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector". In: *J. Inst.* 17.11 (Nov. 1, 2022), P11042. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/11/P11042.

[15] Cms Tracker Collaboration. "Stand-Alone Cosmic Muon Reconstruction before Installation of the CMS Silicon Strip Tracker". In: *J. Inst.* 4.05 (May 12, 2009), P05004–P05004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/4/05/P05004.

[16] R. Leboucher, K. Adamczyk, L. Aggarwal, et al. "Measurement of the Cluster Position Resolution of the Belle II Silicon Vertex Detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1033 (June 2022), p. 166746. ISSN: 01689002. DOI: 10.1016/j.nima.2022.166746.

[17] Belle II Collaboration, F. Abudinén, I. Adachi, et al. "Search for Lepton-Flavor-Violating $\tau^- \to \ell^- \phi$ Decays in 2019-2021 Belle II Data". Version 3. In: (2023). DOI: 10.48550/ARXIV.2305.04759.

[18] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, Aug. 13, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785.

[19] F. Abudinén, M. Bertemes, S. Bilokin, et al. "Punzi-Loss: A Non-Differentiable Metric Approximation for Sensitivity Optimisation in the Search for New Particles". In: *Eur. Phys. J. C* 82.2 (Feb. 2022), p. 121. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-022-10070-0.

[20] Avrim Blum, Adam Kalai, and John Langford. "Beating the Hold-out: Bounds for K-fold and Progressive Cross-Validation". In: *Proc. Twelfth Annu. Conf. Comput. Learn. Theory.* COLT99: The 12th Annual Conference on Computation Learning Theory. Santa Cruz California USA: ACM, July 6, 1999, pp. 203–208. ISBN: 978-1-58113-167-3. DOI: 10.1145/307400.307439.

[21] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". Version 1. In: (2019). DOI: `10.48550/ARXIV.1907.10902`.

[22] Suyong Choi and Hayoung Oh. "Improved Extrapolation Methods of Data-Driven Background Estimations in High Energy Physics". In: *Eur. Phys. J. C* 81.7 (July 2021), p. 643. ISSN: 1434-6044, 1434-6052. DOI: `10.1140/epjc/s10052-021-09404-1`.

[23] A L Read. "Presentation of Search Results: The CL_s Technique". In: *J. Phys. G: Nucl. Part. Phys.* 28.10 (Oct. 1, 2002), pp. 2693–2704. ISSN: 0954-3899. DOI: `10.1088/0954-3899/28/10/313`.

[24] Robert Zitoun. *Introduction à la physique des particules*. 2e éd. Paris: Dunod, 2009. ISBN: 978-2-10-048778-3.

[25] *Physique des particules: introduction aux concepts et au formalisme du modèle standard*. In collab. with Benoît Clément. 2e édition. Paris: Dunod, 2017. ISBN: 978-2-10-076171-5.

[26] Matthew Dean Schwartz. *Quantum Field Theory and the Standard Model*. New York: Cambridge University Press, 2014. 850 pp. ISBN: 978-1-107-03473-0.

[27] Paul Langacker. *The Standard Model and Beyond*. S.l.: CRC PRESS, 2020. ISBN: 978-0-367-57344-7.

[28] Léonard 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". Marseille, France: Aix-Marseille Université, 2020. URL: `https://docs.belle2.org/record/3380?ln=en`.

[29] Laura Zani. "Search for an Invisible Z' in $\mu^+\mu^-$ plus Missing Energy Events at Belle II". Pisa, Italia: University of Pisa, 2020. URL: `https://docs.belle2.org/record/1831?ln=en`.

[30] Alberto Martini. "Search for Charged Lepton Flavour Violation in $\tau \to 3\mu$ Channel at Belle II". Roma, Italia: Università degli Studi Roma Tre, 2021. URL: `http://www.matfis.uniroma3.it/Allegati/Dottorato/TESI/amartini/Search_for_charged_lepton_flavour_violation_in_tau3mu_at_BelleII.pdf`.

[31] Jean Iliopoulos. "Introduction to the STANDARD MODEL of the Electro-Weak Interactions". Version 1. In: (2013). DOI: `10.48550/ARXIV.1305.6779`.

[32] *Standard Model*. Wikipedia. URL: `https://en.wikipedia.org/wiki/Standard_Model`.

