Aix*Marseille université
Socialement engagée

NNT : 2023AIXM0001

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 28 octobre 2024

Arthur Thaller

Search for the lepton flavour violating decays $\tau^- \to e^\pm \ell^\mp \ell^{-\prime}$ and tracking studies at the 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)
Conseil Européen de la Recherche (ERC)

Composition du jury

Alessandro GAZ Université de Padoue, Padoue	Rapporteur
Marie-Hélène SCHUNE IJCLab, Orsay	Rapporteure
Sacha DAVIDSON IP2I, Villeurbanne	Examinatrice
Cristinel DIACONU CPPM, Marseille	Président du jury
Justine SERRANO CPPM, Marseille	Directrice de thèse

Affidavit

Je soussigné, Arthur Thaller, 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 1er septembre 2024

Liste de publications et participation aux conférences

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

1. A. Thaller, J. Serrano, G. Mancinelli, L. Zani et la collaboration Belle II. "Search for lepton-flavour-violating $\tau \to e\ell\ell'$ at Belle II". Publication en cours de revue interne. Soumission prévue au Journal of High Energy Physics (JHEP).

2. A. Thaller, J. Serrano, G. Mancinelli, et L. Zani. "Search for lepton-flavour-violating $\tau \to e\ell\ell'$ at Belle II", Note interne à Belle II, Avril 2024, BELLE2-NOTE-PH-2024-021.

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

Conférences :

1. **XIII International Conference on New Frontiers in Physics,** août 2024, Kolymbari, Grèce. A. Thaller pour la collaboration Belle II. "Recent measurements of τ lepton decays at Belle II". Présentation orale. Un proceeding sera publié.

2. **48th Belle II General Meeting,** juin 2024, Tsukuba, Japon. A. Thaller. "Search for LFV decays $\tau^- \to e^\mp \ell^\pm \ell'^-$ ". Présentation orale.

3. **Annual workshop of Groupement de Recherche "Intensity Frontier",** novembre 2023, Mont Sainte-Odile, France. A. Thaller, C. Lemettais et K. Lautenbach pour la collaboration Belle II. "Search of LFV decays at Belle II". Présentation orale.

4. **Annual workshop of Groupement de Recherche "Intensity Frontier",** novembre 2022, Lyon, France. A. Thaller pour la collaboration Belle II. "Estimation of the fake tracks vs time in Belle II". Présentation orale.

Écoles d'été :

1. **Trans-European School of High Energy Physics 2023,** juillet 2023, Rzeszow, Pologne.
2. **IN2P3 School of Statistics 2022**, Mai 2022, Carry-le-Rouet, France.
3. **2021 Belle II Physics Week : Starter Kit,** novembre 2021, Rome, Italie.

Abstract

This thesis reports the search for the Lepton Flavour Violating (LFV) decays $\tau^- \to e^{\mp}\ell^{\pm}\ell'^{-}$ ($\ell^{(\prime)} = e, \mu$) using the Run 1 dataset of the Belle II experiment, corresponding to an integrated luminosity $\mathcal{L} = 424$ fb^{-1}. The analysis is based on an untagged reconstruction of the signal candidates in $e^+e^- \to \tau^+\tau^-$ events, where only the signal τ is reconstructed while the other is left unconstrained. The background suppression is performed on data sidebands due to the important contribution of QED processes ($e^+e^- \to \ell^+\ell^-\ell^{+\prime}\ell^{-\prime}$, $\ell^{(\prime)} = e, \mu$) that are mis-modeled in the simulations. The analysis relies on a combination of selection criteria and a BDT classifier to achieve an almost background-free scenario, while maintaining a signal reconstruction efficiency in the range [15-24]%. The branching fraction is extracted with an unbinned maximum likelihood fit to the $M_{e\ell\ell'}$ distribution, and the upper limits are computed using the CL_S method. The systematic uncertainties affecting the signal efficiency and the background modelling are evaluated. No signal events are observed across the five modes, and the observed upper limits in four of the modes ($e^-e^+e^-$, $e^-e^+\mu^-$, $\mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$) are more stringent than the previous Belle's results obtained with 762 fb^{-1} while the fifth mode's limit ($e^-\mu^+e^-$) is comparable to Belle's measurement.

In addition, three studies of the fake tracks rate are included. The first one implements a track time variable based on measurement from the silicon vertex detector (SVD), the second one uses a classifier to reject detector hits based on the time information in degraded data acquisition conditions and the third one studies the fake tracks rate in the proposed upgrade of the vertex detector (VTX).

Résumé

Cette thèse décrit la recherche des désintégrations violant la saveur leptonique (LFV) $\tau^- \to e^\mp \ell^\pm \ell'^-$ ($\ell^{(\prime)} = e, \mu$) en utilisant les données du Run 1 de l'expérience Belle II, correspondant à une luminosité intégrée de $\mathcal{L} = 424$ fb^{-1}. L'analyse est basée sur une reconstruction non étiquetée des candidats signal dans les événements $e^+e^- \to \tau^+\tau^-$, où seul le τ d'intérêt est reconstruit tandis que l'autre n'est pas contraint. La suppression des événements de bruit de fond est réalisée sur les bandes latérales des données en raison de la contribution importante de processus de QED ($e^+e^- \to \ell^+\ell^-\ell^{+\prime}\ell^{-\prime}$, $\ell^{(\prime)} = e, \mu$) qui sont mal modélisés dans les simulations. L'analyse repose sur une combinaison de coupures et un arbre de décision boosté pour atteindre un scénario pratiquement exempt de tout bruit de fond, tout en maintenant une efficacité de reconstruction du signal entre 15 et 24%. Le rapport d'embranchement est obtenu par un ajustement de vraisemblance maximale non biné à la distribution $M_{e\ell\ell'}$ des candidats, et les limites supérieures sont calculées en utilisant la méthode CL_S. Les incertitudes systématiques affectant la mesure du rapport d'embranchement sont évaluées. Selon les limites attendues évaluées à partir des données dans les régions latérales, l'analyse améliore la sensibilité dans trois des canaux de désintégration ($e^-e^+e^-$, $\mu^-\mu^+e^-$ et $\mu^-e^+\mu^-$) par rapport aux résultats précédents de Belle obtenus avec 762 fb^{-1}, un canal ($e^-\mu^+e^-$) est comparable aux résultats de Belle et un canal ($e^-e^+\mu^-$) est moins sensible.

De plus, trois études sur le taux de fausses traces sont incluses. La première implémente une variable de temps d'impact des traces basée sur la mesure du détecteur de vertex en silicium (SVD), la deuxième utilise un classificateur pour rejeter les impacts de détecteur basés sur les informations de temps dans des conditions dégradées d'acquisition des données, et la troisième étudie le taux de fausses traces dans la mise à niveau proposée du détecteur de vertex (VTX).

Remerciements

Je tiens à remercier tout d'abord ma directrice de thèse Justine Serrano, ainsi que mon co-directeur non-officiel Giampiero Mancinelli, pour m'avoir accordé leur confiance, m'avoir encadré, accompagné et soutenu tout au long de ces trois années de thèse. Leur bienveillance et leurs conseils, toujours justes et pertinents, m'ont permis de m'épanouir pleinement durant ces trois années. Je suis fier d'avoir appris à leur contact et reconnaissant pour tout ce qu'ils m'ont apporté.

Je tiens aussi à remercier les membres de mon jury, Alessandro Gaz, Marie-Hélène Schune et Sacha Davidson, pour avoir accepté de juger mon travail et pour leurs précieux conseils et retours.

Des remerciements s'imposent bien évidemment pour le reste des membres du groupe Belle II, notamment Robin et Laura, pour leur apport conséquent à cette thèse, et les nombreuses discussions que l'on a pu avoir. J'ai aussi une pensée pour les membres du groupe Belle II de Strasbourg, qui m'ont permis en premier lieu de rejoindre la collaboration, et bien évidemment merci au groupe des étoiles de Belle II.

Je souhaite aussi remercier tout le CPPM : Cristinel Diaconu, qui a été président de mon jury, les footeux, les jeunes depuis longtemps, les joueurs d'échecs, et tous les autres.

J'en arrive maintenant à l'équipe de doctorants : Marie, Maxime, Tyann, Gaya, Jean, Luc et Godefroy. Merci pour ces trois années et tous ces bons moments partagés, qu'il s'agisse des manifs, des foots, des pauses cafés, des soirées et autres. Vous avez énormément contribué à rendre cette thèse cool, sympa et mémorable. Même si on est maintenant dispersés aux quatre coins du monde, on sait tous que l'on se reverra bientôt.

J'adresse aussi ma gratitude à mes amis de longue date et à ma famille, venus de la France entière pour ma soutenance, ainsi que tous ceux qui n'ont pas pu être présents quelle qu'en soit la raison. Je ne vous citerai pas, mais sachez que de par votre présence dans ma vie, vous avez contribué à la réussite de cette thèse.

Je veux néanmoins m'adresser à ma mère : merci Maman. Merci pour m'avoir encouragé, poussé et secoué tout au long de ma vie et de mes études. Merci pour m'avoir donné les moyens de réussir, et pour m'avoir toujours soutenu dans mes choix. C'est sans l'ombre d'un doute grâce à toi que j'ai pu (et voulu !) faire d'aussi longues études, et que j'en suis là aujourd'hui. On pense bien évidemment énormément à Papa, le premier de la famille à avoir eu une thèse et qui, je crois que sans même le savoir, est le premier qui m'a donné le goût de la physique. Il nous manque.

Enfin, mes derniers mots sont pour Yasmine. Tu es la deuxième de ma famille à avoir une thèse et la fierté que j'ai ressenti lors de ta propre soutenance m'a énormément motivé pour cette dernière année. Merci pour ton soutien, ta patience, ta compréhension durant ces trois ans, et pour tout ce que tu m'apportes au quotidien. Tu as été un pilier essentiel de cette thèse, et de ma vie en général. "Être entouré c'est bien, mais être entouré par toi c'est mieux". Je t'aime.

Résumé long

Le Modèle Standard de la physique des particules

La physique est la science qui étudie les lois de la nature et les phénomènes qui en découlent. Elle se divise en plusieurs branches, chacune se concentrant sur un aspect spécifique, en fonction de différentes échelles de temps, d'espace, d'énergie ou de taille des objets étudiés. On peut citer la cosmologie, qui s'intéresse à l'univers dans son ensemble, la géophysique, qui étudie la Terre, la physique de la matière condensée, qui s'intéresse aux propriétés des solides et des liquides, et bien d'autres. Les séparations entre ces différentes branches ne sont pas toujours nettes, et il est fréquent que des phénomènes étudiés dans un domaine se retrouvent dans un autre. Néanmoins, une de ces disciplines, la physique des particules, se distingue en cela qu'elle s'efforce de comprendre les constituants élémentaires de la matière et les interactions qui les gouvernent. Cela n'implique bien évidemment pas qu'elle soit plus "fondamentale" que les autres branches de la physique, mais cela la rend, à mes yeux du moins, la plus fascinante.

La formulation moderne de la physique des particules est le Modèle Standard [1, 2, 3, 4] (MS), qui est une théorie quantique de champs relativiste. Son élaboration remonte principalement au 20^{e} siècle, avec l'apparition des théories de la relativité et de la mécanique quantique. Il décrit les propriétés des fermions, les particules de matière, que sont les six types de quarks (up, down, charm, strange, bottom - ou beauty - et top), les six leptons (électron, muon et tau, et les trois neutrinos associés), ainsi que les anti-fermions, les particules d'anti-matière correspondantes. Ces fermions interagissent par le biais des bosons, le photon, les bosons W et Z, et les gluons, responsables respectivement des forces électromagnétiques, nucléaires faibles et fortes. Il est par ailleurs bâti sur le théorème de Noether [5], qui stipule qu'à chaque symétrie (invariance sous une transformation mathématique) est associée une grandeur conservée : par exemple, la conservation de l'énergie est liée à l'invariance des lois physiques dans le temps. L'aboutissement du MS est la découverte en 2012 du boson de Higgs par les expériences ATLAS [6] et CMS [7], la particule qui "confère" aux autres leur masse.

Toutefois, le consensus actuel est que le MS n'est pas la théorie "ultime" de la physique des particules. En effet, il n'incorpore nullement la gravité, phénomène autrement validé par l'observation, ni n'explique la matière noire ou l'énergie noire, qui constituent la majorité du contenu énergétique de l'univers. De plus, il ne peut pas expliquer pourquoi les masses des particules sont ce qu'elles sont, pas plus qu'il n'explique pourquoi il y a davantage de matière que d'antimatière dans l'univers. Dans l'objectif d'aller "au-delà" du MS, de nombreuses expériences sont menées, notamment au LHC (Large Hadron Collider, Grand Collisionneur de Hadrons) du CERN (Centre Européen de Recherche Nucléaire), pour tenter de mettre en évidence de la "nouvelle physique", c'est-à-dire des phénomènes qui ne sont pas prédits par le MS. Ces expériences sont de deux types : les expériences de "recherche directe", qui consistent à produire et observer directement des particules non prédites par le MS, et les expériences de "recherche indirecte", qui consistent à chercher des effets de la nouvelle physique dans des processus bien connus du MS. C'est dans cette dernière catégorie que s'inscrit le travail présenté dans ce manuscrit.

Un des concepts importants du MS est la saveur des particules, qui est un nombre quantique associé à chaque particule. Dans le cas des leptons, il existe trois saveurs qui correspondent aux trois familles : la saveur électronique, muonique et tauique. Par exemple, l'électron et le neutrino électronique sont porteurs de la saveur électronique. L'interaction électromagnétique conserve la saveur leptonique, c'est-à-dire que les leptons ne changent pas de saveur lorsqu'ils interagissent par le biais du photon. Cependant, diverses observations [8, 9, 10, 11] ont mis en évidence le phénomène d'oscillation des neutrinos (ce phénomène nécessite que les neutrinos aient une masse, allant à l'encontre de la prédiction du MS), qui est une preuve de la violation de la conservation de la saveur leptonique. Cela implique que dans le cadre de désintégrations des leptons chargés (électron, muon et tau), il est possible que la saveur leptonique change, dans des processus tels que $\mu^- \to e^-\gamma$ ou $\tau^- \to e^-e^+e^-$ (les désintégrations conjuguées de charge sont systématiquement incluses tout au long de ce manuscrit). Ces processus sont néanmoins très rares, avec des taux d'occurrence (appelées rapports d'embranchement) attendus de l'ordre de 10^{-55} [12, 13], ce qui est infiniment loin des sensibilités actuelles des expériences. L'observation de telles transitions serait alors une preuve incontestable d'existence de nouvelle physique. Diverses théories allant au-delà du MS prédisent de tels processus à des taux de l'ordre de 10^{-8}, ce qui est accessible aux expériences actuelles.

Mode	Belle	Babar
$e^-e^+e^-$	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

Table 0.1.: Limites supérieures à 90% de niveau de confiance obtenues par les expériences BaBar [14] et Belle [15] pour les désintégrations $\tau^- \to e^\mp\ell^\pm\ell'^-$. Les valeurs sont exprimées en multiples de 10^{-8}.

Le travail présenté dans ce manuscrit se concentre sur la recherche des désintégrations du lepton tau en trois leptons légers, $\tau^- \to e^\mp \ell^\pm \ell'^-$, où ℓ et ℓ' sont des électrons ou des muons. De telles désintégrations ont déjà été recherchées par les expériences BaBar [14] et Belle [15], mais aucune observation des désintégrations n'a été faite et des limites supérieures sur les rapports d'embranchement ont été obtenues, de l'ordre de 10^{-8}. Les valeurs de ces dernières sont répertoriées dans le Tableau 0.1.

L'expérience Belle II

Ce travail est mené au sein de la collaboration Belle II [16, 17], une collaboration internationale qui se matérialise par une expérience située au Japon, dans le laboratoire KEK à Tsukuba. Elle se décline en deux éléments principaux : le collisionneur SuperKEKB [18] et le détecteur Belle II.

SuperKEKB est un collisionneur électron-positon asymétrique, avec une énergie de collision dans le centre de masse de 10.58 GeV. Cette énergie de collision correspond à la masse du $\Upsilon(4S)$, un état lié d'une paire de quark et anti-quark bottom et qui se désintègre en paire de méson B (des systèmes à deux quarks dont l'un est un bottom et l'autre un up ou un down), mais les collisions produisent en quantité similaire des paires de leptons τ. La particularité de SuperKEKB est sa luminosité instantanée (ce qui correspond au nombre total de collisions produites), qui est prévue à terme à 6×10^{35} cm^{-2}s^{-1}, soit environ 30 fois plus élevée que celle de son prédécesseur, KEKB. La luminosité instantanée maximale atteinte à ce jour est de 4.7×10^{3} cm^{-2}s^{-1}, la plus haute jamais atteinte. Cette luminosité accrue permet d'augmenter considérablement la taille des échantillons de données collectés, et donc d'améliorer la sensibilité des expériences, notamment dans la recherche de processus rares. Cette hausse de la luminosité est rendue possible par l'augmentation du courant des faisceaux, et par une réduction de la taille des faisceaux eux-mêmes (au niveau du point d'interaction), ce qui permet d'augmenter la probabilité de collision des particules. L'expérience Belle II a pour objectif de collecter un échantillon de données correspondant à 50 ab^{-1} sur les dix prochaines années.

Le détecteur Belle II est, lui aussi, en grande partie une amélioration de son prédécesseur, le détecteur Belle. La plupart des sous-détecteurs ont été modifiés pour faire face à l'augmentation de la luminosité qui se traduit aussi par une hausse du bruit de fond (comprendre : des radiations réduisant les performances et endommageant le détecteur). Le détecteur suit la structure typique des détecteurs de physique des particules, avec plusieurs couches servant différents objectifs comme l'identification des particules, la mesure des énergies et impulsions ou la reconstruction des trajectoires. Le détecteur couvre un large angle solide (de 17° à 150°) tout autour du point d'interaction et est hermétique : l'énergie manquante peut donc être inférée en sommant les énergies mesurées dans le détecteur. Les sous-détecteurs principaux sont :

- le détecteur de vertex (Vertex detector, VXD), lui-même composé de deux couches de capteurs pixelisés (formant le Pixel Detector, PXD [19]) et de quatre couches de détecteur en silicone (Silicone Vertex Detector, SVD [20]). Le détecteur de vertex sert à reconstruire, comme son nom l'indique, les vertex, c'est-à-dire les points de production (le point d'interaction) et de désintégration des particules comme les mésons B. Il sert aussi à la trajectographie et à l'identification des particules,

- la chambre à dérive (Central Drift Chamber, CDC), remplie de gaz et de fils sous tension. C'est le système principal pour la trajectographie, et permet lui aussi d'identifier les particules,

- les deux systèmes d'identification des particules, un compteur de temps de propagation (Time Of Propagation counter, TOP) et un détecteur Tcherenkov à aerogel (Aerogel Ring Imaging Cherenkov, ARICH [21]). Ces deux détecteurs utilisent l'effet Tcherenkov pour identifier la nature des particules : le TOP mesure le temps de propagation des photons Tcherenkov, qui est lié à la masse des particules. L'ARICH mesure lui l'angle d'ouverture du cône de photons, qui varie selon la particule,

- le calorimètre électromagnétique (Electromagnetic Calorimeter, ECL) fait de cristaux à scintillations, qui mesure l'énergie des électrons, des photons et des hadrons, dont il arrête la course,

- et le détecteur de K-long (K_L^0) et muons, à base de scintillateurs et de plaques en fer, pour arrêter ces particules qui sont les seules particules susceptibles de l'atteindre.

Une mise à niveau du détecteur est prévue pour le futur [22], et en particulier pour le VXD : un nouveau détecteur à vertex (VTX) est proposé pour le remplacer, pour obtenir de meilleures performances et une meilleure résistance au bruit de fond. Ce détecteur est actuellement en phase de R&D, mais conserverait la structure en couches du VXD, à base de capteurs pixelisés CMOS appelés OBELIX (Optimised BELle II monolithic pIXel). Cette amélioration du détecteur accompagnera une augmentation de la luminosité du collisionneur, qui est prévue pour atteindre la luminosité nominale de 6×10^{35} cm^{-2}s^{-1}.

Les informations mesurées par les différents composants du détecteur sont combinés par un logiciel interne à Belle II, (Belle II Analysis Software, basf2) afin de reconstruire les trajectoires des différentes particules et de mesurer leurs énergies et quadri-impulsions. Des simulations Monte-Carlo des collisions, du transport des particules et de la réponse du détecteur sont produites et servent à étudier les différents processus lors des analyses. La collaboration a collecté 424 fb^{-1} lors du Run 1 entre 2019 et 2022 ; cela correspond aux données analysées dans cette thèse.

Études des fausses traces

Une première partie du travail réalisé dans le cadre de la thèse a consisté en l'étude des fausses traces. Les fausses traces peuvent soit être :

- des erreurs de reconstruction, provenant de combinaisons erronées de points d'impact dans le détecteur ou de mauvaises associations d'impacts provenant d'une même particule et menant à un dédoublement de la trace, ou
- des particules réelles, mais qui ne proviennent pas de la collision étudiée. En effet, les faisceaux étant extrêmement collimatés, les électrons d'un même faisceau interagissent parfois entre eux, se faisant alors éjecter du faisceau, heurtent les parois et produisent des gerbes qui peuvent atteindre le détecteur. Ces particules, qui ne viennent pas de la collision elle-même, sont parfois observées par le détecteur : on les appelle aussi fausses traces, bien qu'elles soient correctement reconstruites.

Ce bruit de fond induit par le faisceau est la source principale de fausses traces et est proportionnel à la luminosité du collisionneur. Or, la luminosité de Belle II est bien plus élevée que celle de Belle, et il est prévu qu'elle continue d'augmenter : il est donc important de réduire le taux de fausses traces au maximum. Trois études différentes sont réalisées. La première consiste à implémenter une variable de temps dans le logiciel de reconstruction des traces, qui associe à chaque trace son temps d'impact dans le SVD et à en étudier le pouvoir de discrimination. En effet, les traces venant de la collision doivent être toutes proches dans le temps du moment de l'interaction, alors que les fausses traces ont un temps d'impact beaucoup plus erratique. Il est montré dans des simulations, que cette variable permet de rejeter la moitié des fausses traces tout en conservant une excellente efficacité sur les "vraies" traces. Cette variable est implémentée dans le logiciel et rendue disponible pour les analyses ; de plus amples études sont en cours pour en évaluer l'efficacité sur les données réelles.

La deuxième étude considère un scénario d'acquisition des données dégradé, où la résolution en temps du SVD est moins bonne et la variable précédemment discutée moins efficace. La possibilité de réduire les fausses traces en rejetant directement les impacts du détecteur est alors étudiée. Un algorithme de classification est implémenté (arbre de décision optimisé, BDT), qui permet de rejeter les impacts du détecteur qui sont les plus susceptibles d'être provoqués par des fausses traces. Cet algorithme est testé sur des données simulées et montre une réduction significative du nombre d'impacts (par un facteur 4) causé par des fausses traces.

La dernière étude s'est intéressée au taux de fausses traces dans le VTX, le futur détecteur de vertex, dans les conditions de haute luminosité prévues pour le futur. Le taux de fausses traces est estimé avec une méthode dite *tag-and-probe* dans des simulations d'événements $e^+e^- \to e^+e^-$: les paires d'électrons sont reconstruites (*tag*), avec la présence éventuelle d'une troisième trace (*probe*) qui est alors interprétée comme une fausse trace. Le taux de fausses traces est alors donné par :

$$r_f = \frac{N_3}{N_2 + N_3},$$

où N_2 et N_3 sont respectivement le nombre d'événements avec deux et trois traces reconstruites. Le taux de fausses traces dans les simulations est de l'ordre de 40% dans les conditions de luminosité nominale ; il est toutefois montré que ce taux peut être réduit de moitié en incorporant un algorithme de classification des traces. Cette étude met aussi en lumière des problèmes de reconstruction des traces à haute luminosité, liées à la saturation de la CDC à cause du bruit de fond.

Ces trois études sont des travaux prospectifs menés dans des simulations et qui doivent donc être prolongés par des études sur les données réelles. Elles sont néanmoins cruciales pour garantir la qualité des données collectées par Belle II dans le futur, et pour garantir la sensibilité des analyses de physique.

Recherche de désintégrations ne conservant pas la saveur leptonique $\tau^- \to e^\mp \ell^\pm \ell'^-$

Lors des collisions, les paires de τ sont produites dos à dos dans le centre de masse, et sont boostées dans la direction de leur mouvement. Il y a donc une claire séparation géométrique entre les deux τ : on détermine dans un premier temps la direction de la "poussée" ("thrust"), l'axe selon lequel les τ se déplacent, ou plus formellement la direction qui maximise la somme des projections des impulsions de toutes les particules de l'événement :

$$T = \max_{\mathbf{n}_T} \left(\frac{\sum_i |\mathbf{p}_i^* \cdot \mathbf{n}_T|}{\sum_i |\mathbf{p}_i^*|} \right). \tag{0.1}$$

Une fois cet axe identifié, le plan qui lui est perpendiculaire sépare alors l'événement en deux hémisphères, un pour chaque τ et ses produits de désintégration, comme représenté sur la Figure 0.1.

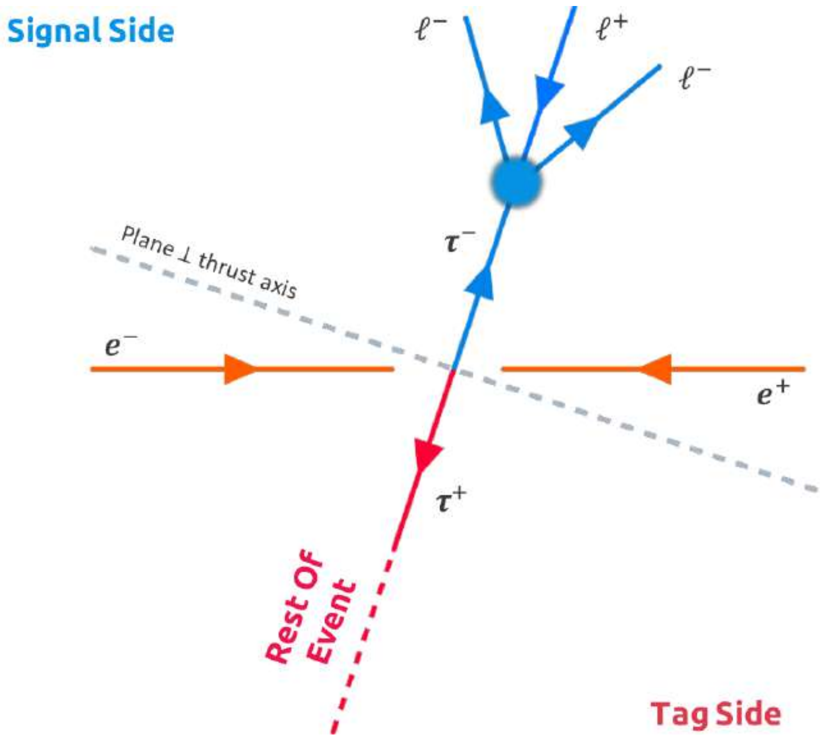


Figure 0.1.: Représentation schématique de la production de la paire de τ et de la séparation en hémisphères.

L'analyse suit la stratégie utilisée pour les recherches des canaux $\tau^- \to \ell^-\phi$ [23] et $\tau^- \to \mu^-\mu^+\mu^-$ [24] réalisées par Belle II. Elle se base sur la reconstruction non étiquetée des cinq désintégrations $\tau^- \to e^\mp \ell^\pm \ell^{-\prime}$ ($\ell, \ell' = $ e,μ). Les candidats sont reconstruits en combinant trois leptons, dont l'un au moins est un électron et en requérant que les trois leptons soient dans le même hémisphère. Les électrons et les muons sont reconnus grâce aux variables d'identification (electronID et muonID). Le plan ($M_{e\ell\ell'}$, $\Delta E_{e\ell\ell'}$) est utilisé pour restreindre les candidats signal, où $M_{e\ell\ell'}$ est la masse invariante du τ candidat et $\Delta E_{e\ell\ell'}$ est la différence entre l'énergie du faisceau et l'énergie du τ dans le centre de masse. En l'absence de neutrinos, les distributions présentent un pic autour de $M_{e\ell\ell'} \sim 1.777$ GeV/c^2 et $\Delta E_{e\ell\ell'} \sim 0$ GeV, comme montré sur la Figure 0.2 pour le mode $\tau^- \to e^-\mu^+e^-$. Ce plan est aussi utilisé pour masquer les données dans la région proche du signal, afin de réaliser l'analyse "à l'aveugle" et ne pas biaiser les résultats.

L'autre τ n'est pas contraint, c'est-à-dire qu'on ne reconstruit pas explicitement un ou plusieurs modes de désintégration possibles pour ce τ, à savoir typiquement en un électron, ou un muon, ou un ou trois pions (plus les neutrinos). Toutes les particules qui ne sont pas associées à un candidat signal sont combinées en un seul objet appelé "reste de l'événement", dont les propriétés et la cinématique sont utilisées pour rejeter les événements de bruit de fond.

Une part importante du bruit de fond provient des processus $e^+e^- \to \ell^+\ell^-\ell'^+\ell'^-$

($\ell^{(\prime)} = e$ ou μ). Ces événements peuvent s'accompagner de radiations dans l'état initial et dans l'état final, qui ne sont pas simulées correctement. D'importants écarts entre les simulations et les données sont observés, qui se caractérisent notamment par une contribution en $\Delta E_{\ell\ell'} \sim 0.1$ GeV bien plus importante dans les données que dans les simulations et à haute valeur de $M_{\ell\ell'}$. Cet excès est montré sur la Figure 0.2, pour le mode $\tau^- \to e^-\mu^+e^-$.

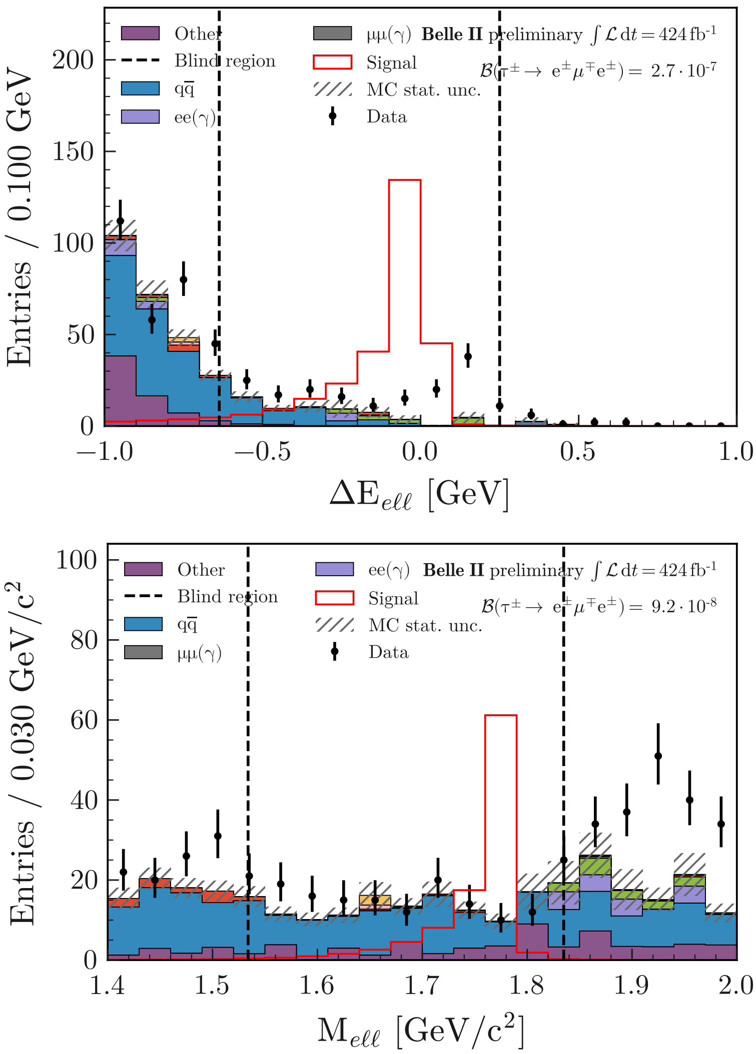


Figure 0.2.: Distribution de $\Delta E_{\ell\ell'}$ (haut) et $M_{\ell\ell'}$ (bas) pour les données et les simulations pour le mode $\tau^- \to e^-\mu^+e^-$. L'histogramme rouge correspond à la simulation du signal, et est normalisé arbitrairement. Les traits noirs pointillés correspondent à la région masquée.

En conséquence, l'optimisation de la sélection visant à rejeter le bruit de fond est réalisée sur les données, en utilisant une combinaison de coupures et un arbre de décision boosté (BDT). Les coupures sont choisies pour rejeter les composantes du bruit de fond "évidentes", qui exhibent des structures particulières dans les distributions des variables cinématiques, telles que la masse invariante des paires de leptons ou l'énergie totale visible dans l'événement. Le BDT est utilisé pour rejeter les événements de bruit de fond restant. Il est entrainé sur les simulations du signal et le bruit de fond des données. Il produit un score qui est attribué à chaque événement, et qui traduit la probabilité pour qu'un événement soit du signal ou du bruit de fond. Les distributions du score du BDT pour les données et les simulations sont montrées sur la Figure 0.3 pour le mode $\tau^- \to e^-\mu^+e^-$.

Table 0.2.: Valeur de la coupure sur le score du BDT s^{BDT}, efficacité finale de reconstruction du signal ϵ_{sig} et nombre d'événements observés N_{SB} dans les régions latérales pour les différents modes.

	s^{BDT}	N_{SB}	ϵ_{sig}^{tot}
$e^-e^+e^-$	0.95	3	15.0%
$e^-e^+\mu^-$	0.8	6	20.4%
$e^-\mu^+e^-$	0.5	6	23.5%
$\mu^-\mu^+e^-$	0.8	12	20.1%
$\mu^-e^+\mu^-$	0.5	4	24.1%

Le Tableau 0.2 compile les valeurs de coupure sur le score du BDT pour chaque mode, avec les efficacités de sélection du signal et le nombre d'événements de bruit de fond restant dans les régions latérales, autour de la région masquée.

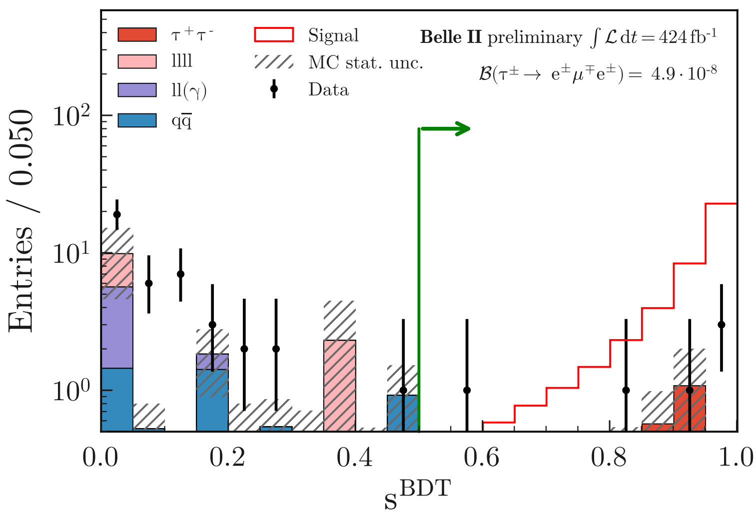


Figure 0.3.: Distribution du score du BDT dans les régions latérales pour les données et simulations, ainsi que pour l'entièreté du signal $\tau \to e^-\mu^+e^-$. Le signal est prédit à de hautes valeurs alors que les bruits de fond sont à basse valeur : le BDT crée une claire séparation entre les deux.

Une fois la sélection optimisée, la grandeur que l'on veut mesurer est le rapport d'embranchement :

$$\mathcal{B}(\tau^- \to e^\mp \ell^\pm \ell'^-) = \frac{N_{sig}}{2 \times \mathcal{L} \times \sigma_{\tau^+\tau^-} \times \epsilon_{sig}}, \tag{0.2}$$

où N_{sig} est le nombre d'événements de signal observés, $\mathcal{L}$ est luminosité intégrée totale, $\sigma_{\tau^+\tau^-}$ est la section efficace de production des paires de τ et ϵ_{sig} est l'efficacité de reconstruction du signal. Ce rapport d'embranchement est directement obtenu par un ajustement par maximum de vraisemblance non-biné à la distribution de $M_{e\ell\ell'}$. La fonction de densité de probabilité (p.d.f) utilisée est :

$$PDF_{tot}(M_{e\ell\ell'}) = \mathcal{B}(\tau^- \to e^\mp \ell^\pm \ell'^-) \cdot \mu \cdot PDF_{sig}(M_{e\ell\ell'}) + N_{bg} \cdot PDF_{bg}(M_{e\ell\ell'}) \tag{0.3}$$

avec $\mu = 2 \cdot \epsilon_{sig} \cdot \sigma_{\tau\tau} \cdot \mathcal{L}$. Une distribution exponentielle est utilisée pour modéliser le bruit de fond et paramétrée par un coefficient C_{bg} :

$$PDF_{bg}(M_{e\ell\ell'}) = \exp(C_{bg} \cdot M_{e\ell\ell'}), \tag{0.4}$$

alors qu'une distribution Crystal Ball asymétrique est utilisée pour le signal.

Différentes incertitudes systématiques sont estimées, notamment sur l'efficacité de reconstruction du signal causée par l'imparfaite simulation des variables d'identification des leptons, du "trigger" (des conditions déclenchant l'enregistrement des événements) et de la reconstruction des traces. L'influence de la méconnaissance du champ magnétique sur la modélisation du bruit de fond est prise en compte, et les incertitudes sur la luminosité et la section efficace de production des paires de τ sont aussi considérées. Les incertitudes systématiques sur les différentes quantités sont résumées dans le Tableau 0.3. La première source d'incertitude est néanmoins statistique et découle du très faible nombre d'événements pour l'ajustement.

Pour estimer la sensibilité de l'analyse, le nombre d'événements de bruit de fond attendus est estimé en ajustant la distribution de $M_{e\ell\ell'}$ dans les régions latérales, et en extrapolant la distribution dans la région masquée. Cet ajustement est montré sur la Figure 0.4 pour le canal $e^-\mu^+e^-$. Les limites supérieures à 90% de niveau de confiance sur le rapport d'embranchement sont alors estimées en utilisant la méthode des CL_S (intervalle de confiance) par la librairie RooStat. Les résultats de la méthode CL_S sont montrés sur la Figure 0.5. Les limites supérieures attendues sont présentées dans le Tableau 0.4 pour les différents modes.

Pour les canaux $e^-e^+e^-$, $\mu^-\mu^+e^-$ et $\mu^-e^+\mu^-$ la limite attendue est meilleure que les limites actuelles obtenues par Belle, similaire pour le canal $e^-\mu^+e^-$, et légèrement plus élevée pour le canal $e^-e^+\mu^-$. Les données sont encore dissimulées, mais pour les quatre premiers modes il est donc probable d'observer une amélioration des limites supérieures sur le rapport d'embranchement par rapport à Belle, en dépit d'une quantité de donnée plus faible (424 fb^{-1} contre 762 fb^{-1}).

Quantité	Incertitude	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
$\mathcal{L}$	$\sigma_{\mathcal{L}}$	0.6%	0.6%	0.6%	0.6%	0.6%
$\sigma_{\tau\tau}$	$\sigma_{\tau\tau}$	0.3%	0.3%	0.3%	0.3%	0.3%
ϵ_{sig}	$\sigma_{tracking}$	1.0%	1.0%	1.0%	1.0%	1.0%
	σ_{BDT}	0.7%	2.5%	0.7%	1.5%	0.4%
	σ_{LID}	1.0%	1.2%	1.3%	1.7%	1.8%
	$\sigma_{trigger}$	0.7%	0.7%	0.7%	0.7%	0.7%
	σ_{ϵ}^{tot}	1.8%	2.9%	1.8%	2.6%	2.2%
C_{bg}	$\sigma_{C_{bg}}$	0.2%	2.1%	0.6%	17.9%	190.7%

Table 0.3.: Incertitudes systématiques relatives sur les différentes quantités impliquées dans l'ajustement. Les incertitudes sur l'efficacité de reconstruction du signal sont sommées en quadrature pour obtenir l'incertitude totale.

Table 0.4.: Valeurs attendues pour les nombres totaux d'événements de bruit de fond et les limites supérieures à 90% de niveau de confiance sur le rapport d'embranchement pour les différents modes. Les valeurs observées par Belle sont données pour comparaison.

	$N_{attendu}$	$\mathcal{B}_{LS}^{90,attendu} \times 10^{-8}$	$\mathcal{B}_{LS}^{90,Belle} \times 10^{-8}$
$e^-e^+e^-$	$6.08^{+4.29}_{-2.87}$	2.50	2.7
$e^-e^+\mu^-$	$12.11^{+5.70}_{-4.31}$	2.00	1.8
$e^-\mu^+e^-$	$10.51^{+5.28}_{-3.90}$	1.54	1.5
$\mu^-\mu^+e^-$	$20.70^{+6.64}_{-5.45}$	1.80	2.7
$\mu^-e^+\mu^-$	$7.53^{+4.51}_{-3.20}$	1.54	1.7

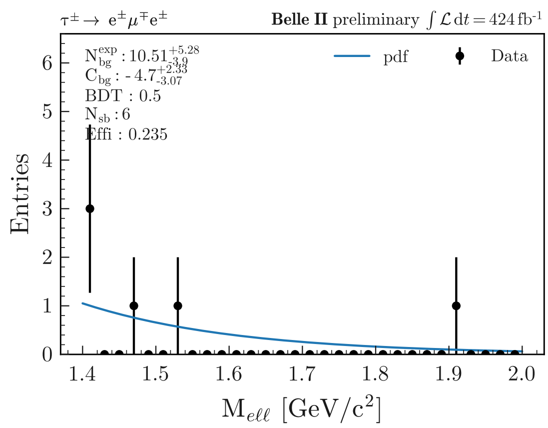


Figure 0.4.: Ajustement de la p.d.f. à la distribution de $M_{\ell\ell'}$ dans les régions latérales pour le canal $e^-\mu^+e^-$.

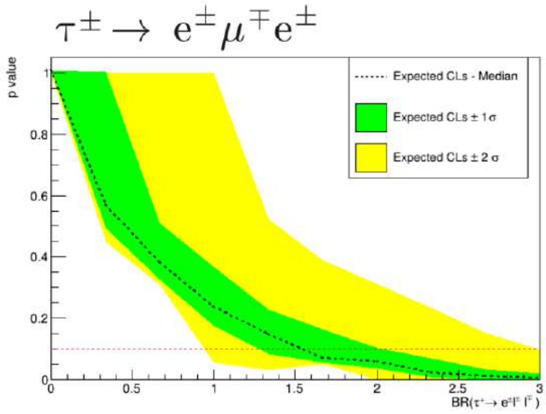


Figure 0.5.: CL_S attendus en fonction du rapport d'embranchement pour le canal $e^{-}\mu^{+}e^{-}$. La ligne rouge en pointillé correspond au niveau de confiance à 90%, et les bandes jaunes et vertes aux intervalles à +1 et +2 sigma.

Contents

Introduction

Particle physics is the field that studies the fundamental constituents of matter and their interactions. Its modern formulation is the Standard Model (SM), a quantum field theory developed during the 20^{th} century. Its most recent achievement has been the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments [6, 7] at the Large Hadron Collider (LHC) of CERN. It accurately describes the fermions, comprising the six quarks (up, down, charm, strange, top, bottom) and the six leptons (electron, muon, tau, and their corresponding neutrinos), and the gauge bosons (photon, $W^{\pm}$, Z^0 and gluon) that mediate the interactions between them.

Despite the fact that the Standard Model continues to withstand experimental tests, it is considered to be incomplete. It fails to account for gravity, the existence of dark matter and dark energy, and faces other issues such as the matter-antimatter asymmetry or the origin of the neutrino masses. Various theoretical and experimental efforts are ongoing to challenge the SM and solve its shortcomings. Two different roads exists: attempting to observe new phenomena forbidden by the SM, or performing precision measurements to compare to theoretical predictions. The search for charged Lepton Flavour Violating (LFV) decays falls in the former. The SM allows this kind of processes at loop level, but they require neutrino oscillations and are therefore highly suppressed to rates around 10^{-55}, well below any current or future experimental sensitivity. The observation of such decays would therefore be a clear sign of new physics beyond the SM.

In that regard, LFV τ decays have garnered a lot of attentions in the past decades, with more than 50 decays channels searched for in leptonic colliders as CLEO, BaBar and Belle.

This thesis outlines the work accomplished during my doctorate at CPPM, within the Belle II collaboration. It is divided in four chapters, with the first two focusing on the theoretical background and experimental setup, while the other two present my personal contributions :

- in Chapter 1, the SM is presented together with various possible new physics scenarios that could enhance the LFV decays rates. The chapter is concluded with a summary of the existing LFV searches.
- The experimental apparatus, comprising the Belle II detector and the SuperKEKB collider, is described in Chapter 2.
- Chapter 3 presents three studies of the fake tracks rate, in the Belle II vertex

detector and its proposed upgrade, the VTX. These studies correspond to the qualification task required to gain authorship of the Belle II publications.

- Lastly, Chapter 4 presents the search for the $\tau \to e\ell\ell'$ decays which I carried out in its entirety.

1. The Standard Model of particle physics

Content

1.1. Introduction

The Standard Model (SM) of particle physics is the theory describing the elementary particles and their interactions. It has been tested to an extraordinary precision but is known to be incomplete; many efforts are made to find its limitations and to build a more complete theory. This chapter provides a brief overview of the SM, and introduces the concept of lepton flavour violation (LFV) in the SM and present several beyond the SM (BSM) scenarios that predict LFV. This chapter is concluded by presenting the current experimental status of the search for the $\tau^- \to e^\mp \ell^\pm \ell'^-$ decays ($\ell, \ell' = e, \mu$).

1.2. Standard Model of particle physics

The SM is a quantum field theory [1, 2, 3, 4], in which the particles are described as quanta of their respective fields. There is a total of seventeen fundamental particles (or fields), which are divided into two categories:

- Fermions, which have a half-integer spin, and are the building blocks of matter. To each fermion is associated its anti-fermion, which has the same mass and opposite quantum numbers (charge, baryon/lepton number, color, etc.).

- Bosons, which have an integer spin, and are responsible for the interactions between said fermions. The bosons are their own anti-particles (note that the W^+ is the antiparticle of the W^-; they are the only two charged bosons.).

The fermions consist of the six quarks: up, down, charm, strange, top and bottom (respectively referred as u, d, c, s, t and b), and of the six leptons: electron, muon, tau (respectively referred to as e, μ and τ), and their corresponding neutrinos (ν_e, ν_μ and ν_τ). The bosons are the photon, the $W^\pm$ and Z, the gluons, and the Higgs. Without considering the anti-particles and the eight different gluons, they add up to seventeen particles, which are shown in Figure 1.1. The fermions come in three generations of two leptons and two quarks and are represented by spinors. The families exhibit a mass hierarchy: each generation is heavier than the previous one, with the top quark (third generation) being the heaviest fermion (and fundamental particle) by a margin.

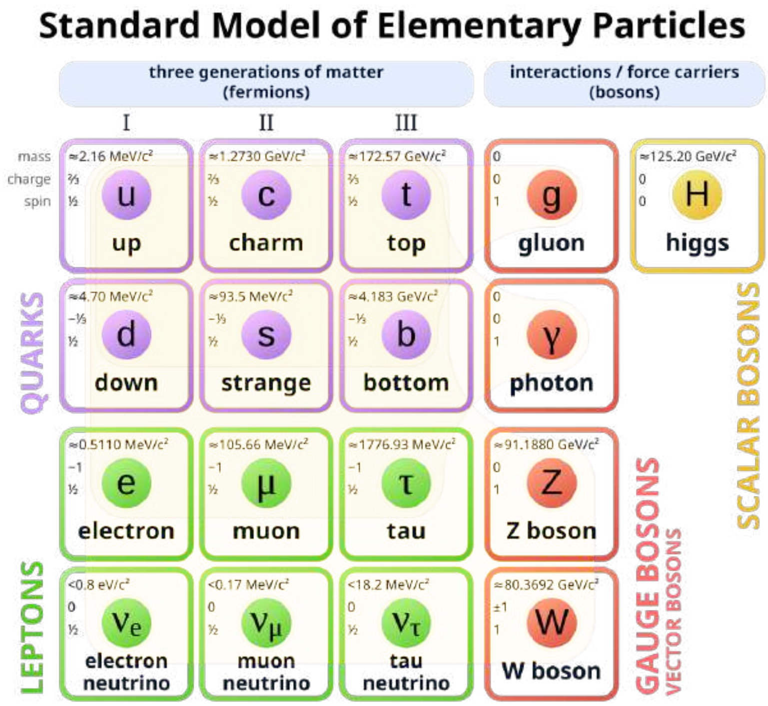


Figure 1.1.: The fundamental particles of the Standard Model. Source: [25].

It is noteworthy to understand that while the whole formulation of the Standard Model, described below, is non-trivial and quite hard to grasp, **it is the underlying theory that describes the world we live in.** The first generation of fermions are the building blocks from which atoms are made; one of the electroweak bosons is what allow us to *see* the world. To properly and consistently understand these four particles and their behaviour, we *have to* incorporate the other generations of fermions and other bosons.

The SM is based on Noether's theorem [5], which states that for every continuous symmetry of the Lagrangian of a physical system, there is a corresponding conserved quantity. In the case of the SM, the symmetries are the gauge symmetries of the Lagrangian, and the conserved quantities are the charges associated with the gauge symmetries. The gauge group of the SM is $SU(3)_C \times SU(2)_L \times U(1)_Y$, where $SU(3)_C$ is the gauge group of the strong interaction, and $SU(2)_L \times U(1)_Y$ is the gauge group of the electroweak interaction. The subscripts C, L and Y stand for "color", "left" and "hypercharge", respectively. The $SU(2)_L$ group is the weak isospin group, and the $U(1)_Y$ group is the hypercharge group. The Lagrangian of the SM, which must be invariant under this gauge group, is then given by:

$$\mathcal{L}_{SM} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}, \tag{1.1}$$

where $\mathcal{L}_{\text{gauge}}$ corresponds to the gauge fields self-interactions, $\mathcal{L}_{\text{QCD}}$ is the Lagrangian of the strong interaction, $\mathcal{L}_{\text{EW}}$ is the Lagrangian of the electroweak interaction. The terms $\mathcal{L}_{\text{Higgs}}$ and $\mathcal{L}_{\text{Yukawa}}$ corresponds to the Higgs mechanism, respectively for the mass of the gauge bosons and of the fermions.