[33] Jeremy Bernstein. "Spontaneous Symmetry Breaking, Gauge Theories, the Higgs Mechanism and All That". In: *Rev. Mod. Phys.* 46.1 (Jan. 1, 1974), pp. 7–48. ISSN: 0034-6861. DOI: `10.1103/RevModPhys.46.7`.

[34] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10.12 (June 15, 1963), pp. 531–533. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.10.531.

[35] Makoto Kobayashi and Toshihide Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652.

[36] Q. R. Ahmad, R. C. Allen, T. C. Andersen, et al. “Measurement of the Rate of $\nu_e + d \top + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory”. In: *Phys. Rev. Lett.* 87.7 (July 25, 2001), p. 071301. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.87.071301.

[37] Y. Fukuda, T. Hayakawa, E. Ichihara, et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81.8 (Aug. 24, 1998), pp. 1562–1567. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.81.1562.

[38] G. Danby, J-M. Gaillard, K. Goulianos, et al. “Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos”. In: *Phys. Rev. Lett.* 9.1 (July 1, 1962), pp. 36–44. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.9.36.

[39] K. Eguchi, S. Enomoto, K. Furuno, et al. “First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance”. In: *Phys. Rev. Lett.* 90.2 (Jan. 17, 2003), p. 021802. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.90.021802.

[40] C. Giganti, S. Lavignac, and M. Zito. “Neutrino Oscillations: The Rise of the PMNS Paradigm”. In: *Progress in Particle and Nuclear Physics* 98 (Jan. 2018), pp. 1–54. ISSN: 01466410. DOI: 10.1016/j.ppnp.2017.10.001.

[41] Ivan Esteban, M. C. Gonzalez-Garcia, Michele Maltoni, et al. “The Fate of Hints: Updated Global Analysis of Three-Flavor Neutrino Oscillations”. Version 1. In: (2020). DOI: 10.48550/ARXIV.2007.14792.

[42] *Neutrino Oscillation*. Wikipedia. URL: https://en.wikipedia.org/wiki/Neutrino_oscillation.

[43] Patrick Blackstone, Matteo Fael, and Emilie Passemar. “$\tau \to \mu\mu\mu$ at a Rate of One out of 10^{14} Tau Decays?” In: *Eur. Phys. J. C* 80.6 (June 2020), p. 506. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-020-8059-7.

[44] Tong Li, Michael A. Schmidt, Chang-Yuan Yao, et al. “Charged Lepton Flavor Violation in Light of the Muon Magnetic Moment Anomaly and Colliders”. In: *Eur. Phys. J. C* 81.9 (Sept. 2021), p. 811. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-021-09569-9.

[45] Benjamin M Dassinger, Thorsten Feldmann, Thomas Mannel, et al. “Model-Independent Analysis of Lepton Flavour Violating τ Decays”. In: *J. High Energy Phys.* 2007.10 (Oct. 8, 2007), pp. 039–039. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/10/039.

[46] Alejandro Celis, Vincenzo Cirigliano, and Emilie Passemar. “Model-Discriminating Power of Lepton Flavor Violating τ Decays”. In: *Phys. Rev. D* 89.9 (May 13, 2014), p. 095014. ISSN: 1550-7998, 1550-2368. DOI: `10.1103/PhysRevD.89.095014`.

[47] Yoshitaka Kuno and Yasuhiro Okada. “Muon Decay and Physics beyond the Standard Model”. In: *Rev. Mod. Phys.* 73.1 (Jan. 12, 2001), pp. 151–202. ISSN: 0034-6861, 1539-0756. DOI: `10.1103/RevModPhys.73.151`.

[48] M. Raidal, A. Van Der Schaaf, I. Bigi, et al. “Flavor Physics of Leptons and Dipole Moments”. In: *Eur. Phys. J. C* 57.1-2 (Sept. 2008), pp. 13–182. ISSN: 1434-6044, 1434-6052. DOI: `10.1140/epjc/s10052-008-0715-2`.

[49] Bhubanjyoti Bhattacharya, Alakabha Datta, Jean-Pascal Guévin, et al. “Simultaneous Explanation of the R_K and $R_{D^{(*)}}$ Puzzles: A Model Analysis”. In: *J. High Energ. Phys.* 2017.1 (Jan. 2017), p. 15. ISSN: 1029-8479. DOI: `10.1007/JHEP01(2017)015`.