1.2.1. Quantum chromodynamics

Quantum chromodynamics (QCD) is the theory of the strong interaction, which describes the interactions between quarks and gluons. The quarks are the fermions that carry the color charge, and the gluons are the bosons that mediate the strong interaction. The QCD Lagrangian is given by:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a + \sum_i \bar{\psi}_i (i\gamma^\mu D_\mu) \psi_i, \tag{1.2}$$

with D_μ the covariant derivative:

$$D_\mu = \partial_\mu - i g_s T^a G^a_\mu. \tag{1.3}$$

The $G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} G^b_\mu G^c_\nu$ are the gluon strength tensor fields, γ^μ the Dirac matrices, g_s is the strong coupling constant, the ψ_i are the quark fields, and the G^a_μ are the gluon fields. The T^a are the generators of the $SU(3)_C$ group, and the f^{abc} are the structure constants of the $SU(3)_C$ group. The $SU(3)_C$ group has eight generators, which correspond to the eight different gluons. The color charge of the quarks, the source of the strong interaction, is confined. This means that the color charge is never observed in isolation. The confinement of the color charge is due to the fact that the strong coupling constant increases with the distance between the quarks, which leads to the creation of quark-antiquark pairs, and to the creation of mesons and baryons.

1.2.2. Electroweak interaction

The electroweak interaction is the unification of the electromagnetic and weak interactions, whose Lagrangian is:

$$\mathcal{L}_{\text{EW}} = -\frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \sum_i \bar{\psi}_i (i\gamma^\mu D_\mu) \psi_i, \tag{1.4}$$

with D_μ the covariant derivative:

$$D_\mu = \partial_\mu - i g T^a A^a_\mu - i g' Y B_\mu, \tag{1.5}$$

where $W^a_{\mu\nu}$ ($B_{\mu\nu}$) are the strength tensors of the $SU(2)_L$ ($U(1)_Y$) group, the T_a (Y) correspond to the weak isospin charge (hypercharge), the A^a_μ (B_μ) are the gauge fields of the $SU(2)_L$ ($U(1)_Y$) group and g (g') is the associated coupling constant. The ψ fields are the doublets and singlets of the $SU(2)_L$ group, which correspond to the left-handed and right-handed fermions:

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \tag{1.6}$$

As stated above, there are three generations of fermions for both leptons and quarks, ordered as doublet and singlet of the weak interaction:

$$\begin{pmatrix} e_L \\ \nu_L \end{pmatrix}_L, e_R, \begin{pmatrix} u_L \\ d_L \end{pmatrix}_L, u_R, d_R, \tag{1.7}$$

where e runs over three the charged leptons, ν are associated neutrinos and $u = (u, c, t)$ and $d = (d, s, b)$ are the up-type and down-type quarks. It is important to notice that there are no right-handed neutrinos in the SM. This implies that they don't couple to the Higgs field and therefore are massless, as explained in section 1.2.3. The fermions also have an electric charge related to the weak isospin and hypercharge: $Q = T^3 + Y/2$.

1.2.3. Higgs mechanism

At this stage, the total Lagrangian built from $\mathcal{L}_{\text{QCD}}$ and $\mathcal{L}_{\text{EW}}$ describes massless fermions, as a mass term (typically $m\bar{\psi}\psi$) would break the gauge invariance of the Lagrangian. The Higgs mechanism solves this issue by introducing a field breaking the $SU(2)_L \times U(1)_Y$ symmetry. This symmetry breaking leads to the appearance of the Higgs boson and the generation of the masses of the $W^{\pm}$ and Z^0 bosons, and of the fermions while keeping the photon massless. Consider the following complex scalar field, called the Higgs field:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.8}$$

The Lagrangian for the Higgs field interaction with the electroweak bosons can be written as:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2, \tag{1.9}$$

where D_μ is the covariant derivative of the electroweak Lagrangian from Eq 1.5, μ^2 is the mass term of the Higgs field, and λ is the self-interaction term of the Higgs field. The Higgs field has a non-zero vacuum expectation value (vev):

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.10}$$

where $v = \frac{1}{\sqrt{\sqrt{2}G_F}} = 246$ GeV is the vev of the Higgs field. The gauge bosons mass terms are obtained by evaluating the Lagrangian 1.9 at the vev of the Higgs field (considering only the relevant terms):

$$\mathcal{L}_{\text{Higgs-gauge}} = \frac{1}{2}\begin{pmatrix}0 & v\end{pmatrix}(\partial_\mu - igT^a A^a_\mu - ig'YB_\mu)(\partial^\mu - igT^b A^{b\mu} - ig'YB^\mu)\begin{pmatrix}0\\ v\end{pmatrix}, \tag{1.11}$$

There are three massive vector bosons, that can be written as:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(A^1_\mu \mp iA^2_\mu) \tag{1.12}$$

$$Z^0_\mu = \frac{1}{\sqrt{g^2+g'^2}}(gA^3_\mu - g'B_\mu) \tag{1.13}$$

whose respective masses are $m_W = g\frac{v}{2}$ and $m_Z = \sqrt{g^2+g'^2}\frac{v}{2}$. The fourth vector field corresponds to the photon and remains massless:

$$A_\mu = \frac{gA^3_\mu + g'B_\mu}{\sqrt{g^2+g'^2}} \tag{1.14}$$

The fermions also acquire masses through the Higgs mechanism:

$$\mathcal{L}_{\text{Yukawa}} = Y^u_{ij}\bar{q}_{L_i}\tilde{\phi}u_{R_j} + Y^d_{ij}\bar{q}_{L_i}\phi d_{R_j} + Y^e_{ij}\bar{l}_{L_i}\phi e_{R_j} + h.c. \tag{1.15}$$

where i and j are running over the three generations of fermions. The right-handed fields are involved in the mass terms; since there are no right-handed neutrinos, they remain massless in the SM.

1.2.4. Discrete symmetries of the Standard Model

In addition to the symmetry group of the SM, there exists additional discrete symmetries. The most important ones are :

- the charge conjugation (C), which maps particle on anti-particle and vice-versa,
- the parity (P), which inverts the spatial coordinates,
- the time reversal (T), which inverts the time coordinate.

These three symmetries are *not* conserved independently in the SM (for instance parity has famously been shown to be violated in the weak interaction by the observation of the β decay of the Cobalt-60 nucleus by the Wu experiment in 1957 [26]). Yet, the three together are conserved i.e. CPT is a symmetry of the SM. CP alone is known to be violated in the weak interaction, so T must also not be conserved. The interesting point is that CP symmetry maps right-handed fermions onto left-handed anti-fermions and therefore can introduce matter-antimatter asymmetry from a universe with as much matter as anti-matter. There also are other discrete symmetries, such as the baryon number (B) and the lepton number (L), which are conserved in the SM. Lastly, the flavour must be mentioned: to each lepton family is associated a quantum number called lepton flavour, with both the lepton and its associated neutrino having the same lepton flavour. Quarks also carry a flavour. As will be discussed in the following parts of this chapter, the flavour of the quarks is not conserved by the weak interaction, nor is the lepton flavour (in the case of massive neutrinos).

1.2.5. Quark mixing and the CKM matrix

Now that fermions have acquired mass, it is important to observe that their mass eigenstates are not the same as their weak eigenstates, which implies that the weak interaction can change the flavour of the fermions. For the quarks, the matrix to rotate from one basis to the other is called the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.16}$$

It can be parametrised by four parameters, three rotation angles and a complex CP-violating phase. This rotation matrix induces that the weak interaction can change the flavour of the quarks, which is called quark mixing. This mixing is possible only through charged currents: up-type quarks can change into down-type quarks (or vice versa) through the exchange of a W boson.

I concluded the previous section by stating that the absence of right-handed neutrinos implied massless neutrinos: therefore such rotation matrix cannot exist for the leptons and the lepton flavour has to be conserved in the SM.

1.3. Lepton flavour violation in the SM

1.3.1. Neutrino oscillations

In 1976, the Homestake experiment (and followed in 1998 by the Super-Kamiokande experiment) observed for the first time neutrino oscillations, which is the mixing of the lepton flavour of the neutrinos. This process was then observed again in multiple contexts (using atmospheric [8], solar [9], reactor [10] and accelerator [11] neutrinos), and is now experimentally well established. The straightforward explanation to this lepton flavour mixing is to introduce a CKM-like[1] matrix for the leptons, and non-zero masses for the neutrinos.

This is formalised by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix. The PMNS matrix is a unitary matrix diagonalising the neutrino mass matrix, and,

[1] Neutrino oscillations were actually predicted in 1957 by Bruno Pontecorvo[27, 28], before the introduction of the CKM matrix.

similarly to the CKM matrix for quarks, is parametrised by three mixing angles and a CP-violating phase.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}. \tag{1.17}$$

Each element of the matrix represent the mixing between the flavour and mass eigenstates: $|\nu_\alpha\rangle = \sum_i U_{\alpha i} |\nu_i\rangle$.

The probability for a neutrino of flavour α to be detected as a neutrino of flavour β after travelling a distance L in vacuum is then given by:

$$\begin{aligned} P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} - 4 \sum_{i>j} \mathrm{Re}(U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}) \sin^2 \left(\frac{\Delta m^2_{ij} L}{4E} \right) \\ + 2 \sum_{i>j} \mathrm{Im}(U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}) \sin \left(\frac{\Delta m^2_{ij} L}{2E} \right) \end{aligned} \tag{1.18}$$

where $\Delta m^2_{ij} = m^2_i - m^2_j$, E is the energy of the neutrino, i and j are running over the neutrino mass states and L is the distance travelled by the neutrino.

It is important to note that neutrino oscillations and quark mixing are the exact same process; the main difference being that the neutrino are always considered in the *weak eigenstates* instead of the mass eigenstates.

Neutrino oscillations are a clear manifestation of lepton flavour violation (LFV), although they imply that neutrinos have a non-zero mass. Therefore, this raises the question of the origin of their mass; one straightforward way to answer it is to add a right-handed neutrino to the Standard Model through the see-saw mechanism (there exist different ways to introduce these right-handed neutrinos, for instance as singlet or triplet). In such scenario, the right-handed neutrinos acquire an extremely large mass M_R, and the left-handed neutrinos acquire a mass $m_\nu \approx \frac{m_D^2}{M_R}$, where m_D is the Dirac mass term of the neutrinos. It is important to note that no heavy right-handed neutrinos have been observed yet: they could be sterile (i.e. not interacting with the SM particles[1]) or simply not be the answer to the neutrino mass question.

1.3.2. Charged lepton flavour violation

In the SM, the charged lepton flavour violation (cLFV) is forbidden at tree level (and similarly there is no transition between up-type quarks or between down-type quarks at tree level). However, cLFV can occur at loop level, through the exchange of W bosons and neutrino oscillations leading to processes such as $\mu \to e\gamma$, $\tau \to \ell\gamma$ or $\tau \to \ell\ell\ell$ ($\ell = e, \mu$). The diagram for $\mu \to e\gamma$ is shown on Figure 1.2 and

[1] Note that in this case, they would not contribute to the neutrinos mass m_ν, which therefore require more complicated explanations

illustrates the transition between any two charged leptons. For the energy to be conserved, a photon must be emitted. $\tau \to \ell\ell\ell$ decays follow the same principle (see Figure 1.3), where the radiative neutral boson (it can also be an off-shell Z) decays into a pair of charged leptons. Box diagrams are also possible and allow decays such as $\tau^- \to e^- e^- \mu^+$ where the lepton flavour is not conserved by two units.

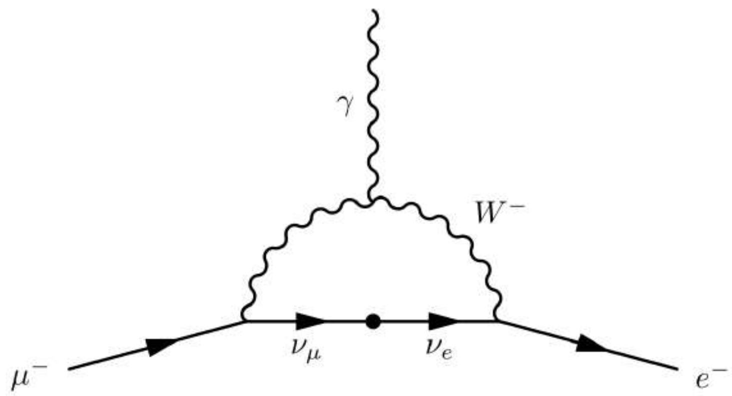


Figure 1.2.: Standard Model diagram for $\mu \to \ell\gamma$. The dot corresponds to the oscillation of the ν_μ into a ν_e. Source: [29].

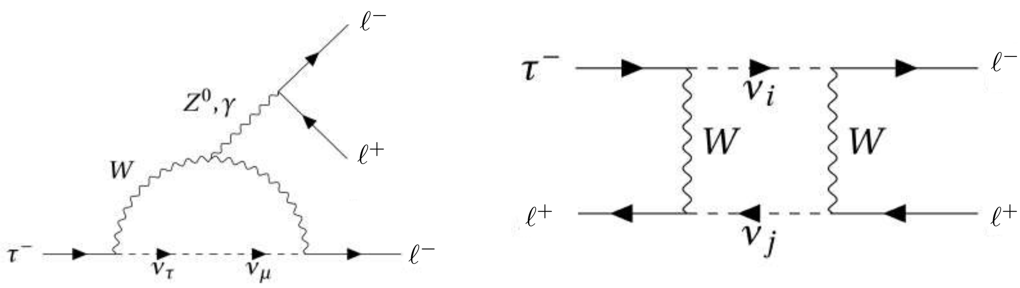


Figure 1.3.: Standard Model diagram for $\tau \to \ell\ell\ell$. Left is the Z^0 penguin contribution, right is the $W^\pm$ box contribution. Source: [30].

The amplitude for $L^- - \ell^-$ transition can be approximated by:

$$\mathcal{A}_{L-\ell} \propto \sum_i U_{Li} U^*_{\ell i} m^2_{\nu_i}, \tag{1.19}$$

where U_{Li} and $U_{\ell i}$ are the elements of the PMNS matrix, and m_{ν_i} are the neutrino masses. This process is consequently heavily suppressed by the neutrino masses; the branching ratio for $\tau \to \ell\ell\ell$ is of the order of 10^{-55}[12, 13] in the SM[1]. While

[1] I initially introduced the SM with no neutrino masses, thus no LFV, and now claim that

the exact value is not necessarily precisely known (as we only have upper limits on the neutrino masses), it remains an extremely clear prediction from the Standard Model:
There will be no cLFV observed in current, or any foreseeable future, experiments.
This in turn makes cLFV a very interesting probe of new physics, as any observation of cLFV would be a clear sign of new physics beyond the Standard Model (BSM).

1.4. cLFV in BSM scenarios

1.4.1. The limitations of the Standard Model

Despite the numerous successes of the SM (the computation of the anomalous magnetic dipole moment of the electron, the prediction of the electroweak bosons, the Higgs boson, the top quark, etc.), it is known to be incomplete. For one, it does not include gravity and dark matter (assuming that gravity can be described as a quantum phenomenon and that dark matter has a particle nature), does not answer the question of the nature of neutrinos and the source of their mass, nor does it explain the apparent matter/antimatter asymmetry (the amount of CP violation is not enough to explain it). A few anomalies have also been observed in the past, such as discrepancies in the measured muon g-2 with respect to the SM value [31] or tensions in lepton flavour universality from decays of b-mesons, that hint at possible new physics. In particular, the measurements of the ratio $R(D^{(*)})$ [32, 33] defined as:

$$R(D^{(*)}) = \frac{\mathcal{B}(B \to D^{(*)}\tau\nu)}{\mathcal{B}(B \to D^{(*)}\ell\nu)}$$

shows a deviation as high as 3 standard deviations from the SM prediction.

Various BSM models try to tackle these shortcomings or to explain the anomalies. These models usually allow cLFV at much higher rates than in the SM, a few of which are presented in the following section. The experimental constraints on cLFV are first presented, as they dictate the possible phasepace for the BSM models.

1.4.2. Type II Seesaw

The seesaw mechanism is a generic class of models that explain the smallness of the neutrino mass by introducing new particles with a large mass scale, which can also induce cLFV [34, 35, 36, 37] through its couplings to the leptons. There

it actually predicts LFV. This slight confusion stems from how exactly one want to define the SM: either without neutrino masses (contrary to observations) or with neutrino masses (with no explanation for them). Wether neutrino oscillations are part of the SM is left to the reader's interpretation.

exist different types of seesaw; this section will present type II [38, 39, 40] seesaw model. In such scenario, the SM is extended with a SU(2) triplet scalar field[1] Δ with hypercharge $Y = 1$. The coupling of the leptons to the triplet is given by the Lagrangian:

$$\mathcal{L}_{\Delta-leptons} = f_{\alpha\beta}\overline{\ell^c_\alpha}\left(i\tau_2\right)\tau_I\ell_\beta\Delta^I, \tag{1.20}$$

where f is a 3×3 matrix proportional to the neutrino masses, ℓ are the left-handed lepton doublets, τ are the Pauli matrices, and Δ^I are the components of the triplet. This coupling allows neutrinos to have masses of the order of $\frac{f_{\alpha\beta}\lambda_H v^2}{M_\Delta}$ where λ_H is the coupling of the scalar triplet to the Higgs, v^2 the Higgs vev and M_Δ the mass of the triplet ($\sim$ TeV). More interestingly for the present work, it also introduces cLFV; the branching ratios for $\tau \to \ell\ell\ell$ are given by [38]:

$$BR\left(\tau \to l_L\bar{l}l\right) = 0.18\left\{2\left|C^{l\tau\ell\ell}_{V,LL} + 4eC^{l\tau}_{D,R}\right|^2 + \left(64\ln\frac{m_\tau}{m_l} - 136\right)\left|eC^{l\tau}_{D,R}\right|^2 + \left|C^{l\tau\ell\ell}_{V,LR} + 4eC^{l\tau}_{D,R}\right|^2\right\} \tag{1.21}$$

$$BR\left(\tau \to \ell'_L\bar{l}l\right) = 0.18\left\{\left|C^{\ell'\tau\ell\ell}_{V,LL} + 4eC^{\ell'\tau}_{D,R}\right|^2 + \left(64\ln\frac{m_\tau}{m_l} - 136\right)\left|eC^{\ell'\tau}_{D,R}\right|^2 + \left|C^{\ell'\tau\ell\ell}_{V,LR} + 4eC^{\ell'\tau}_{D,R}\right|^2\right\} \tag{1.22}$$

$$BR(\tau \to \ell_L\bar{\ell}'\ell) = 0.18\left|C^{\ell\tau\ell'\ell}_{V,LL}\right| \tag{1.23}$$

where the C are the coefficients from the effective field theory formulation of the Lagrangian and the D and V subscript stand for dipole and vector. These coefficients depend on the elements of the neutrino mass matrix; it is therefore possible to consider situations where τ LFV decays are within reach of experiments despite the tight constraints from $\mu \to 3e$.

1.4.3. Supersymmetry

Considering the success of the SM as a theory built on symmetries, supersymmetry comes as a natural extension of the SM. It introduces, for every existing particle, its supersymmetric partner with same quantum numbers except for spin: bosons are mapped onto fermions, and fermions onto bosons. There are many different supersymmetric models; I consider here one of the simplest called the Minimal Supersymmetric Standard Model (MSSM) [41, 42, 43], whose symmetry group is SO(10) including right-handed neutrinos.

Similarly to the SM, LFV is caused by the mixing of sleptons (superpartners of the leptons) and sneutrinos (neutrinos), in particular because of the inclusion of right-handed neutrinos. The main difference from the SM is that the mixing

[1] Note that in other seesaw types, a heavy sterile neutrino is added.

is not driven by the neutrino masses ($\propto 1/M_R$, the mass of possible right-handed neutrinos), but by $1/M_{Susy}$ with M_{Susy} the energy scale of supersymmetry and $M_{Susy} \ll M_R$.
Another source of LFV in the MSSM is the Higgs sector: two Higgs doublets H_U and H_D are added, whose coupling to the slepton induce flavour-changing neutral currents. Under some assumptions [41], the branching ratio of $\tau \to \mu\mu\mu$ can be written as:

$$\mathrm{Br}(\tau \to 3\mu) \simeq \left(1 \times 10^{-7}\right) \times \left(\frac{\tan\beta}{60}\right)^6 \times \left(\frac{100\mathrm{GeV/c}^2}{m_A}\right)^4, \tag{1.24}$$

where $\tan\beta$ is the ratio of the vacuum expectation values of the two Higgs doublets, and m_A is the mass of the Higgs. In the high $\tan\beta$, this branching ratio can be as high as 10^{-8}. This decay involves the $\bar{\tau}_R\mu_L$-Higgs coupling but other couplings such as $\bar{\tau}_R e_L$-Higgs and $\bar{\mu}_R e_L$-Higgs can produce $\tau \to e\mu\mu$ and $\tau \to ee\mu$.

1.4.4. Littlest Higgs with T-parity

Electroweak precision measurements have shown no sign of new physics up to $\propto$ 10 TeV, while stabilizing the mass of the Higgs requires new physics at $\propto$ 1 TeV. This discrepancy is called the hierarchy problem, to which a possible solution is the Littlest Higgs with T-parity (LHT) [44, 45, 46] model, where the discrete symmetry forbids tree-level correction to electroweak observables. It introduces mirrors gauge bosons that are T-odd, while the SM particles (and the mirror top quark) are T-even. Mixing from the mirror leptons can drastically enhance the LFV branching ratios, with $\mathcal{B}(\tau \to \ell\ell\ell') \propto 10^{-8}$.

1.4.5. Axion-like particles

Axion-like particles (ALPs) [47, 48] are hypothetical pseudoscalar bosons similar to the QCD axion (that was proposed to solve the strong CP problem [49]). They have a large phasepace in mass and coupling to the SM in which they can exist and also provide a candidate to dark matter.

If the ALP is taken as a pseudo-Goldstone boson X, its coupling to SM charged leptons is given by:

$$\mathcal{L}_X = -i\frac{X}{\Lambda}\bar{\ell}_\alpha G_{\alpha\beta}((m_\alpha - m_\beta) + (m_\alpha + m_\beta)\gamma^5)\ell_\beta, \tag{1.25}$$

where $\ell_{\alpha,\beta} = e, \mu, \tau$ are the SM charged leptons, $G_{\alpha\beta}$ is the (real) coupling matrix and Λ is the effective scale of the model. The width of the decays $\ell_\alpha \to \ell_\beta X$ is then given by:

$$\Gamma(\ell_\alpha \to \ell_\beta X) = \frac{m_\alpha^3}{8\pi}\sqrt{\left(1-\left(\frac{m_\beta+m_X}{m_\alpha}\right)^2\right)\left(1-\left(\frac{m_\beta-m_X}{m_\alpha}\right)^2\right)} \tag{1.26}$$

$$\times g_{\alpha\beta}^2\left[\left(1-\frac{m_\beta^2}{m_\alpha^2}\right)^2-\frac{m_X^2}{m_\alpha^2}\left(1+\frac{m_\beta^2}{m_\alpha^2}\right)\right] \tag{1.27}$$

and the two-body decays of X into two charged leptons:

$$\Gamma\left(X \to l_\beta^- l_\beta^+\right) = \frac{m_X m_\beta^2}{2\pi} g_{\beta\beta}^2 \sqrt{1-\frac{4m_\beta^2}{m_X^2}} \tag{1.28}$$

$$\Gamma\left(X \to \ell_\alpha^- \ell_\beta^+\right) \simeq \frac{m_X m_\alpha^2}{2\pi} g_{\alpha\beta}^2 \left(1-\frac{m_\alpha^2}{m_X^2}\right)^2 \tag{1.29}$$

where the second one is obtained in the limit $m_\alpha \gg m_\beta$ (a quite safe assumption in the context of a decay into μ and e). Measurement of the branching fraction of $\tau \to \ell\ell\ell'$ can be used to constrain the elements of the coupling matrix $G_{\alpha\beta}$ and the mass of the ALP. A key point of this model is that the ALP can be long-lived and therefore may travel even a few meters before decaying into a pair of leptons. It therefore can be searched for by looking at displaced vertices (assuming it decays within the detector, otherwise it obviously cannot be observed) and would yield a unique and clean experimental signature making it almost background-free. However, such a search is not performed in the work presented in this thesis; the prompt search (i.e. non-displaced) nonetheless can probe ALP phasespace, with a lower sensitivity.

1.4.6. Leptoquarks

A key feature of the SM is lepton flavour universality (LFU), the fact that leptons of different flavours all couple to the gauge bosons in the same way. It has noticeably been studied in semi-leptonic B decays, and as mentioned above, tensions in $b \to c\tau\nu$ have persisted. One possible way to solve these discrepancies is to introduce leptoquarks, which are particles that couple to both quarks and leptons. They carry a new quantum number, $F = 3B + L$, where B and L are respectively the baryon and lepton number and can be either scalars or vectors. Leptoquarks models also predict cLFV in τ decays; while it would mainly enhance semi-leptonic decays such as $\tau \to \ell\phi$, loop contributions could also lead to cLFV in $\tau \to \ell\ell\ell'$ decays. For $\tau \to 3\mu$ and $\tau \to 3e$ the branching ratio can be up to 10^{-10} [50]. This value is still out of our current reach but should be attained by the Belle II experiment with its expected dataset of 50 ab^{-1}.