[50] Ferruccio Feruglio, Paride Paradisi, and Andrea Pattori. “On the Importance of Electroweak Corrections for B Anomalies”. Version 2. In: (2017). DOI: `10.48550/ARXIV.1705.00929`.

[51] Ian Low. “T Parity and the Littlest Higgs”. In: *J. High Energy Phys.* 2004.10 (Oct. 29, 2004), pp. 067–067. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2004/10/067`.

[52] Monika Blanke, Andrzej J Buras, Björn Duling, et al. “Charged Lepton Flavour Violation and $(g-2)_\mu$ in the Littlest Higgs Model with T-Parity: A Clear Distinction from Supersymmetry”. In: *J. High Energy Phys.* 2007.05 (May 8, 2007), pp. 013–013. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2007/05/013`.

[53] Monika Blanke, Andrzej J. Buras, Bjoern Duling, et al. “FCNC Processes in the Littlest Higgs Model with T-Parity: An Update”. Version 2. In: (2009). DOI: `10.48550/ARXIV.0906.5454`.

[54] R. Barbier, C. Bérat, M. Besançon, et al. “R-Parity-violating Supersymmetry”. In: *Physics Reports* 420.1-6 (Nov. 2005), pp. 1–195. ISSN: 03701573. DOI: `10.1016/j.physrep.2005.08.006`.

[55] Chongxing Yue, Yanming Zhang, and Lanjun Liu. “Non-Universal Gauge Bosons Z’ and Lepton Flavor-Violating Tau Decays”. In: *Physics Letters B* 547.3-4 (Nov. 2002), pp. 252–256. ISSN: 03702693. DOI: `10.1016/S0370-2693(02)02781-8`.

[56] K. S. Babu and Christopher Kolda. “Higgs-Mediated $\tau \to 3\mu$ in the Supersymmetric Seesaw Model”. In: *Phys. Rev. Lett.* 89.24 (Nov. 22, 2002), p. 241802. ISSN: 0031-9007, 1079-7114. DOI: `10.1103/PhysRevLett.89.241802`.

[57] S.T Petcov, S Profumo, Y Takanishi, et al. “Charged Lepton Flavor Violating Decays: Leading Logarithmic Approximation versus Full RG Results”. In: *Nuclear Physics B* 676.1-2 (Jan. 2004), pp. 453–480. ISSN: 05503213. DOI: `10.1016/j.nuclphysb.2003.10.020`.

[58] Takeshi Fukuyama, Tatsuru Kikuchi, and Nobuchika Okada. "Lepton Flavor Violating Processes and Muon g - 2 in the Minimal Supersymmetric SO(10) Model". In: *Phys. Rev. D* 68.3 (Aug. 26, 2003), p. 033012. ISSN: 0556-2821, 1089-4918. DOI: `10.1103/PhysRevD.68.033012`.

[59] Andrea Brignole and Anna Rossi. "Lepton Flavour Violating Decays of Supersymmetric Higgs Bosons". In: *Physics Letters B* 566.3-4 (July 2003), pp. 217–225. ISSN: 03702693. DOI: `10.1016/S0370-2693(03)00837-2`.

[60] G. Cvetič, C. Dib, C. S. Kim, et al. "Lepton Flavor Violation in Tau Decays". In: *Phys. Rev. D* 66.3 (Aug. 13, 2002), p. 034008. ISSN: 0556-2821, 1089-4918. DOI: `10.1103/PhysRevD.66.034008`.

[61] Abdesslam Arhrib, Rachid Benbrik, and Chuan-Hung Chen. "Lepton Flavor Violating τ Decays in the Type-III Seesaw Mechanism". In: *Phys. Rev. D* 81.11 (June 15, 2010), p. 113003. ISSN: 1550-7998, 1550-2368. DOI: `10.1103/PhysRevD.81.113003`.

[62] Yi Cai, Tao Han, Tong Li, et al. "Lepton Number Violation: Seesaw Models and Their Collider Tests". In: *Front. Phys.* 6 (May 9, 2018), p. 40. ISSN: 2296-424X. DOI: `10.3389/fphy.2018.00040`.

[63] C. Hati, J. Kriewald, J. Orloff, et al. "The Fate of $\mathbf{V_1}$ Vector Leptoquarks: The Impact of Future Flavour Data". In: *Eur. Phys. J. C* 81.12 (Dec. 2021), p. 1066. ISSN: 1434-6044, 1434-6052. DOI: `10.1140/epjc/s10052-021-09824-z`.

[64] Nima Arkani-Hamed, Andrew G Cohen, Emanuel Katz, et al. "The Minimal Moose for a Little Higgs". In: *J. High Energy Phys.* 2002.08 (Aug. 13, 2002), pp. 021–021. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2002/08/021`.

[65] Hsin-Chia Cheng and Ian Low. "TeV Symmetry and the Little Hierarchy Problem". In: *J. High Energy Phys.* 2003.09 (Sept. 21, 2003), pp. 051–051. ISSN: 1029-8479. DOI: `10.1088/1126-6708/2003/09/051`.

[66] CMS Collaboration. *Search for the Lepton Flavor Violating Decay Tau to 3mu in Proton-Proton Collisions at Sqrt(s) = 13 TeV.* 2023. URL: `http://cds.cern.ch/record/2860087`. preprint.

[67] Swagato Banerjee, Vincenzo Cirigliano, Mogens Dam, et al. *Snowmass 2021 White Paper: Charged Lepton Flavor Violation in the Tau Sector.* May 26, 2022. arXiv: `2203.14919 [hep-ex, physics:hep-ph]`. URL: `http://arxiv.org/abs/2203.14919`. preprint.

[68] T. Abe, I. Adachi, K. Adamczyk, et al. "Belle II Technical Design Report". Version 1. In: (2010). DOI: `10.48550/ARXIV.1011.0352`.

[69] A. J. Bevan, B. Golob, Th. Mannel, et al. "The Physics of the B Factories". In: *Eur. Phys. J. C* 74.11 (Nov. 2014), p. 3026. ISSN: 1434-6044, 1434-6052. DOI: `10.1140/epjc/s10052-014-3026-9`.

[70] Y. Ohnishi, T. Abe, T. Adachi, et al. “Accelerator Design at SuperKEKB”. In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 26, 2013), 3A011–. ISSN: 2050-3911. DOI: 10.1093/ptep/pts083.

[71] 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 (Feb. 2002), pp. 117–232. ISSN: 01689002. DOI: 10.1016/S0168-9002(01)02013-7.

[72] B. Aubert, D. Boutigny, I. De Bonis, et al. “Measurement of CP -Violating Asymmetries in B^0 Decays to CP Eigenstates”. In: *Phys. Rev. Lett.* 86.12 (Mar. 19, 2001), pp. 2515–2522. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.86.2515.

[73] B. Aubert, A. Bazan, A. Boucham, et al. “The BABAR Detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 479.1 (Feb. 2002), pp. 1–116. ISSN: 01689002. DOI: 10.1016/S0168-9002(01)02012-5.

[74] K. Abe, K. Abe, R. Abe, et al. “Observation of Large CP Violation in the Neutral B Meson System”. In: *Phys. Rev. Lett.* 87.9 (Aug. 14, 2001), p. 091802. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.87.091802.

[75] SuperB Collaboration. *SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report.* Sept. 6, 2007. arXiv: 0709.0451 [hep-ex]. URL: http://arxiv.org/abs/0709.0451. preprint.

[76] SuperB Collaboration, M. Baszczyk, P. Dorosz, et al. “SuperB Technical Design Report”. Version 1. In: (2013). DOI: 10.48550/ARXIV.1306.5655.

[77] 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 (Jan. 2019), pp. 69–144. ISSN: 01689002. DOI: 10.1016/j.nima.2018.05.071.

[78] F. Abudinén, I. Adachi, P. Ahlburg, et al. “Measurement of the Integrated Luminosity of the Phase 2 Data of the Belle II Experiment *”. In: *Chinese Phys. C* 44.2 (Feb. 1, 2020), p. 021001. ISSN: 1674-1137, 2058-6132. DOI: 10.1088/1674-1137/44/2/021001.

[79] I. Adachi, P. Ahlburg, H. Aihara, et al. “Search for an Invisibly Decaying Z’ Boson at Belle II in $e^+e^- \to \mu^+\mu^-(e^\pm\mu^\mp)$ Plus Missing Energy Final States”. In: *Phys. Rev. Lett.* 124.14 (Apr. 6, 2020), p. 141801. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.124.141801.