1.5. Experimental status

1.5.1. Experimental constraints

The main message from the previous section is that there exist many models allowing cLFV. Until now, no cLFV has been observed, but the experimental constraints are not tight enough to rule out any of the models. The most stringent limits come from muon decays:

- in $\mu \to e\gamma$, the upper limit at 90% confidence level (CL) is $\mathcal{B}(\mu \to e\gamma) < 4.2 \cdot 10^{-13}$ and was set by the MEG collaboration [51],

- in $\mu \to eee$, the upper limit at 90% CL is $\mathcal{B}(\mu \to eee) < 1.0 \cdot 10^{-12}$ and was set by the SINDRUM collaboration [52],

- in $\mu N \to eN$, from target experiments such as COMET [53], DeeMe [54] or Mu2e [55].

These three modes are the so-called "golden modes" for cLFV searches as they are the ones with the best sensitivity. They typically limit the observations of other cLFV processes such as $\tau \to \ell\ell\ell$ (although some models can predict a higher rate for τ decays than for μ decays). There also exists less constraining limits from other decays:

- in $Z \to \ell\ell'$ ($\ell, \ell' = e, \mu, \tau$), the upper limits at 95% CL are $\mathcal{B}(Z \to e\mu) < 2.62 \cdot 10^{-7}$ [56], $\mathcal{B}(Z \to e\tau) < 8.1 \cdot 10^{-6}$ and $\mathcal{B}(Z \to \mu\tau) < 9.5 \cdot 10^{-6}$ [57], by the ATLAS collaboration[1],

- in B decays, the upper limits $\mathcal{B}(B^0 \to \tau\mu) < 1.4 \cdot 10^{-5}$ and $\mathcal{B}(B^0_s \to \tau\mu) < 4.2 \cdot 10^{-5}$ at 95% CL have been set by the LHCb collaboration[2] [60].

1.5.2. Searches for τ LFV decays

While an observation of a single cLFV process would be an evidence of new physics, it would not necessarily tell us which model in particular is correct. The various models however predict different correlations between the various cLFV processes; the measurement of multiple cLFV processes[3] would therefore discriminate between the different models. In that respect, τ LFV decays are extremely interesting as they allow to probe a wide phasepace of possible new physics with semi-leptonic decays (such as $\tau \to \ell V^0$, $V^0 = \phi, \rho, K^*$...), fully hadronic decays (such as $\tau \to \Lambda\pi$ - also not conserving the baryon number) or fully leptonic decays. Over the years, roughly 50 different τ LFV decays [61] have been searched for, with the most

[1] There also exists limits from Higgs decays, but these are even less constraining [58, 59]
[2] Other B decays such as $b \to s\ell\ell'$ ($\ell, \ell' = e, \mu, \tau$) are used to search for LFV.
[3] Or the measurement of one decay coupled to an extremely stringent limit on other processes

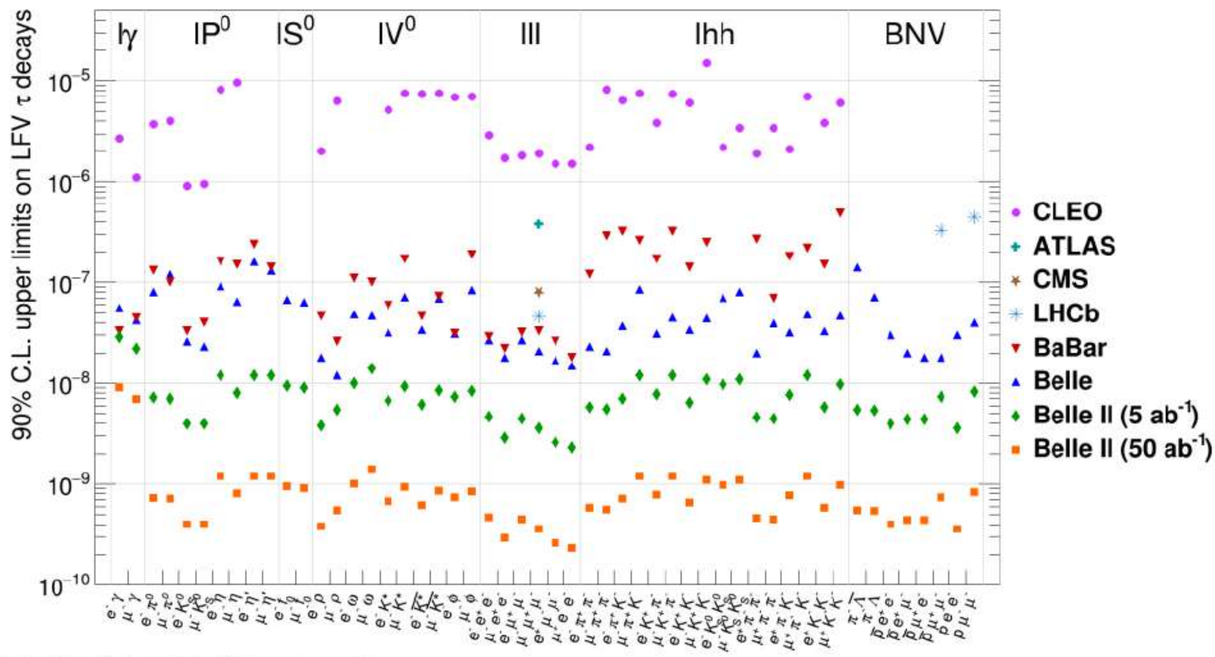
stringent limits set by the Belle experiment. The various measurements are shown on Figure 1.4, with the expected sensitivity of Belle II. The analysis presented in this thesis is the search of the five decays $\tau \to e\ell\ell'$, with $\ell, \ell' = e, \mu$.

The 3μ final state was already studied at Belle II [24] and the limit was improved with respect to the Belle's result [15]. The $\tau \to \ell\phi$ channel was also studied [23], although with a smaller dataset (190fb^{-1}, from the 2019-2021 data-taking period). The values for $\tau \to \ell\ell\ell$ are reported in Table 1.1.

Mode	Belle	Babar	Belle II
$e^-e^+e^-$	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	-
$\mu^-\mu^+\mu^-$	2.1	3.3	1.9

Table 1.1.: Current observed upper limits at 90% C.L. on $\tau \to \ell\ell\ell$ branching fractions obtained by BaBar [14], Belle [15], Belle II [24], CMS [62] and LHCb [63] experiments. Values are given multiplied by 10^8.

Now that the relevance of searching for τ LFV decays has been established, we need to 1) build a machine that will provide as many τ as possible and 2) build a detector that will allow us to observe the decays of the τ's.



Banerjee et al., 2022a; Kou et al., 2019a

Figure 1.4.: Current limits on τ LFV decays and expected sensitivity of Belle II [64]. The expected sensitivity is given for 5 and 50 ab^{-1}, and is obtained by rescaling the results obtained by Belle.

2. The Belle II experiment

Content

Belle II [16, 17] is a particle physics experiment located at the collision point of the SuperKEKB [18] collider, at the KEK laboratory in Tsukuba, Japan. Both the collider and the detector are upgrades of the previous KEKB collider and Belle detector, respectively. This section describes the collider and the Belle II detector.

2.1. SuperKEKB

The SuperKEKB collider is a two-ring collider, with an electron ring and a positron ring. The electron ring has an energy of 7 GeV (HER), and the positron ring has an energy of 4 GeV (LER). The beams' energy was chosen to obtain a collision energy in the centre of mass of $\sqrt{s} = 10.58\,\text{GeV}$, corresponding to the mass of the $\Upsilon(4S)$ resonance that mainly decays in a pair of b-flavoured mesons (B^+B^- and

$B^0\bar{B}^0$). The energy asymmetry leads to a Lorentz boost in the laboratory frame:

$$\beta\gamma = \frac{\mathbf{p_{e^-}} - \mathbf{p_{e^+}}}{\sqrt{s}} \simeq 0.28. \tag{2.1}$$

This boost allows the B mesons to fly an average distance of 130 μm before decaying, which is important for time-dependent CP analyses.

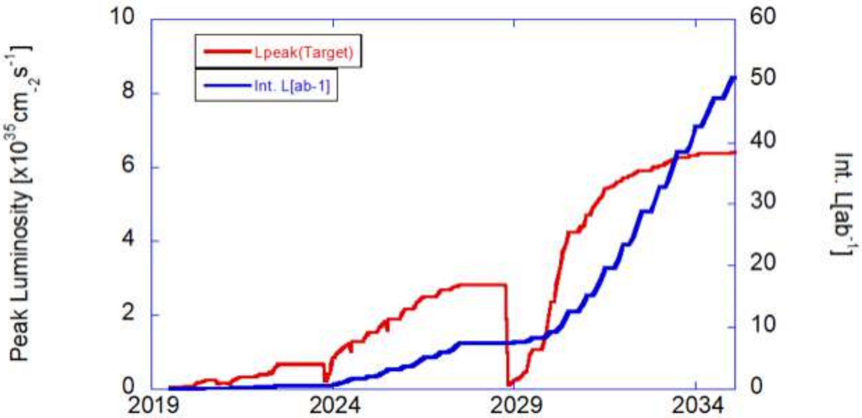


Figure 2.1.: Projected peak luminosity for SuperKEKB.

The SuperKEKB collider is designed to operate at a luminosity of $6\times10^{35}\text{cm}^{-2}\text{s}^{-1}$. This represents a luminosity 30 times higher than that of the KEKB collider, in order to collect a total of 50 ab^{-1} and to increase the number of B mesons produced to push the so-called "intensity" frontier (as opposed to the "high-energy" frontier, i.e. the strategy followed by the LHC). The expected instantaneous luminosity and integrated luminosity are shown in Figure 2.1.

There are two essential elements to attain this goal:

- A twofold increase in beam current. Figure 2.3 shows the beam current during data taking in June 2024.

- A reduction of the beam size at the interaction point (IP) from about 1μm to 50 nm by using a nano-beam scheme (initially designed for the canceled Italian Super B) [65].

This is understood when considering the instantaneous luminosity:

$$\mathcal{L} = \frac{\gamma^{\pm}\xi_y n^{\pm} f_c}{2r_e\beta_y}\left(1 + \frac{\sigma_y}{\sigma_x}\right), \tag{2.2}$$

where the sign + (−) stands for the positron (electron), γ is the Lorentz factor, ξ_y is vertical beam-beam parameter, n is the number of particles per bunch, f_c is the collision frequency, r_e is the classical electron radius, β_y is the vertical beta function at the IP, σ_y and σ_x are the vertical and horizontal beam sizes at the IP.

Considering that $n^{\pm} f_c \propto \frac{I^{\pm}}{e}$ and $\frac{\sigma_y}{\sigma_x} \propto 10^{-3}$, Eq 2.2 can be rewritten as:

$$\mathcal{L} \propto \frac{I \xi_y}{\beta_y}. \tag{2.3}$$

Minimising β_y requires minimising the longitudinal beam overlap d, which can be done by introducing a crossing angle $2\phi = 83$ mrad as shown on Figure 2.2. The beam overlap and the crossing angle are inversely proportional:

$$d = \frac{\sigma_y}{\phi}. \tag{2.4}$$

The value for ϕ is dictated by the magnet design and detector background level; increasing it further would deteriorate ξ_y, thus the beam sizes are reduced instead. Another improvement was to adopt the so-called "crab-waist" scheme [66], consisting in focusing one of the beams (minimising its vertical beta function) along the trajectory of the other beam using a crab sextupole. The list of the updated parameters compared to KEKB is given in Table 2.1.

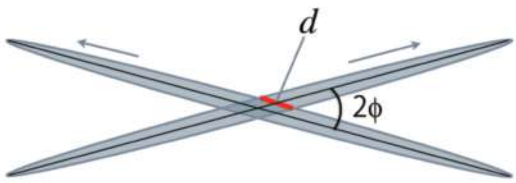


Figure 2.2.: Schematic view of the beam colliding in SuperKEKB. Source: [16].

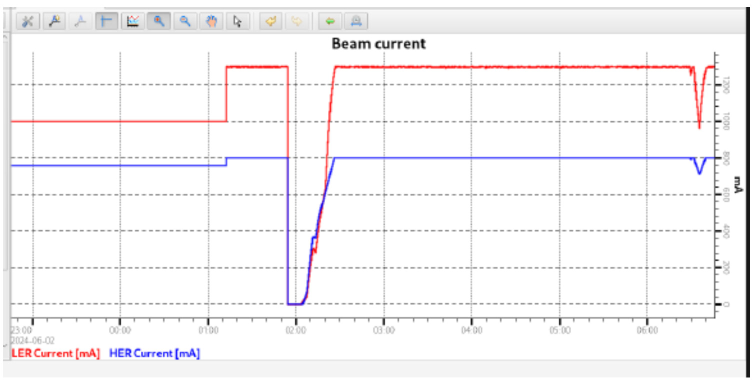


Figure 2.3.: Beam currents during a control room shift in June 2024. The nominal values have not been reached yet but the beam currents are slowly ramping up. The small drop in the beam currents before 7 am corresponds to an earthquake.

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} cm^{-2} s^{-1}	2.1	4.65	80

Table 2.1.: Beam parameters for KEKB and SuperKEKB.

The beams also offer some flexibility in the choice of the collision energy allowing to perfom various off-resonance data-taking. SuperKEKB is by design a b-factory, but many other collision outcomes are possible, such as light quark pairs production and QED processes, for which the production cross sections are given in Table 2.2.

SuperKEKB can be considered as a τ-factory: the pair production cross section is of the same magnitude than the $b\bar{b}$ one, and doesn't require the $\Upsilon(4S)$ resonance. The analysis presented in Chapter 4 therefore uses the full Run 1 data set collected by Belle II, including the off-resonance data, of 424 fb^{-1}, consisting of roughly 380 million τ pairs.

In addition to these physics processes, other phenomena are also possible coming from the beams themselves instead of their collision:

- Synchrotron radiation: electrons and positrons emit synchrotron radiation when they are bent by the magnetic fields of the collider,
- Beam-gas interactions: the beams can interact with the residual gas in the beam pipe,
- Touschek scattering: Coulomb scattering from electrons within the same bunch. Since the luminosity increase is achieved through the reduction of the beam size, this effect is expected to be twenty time more important than in KEKB.

Beam-gas and Touschek scattered particles are lost by hitting the beam pipe walls; the resulting showers can reach the detector. These beam induced background effects result therefore in additional tracks that don't come from a collision; they produce so-called fake tracks degrading the tracking performances. Chapter 3 presents study of these beam-induced fake tracks.

Process	Cross section (nb)
$\Upsilon(4S)$	1.110
$u\bar{u}$	1.61
$d\bar{d}$	0.40
$s\bar{s}$	0.38
$c\bar{c}$	1.30
$\tau^+\tau^-$	0.919
$\mu^+\mu^-(\gamma)$	1.148
$e^+e^-(\gamma)$	300
$e^+e^-e^+e^-$	39.7
$e^+e^-\mu^+\mu^-$	18.9

Table 2.2.: Cross sections for various processes at SuperKEKB at $\sqrt{s} = 10.58$ GeV. Bhabha scattering, $e^+e^-e^+e^-$ and $e^+e^-\mu^+\mu^-$ are dominant processes that unfortunately for the analysis of Chapter 4 are relevant sources of background.

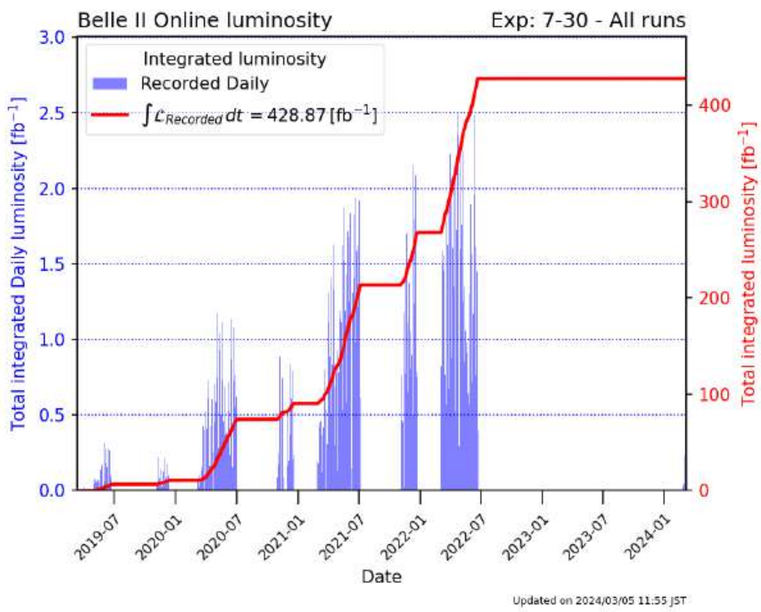


Figure 2.4.: Luminosity recorded by Belle II during Run 1. The actual luminosity from *good* runs is 424 fb^{-1}.

SuperKEKB has already recorded 424fb^{-1} during Run 1 as shown in Figure 2.4. An additional $\sim$ 100fb^{-1} has been recorded since data taking resumed, but is not part of the analysis presented in this thesis.

2.2. Detector

The Belle II detector [16] (see Figure 2.5) is located at, or around, the IP of SuperKEKB. It is an eight meter tall cylinder, most of whose components are either new of upgraded from the previous Belle detector. The detector is composed of several subdetectors, each with a specific role in the reconstruction of the particles produced in the collision. The full detector must satisfy the following requirements for Belle II physics programm:

- Excellent vertex resolution ($\sim 50\ \mu m$)
- High reconstruction efficiency for charged particles down to a few hundred MeV/c
- Very good momentum resolution up to $\sim$ 8 GeV/c
- Precise measurement of photons energy from a few tens of MeV to $\sim$ 8 GeV
- Efficient particle identification for pions, kaons, protons, electrons and muons
- Coverage of the (almost) full solid angle
- Fast trigger system

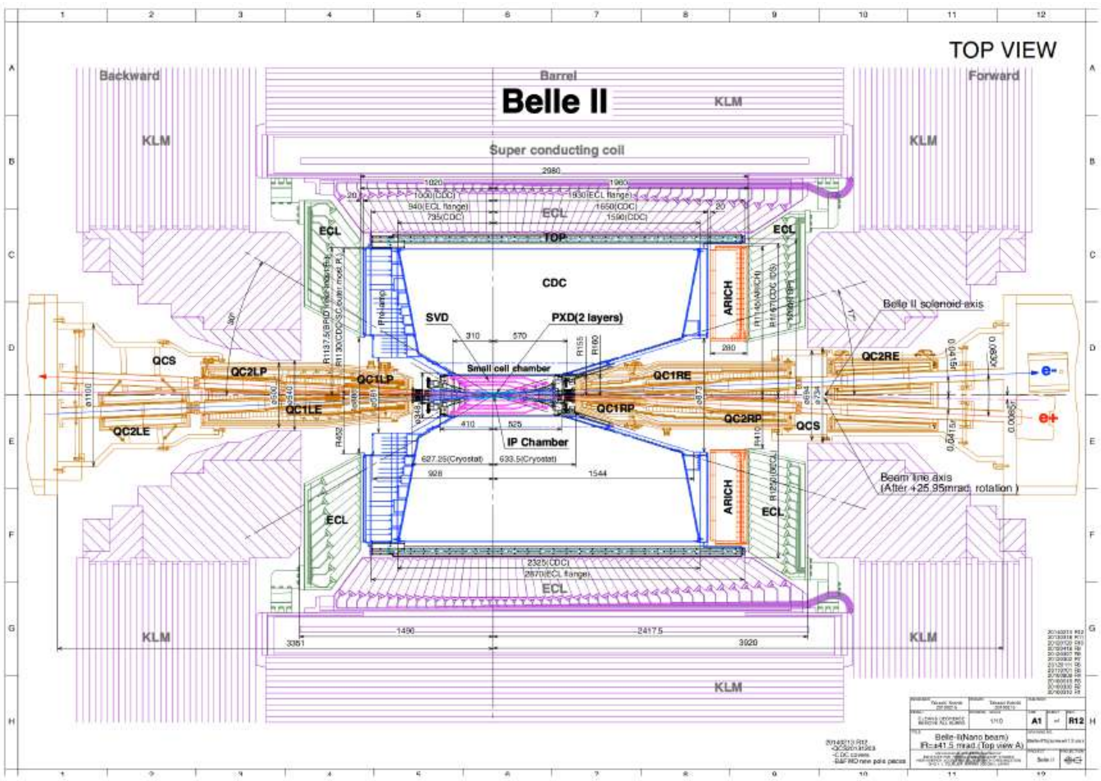


Figure 2.5.: Top view of the Belle II detector. Source: [17].

2.2.1. VXD

The vertex detector (VXD) is the innermost part of the Belle II detector. It is designed to provide precise measurements of the tracks produced in the collision, and to reconstruct the decay vertices of the particles. It is also the first detector to be crossed by the particles, and must therefore be able to withstand the high radiation levels produced by the beams. The VXD is composed of two subdetectors: the pixel detector (PXD) [19] and the silicon vertex detector (SVD) [20]. A schematic view of the VXD is shown in Figure 2.6.

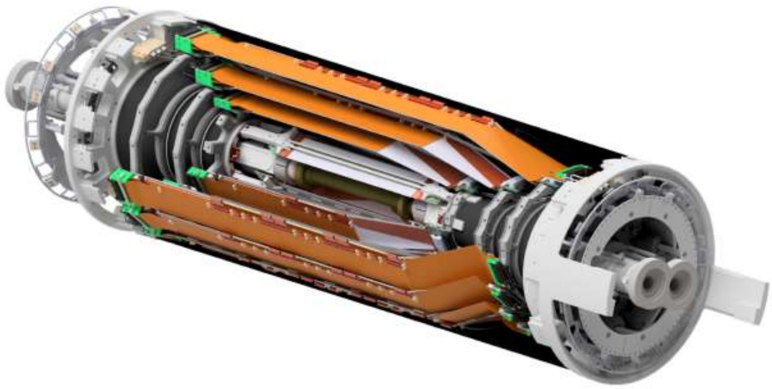


Figure 2.6.: The vertex detector (VXD), composed of the pixel detector (PXD) and silicon vertex detector (SVD). Source: [20].

PXD

The PXD is has two layers of DEPFET pixels (DEPleted Field Effect Transistor) with radii of 14 and 22 mm. Its detection modules are made of 768 × 250 pixels and are glued in pairs to form a ladder. The inner layer is composed of eight ladders, and the outer layer of twelve ladders.

During Run 1, only the first layer of the PXD and two ladders of the second were installed. The installation was completed during the Long Shut Down 1 (LS1) in 2023. When operation resumed for Run 2, the PXD was fully operational but shut off to protect it from radiations caused by suden beam loss, and will be turned on again when these beam loss events are understood.

SVD

The SVD is composed of four layers (see Figure 2.7) of silicon strips, with radii of 39, 80, 104 and 135 mm and a total of 172 double sided strips over the four layers. The sensing strips are implanted in a N-type bulk either with acceptors or donors, depending on the sensor side. The acceptor-implanted side is called the “u/P-side”, while the other side is the “v/N-side”. The strips are arranged perpendicularly depending on their side; the v/N side measures the z direction, while the u/P side measures the r-ϕ direction.

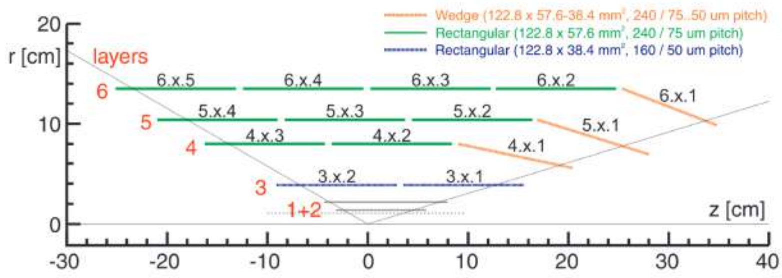


Figure 2.7.: A schematic layout of the longitudinal cross-section of the SVD. The layers are numbered; layers 1 and 2 corresponds to the PXD, while the layers 3 to 6 are the SVD. Source: [20].

To cope with the beam background and limit hit occupancy in the innermost layers, APV25 chips [67] are used for the readout, as such chips have a shaping time of 50 ns and are sufficiently resistant to radiations.

The shaped output is written to a pipeline of 192 cells at the clock frequency, which is 31.805 MHz. Six consecutive samples along the shaped waveform are therefore recorded, allowing to reconstruct both the amplitude and the time of the hit. Further details on the time measurement are given in section 2.5.1 on tracking.

The vertex detector covers all the azimuthal angles ϕ, and polar angles θ going from 17° to 150°, respectively in the forward and backward direction, where the asymetry comes from the Lorentz boost of the collisions.

The position resolution, measured with different techniques [68], was found to be between 7 and 11μm for the azimuthal resolution and 23 and 35 μm for the longitudinal resolution.

2.2.2. CDC

The Central Drift Chamber (CDC) plays three different roles. First, it's the main subdetector for reconstructing charged track and their momenta. Second, it provides particle identification through measurements of energy loss within the gas. Lastly, it provides efficient triggers for the charged particles.

Its volume lies from 11 to 113 cm filled with $He - C_2H_6$ (helium-ethane) gas mixture. There are a total of 14,336 wires with radii ranging from 168 mm for the innermost wires to 1111.4 mm for the outermost ones. The wires are arranged in 56 layers, themselves divided in 9 superlayers, with the wires in each superlayer being arranged in axial or stereo (where the wires are skewed by an angle of 80mrad at most, see Figure 2.8) configurations. The spatial resolution of the CDC for single hits is aroud 100 μm and its acceptance, as for the VXD goes from 17° to 150° in polar angle.

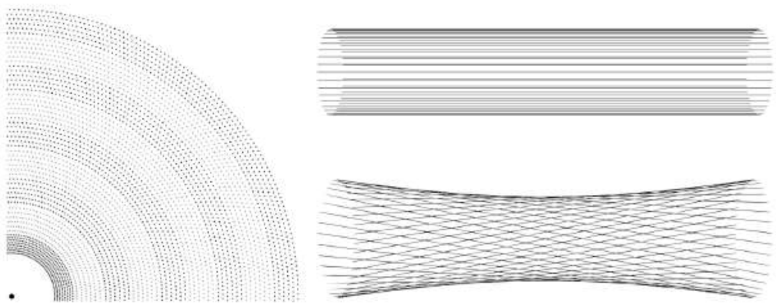


Figure 2.8.: Left: Orthogonal view of a CDC quadrant. Right: Scheme of axial (up) and stereo (down) wires. Source: [69].

2.2.3. PID system: TOP & ARICH

The particle identification system (PID) is composed of two subdetectors: the Aerogel Ring Imaging Cherenkov (ARICH) detector, in the forward end-cap of the detector, and the Time Of Propagation (TOP) detector, in the barrel. They both exploit the Cherenkov radiation produced by charged particles to identify them.

TOP

The TOP consists of quartz radiators arranged in 16 modules, surrounding the outer walls of the CDC. The radiators are made of two quartz bars joined together for a total length of 2.5 m, a width of 44 cm and a thickness of 2cm (see Figure 2.9 for a conceptual overview).

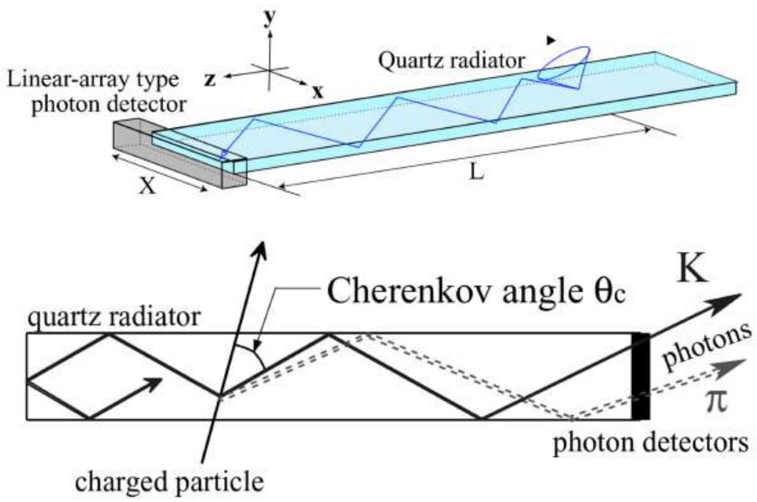


Figure 2.9.: Up: Conceptual view of a TOP radiator. Bottom: Schematic side view of the working principle. Source: [16].

The Cherenkov image is reconstructed from the 3-dimensional information provided by the (x,y) coordinate and precise timing of the photons, which is measured by micro-channel plate photomultipliers (MCP-PMTs) at the backward end. A mirror is placed at the forward end to reflect the Cherenkov light to the MCP-PMTs. A focusing system is used to separate the photons according to their wavelength onto different PMT channels, allowing to improve the time resolution (and consequently the PID efficiency). Due to the small transit time of ∼30 ps of MCP-PMTs, the time resolution of the TOP is of 50 ps. This is enough to separate pions from kaons, as their photons arrival time differs by ∼ 100 ps at 2 GeV/c.

ARICH

The ARICH counter is a radiator made of aerogel together with an array of photon detectors, separated by a 20 cm expansion volume to allow Cherenkov photons to form rings on the detectors as shown on Figure 2.10.

A key parameter of the ARICH is the number of detected photons: it is increased by using two layers of aerogel with different refractive indices $n_1 = 1.045$ and $n_2 = 1.055$, respectively corresponding to the upstream and downstream aerogels, (Figure 2.10) without degrading the Cherenkov angle resolution. The single photon resolution is $\sigma_\theta = 14.3$ mrad, which gives a per track resolution of $\sigma_{track} = \sigma_\theta / \sqrt{N_\gamma} = 4.5$ mrad, assuming an average of N_γ=12.7 detected photons per track.

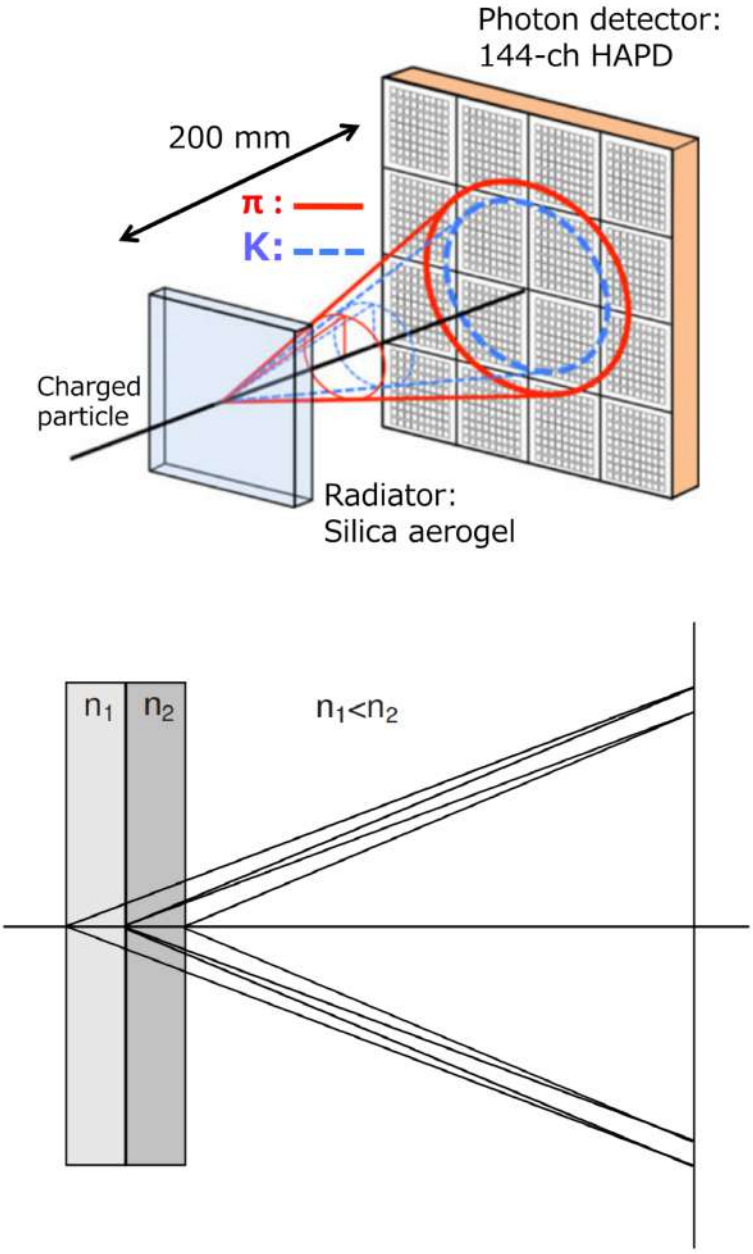


Figure 2.10.: UP: Schematic view of the ARICH counter. Bottom: Principle of a non-homogenous aerogel radiator. Source: [21].

2.2.4. ECL

The Electromagnetic Calorimeter (ECL) is used for the detection of photons, electron identification, various trigger decisions, as well as K_L^0 detection and luminosity measurements. It is a 3 m long barrel, with an inner radius of 1.25 m, that covers polar angles between 12.4° and 155.1°, although with a 1° gap between the barrel and endcaps. The ECL is composed of 8736 CsI(Tl) crystals, with each crystal being a truncated pyramid of average size of 6×6 cm^2 in cross-section and 30 cm in length with a total mass of about 43 tons. The scintillation light readout is performed by two photodiodes, with a preamplifier attached to each of them to provide two independent output signals that are subsequently summed.

The crystals have an average output signal of about 5000 photoelectrons per MeV. The energy resolution in GeV can be approximated by:

$$\sigma_E/E = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34\%)^2}. \quad (2.5)$$

2.2.5. Solenoid magnet

A superconducting solenoid provides a 1.5 T magnetic field in a cylindrical volume with a 3.4 diameter and 4.4 m length. The magnetic field bend the charged particles trajectories, therefore allowing to reconstruct their momenta. The main coil is made of NbTi/Cu cooled with liquid helium cryogenic system and is powered by a 4400 A current. Figure 2.11 shows a contour plot of magnetic field, made out of 100 000 measured points.

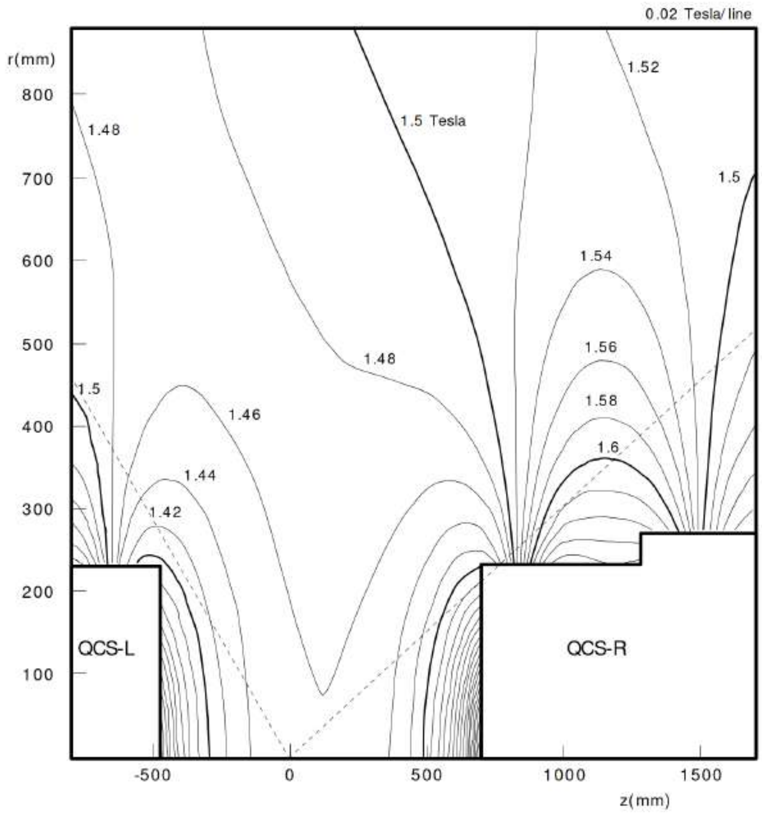


Figure 2.11.: Contour plot of the magnetic field in the Belle II solenoid. Source: [16].

2.2.6. KLM

The outermost part of Belle II is the K_L^0 and muon detector (KLM), used to detect muons and neutral kaons. It is composed of 15 iron plates (14 in the end-cap) alternated with 14 detector layers, located outside the solenoid. The iron plates serve as the magnetic flux return for the solenoid and provide a 3.9 interaction length in which the K_L^0 can shower hadronically. The KLM uses glass electrode resistive plate chambers (RPCs) filled with gas that get ionized by passing charged particles. RPC are paired to form a superlayer with external pick up strips on both sides. RPC suffer from long dead time, rendering them unsuitable for the machine background in the endcap regions. The endcaps detectors are therefore made of 16,800 scintillator strips with a cross-section of (7-10)$\times$10 mm^2 and a length up to 2.8 cm. The muon detection efficiency is 89% for muons with a momentum above 1 GeV/c, and 80% at 3 GeV/c for K_L^0.

2.2.7. VTX

All the subdetectors presented until this point form the Belle II detector in its current form. However, an upgrade [22] will further improve the physics performances of the detector and cope with the background levels. While most of Belle II and SuperKEKB are concerned by the upgrade projects, this section focuses on the proposed upgrade to replace the entire VXD, a CMOS-based vertex detector called the VTX. This new vertex detector would roughly occupy the same volume of the current VXD (between 14 and 135 mm in radius). It would keep a layered structure with five layers, although a 6 layers scenario is considered (and a 7 layers one used to be studied as well).

There are two different designs for the ladders depending on their radii:

- The two inner layers at 14 and 22 mm follow the so-called iVTX design made of silicon and with a material budget of 0.2 X_0 per layer
- The outermost layers at 39, 89 and 140 mm are made of a carbon-fiber support structure. The material budget should be 0.3%, 0.5% and 0.8% X_0, respectively.

The VTX layout is shown on figure Figure 2.12.

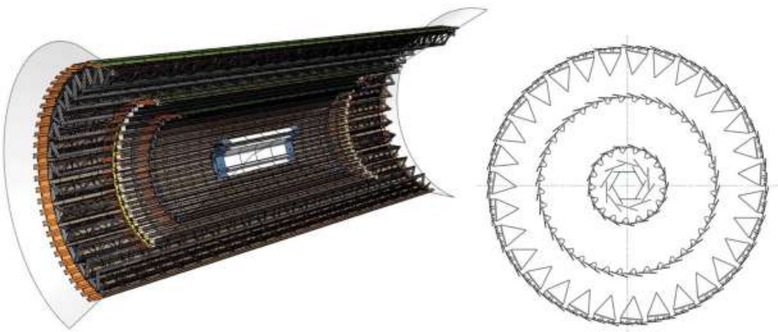


Figure 2.12.: 3D cut view (left) and end view (right) of the VTX layout. Source: [22].

All layers will use the same sensor, a monolithic active pixel CMOS sensor (MAPS) called OBELIX (Optimised BELle II monolithic pIXel). The layout of the sensor is shown in Figure 2.13. It will use a matrix of 896×464 pixels, with a pixel pitch of 33×33 μm^2. Table 2.3 summarizes the main parameters of the VTX.

Layer	1	2	3	4	5	Total
Radius (mm)	14.1	22.1	39.1	89.5	140.0	
# Ladders	6	10	17	40	31	104
# Sensors/ladder	4	4	7	16	48	2311
Mat. budget (% X_0)	0.2%	0.2%	0.3%	0.5%	0.8%	2%

Table 2.3.: VTX detector mains parameters.

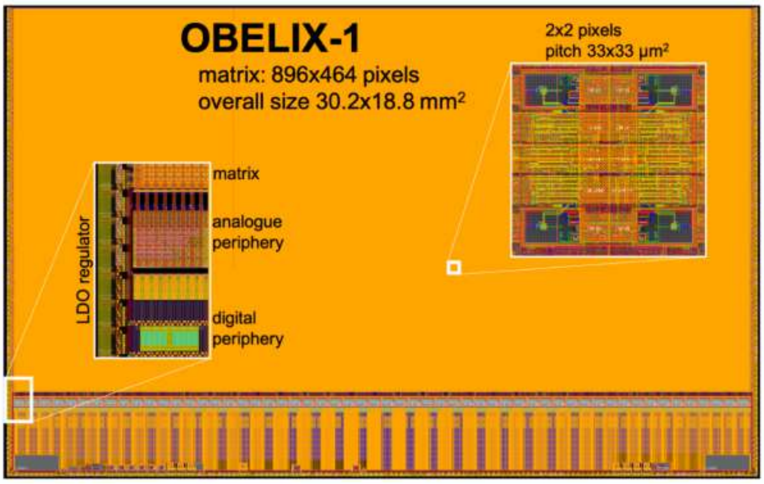


Figure 2.13.: Layout of the OBELIX sensor. Source: [22].

While the OBELIX sensor doesn't exist yet, its requirements are similar to the ones used in the development of the upgrade of the inner track of the ATLAS experiment. In particular the TJ-monopix sensors, depleted monolithic active pixel CMOS sensors (MAPS), match the Belle II requirements; the second version of the sensors, TJ-Monopix2 [70], was therefore chosen as a starting point for the OBELIX sensor and can be considered as the first prototype of the VTX detector.

The first version of the OBELIX sensor is currently being designed with involvement from France through CPPM (Marseille) and IPHC (Strasbourg).

2.3. Trigger

The trigger system allows the proper collection of data for physics events of interest. The Belle II system has to cope with a large event rate of 20 kHz at the nominal luminosity. While triggers designated for $B\bar{B}$ events can easily discriminate low-multiplicity processes (such as Bhabha scattering or other QED processes) using the number of charged tracks, in the case of τ physics such separation is harder to achieve. To this end, the trigger system consists of two levels:

- A hardware-based low-level trigger ("L1") that can reach a maximum output of 30kHz and 5μs latency, meant to suppress Bhabha scattering and other QED processes,

- A software based high-level trigger ("HLT") to bring the output rate down to 20kHz, with a latency of 100ms.

The L1 trigger combines each sub-detector information using FPGA-based global reconstruction logic (GRL) to perform low-level particles and event reconstruction with 3D tracking algorithms. The output of the GRL is sent to the global decision logic (GDL) responsible for issuing the trigger signal. The CDC trigger is able to reject events caused by Touschek scattering for which tracks typically exhibit a displacement along the z-axis with respect to the IP. The ECL trigger uses two complementary trigger schemes: a total energy trigger and an isolated cluster counting trigger. The ECL trigger is also able to reject the back-to-back topology from Bhabha events. The L1 trigger is divided in various trigger lines (or bits) that are used as a first step of event filtering and background rejection for analysis.

The definitions of the trigger lines relevant to the analysis of Chapter 4 (ECL-based, and a few CDC-based to estimate the trigger systematic uncertainty) are given in Table 2.4 and Table 2.5.

The events passing the L1 triggers are processed by the full offline reconstruction (except for the PXD, for which the high amount of pixels prevents a full readout for each event). The output is a physics trigger that classifies the event category, and enables the calculation of a region of interest in the PXD to complete the event reconstruction only with the relevant PXD pixels. The HLT was not used in the early data taking, and turned on in 2021.

2.4. The Belle II analysis software framework

The Belle II analysis software framework (basf2) [71, 72, 73] is the core software of Belle II, developed and maintained by members of the collaboration. It provides (most of) the tools needed to perform a physics analysis, from the raw data to the final results. The framework is written in C++ and Python, and is based on the ROOT framework [74] for the data storage and analysis. Data are processed in a modular way, with each module performing a specific task, event by event (à la ROOT). Reconstruction and tracking are typically performed centrally, providing to analysts only the high level objects needed. In the case of a physics analysis, basf2 is used to select the relevant particles, combine them to reconstruct the decay chains of interest and store the results. There are many different releases (versions)[1] of basf2; with each new release data are reprocessed, and new simulations are produced.

[1] There are seven releases at the time where this manuscript is written, the most recent being release 08 (release 07 was skipped). On top of that are so-called "light-releases", meant for analysis only and named after cats races (ragdoll, quaxo, persian for the latest).

Table 2.4.: Definitions of L1 low-multiplicity trigger lines (lmlX) used in the analysis of Chapter 4. The center of the mass frame is referred as c.m.

Trigger line	ECL clus-ter(s)	Energy of cluster(s) E	Polar angle of cluster(s)	Opening angle in c.m $\Delta\phi^{\mathrm{c.m}}$	Other requirements
*lml*6	1	$E^{\mathrm{c.m}} > 1\mathrm{GeV}$	$[32.2°; 128.7°]$	-	no other cluster with $E > 300\mathrm{MeV}$
*lml*7	1	$E^{\mathrm{c.m}} > 1\mathrm{GeV}$	$[18.5°; 31.9°]$ or $[128.7°; 139.3°]$	-	no other cluster with $E > 300\mathrm{MeV}$
*lml*8	2	$E > 250\mathrm{MeV}$	-	$[170°; 190°]$	no cluster with $E^{\mathrm{c.m}} > 2\mathrm{GeV}$
*lml*9	2	$E > 250\mathrm{MeV}$ $E < 250\mathrm{MeV}$	-	$[170°; 190°]$	no cluster with $E^{\mathrm{c.m}} > 2\mathrm{GeV}$
*lml*10	2	-	$\sum \theta^{\mathrm{c.m}} \in [160°; 200°]$	$[160°; 200°]$	no cluster with $E^{\mathrm{c.m}} > 2\mathrm{GeV}$
*lml*12	≥ 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 2.5.: Definitions of L1 CDC trigger lines used in Chapter 4. The centre of the mass frame is referred as c.m.

Trigger line	Full track(s)	Short track(s)	Neuro 3D track(s) with $\|z\| < 20\mathrm{c.m}$	Opening angle $\Delta\phi$	Other requirements
ffy	≥ 3	-	≥ 1	-	-
fyo	≥ 2	-	≥ 1	$> 90°$	not an ECL Bhabha

2.5. Tracking

The tracking algorithm is responsible for converting the signal hits from the VXD and CDC into tracks. It uses as an input digital objects (digits) called spacepoints that are reconstructed from the raw data (the direct response of the sub-detector).

2.5.1. SVD reconstruction

Let's consider as an example how SVD detector hits are turned into tracking spacepoints. The SVD readout is done by the APV25 chips, which provide the amplitude of the signal and the time of the hit for the SVD strips. For each strip a set of six samples is acquired; adjacent strips are then grouped together in clusters [1]; depending on which side the strips are (2.2.1), the clusters are called either U or V clusters.

Charge reconstruction

The charge of a cluster is then computed using the MaxSample algorithm: for each strip, the sample with the highest amplitude is selected to represent the strip charge, and the cluster charge is the sum of the charge of all the strips of the cluster.

Two alternative algorithms are also implemented in the tracking chain and can be used instead:

- SumSample: the strip charge is the sum of all the samples of the strip, and the cluster charge is the sum of the charge of all the strips.

- ELS3: the strips are summed sample by sample and the best consecutive three summed-samples (such that their sums is the maximum) are selected. The charge is computed as the amplitude of the theoretical CR-RC waveform (with $\tau = 55\text{ns}$):

$$a(t) = A\frac{t - t_{raw}}{\tau} exp\left(1 - \frac{t - t_{raw}}{\tau}\right), \tag{2.6}$$

with

$$A = \frac{\left(e_\tau^{-1} - e_\tau^3\right) a_1 + \left(2 + e_\tau^2\right) a_2 - \left(1 + 2e_\tau^2\right) a_0}{\Delta t/\tau \exp\left(1 + t_{\mathrm{raw}}/\tau\right)\left(1 + 4e_\tau^2 + e_\tau^4\right)}, \tag{2.7}$$

where $e_\tau = e^{\frac{-\Delta t}{\tau}}$, $\Delta t \approx 31.44$ ns is the sampling period of the APV readout, t_{raw} is the raw ELS3 time (see 2.5.1), and a_i is the amplitude of the i-th summed-sample.

[1] For such a cluster to be kept, at least one of its strips must have a signal over noise ratio of 5.

Time reconstruction

The time associated to the cluster (at which the incoming particle deposited energy) is reconstructed by the CoG3 (Centre of Gravity) algorithm: similarly to the ELS3 algorithm, the strips are summed sample by sample and the three best consecutive summed-samples are selected. The raw time is then computed as the sum of the time of the three best samples weighted by their charge:

$$t^{raw}_{cluster} = \Delta t \cdot \frac{\sum_{i=0}^{2} t_i a_i}{\sum_{i=0}^{2} a_i}, \tag{2.8}$$

where t_i is the time of the i-th sample in ns, a_i is the amplitude of the i-th sample, and Δt is the sampling period of the APV readout.

Two additional algorithms are available:

- CoG6: The six summed-samples are used to compute the time instead of the three best ones with the CoG algorithm.

- ELS3: As for CoG3, the three best summed-samples are selected, but the time is computed as the time of the theoretical CR-RC waveform:

  $$t^{raw}_{cluster} = -\Delta t \cdot \frac{2e_\tau^4 + \omega e_\tau^2}{1 - e_\tau^4 - \omega(2 + e_\tau^2)}, \tag{2.9}$$

  where $\omega = \frac{a_0 - 2e_\tau^{-2} a_2}{2a_0 + e\tau^{-1} a_1}$.

The time is computed in the trigger reference frame (t=0 when the trigger is fired) and is then calibrated using a function that depends on the algorithm used.

Spacepoints creations

All U-clusters are combined with all V-clusters to form spacepoints (SP). Combinations are filtered using two criteria:

- Spacepoints with at least one cluster too far in time from the trigger are rejected
  $$|t_{U,V} < t_{max}|, \tag{2.10}$$

- Spacepoints with a too large time difference between the U and V clusters are rejected
  $$|t_U - t_V| < \Delta t_{max}. \tag{2.11}$$

where t_{max} and Δt_{max} are 100 ns and 20 ns respectively. In higher luminosity scenario, t_{max} will be reduced to 50 ns.

With the increase in trigger rate that will accompany the higher luminosity, the acquisition will transition from 6 to 3 samples per strip and the time resolution will therefore be degraded. This time information is used to reject both bad clusters (see above) and fake tracks, as presented in Chapter 3; this chapter also explores possible a way to mitigate this loss of resolution.

2.5.2. Track reconstruction

The track reconstruction at Belle II [75] relies on the information from the tracking devices (CDC and VXD). A general overview of the tracking steps in given in Figure 2.14.