[80] *Measurement of the Integrated Luminosity of the Early Phase 3 Data of the Belle II Experiment.* BELLE2-NOTE-PH-2022-056. Dec. 3, 2022. URL: https://docs.belle2.org/record/3346?ln=en.

[81] Francesco Forti. *Snowmass Whitepaper: The Belle II Detector Upgrade Program.* Mar. 21, 2022. arXiv: `2203.11349 [hep-ex, physics:physics]`. URL: `http://arxiv.org/abs/2203.11349`. preprint.

[82] Leo Piilonen. "The Belle II Upgrade Program". 31st International Symposium on Lepton Photon Interactions at High Energies (Melbourne, Australia). July 19, 2023. URL: `https://indico.cern.ch/event/1114856/contributions/5360199/`.

[83] 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 (Jan. 2021), p. 164875. ISSN: 01689002. DOI: `10.1016/j.nima.2020.164875`.

[84] K. Adamczyk, L. Aggarwal, H. Aihara, et al. "The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector". In: *J. Inst.* 17.11 (Nov. 1, 2022), P11042. ISSN: 1748-0221. DOI: `10.1088/1748-0221/17/11/P11042`.

[85] 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 (June 2019), pp. 132–141. ISSN: 01689002. DOI: `10.1016/j.nima.2019.03.072`.

[86] Valerio Bertacchi, Tadeas Bilka, Nils Braun, et al. "Track Finding at Belle II". In: *Computer Physics Communications* 259 (Feb. 2021), p. 107610. ISSN: 00104655. DOI: `10.1016/j.cpc.2020.107610`.

[87] Umberto 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* 952 (Feb. 2020), p. 162208. ISSN: 01689002. DOI: `10.1016/j.nima.2019.05.049`.

[88] S. Iwata, I. Adachi, K. Hara, et al. "Particle Identification Performance of the Prototype Aerogel RICH Counter for the Belle II Experiment". In: *Prog. Theor. Exp. Phys.* 2016.3 (Mar. 2016), 033H01. ISSN: 2050-3911. DOI: `10.1093/ptep/ptw005`.

[89] 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* 876 (Dec. 2017), pp. 265–268. ISSN: 01689002. DOI: `10.1016/j.nima.2017.04.043`.

[90] V. Aulchenko, A. Bobrov, B.G. Cheon, et al. "Development of Data Acquisition System for Belle II Electromagnetic Calorimeter". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1030 (May 2022), p. 166468. ISSN: 01689002. DOI: 10.1016/j.nima.2022.166468.

[91] J.-F. Krohn, F. Tenchini, P. Urquijo, et al. "Global Decay Chain Vertex Fitting at Belle II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 976 (Oct. 2020), p. 164269. ISSN: 01689002. DOI: 10.1016/j.nima.2020.164269.

[92] Alexander Glazov, Petar Rados, Ami Rostomyan, et al. *Measurement of the Tracking Efficiency in Phase 3 Data Using Tau-Pair Events.* BELLE2-NOTE-PH-2020-006. July 20, 2020. URL: https://docs.belle2.org/record/1867/files/BELLE2-NOTE-PH-2020-006%20%28ICHEP%29%20-%20reviewed.pdf.

[93] Y. Iwasaki, ByungGu Cheon, Eunil Won, et al. "Level 1 Trigger System for the Belle II Experiment". In: *IEEE Trans. Nucl. Sci.* 58.4 (Aug. 2011), pp. 1807–1815. ISSN: 0018-9499, 1558-1578. DOI: 10.1109/TNS.2011.2119329.

[94] S Lee, R Itoh, T Higuchi, et al. "Belle-II High Level Trigger at SuperKEKB". In: *J. Phys.: Conf. Ser.* 396.1 (Dec. 13, 2012), p. 012029. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/396/1/012029.

[95] Charged PID Group. *Muon and Electron Identification Efficiencies and Hadron-Lepton Mis-Identification Rates at Belle II for Moriond 2022.* BELLE2-CONF-PH-2022-003. Mar. 15, 2022. URL: https://docs.belle2.org/record/2895?ln=en.

[96] *Belle II Software Documentation.* URL: https://software.belle2.org/development/sphinx/index.html.

[97] T. Kuhr, C. Pulvermacher, M. Ritter, et al. "The Belle II Core Software: Belle II Framework Software Group". In: *Comput Softw Big Sci* 3.1 (Dec. 2019), p. 1. ISSN: 2510-2036, 2510-2044. DOI: 10.1007/s41781-018-0017-9.