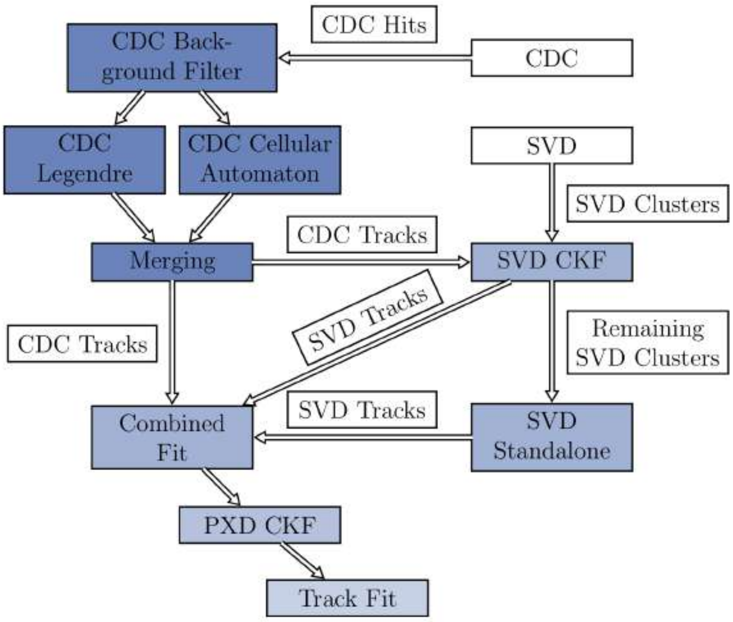


Figure 2.14.: Overview of the tracking steps in Belle II. Source: [69].

Due to the differences between the sub-detectors, two algorithms are used for the track reconstruction. The first step is to reconstruct tracks candidates within the CDC; to this end two algorithms are used:

- A global Legendre algorithm [76] that reconstructs circles by mapping the spatial coordinates into the (θ, ρ) space, where θ is the tangent to the circle at the IP and ρ the curvature. A track is then a set of hits belonging to the same circle.

- A local algorithm designed to detect short tracks and tracks displaced from the IP using a cellular automaton [77, 78, 79, 80]. It acts on acyclic graphs, where the vertices are sets of three hits, and the edges are the relations of the vertices based on a figure of merit (χ^2 from a fit). The tracking can then be seen as a maximisation of the weights of the edges and vertices.

The tracks found are then merged in a single set and are assigned various quality indicators based on multivariate estimators to increase hit purity, improve the track parameter resolution and reduce the fake tracks rate. They are then fitted using a deterministic annealing filter (DAF) [81] to extract the track parameters.

A standalone algorithm for the SVD is then used to reconstruct tracks with low momentum (under 100 MeV/c) that don't deposit enough hits in the CDC. It uses a cellular automaton that operates on a map (called Sector Map) dividing the detector in geometrical regions and holds information on how hits from different regions can (or cannot) be related.

Lastly, a combinatorial Kalman filter (CKF) [82, 83, 84, 85] is used to match CDC tracks to SVD spacepoints as well as SVD tracks from the SVD standalone. The final step is to fit the tracks using a DAF; the results of this fit are the output of the tracking and stored to be used in physics analysis. One should also note that (at least technically speaking), the tracks are *not* particles; only once a particle hypothesis is assigned to a track (done at analysis level, though some tracking steps make the assumption that the track is produced by a pion), it "becomes" a particle.

The tracking algorithm is not perfect; some tracks reconstructed can contain additional hits, can be a clone of another track (for instance if the merging was not properly performed) or not correspond to any actual particle that has interacted with the detector. In the latter situation, they are called fake tracks; the term "fake" is in general even used in a loose way to designate any tracks that are not produced by the e^+e^- collision and therefore also contains beam-induced background tracks. The rate of fake tracks (the ratio of events with a fake track to the total number of events) is measured in data [86] using $e^+e^- \to \tau^+(\to e^+\nu_\tau\bar{\nu}_e)\tau^-(\to \pi^-\pi^+\pi^-\bar{\nu}_\tau)(+c.c.)$ events at 1% (and more recently at 1.6%, in higher background conditions). Chapter 3 presents a set of three studies on fake tracks (mainly the beam-induced ones) with the goal of reducing their rate.

The tracking efficiency has also been measured [86]: it ranges from 75% for low p_T particles up to 95% for momenta around 4 GeV/c as shown on Figure 2.15. It degrades at low and high polar angles (closer to the beam axis) while remaining constant across the azimuthal angles.

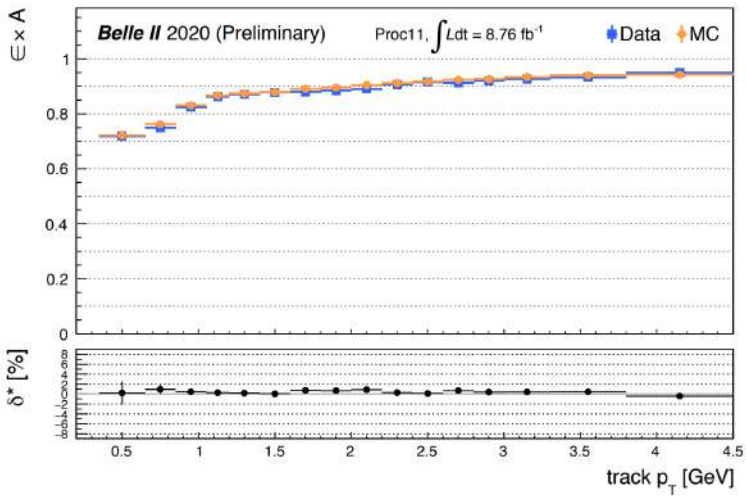


Figure 2.15.: Tracking efficiency as a function of the transverse momentum of the particle. Source: [86].

2.6. Particle identification

The charged Particle identification (PID) variables are computed using the information from all the sub-detectors[1]. Each sub-detector S provides a likelihood $\mathcal{L}_i^S$ for a given particle to be of type i (i $\in$ e, μ, π, K, p and d, where p and d are proton and deuteron); the total likelihood for a particle hypothesis $\mathcal{L}_i$ is then the sum of the sub-detectors likelihood. The PID for a given particle hypothesis is:

$$\text{PID(i)} = \frac{\mathcal{L}_i}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}. \tag{2.12}$$

Another method [87] has been developed for the identification of leptons, combining information from the various sub-detector such as ECL clusters spatial structure and showers shape. A classifier is then trained to separate leptons from other particles either in multi-class modes or against pions only. The classifier is trained using a gradient boosting algorithm, and the output is interpreted as a probability and is used as the PID variable for leptons.

The efficiency and mis-identification rate have been measured using $e^+e^- \to \ell^+\ell^-(\gamma)$, $e^+e^- \to e^+e^-\ell^+\ell^-$, $\mathrm{J}/\psi \to \ell^+\ell^-$, $K_S^0 \to \pi^+\pi^-$ and $e^+e^- \to \tau^+\tau^-$. For example, the identification efficiency of electrons and muons is 86% and 88.5% respectively with misidentification rates of 0.4% and 7.3% for a PID threshold of 0.9 (for their respective PID variables).

[1]Due to calibration issues, the muon ID variable used in the analysis (Chapter 4) is computed excluding the SVD.

2.7. Data and simulations

Simulations

Monte-Carlo (MC) simulations are used to generate the physical process that are produced at the IP and the interaction of the produced particles with the detector until the readout. They are used to optimise the selection criteria, to estimate the signal efficiency, etc. Various physics generators are used for the different processes: the $e^+e^- \to \tau^+\tau^-$ process is generated using the KKMC generator [88]. The τ decays are simulated by the TAUOLA generator [89] and their final state radiated photons by PHOTOS [90]. KKMC is also used to simulate $\mu^+\mu^-(\gamma)$ and $q\bar{q}$ production; PYTHIA [91] for the fragmentation of the $q\bar{q}$ pair; PYTHIA interfaced with EvtGen [92] for the production and decay of $e^+e^- \to b\bar{b}$ events; BabaYaga@NLO [93, 94, 95, 96, 97] for $e^+e^- \to e^+e^-(\gamma)$ events; and AAFH [98, 99, 100] and TREPS [101] for the production of non-radiative four-leptons and $e^+e^-h^+h^-$ final states. Simulations of the detector response are performed using the GEANT4 toolkit [102], while the (L1) trigger is simulated with the software TSIM [103] and the magnetic field by Opera3D/TOSCA [104].

The fact that the four-leptons generators don't include ISR and FSR (initial and final state radiation) is an important issue for the analysis, since radiative four-leptons events are a significant background. The usage of WHIZARD [105, 106] for the generation of the four-leptons events is being considered to address this issue, although the analysis has been performed without it.

There are different MC campaigns, labelled as MCX_ri or MCX_rd where X corresponds to the campaign number, while "ri" and "rd" stand for run-independent and run-dependent. Run-independent use simulated beam background and static detector conditions, while run-dependent use background obtained with random triggers and detector condition matching the data taking.

2.7.1. Experimental data

The data sample consists of a total of 424 fb^{-1} collected from March 2019 to end of 2022 with 362 fb^{-1} of on-resonance data (at the $\Upsilon(4S)$ energy), 42 fb^{-1} of off-resonance and 19 fb^{-1} of $\Upsilon(5S)$ scan. At the beginning of summer 2024, the data taking resumed and around 100 fb^{-1} have already been collected; they are however not part of the work described here.

2.8. Skimming

In addition to the trigger system, a skimming is performed on data (and simulations) to reduce the size of data to be processed. These skims are a set of selections with a particular (or a set of) analysis in mind; they act as a flag to easily select some events with a specific signature. Data and simulations are processed in skimmed samples available to reduce the overall CPU usage. They must have an extremely high efficiency on the signal targeted by the analysis, while removing events that are not of interest; the size of the skimmed samples is typically reduced to 10% of the original size. For example, the Tau-LFV skim used in the analysis of Chapter 4 select events with at most 7 tracks and some loose requirements on the mass and energy of the candidate τ lepton.

3. Fake track studies

Content

3.1. Introduction

As mentioned in Chapter 2, the tracking system can reconstruct tracks not coming from the interaction, called fake tracks. The fake track rate is an important figure of merit of tracking algorithm performances that needs to be properly measured and minimized. The word "fake" is used loosely here, and refers to different types of tracks:

- **Ghost tracks**: tracks not associated to any real particle, and reconstructed from random hits in the detector.
- **Combinatorial tracks**: tracks reconstructed from real hits coming from different particles.
- **Clone tracks**: multiple tracks reconstructed from the same particle.
- **Beam induced background**: tracks that are reconstructed from real particles, where said particles are not coming from an e^+e^- collision but instead from the beam.

Beam induced tracks are the main component of fake tracks due to their sheer amount, and are therefore the main target of the studies presented in this chapter. Indeed, the rate of these fake tracks grows with the luminosity and is expected to increase dramatically as the accelerator will reach its nominal luminosity.

The method used to measure the fake track rate is a tag and probe method: physical events are completely reconstructed (as tag) and any additional particle is thus considered as a fake track (probe). Bhabha scattering events ($e^+e^- \to e^+e^-$) are used: events with two tracks produced back-to-back, with one being identified as an electron and the other as a positron, are selected. Any third track found is considered to be a fake track. The invariant mass of the dilepton system is required to be greater than 8 GeV, to reduce contamination that could arise from radiative $e^+e^- \to e^+e^-\gamma$ events where the photon could undergo pair conversion, from which electrons could be mistaken for fake tracks. All the tracks are required to come from the interaction point, with a transverse distance of closest approach to the interaction point less than 0.5 cm and a longitudinal distance of closest approach less than 2 cm. This requirement already rejects fake tracks from beam background, but such tracks are not interesting as they would not be reconstructed in any analysis. Two different reconstruction schemes can be considered:

- In the first case, the electron ID variable is used, and the two electrons are identified by requiring a probability to be an electron greater than 0.9, while the fake track is identified by requiring a probability lower than 0.9. These requirements are made to ensure that there is no overlap between the electron candidates and the fake tracks.
- In the second case, the electron ID variable is not used at all, as it was realised that it could lead to a bias in the fake rate. The electron ID is not properly defined if no ECL cluster is associated to the track, which can be the case for low-momentum track (that would not leave the vertex detector for instance).

The case without any electron ID used can be easily applied here because I am working with simulated Bhabha events only; the electrons are high p_{T} tracks and need no particle identification variable to be properly identified. A study on data (or on simulation but in data conditions) would require additional selection to ensure the purity of the reconstructed events.

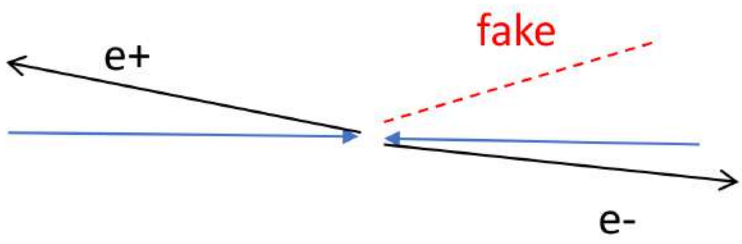


Figure 3.1.: Illustration of the tag and probe method in a Bhabha event. The two outgoing electrons are used to tag the event. Any additional track is considered as a fake track. Contrary to the schema, fake tracks are not necessarily coming from the IP.

Two sets are therefore obtained: N_2 and N_3, respectively the sets with exactly 2 and 3 tracks reconstructed. The case with four tracks reconstructed is not considered as it would be small compared to N_3. The fake track rate is then defined as:

$$r_f = \frac{N_3}{N_2 + N_3}, \tag{3.1}$$

where $N_{2,3}$ refer to both the sets and the number of events within each set. The fake track rate thus corresponds to the rate of events with a fake track (and not the relative amount of fake among the tracks, which would be $\frac{3 \cdot N_3}{2 \cdot N_2 + 3 \cdot N_3}$).

This method has two main advantages: it does not use Monte-Carlo matching (i.e. matching reconstructed track with the generated Monte-Carlo particles) which can be used only in simulations and is easy to implement, in particular for development purposes (Bhabha events are easy to simulate with high statistics). However, one drawback is the important bias coming from the trigger that vetoes Bhabha events; events with a fake track are more likely to pass the trigger.

Section 3.2 presents the first study to quantify the fake track rate in the context of VTX (replacing the SVD) and nominal luminosity. Section 3.3 presents the implementation of the track time variable, and its use for reduction of fake tracks, while section 3.4 explores the possibility of directly rejecting fake clusters in the SVD.

I begin with the study on the potential *future* vertex detector before addressing the *current* vertex detector. Although this order might seem counterintuitive, it should provide a clearer introduction to the properties of fake tracks for the reader and also follows the chronological order in which I conducted the studies.

3.2. Fake tracks for the VTX

The VTX is the proposed upgrade for the VXD, the vertex detector of Belle II. It is expected to operate when the nominal luminosity of the accelerator will be reached and will therefore face an important beam-background. To make matters worse, the VTX efficiency will be improved for low momentum tracks and consequently will be more sensitive to fake tracks. At the time of this study, two geometries were considered for the VTX with either five or seven layers (respectively labelled vtx5 and vtx7 in the following). This study aims measuring the fake track rate in the VTX, to study the properties of these fake tracks and to provide an assessment of the tracking performance in the VTX.

I start by presenting the measurement of the fake rate for the two geometries, and in two scenarios: without any background, and with background corresponding to the expected nominal luminosity. The fake rates in the different configurations are shown in table 3.1.

From now on, I will focus on the five layers geometry and nominal background case, since the five layers geometry is the preferred one (as stated in Chapter 2) and what matters is the fake rate in a realistic case, i.e. with background. The reconstruction using electron ID underestimates the fake rate, hence it will not be

	VTX5		VTX7	
	No background	Nominal background	No background	Nominal background
r_f, no electron ID	6.90 ± 0.09 %	38.70 ±0.2 %	4.21±0.07 %	44.29 ±0.20 %
r_f, electron ID	2.44±0.06 %	9.53± 0.21%	0.90±0.04%	5.90±0.17%

Table 3.1.: Fake track rate for the two VTX geometries, with and without background, and with and without electron ID. The uncertainty is statistical only.

considered further here.

The p_T and number of VTX hits distributions for the fake and the two electrons is shown in Figure 3.2: the fake tracks are low p_T tracks, which is expected for beam induced tracks. The majority of them have only 3 hits in the VTX, which is the minimum number of hits required to reconstruct a track in the VTX (the very few tracks with two (zero) VTX hits come from CDC tracks (not) extrapolated in the VTX). Figure 3.2 also shows the number of CDC hits, with a majority of tracks without any CDC hits, even for the two electrons. This behaviour is understood for the fake tracks as they are low p_{T} tracks and therefore don't reach the CDC, but not expected for the two electrons. This in fact is a rather well-known issue and regularly "rediscovered" in any high background tracking studies: the CDC suffers from the high background and its reconstruction efficiency drops significantly. Lastly, 3.2 shows the angle between the fake and the closest electron to showcase a property of clone tracks; they are the main contribution to the peak at 0 degree.

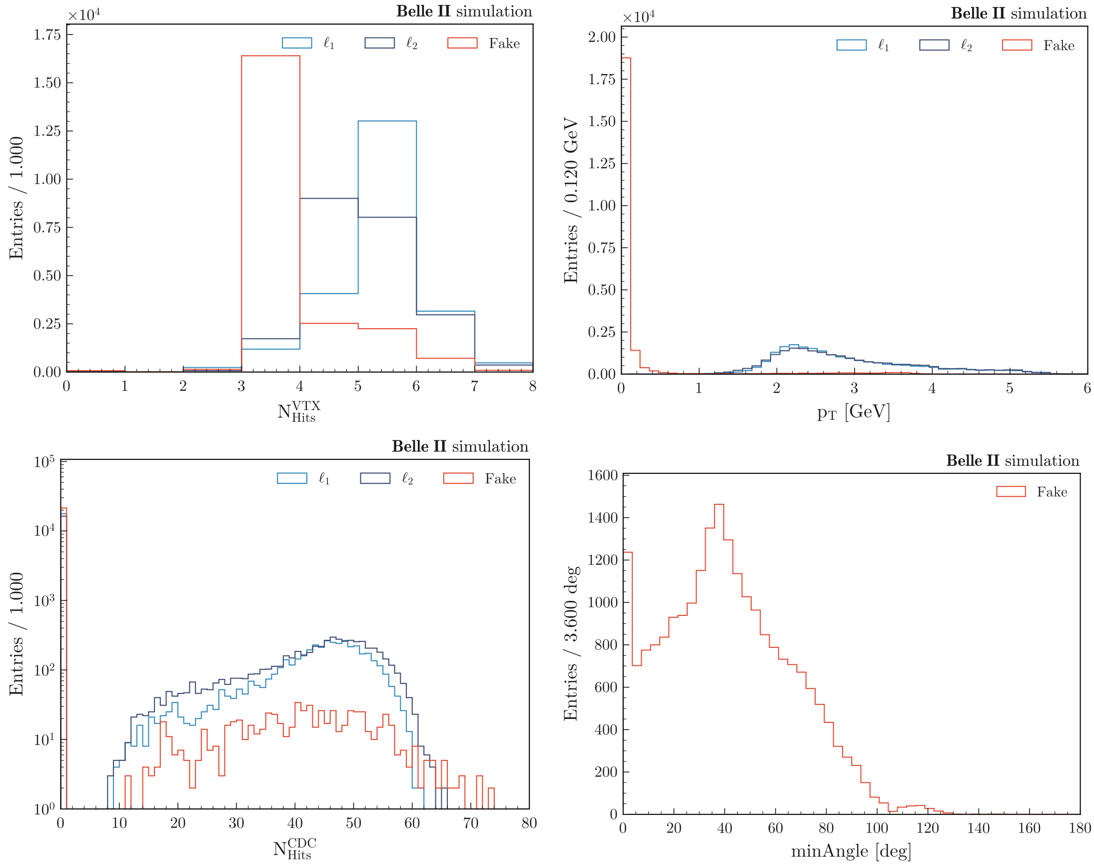


Figure 3.2.: Distributions of the number of VTX hits (top right), p_T (top left), number of CDC hits (bottom left) for the fake tracks and the two electrons ($\ell_{1,2}$) and angle between the fake and the closest lepton (bottom right) for the five layers geometry and nominal background case.

To address the large proportion of fake tracks with only 3 VTX hits, a background suppression tool in the form of a classifier using the VTX hits shape information only (charge, size, position), was developed and trained to remove background tracks produced by the VTX track finding algorithm.

The fake rate after the application of the classifier is shown in table 3.2, with a comparison to the original fake rate. It is reduced by 40%, mainly caused by the reduction in the number of the 3-hits fake tracks. Figure 3.3 shows the number of VTX hits for the fake tracks after the application of the classifier.

	no BST	with BST
r_f	38.70 ± 0.2 %	22.75 ± 0.16 %

Table 3.2.: Fake rate before and after the application of the background suppression tool (BST) classifier.

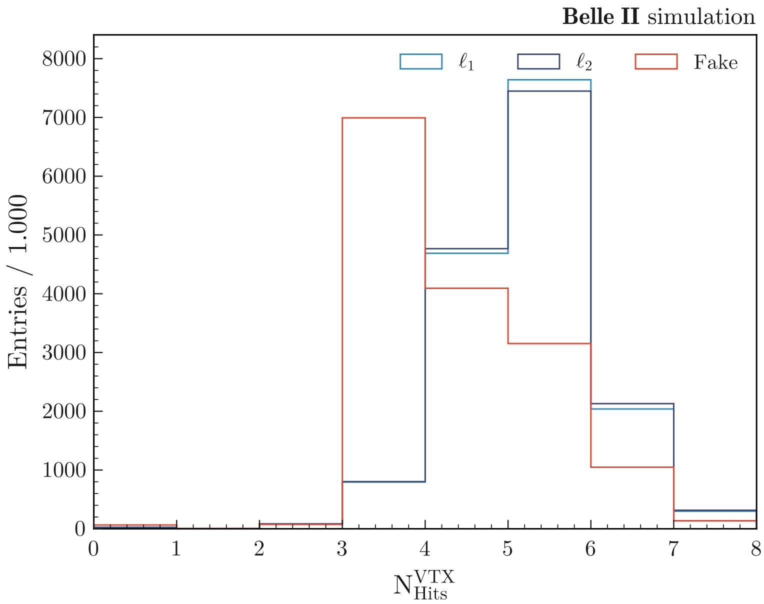


Figure 3.3.: Number of VTX hits after the application of the background suppression tool for the two leptons and the fake.

The issue with the CDC reconstruction efficiency goes beyond the scope of this work and is not really expected to reduce the fake rate. However, a modification of the tracking can be considered. For now, the tracking goes "inward": it starts by reconstructing tracks from the CDC, and then extrapolates them to the VTX (or the SVD, the same tracking is performed for SVD-based Belle II) to match the CDC tracks to VTX hits and a standalone VTX tracking algorithm is then used to reconstruct VTX tracks with the remaining VTX hits. There however exists in the tracking chain an algorithm that can be used to go "outward" by extrapolating tracks from the VTX to the CDC, to match VTX tracks to CDC hits. The fake rate obtained after activating this "to-CDC" algorithm is 19.76 ±0.15 % ; the reason for the small decrease is not clear and would require further investigation. Figure 3.4 shows the number of CDC hits after the "to-CDC" algorithm: as one would expect, most of the leptons are now matched to CDC hits while the fakes are not.

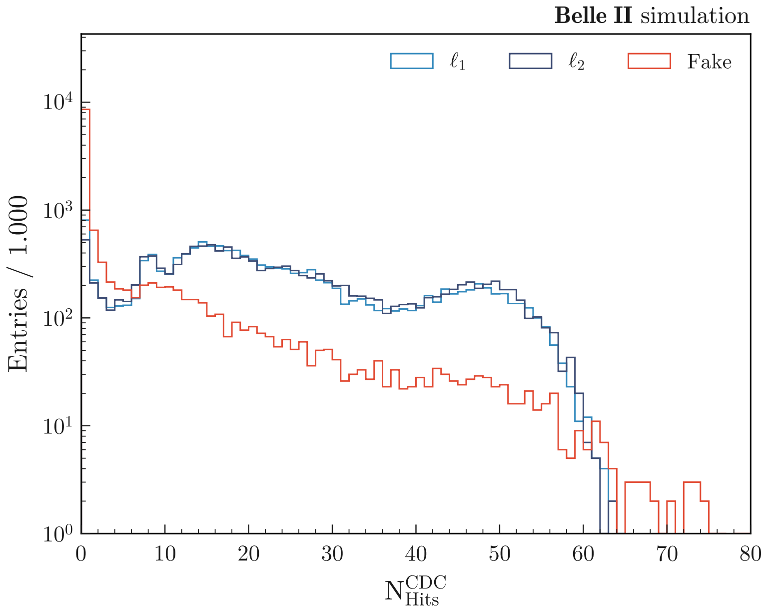


Figure 3.4.: Number of CDC hits after the application of the "to-CDC" algorithm for the two leptons and the fake.

This study of the fake tracks in the VTX is a first step towards the understanding of the tracking performance in the VTX. The fake rate is high, but some tools have already been developed to greatly reduce it. However, more general issues within the tracking (in particular with the CDC) needs to be addressed; the entire tracking chain might need to change from its current version to something different. Various efforts are ongoing to improve the tracking for the upgrade, such as the mentioned "outward" tracking.

3.3. Fake tracks in the SVD and track time variable

The objective of this study is to implement a track time variable, using the SVD time information attached to every SVD hit, and to study its impact for the rejection of fake tracks. This time information is currently unused in tracking but could be used to easily reduce the rate of fake tracks, as beam induced tracks are expected to have a different time distribution than tracks coming from the interaction point. For this study, I again use Bhabha simulations in two different background scenarios: one that corresponds to the Run 1 background level, and one that corresponds to the expected nominal background.

The SVD track time variable is therefore defined and computed as the average SVD time of all the SVD clusters attached to the track:

$$t_{\text{track}}^{\text{raw}} = \frac{\sum_{\text{cls}} t_{\text{cls}}}{N_{\text{cls}}}, \tag{3.2}$$

where t_{cls} is the time of the cluster and the sum runs over all the N_{cls} clusters attached to the track.

The time initially comes in the time frame of the trigger, which is lacking precision. It is better to move in frame where the eventT0 = 0 corresponds to the collision time, since all the tracks should be extremely close in time to the collision time. The SVDEventT0 is the collision time measured by the SVD, and is computed as the average of the time of all the SVD clusters matched to a track of the event:

$$\text{SVDEventT}_0 = \frac{\sum_{\text{CLS}} t_{\text{cls}}^{\text{SVD}}}{N_{\text{CLS}}}, \tag{3.3}$$

where the capitalised subscript denotes that the sum runs over all the clusters matched to a track of the event. It was shown in simulation that this method gives a very good resolution on the collision time of ≈ 0.7 ns on $B^0\bar{B}^0$ simulated events. The residual defined as the difference of the eventT0 and the true (generated) EventT0 is shown on Figure 3.5.

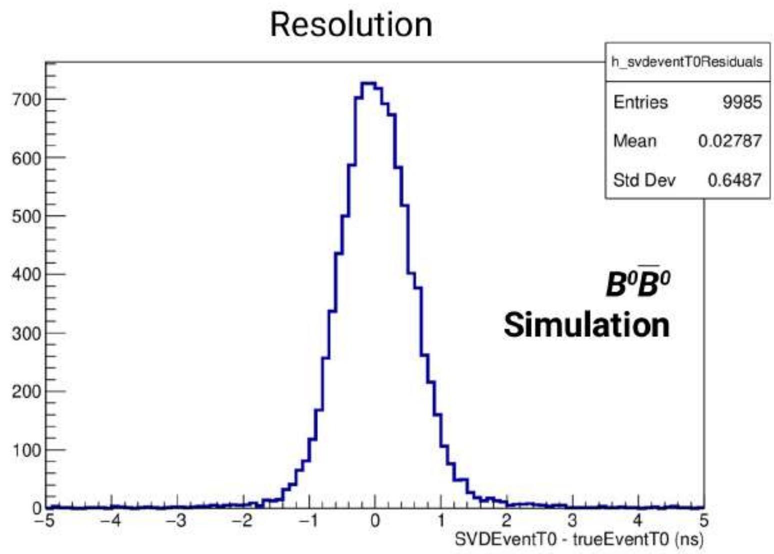


Figure 3.5.: Resolution of the SVD EventT0 in $B^0\bar{B}^0$ simulated events. Source: [107].

The SVD track time is then computed as:

$$t_{\text{track}}^{\text{SVD}} = t_{\text{track}}^{\text{raw}} - \text{SVDEventT}_0. \tag{3.4}$$

The distribution of $t_{\text{track}}^{\text{SVD}}$ for the two background scenarios is shown on Figure 3.6. The $t_{\text{track}}^{\text{SVD}}$ of the leptons shows a clear peak at 0 ns as expected since they are seen by the SVD right after the collision time. The fake tracks also exhibit a peak at 0 ns, accompanied by a continuous contribution. This continuum is attributed to the

	Run1 background	Nominal background
r_f	3.42 ± 0.07 %	10.84 ± 0.12 %

Table 3.3.: Fake rate before the application of the track time variable. The uncertainty is statistical only.

beam-background tracks, which are not necessarily synchronised with the collision time. The fake distribution in the nominal background case almost vanishes at $|t_{\mathrm{track}}^{\mathrm{SVD}}|$ above 50 ns ; reconstruction in higher background conditions rejects clusters in a smaller time frame [1]. Despite this, the fake rate is still higher in the nominal background case.

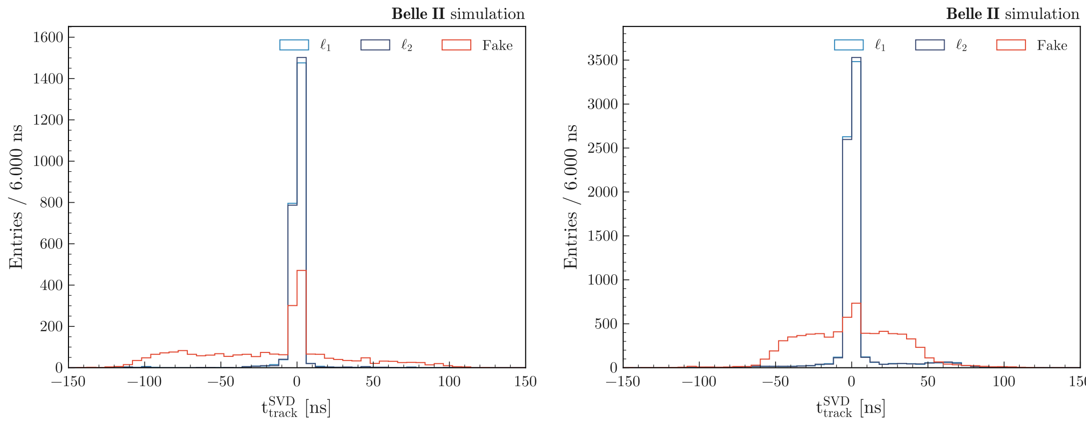


Figure 3.6.: Distribution of the SVD track time in the Run 1 background condition (left) and expected nominal background (right).

The initial fake rate in both background scenarios is shown in Table 3.3. One needs to be cautious when comparing them as they don't correspond to the exact same tracking conditions.

In order to find a suitable selection criterion on the $t_{\mathrm{track}}^{\mathrm{SVD}}$, the fake rate as a function of the cut on the $t_{\mathrm{track}}^{\mathrm{SVD}}$ (in absolute value) is looked at, together with the reconstruction efficiency. This efficiency is defined as the ratio of events with only two leptons at any given cut on the $t_{\mathrm{track}}^{\mathrm{SVD}}$ over the initial number of events with two leptons. The results are shown on Figure 3.7.

[1]See section 2.5.1 of Chapter 2

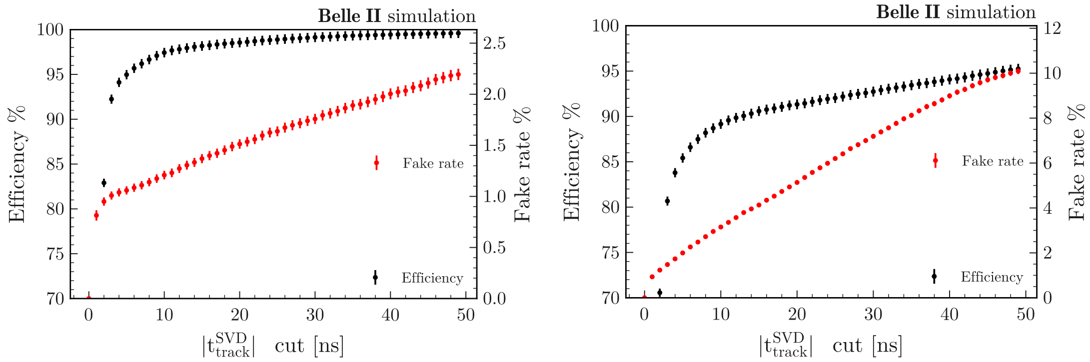


Figure 3.7.: Fake rate and reconstruction efficiency as a function of the cut (in absolute value) on the SVD track time for the Run 1 background condition (left) and expected nominal background (right).

	$\|t^{SVD}_{track}\| <$	40 ns	20 ns	10 ns
Run1 background	r_f Efficiency	2.00 ± 0.05 % 99.4±0.5 %	1.51±0.05 % 98.6±0.5	1.21±0.04 % 97.4±0.5 %
Nominal background	r_f Efficiency	8.99 ± 0.12 % 94.1 ± 0.6 %	5.14±0.09 91.3±0.5 %	3.15±0.07 % 89.2±0.5 %

Table 3.4.: Fake rate and reconstruction efficiency for different cuts on the SVD track time.

A selection criteria of $|t^{SVD}_{track}| < 20$ ns is a good compromise between fake rate reduction and reconstruction efficiency. The corresponding fake rate and reconstruction efficiency are shown in Table 3.4, together with the values corresponding to cuts at 40 and 10 ns. The fake rate is reduced by a factor of two while maintaining a reconstruction efficiency above 98 % for the Run 1 background conditions. However, in the case of the nominal background, the efficiency drop is too important to be acceptable. This drop in efficiency comes from the fact that background hits are attached to the lepton tracks, therefore reducing the t^{SVD}_{track} resolution.
There are ongoing studies to properly use the time information in the tracking in order to build tracks only from hits that are close in time, that could help to reduce the fake rate.

This work implemented the t^{SVD}_{track} variable in the Belle II software and showed that it can be used to easily reduce the fake tracks rate. The long term objective is to provide this t^{SVD}_{track} variable for analysts with a recommendation on the cut to apply. This work was performed on simulation and requires further study on data; such work is now being done in the tracking group using τ pair events.

3.4. Fake clusters rejection

As explained in Chapter 2, the SVD cluster time is computed using the CoG3 algorithm, using a six samples acquisition mode. With the increase of luminosity and consequent increase of trigger rate, the acquisition will need to be using a 3 samples acquisition. This effect can be seen on Figure 3.8, where the background clusters are much closer to the signal (from collision tracks) clusters when using the 3 samples acquisition. This will result in a loss of time resolution both for the clusters and the track time, and the tracking performances will be affected.

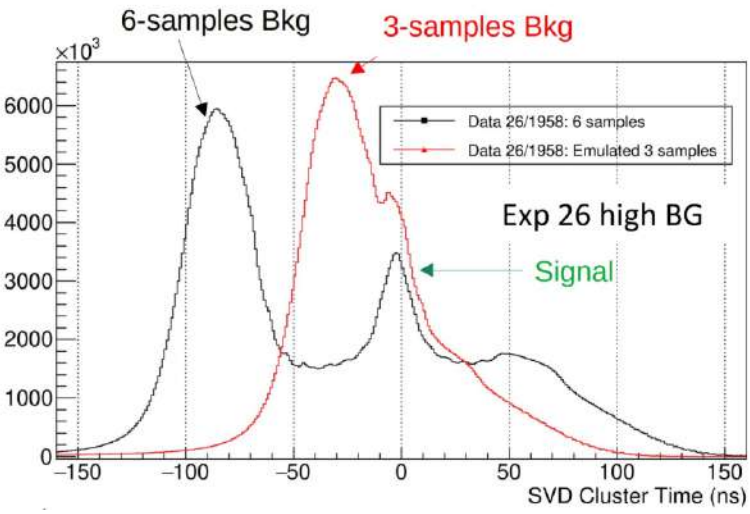


Figure 3.8.: Comparison of the time distribution of the clusters in the SVD for the 3 samples (red) and 6 samples acquisition (black) mode, in data. The central peak corresponds to the signal clusters, while the background clusters are the remaining contributions. Source: [107].

This section is exploring the possibility of rejecting fake clusters in the SVD in the 3 samples acquisition mode, before tracking. I first compare the impact of using the ELS3 algorithm (also presented in Chapter 2) instead of the CoG3, and secondly explore the possibility of rejecting fake clusters using either a set of selection or a classifier.

Therefore, the discussion is not on tracks anymore, but *clusters* and *spacepoints*. Until now the term "clusters" was used in a loose way; it now refers to either U or V clusters from the SVD while spacepoints refers to pairs of U and V clusters and are the input to the tracking.

For this study, generic simulated $B\bar{B}$ events are used, with a 3 samples acquisition mode and nominal background.

As previously mentioned, tracking uses only spacepoints made of clusters with $|t_{\text{cls}}^{\text{raw}}| < 50$ ns. Our first step is to compare the number of such clusters provided by the ELS3 and CoG3 algorithms. The two algorithms retain all the signal clusters

(clusters matched to a simulated particle), but the ELS3 algorithm keeps only 55.2% of the background clusters, while the CoG3 algorithm keeps 85.6% of them. The time distributions of the signal and background clusters for the two algorithms are shown on Figure 3.9. The two algorithms are correctly reconstructing the signal cluster times, but the CoG3 estimates the time of many background clusters around −30 ns, overlapping too much with the signal. Since tracking is roughly a combinatorial problem, the ELS3 is preferred and the CoG3 is discarded for the rest of the study.

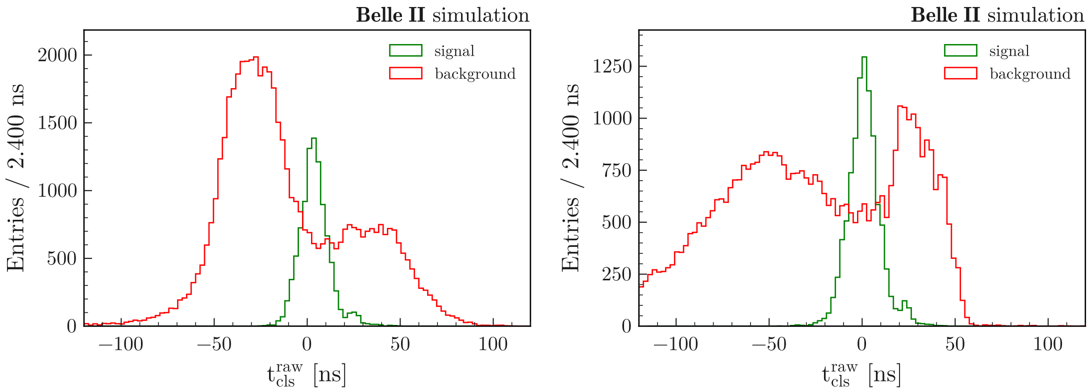


Figure 3.9.: Time distribution for signal clusters (clusters matched to a simulated particle) in green and background clusters in black, for the CoG3 (left) and ELS3 (right) algorithms.

To further reduce the number of fake spacepoints, a set of selection criteria on their clusters is considered. First, the requirement on the $t_{\text{cls}}^{\text{raw}}$ can be tightened without any loss of efficiency, and is set to $-20 < t_{\text{cls}}^{\text{raw}} < 40$ ns. Additionally, the mean charge of the second sample (the mean over all the strips of the cluster) of each cluster is required to be greater than 10 electrons:

$$\mathrm{Q}_2^{\text{mean}} = \sum_{s \in \text{strips}} \frac{Q_2^{\text{strip}}}{N_{\text{strips}}}, \tag{3.5}$$

where Q_2^{strip} is the second out of the tree acquired samples. The distributions of $\mathrm{Q}_2^{\text{mean}}$ for signal and background U and V clusters are shown on Figure 3.10.

These cuts are chosen to be loose enough to have virtually no impact on the signal spacepoints, but to reject a significant amount of background spacepoints. Table 3.5 summarises the signal and background retention of these two cuts on the spacepoints. The signal over background ratio is also shown as it is a good indicator of the composition of the spacepoints inputted to the tracking.

	Signal	Background	$\frac{\text{Signal}}{\text{Background}}$
-25 ns $> t_{\text{cls}}^{\text{raw}} < 40$ ns	99.1%	70.6%	15.5%
$Q_2^{\text{mean}} > 10$ e$^-$	100.0%	80.3%	12.9%
Total	99.1%	57.5%	17.8%

Table 3.5.: Signal, background retention and signal over background ratio for the $t_{\text{cls}}^{\text{raw}}$ and Q_2^{mean} cuts.

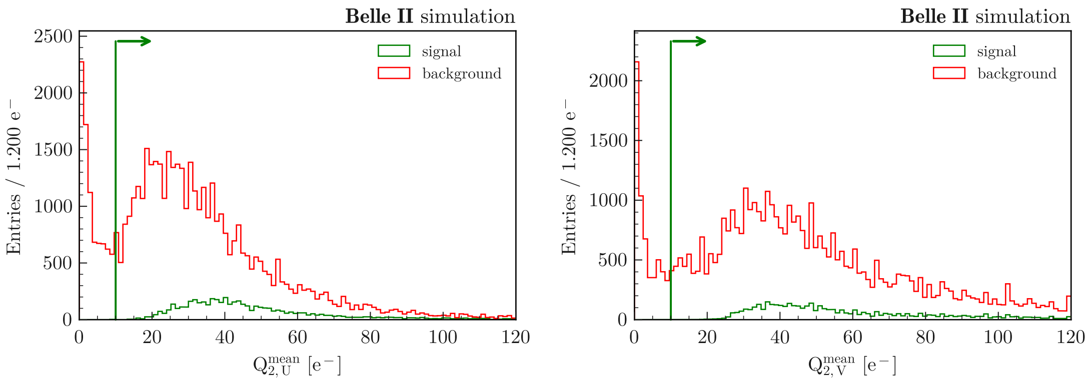


Figure 3.10.: Distribution of the mean charge of the second sample for signal and background U (left) and V (right) clusters. The vertical green lines represent the cuts. The charge is expressed in electrons.

The last step is to train a classifier to reject additional fake spacepoints. A BDT from the XGBoost package is trained, using 11 spacepoints features: the time, the number of strips, the mean charge and the signal over noise ratio of each of their clusters as well as the position of the spacepoint. The score distribution for the training and testing samples is shown on Figure 3.11. There is a clear separation between the signal and background; it is possible to keep a very good signal efficiency while rejecting a significant amount of background. For instance, a cut at 0.02 retains 99.3% of the signal and rejects 55.0% of the background, for a signal over background ratio of 30.4%.

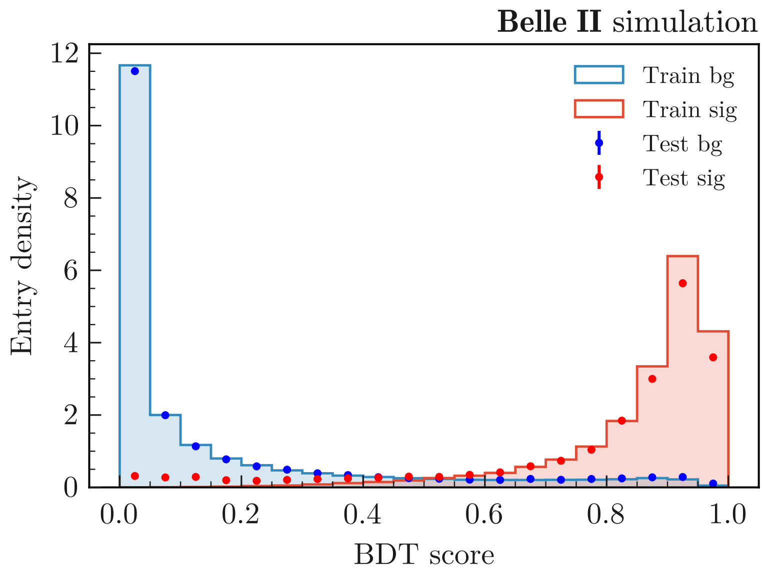


Figure 3.11.: Score distribution for the training (bars) and testing (dots) sample. The background spacepoints are in blue and the signal in red.

This study is a proof of concept that it is possible to reject beam-induced background at the cluster level in the degraded condition of 3 samples acquisition. The next step would be to implement the selection and the classifier in the tracking chain to study the impact on the tracking performances and control the data-MC agreement of the BDT output. The classifier also shows slight overtraining that could be addressed by a careful optimisation.

3.5. Conclusion

This chapter presented various studies on the fake tracks in different detector and background conditions. They will need to be updated and improved: as stated, the impact of the track time variable is now being assessed on data using τ pairs events. Further studies of the fake rate as a function of the track multiplicity should be performed. Even if the fake rate is believed to be driven by an event-based contamination from beam background, the rate might differ in $B\bar{B}$ events, in which typically more than 10 tracks are found. The fake rate in the current situation is low, but the increase of the luminosity will worsen the situation (and there are hints that the nominal background might be underestimated). A high fake rate is a problem for the tracking performances, but a situation with a rate $\sim 20\%$ is not necessarily critical. One should keep in sight the physics performances: the impact of a high fake rate needs to be assessed on the results of physics analysis. Fake tracks might impact tagged analysis in which the (missing) recoil mass is reconstructed, and certain decays involving low-momentum pions and K^0_S (such decays are one of the motivation for the VTX upgrade) might probably suffer from the presence of fakes, but it may be overcome with advanced analysis techniques.

4. Search for $\tau \to e\ell\ell'$

Content

4.1. Analysis overview

This Chapter describes the search for the $\tau^- \to e^{\mp}\ell^{\pm}\ell^{-\prime}$ [1] ($\ell, \ell' =$ e,μ) decays for a total of five final states: $e^-e^+e^-$, $e^-e^+\mu^-$, $e^-\mu^+e^-$, $\mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$.

The observable to be measured is the branching fraction of the $\tau^- \to e^{\mp}\ell^{\pm}\ell'^-$ decays. It is given by:

$$\mathcal{B}(\tau^- \to e^{\mp}\ell^{\pm}\ell'^-) = \frac{N_{sig}}{2 \times \mathcal{L} \times \sigma_{\tau^+\tau^-} \times \varepsilon_{\tau^- \to e^{\mp}\ell^{\pm}\ell'^-}} \tag{4.1}$$

where N_{sig} is the number of observed signal events, $L = \int \mathcal{L}dt = 424\text{fb}^{-1}$ the integrated luminosity of the Run 1 dataset, $\sigma_{\tau^+\tau^-}$ the $\tau^+\tau^-$ production cross-section measured as 0.919 ± 0.003 nb and $\varepsilon_{\tau^- \to e^{\mp}\ell^{\pm}\ell'^-}$ the $\tau^- \to e^{\mp}\ell^{\pm}\ell'^-$ signal selection efficiency. In the case of a branching fraction compatible with 0, the computation of the upper limit is based on the CL_s method [108, 109] described in Section 4.5.

When performing searches, the sensitivity of the analysis is often limited by the number of background events: N_{sig} from the previous equation can be estimated as $N_{tot} - N_{bg}$ where N_{tot} and N_{bg} are the total observed number of events and the number of estimated background events. Therefore, the main goal is to try to reach an almost background-free environment, while having a signal retention as high as possible.

The previous τ-LFV studies performed at B-factories [14, 15, 23, 24] adopted a Poisson counting experiment technique to extract the final yields. Differing from these studies, the final yield is extracted by fitting the $M_{e\ell\ell'}$ distribution after all the selections. This allows to improve the sensitivity and makes the whole analysis more robust to the presence of remaining background events near the signal peak.

The Chapter is organised as follows:

- The signal candidate reconstruction is described in Section 4.2. It is based on an inclusive reconstruction of the signal $\tau \to e\ell\ell'$ decays.
- The background suppression is detailed in Section 4.3 and is performed in two steps, first with a set of cut-based selections, then with a BDT trained on data.
- The validation of the selection is presented in Section 4.4. Various checks are performed to ensure that the BDT doesn't introduce any bias.
- The fitting procedure, through which the branching fraction is measured, is described in Section 4.5.
- The systematic uncertainties are evaluated in Section 4.6.

[1] Charge conjugation is implied throughout the entire Chapter.

- The upper limit computation is detailed in Section 4.7.

The analysis uses the Run 1 dataset, corresponding to an integrated luminosity of 424fb^{-1}. Simulations of the various background processes are included in the analysis, incorporating $e^+e^- \to \ell\ell(\gamma)$ ($\ell = e$ or μ), $e^+e^- \to q\bar{q}$ ($q = u, d, s, c$) and other less contributing final states such as $\tau\bar{\tau}$, $eehh$ ($h = K, \pi$) or $B\bar{B}$. However, the main background source, $e^+e^- \to \ell\ell\ell'\ell'$, is not properly simulated and the entire optimisation of the selection is based on data sidebands. Additionally, 10 million signal events per mode are used. They are simulated with TAUOLA [89], following a three-body phasepace model; some BSM models, such as the ALP model described in Chapter 1, might deviate from it, lowering the sensitivity of the search for such scenarios.

4.2. Reconstruction of the signal candidates

Trigger selections

The trigger selections considered for this study are the low multiplicity (lml) trigger selections and the hie trigger[1], all based on the ECL.

Only the events that fire any of the lml[6-10], lml12 or hie are conserved. The overall trigger efficiency for $\tau^- \to e^\mp\ell^\pm\ell'^-$ is between 96 and 98% on reconstructed signal candidates. The definition of these triggers can be found in Table 2.4 of Chapter 2.

4.2.1. Particle lists

The first step of the reconstruction is to build the particle lists, by selecting the relevant tracks and assigning them the proper mass hypothesis. The charged particles are all required to come from the IP: the transverse (dr) and longitudinal (dz) projections of their point of closest approach must be smaller than 1 and 3 cm respectively[110].

Muons are identified by utilising the likelihood-based muon identification variable, while for electrons the BDT-based electron identification variable is employed. The initial selection value is loose: electrons and muons are required to have respectively an electron ID and muon ID greater than 0.5, and will be tightened at a later stage. The simulation of the lepton identification (LID) variables is not perfect; corrections in the form of weights are provided by the PID performance group, depending on the cut value and in bins of (θ, p_T) for each particle. Every quoted yields for MC events in this chapter correctly account for such corrections. For electrons, energy loss via bremsstrahlung is accounted for and corrected by adding back to their 4-momentum all the photons within 8.6 degrees (with respect to the electron momentum) and with an energy higher than 20 MeV.

[1] See section 2.3 of Chapter 2.

Every track not identified as a lepton is consequently assigned the pion hypothesis without further consideration for PID variables.

The photons are defined as ECL clusters within the CDC acceptance ($\theta \in (17°, 150°)$), not associated with any charged track, with additional requirements on the distance between clusters and tracks, energy and cluster timing. Photons whose energy is greater than 100 MeV are used for π^0 candidates reconstruction. π^0 are obtained by combining pairs of such photons with an invariant mass in the range $[0.115, 0.152]$ GeV/c^2. The remaining photons are required to have an energy of at least 200 MeV, to prevent beam background contamination. The neutral particles are reconstructed to properly measure the total energy of the event.