[98] Rene Brun and Fons Rademakers. "ROOT - An Object Oriented Data Analysis Framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1-2 (Apr. 1997), pp. 81–86. ISSN: 01689002. DOI: 10.1016/S0168-9002(97)00048-X.

[99] *Belle II Luminosity.* Belle II Luminosity. 2023. URL: https://confluence.desy.de/display/BI/Belle+II+Luminosity.

[100] Sw. Banerjee, D. Biswas, T. Przedzinski, et al. "Monte Carlo Event Generator Updates, for Tau Pair Events at Belle II Energies". Version 1. In: (2021). DOI: 10.48550/ARXIV.2111.05914.

[101] David J. Lange. "The EvtGen Particle Decay Simulation Package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1-2 (Apr. 2001), pp. 152–155. ISSN: 01689002. DOI: 10.1016/S0168-9002(01)00089-4.

[102] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, et al. "An Introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. ISSN: 00104655. DOI: 10.1016/j.cpc.2015.01.024.

[103] S. Jadach, B.F.L. Ward, and Z. Waş. "The Precision Monte Carlo Event Generator for Two-Fermion Final States in Collisions". In: *Computer Physics Communications* 130.3 (Aug. 2000), pp. 260–325. ISSN: 00104655. DOI: 10.1016/S0010-4655(00)00048-5.

[104] N. Davidson, G. Nanava, T. Przedziński, et al. "Universal Interface of TAUOLA: Technical and Physics Documentation". In: *Computer Physics Communications* 183.3 (Mar. 2012), pp. 821–843. ISSN: 00104655. DOI: 10.1016/j.cpc.2011.12.009.

[105] C.M. Carloni Calame, C. Lunardini, G. Montagna, et al. "Large-Angle Bhabha Scattering and Luminosity at Flavour Factories". In: *Nuclear Physics B* 584.1-2 (Sept. 2000), pp. 459–479. ISSN: 05503213. DOI: 10.1016/S0550-3213(00)00356-4.

[106] G. Balossini, C. Bignamini, C.M. Carloni Calame, et al. "Photon Pair Production at Flavour Factories with per Mille Accuracy". In: *Physics Letters B* 663.3 (May 2008), pp. 209–213. ISSN: 03702693. DOI: 10.1016/j.physletb.2008.04.007.

[107] F.A. Berends, P.H. Daverveldt, and R. Kleiss. "Complete Lowest-Order Calculations for Four-Lepton Final States in Electron-Positron Collisions". In: *Nuclear Physics B* 253 (Jan. 1985), pp. 441–463. ISSN: 05503213. DOI: 10.1016/0550-3213(85)90541-3.

[108] S. Agostinelli, J. Allison, K. Amako, et al. "Geant4—a Simulation Toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), pp. 250–303. ISSN: 01689002. DOI: 10.1016/S0168-9002(03)01368-8.

[109] *Opera-3D User Guide*. Ver. 15R3. Oxford, England: Cobham Technical Services, Vector Fields Software.

[110] G Casarosa, G Dujany, Ch Finck, et al. *Measurement of the SVD Cluster Position Resolution*. BELLE2-NOTE-TE-2022-005. Mar. 16, 2022.

[111] Jerome Baudot and Pierre Billoir, director. *GDR-Intensity Frontier Lectures: Tracking Detectors and Algorithms*. Nov. 22, 2021. URL: https://indico.in2p3.fr/event/25006/.

[112] Student. "The Probable Error of a Mean". In: *Biometrika* 6.1 (Mar. 1908), p. 1. ISSN: 00063444. DOI: 10.2307/2331554. JSTOR: 2331554.

[113] Thomas Junk. "Confidence Level Computation for Combining Searches with Small Statistics". Version 1. In: (1999). DOI: `10.48550/ARXIV.HEP-EX/9902006`.

[114] Robin Leboucher. *Good Track Selection for Tau Events.* BELLE2-NOTE-PH-2020-029. June 9, 2020. URL: `https://docs.belle2.org/record/1946?ln=en`.

[115] J.-F. Krohn, F. Tenchini, P. Urquijo, et al. "Global Decay Chain Vertex Fitting at Belle II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 976 (Oct. 2020), p. 164269. ISSN: 01689002. DOI: `10.1016/j.nima.2020.164269`.

[116] W. Verkerke and D. Kirkby. *RooFit Users Manual v2.91.* Oct. 14, 2008. URL: `https://root.cern/download/doc/RooFit_Users_Manual_2.91-33.pdf`.

[117] *XGBoost Documentation.* URL: `https://xgboost.readthedocs.io/en/stable/`.

[118] Paolo Calafiura, David Rousseau, and Kazuhiro Terao. *Artificial Intelligence for High Energy Physics.* WORLD SCIENTIFIC, Mar. 2022. ISBN: 9789811234026. DOI: `10.1142/12200`.

[119] R. Aaij, B. Adeva, M. Adinolfi, et al. "Search for Long-Lived Scalar Particles in $B^+ \to K^+\chi(\mu^+\mu^-)$ Decays". In: *Phys. Rev. D* 95.7 (Apr. 14, 2017), p. 071101. ISSN: 2470-0010, 2470-0029. DOI: `10.1103/PhysRevD.95.071101`.

[120] James Bergstra, James Bergstra, Yoshua Bengio, et al. "Algorithms for Hyper-Parameter Optimization". In: Advances in Neural Information Processing Systems 24 (NIPS 2011). Granada, Spain, 2011. ISBN: 978-1-61839-599-3.

[121] Gregor Kasieczka, Benjamin Nachman, Matthew D. Schwartz, et al. "Automating the ABCD Method with Machine Learning". In: *Phys. Rev. D* 103.3 (Feb. 22, 2021), p. 035021. ISSN: 2470-0010, 2470-0029. DOI: `10.1103/PhysRevD.103.035021`.

[122] Swarna Prabha Maharana. *Charged Particle Identification Performances in Belle II.* BELLE2-CONF-PROC-2023-027. July 12, 2023.

[123] Petar Rados, Ami Rostomyan, and Alberto Martini. *Measurement of the Trigger Efficiency in Phase 3 Data Using Tau-Pair Events.* BELLE2-NOTE-PH-2020-028. June 9, 2020. URL: `https://docs.belle2.org/record/1944?ln=en`.

[124] Qi-Dong Zhou. *Correction for Tracking Momentum Bias Based on Invariant Mass Peak Studies.* BELLE2-NOTE-PH-2020-030. June 10, 2020. URL: `https://docs.belle2.org/record/1947`.

[125] Swagato Banerjee, Bolek Pietrzyk, J. Michael Roney, et al. "Tau and Muon Pair Production Cross Sections in Electron-Positron Annihilations at s = 10.58 GeV". In: *Phys. Rev. D* 77.5 (Mar. 13, 2008), p. 054012. ISSN: 1550-7998, 1550-2368. DOI: `10.1103/PhysRevD.77.054012`.

[126] Kyle Cranmer, George Lewis, Lorenzo Moneta, et al. *HistFactory: A Tool for Creating Statistical Models for Use with RooFit and RooStats*. CERN-OPEN-2012-016. 2012. URL: https://cds.cern.ch/record/1456844.

[127] Gary J. Feldman and Robert D. Cousins. “Unified Approach to the Classical Statistical Analysis of Small Signals”. In: *Phys. Rev. D* 57.7 (Apr. 1, 1998), pp. 3873–3889. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.57.3873.

[128] E Kou, P Urquijo, W Altmannshofer, et al. “The Belle II Physics Book”. In: *Prog. Theor. Exp. Phys.* 2019.12 (Dec. 1, 2019), p. 123C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106.

[129] A.Yu. Barnyakov. “The Project of the Super Charm-Tau Factory in Novosibirsk”. In: *J. Phys.: Conf. Ser.* 1561.1 (June 1, 2020), p. 012004. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/1561/1/012004.

[130] Xiao-Rui Lyu and STCF working group. “Physics Program of the Super Tau-Charm Factory”. In: *Proc. 19th Int. Conf. B-Phys. Front. Mach. — PoSBEAUTY2020*. 19th International Conference on B-Physics at Frontier Machines. Kashiwa, Japan (online): Sissa Medialab, Oct. 7, 2021, p. 060. DOI: 10.22323/1.391.0060.

[131] C C Ahdida, I Bezshyiko, A Buonaura, et al. “TauFV: A Fixed-Target Experiment to Search for Flavour Violation in Tau Decays”. In: ().