4.2.2. Event reconstruction

Signal side From these lists of particles, an inclusive (or *untagged*) reconstruction of $e^+e^- \to \tau^+\tau^-$ events is performed, illustrated on Figure 4.1, as was done already for the searches at Belle II for the $\tau \to \ell\phi$ [23] and $\tau \to 3\mu$ [24] LFV decays. Only the signal τ candidate is reconstructed, by combining three leptons with a total charge of ± 1, an invariant mass $M_{e\ell\ell'} \in (1.4,2.0)$ GeV/c^2 and $\Delta E_{e\ell\ell'} \in (-1.0,0.5)$ GeV, with

$$M_{e\ell\ell'} = \sqrt{E_{e\ell\ell'}^{\mathrm{c.m}^2} - P_{e\ell\ell'}^{\mathrm{c.m}^2}}, \tag{4.2}$$

and the difference between the three leptons energy and half of the centre of mass (c.m) energy $\sqrt{s}$:

$$\Delta E_{e\ell\ell'} = E_{e\ell\ell'}^{\mathrm{c.m}} - \frac{\sqrt{s}}{2}, \tag{4.3}$$

where $E_{e\ell\ell'}^{\mathrm{c.m}}$ and $P_{e\ell\ell'}^{\mathrm{c.m}}$ are the sum of the energies and the 3-momenta of the three leptons in the c.m. system.

The signal is expected to be peaking at $M_{e\ell\ell'} \sim m_\tau = 1776.86 \pm 0.12$ MeV/c^2 [61] and $\Delta E_{e\ell\ell'} \sim 0$ GeV due to the neutrinoless final state, as shown on figure 4.2. On the contrary, background events are smoothly distributed without peaking structures in both variables [1].

Finally, and the τ vertex is fitted by the *treeFitter* tool, which updates the momenta and position of the final state particles.

Rest of Event The other τ is not explicitly reconstructed, instead every particle and cluster not used in the signal reconstruction is gathered in a single object

[1]This is not entirely true; as shown later, there is a background that peaks around 100 MeV in $\Delta E_{e\ell\ell'}$, but remains flat in $M_{e\ell\ell'}$. The important point is that no background peaks at (0,1.777) in the ($M_{e\ell\ell'}$,$\Delta E_{e\ell\ell'}$) plane.

called the Rest Of Event (ROE), as illustrated in Figure 4.1. A mask[1] is applied to the ROE to ignore the tracks with a transverse momentum lower than 75 MeV/c or outside the CDC acceptance. The ROE is required to have a charge opposite to the signal τ, so that the total charge of the event is zero[2]. The ROE is also required to be made of at most three charged particles[3].

The ROE properties, such as its mass or its energy are stored and used for background suppression.

It now appears that defining each charged particle as a lepton or a pion is a good approximation. Signal events should have a ROE matching the SM decays of a τ, mainly $\tau \to \ell\nu\nu$, $\tau \to \pi\nu$ and $\tau \to \pi\pi\pi\nu$ (and additional π^0s). On the other hand, background events (as will be shown at the end of this section) are mainly QED processes (2ℓ or 4ℓ events) with a fully leptonic final state.

[1] Some particles are masked: the event is not rejected, but particles not coming from the e^+e^- collisions are not considered in the Rest Of Event

[2] The $\tau \to 3\mu$ analysis observed a small excess in data events with a total charge different from 0. Such events being difficult to interpret as a τ pairs, requiring a total charge of 0 comes naturally. Nonetheless, one might ponder whether to exhibit a *pair* of τ leptons with one undergoing a LFV decay, or focus solely on *a* τ doing so.

[3] It equates to six charged tracks at most in the event, counting the three leptons from the signal candidate. This requirement together with the selection on $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$ corresponds to the τ-LFV skim.

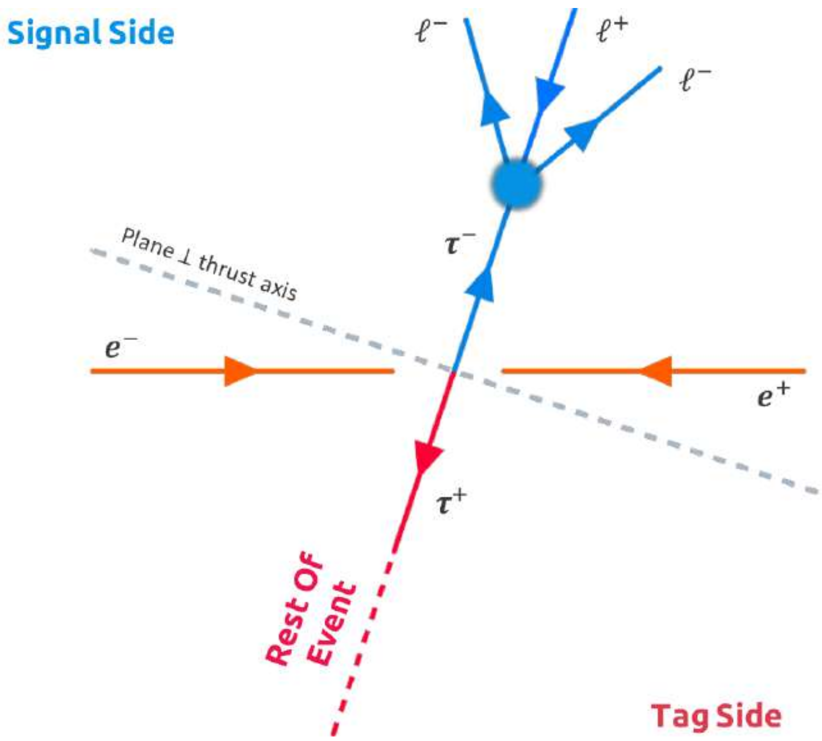


Figure 4.1.: Diagram of the $e^+e^- \to \tau^+\tau^-$ events reconstruction for the $\tau^- \to \ell^-\ell^+\ell^-$ LFV decays with an inclusive approach. The dashed line correspond to the plane perpendicular to the thrust axis, which is used to define two hemispheres. The three leptons from the τ candidate are required to be in the same hemisphere, while the ROE must be in the other one.

Event shape and kinematics

All charged and neutral particles contained in the τ signal candidate and the ROE 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 energies of every input particle. The missing momentum (energy) is calculated by subtracting the sum of the momenta (energies) 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.

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

$$T = \max_{\mathbf{n}_T} \left(\frac{\sum_i |\mathbf{p}_i^* \cdot \mathbf{n}_T|}{\sum_i |\mathbf{p}_i^*|} \right), \tag{4.4}$$

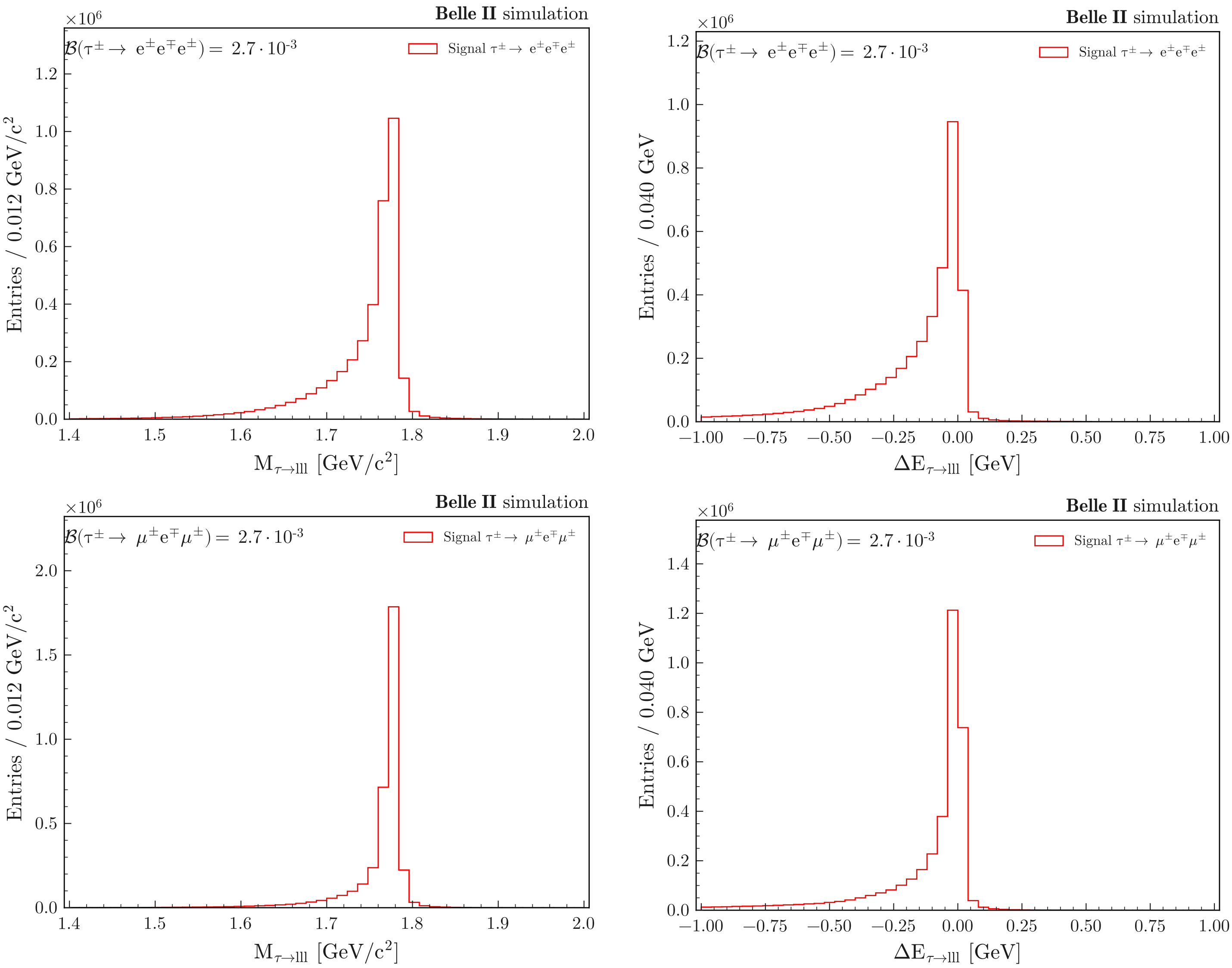


Figure 4.2.: Signal distribution of $M_{e\ell\ell'}$ (left) and $\Delta E_{e\ell\ell'}$ (right) for the $e^-e^+e^-$ (upper) and $\mu^-e^+\mu^-$ (lower) final states. The asymmetry in the distributions is caused by energy loss (initial or final state radiation, or bremsstrahlung; the more electrons are in the final state, the more important is the tail) creating the tails at low $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$. The sharp cut-off is due to the absence of neutrino in the final state.

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

In the $e^+e^- \to \tau^+\tau^-$ process, τ leptons are produced back-to-back in the centre of mass frame and quite boosted, so their decay products are well geometrically separated. Thus, the three signal side leptons must be in the same hemisphere defined by the plane orthogonal to the thrust axis (see Figure 4.1), and the ROE be on the other side.

Signal truth matching

In order to ensure the correctness of the $\tau^- \to e^\mp\ell^\pm\ell'^-$ reconstruction for the signal (simulated) sample, a set of requirements based on MC information is applied :

- the charge of the τ is correctly reconstructed
- the three leptons have a τ in their ancestors,
- the three leptons are correctly identified

These requirements on simulated signal sample is important for background suppression optimization, where signal candidates must be well identified. The absolute truth-matching efficiency for signal events is 99.9% for all five modes, also indicating that the signal events are properly reconstructed.

Signal efficiency and background composition

The signal efficiency at reconstruction, and after the different steps (skim and trigger) defined previously, as well as the expected background yields, are given in Table 4.1. The reconstruction requirements are quite loose, and the yields are therefore dominated by Bhabha events.

4.3. Background suppression

The background suppression strategy operates on two levels, with the first targeting the easier-to-reject backgrounds (showing obvious structures in different distributions) using a cut-based selection, while the second is designed to eliminate any remaining background components (2ℓ and 4ℓ, depending on the mode) using a BDT classifier (one for each mode). Since some background contributions are either missing or mismodeled in the simulations[1] (mainly radiative events, with photo-conversion producing electron pairs), all background suppression is performed on data, and the BDT classifiers are trained on data events. Consequently, data have to be blinded in order not to bias the selection, with the blind region being defined in the $(M_{e\ell\ell'}, \Delta E_{e\ell\ell'})$ plane.

[1]The fact that some simulations are incomplete have been stated a few times now. This section will show data/MC comparison to finally justify this affirmation

4.3.1. Blind region definition

Four widths $\delta^{low}_{\Delta E}$, $\delta^{high}_{\Delta E}$, δ^{low}_{M} and δ^{high}_{M} are defined to construct the box delimited by $[-\delta^{low}_{\Delta E}, \delta^{high}_{\Delta E}]$ in $\Delta E_{e\ell\ell'}$ and $[m_\tau - \delta^{low}_{M}, m_\tau + \delta^{high}_{M}]$ in $M_{e\ell\ell'}$. They are chosen such that the coverage on each side of the distributions is 99% of the respective signal yield :

$$N_{sig}^{\delta^{low}_{\Delta E}} = 0.99 \cdot N_{sig}^{\Delta E,low} \tag{4.5}$$
$$N_{sig}^{\delta^{high}_{\Delta E}} = 0.99 \cdot N_{sig}^{\Delta E,high} \tag{4.6}$$
$$N_{sig}^{\delta^{low}_{M}} = 0.99 \cdot N_{sig}^{M,low} \tag{4.7}$$
$$N_{sig}^{\delta^{high}_{M}} = 0.99 \cdot N_{sig}^{M,high} \tag{4.8}$$

where $N_{sig}^{\delta^{low}_{\Delta E}}$, $N_{sig}^{\delta^{high}_{\Delta E}}$, $N_{sig}^{\delta^{low}_{M}}$, $N_{sig}^{\delta^{high}_{M}}$ are respectively the numbers of signal events in the $\Delta E_{e\ell\ell'}E$ intervals $[\delta^{low}_{\Delta E}$,0], [0, $\delta^{high}_{\Delta E}]$ and $M_{e\ell\ell'}$ intervals [1.777, δ^{high}_{M}] and [δ^{low}_{M}, 1.777].

$N_{sig}^{\Delta E,low}$, $N_{sig}^{\Delta E,high}$ are respectively the number of signal events in the ΔE intervals $[-\infty$,0] and [0,∞] (and similar for $N_{sig}^{M,low}$, $N_{sig}^{M,high}$), but with an "initial state radiation (ISR) mask" applied. This ISR mask hides signal events where the $\Delta E_{e\ell\ell'}$ is negative, while the mass remains near 1.777 GeV :

$$\Delta E_{e\ell\ell'} \cdot cos(\theta_{corr}) - M_{e\ell\ell'} \cdot sin(\theta_{corr}) > -1.76 \tag{4.9}$$

Table 4.1.: Total signal efficiencies and number of simulated background events retained after the reconstruction and after additional requirements' step for $\tau^- \to e^\mp\ell^\pm\ell'^-$. The yields are normalized to 424 fb^{-1}. N^{tot}_{bkg} is the sum of the individual background components. "Other" stands for everything else ($B\bar{B}$ and $\ell\ell hh$ events).

		ε_{sig}	$N_{q\bar{q}}$	$N_{\tau\tau}$	$N_{\ell\ell}$	$N_{\ell\ell\ell\ell}$	Other	N^{tot}_{bkg}
$e^-e^+e^-$	Reconstruction	40.2%	1,211	1,147	1,000,206	41,054	6,265	1,049,883
	Offline	37.9%	1,180	1,104	718,694	25,152	4,934	751,064
$e^-e^+\mu^-$	Reconstruction	39.2%	6,504	1,779	71,295	15,258	711	95,547
	Offline	37.2%	6,359	1,693	54,528	13,748	587	76,914
$e^-\mu^+e^-$	Reconstruction	39.2%	1,034	84	28,290	864	168	30,439
	Offline	37.1%	1,006	75	13,498	469	138	15,186
$\mu^-\mu^+e^-$	Reconstruction	38.3%	6,828	385	94,285	21,305	419	123,221
	Offline	35.9%	6,640	321	13,240	15,863	361	36,424
$\mu^-e^+\mu^-$	Reconstruction	38.4%	3,035	191	1,520	835	221	5,803
	Offline	36.1%	2,960	162	1,283	695	218	5,318

where θ_{corr} is the correlation angle between $\Delta E_{e\ell\ell'}$ and $M_{e\ell\ell'}$. It is obtained by fitting $M_{e\ell\ell'}$ in the range [1.72,1.82] against $\Delta E_{e\ell\ell'}$ in the range $[-0.25,1]$, assuming a linear correlation, as can be seen on Figure 4.3. The values for θ_{corr} are given in Table 4.2, while the ISR mask can be seen on Figure 4.4 with the blind region.

Masking these events for the definition of the blind box is useful : the ISR tail is quite long (particularly for the modes with two and three electrons on the signal side) and would draw too much the blind box at low values of $\Delta E_{e\ell\ell'}$ (and blinding half of the 2D plane would be too much).

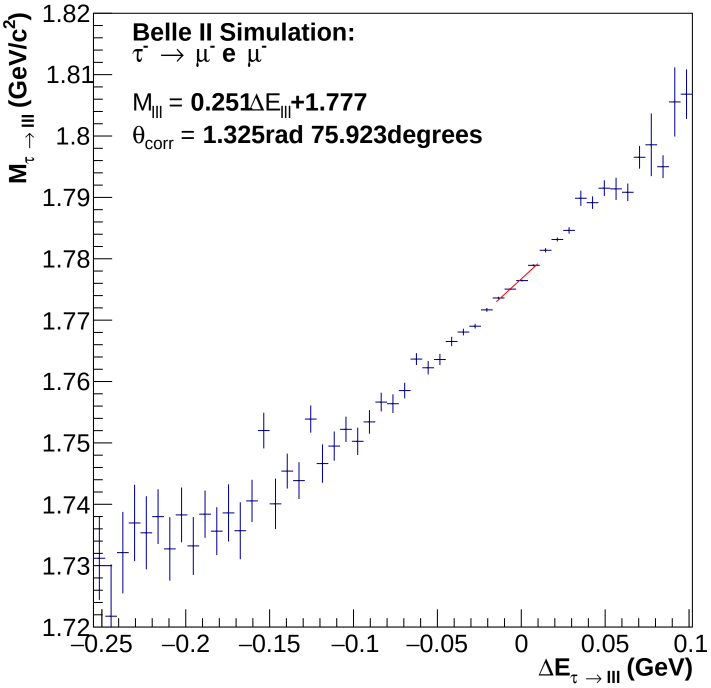


Figure 4.3.: Fit of the correlation angle θ_{corr} between $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$ for $\tau \to e^-\mu^+e^-$ simulated signal events.

Table 4.2 shows the $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$ intervals of the blind region, as well as the relative signal coverage (i.e. the fraction of signal that lies in it). Any data mentioned from this point will indeed be blinded using the box detailed here, as well as background simulation (unless specified otherwise).

Table 4.3 shows the total signal efficiency after reconstruction (from previous section), background yields from simulation and data after reconstruction outside the blinded box.

Simulations show that the main sources of background are QED processes leading to four leptons in the final state originating from either one pair of leptons with ISR or FSR, or two pairs of leptons. $q\bar{q}$ events are also identified as part of the background composition. An important data excess is observed[1] for almost all the modes, understood as two pairs of leptons with ISR or FSR, missing from the simulations (since the corresponding generator does not handle FSR and ISR). The $\mu^- e^+ \mu^-$ mode is the only one with no such data excess; this is explained by the presence of two same sign muons on the signal side, ensuring an almost background free scenario for this mode, at least from QED processes.

Table 4.2.: Blind box definition with $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$ intervals. ϵ_{sig}^{rel} is the relative coverage of the blind region with respect to reconstruction and θ_{corr} is the correlation angle between $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$

	$M_{e\ell\ell'}^{blind}$ (GeV)	$\Delta E_{e\ell\ell'}^{blind}$ (GeV)	ϵ_{sig}^{rel}	θ_{corr} (deg)
$e^- e^+ e^-$	[1.53,1.85]	[-0.65,0.24]	94.0%	72.2
$e^- e^+ \mu^-$	[1.52,1.83]	[-0.62,0.15]	94.5%	73.9
$e^- \mu^+ e^-$	[1.53,1.83]	[-0.64,0.25]	94.3%	75.6
$\mu^- \mu^+ e^-$	[1.56,1.81]	[-0.60,0.21]	94.2%	76.8
$\mu^- e^+ \mu^-$	[1.53,1.82]	[-0.63,0.14]	94.4%	75.9

[1]At last !

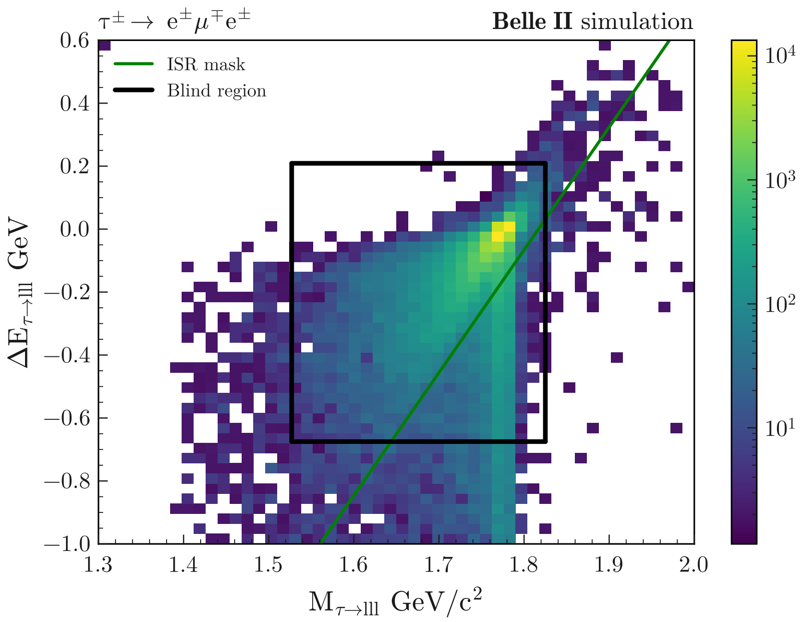


Figure 4.4.: Signal distribution in the ($M_{e\ell\ell'}$,$\Delta E_{e\ell\ell'}$) plane with the blind region (black box), for $\tau \to e^- \mu^+ e^-$. The ISR mask corresponds to the green line.

4.3.2. Cut-based preselection

The first step of background suppression is to refine the LID selection to maintain high signal efficiency and reduce background from misidentification of the leptons (mainly in $q\bar{q}$ events). The performance group only provides simulation corrections for a few threshold values (0.5, 0.9, 0.95, 0.99), hence only these can be used. Cutting at a value of 0.95 for the three leptons would result in a too-low signal efficiency. This is especially true for the modes with 2 muons, and as was done in the $\tau \to 3\mu$ analysis, the leptons LID are ranked : LID_{lead}, LID_{sub} and LID_{third} where *lead* and *sub* refer to the two same flavoured leptons, ranked according to their lepton ID (electron or muon depending on the mode). *Third* refers to the other flavoured lepton, or the third (ranked by PID) electron in the case of the $e^-e^+e^-$ mode. These lepton ID distributions are shown for simulations and data outside the blind region in Figures 4.5, 4.6 and 4.7.

Both LID_{lead} and LID_{third} are required to be greater than 0.9, meaning that every event must have at least one electron and one muon with a high LID value (except for the $e^-e^+e^-$ mode where obviously no muon ID requirement is used). For the $e^-e^+e^-$ mode, LID_{sub} is also required to be greater than 0.9 (while it is not tightened from the initial value of 0.5 for the other modes). This special treatment is due to both the higher initial background level for the 3e mode, combined with the overall better electron ID performance. These values are compiled in Table 4.4, where e$\mu\ell$ stands for every mode other than $e^-e^+e^-$.

Since an important contribution to background comes from 2ℓ and 4ℓ QED processes, in which the leptons are not mis-identified, tighter LID requirements would not be rejecting more background.

Table 4.3.: Total signal efficiency, background yields obtained from simulation and yields in data. The yields are computed outside the blind region. For each mode, the second line correspond to the yields in simulation including the blind region. The MC yields are scaled to 424 fb^{-1}.

	ϵ_{sig}^{tot}	$N_{q\bar{q}}$	$N_{\tau\tau}$	$N_{\ell^+\ell^-}$	$N_{\ell\ell\ell\ell}$	Others	N_{MC}^{tot}	N_{data}
$e^-e^+e^-$	37.9%	876	829	387,481	13,303	3,544	406,033	432,754
	_	1,181	1,104	718,597	25,137	4,931	750,949	_
$e^-e^+\mu^-$	37.2%	4,770	1,313	33,970	7,902	468	48,423	53,338
	_	6,362	1,692	54,506	13,744	586	76,890	_
$e^-\mu^+e^-$	37.2%	793	60	7,504	276	123	8,755	13,160
	_	995	75	13,489	469	137	15,605	_
$\mu^-\mu^+e^-$	36.0%	5,689	280	11,027	9,477	349	26,821	39,629
	_	6,703	320	13,239	15,860	361	36,483	_
$\mu^-e^+\mu^-$	36.2%	2,416	135	921	532	209	4,213	3,801
	_	2,970	161	1,283	695	218	5,327	_

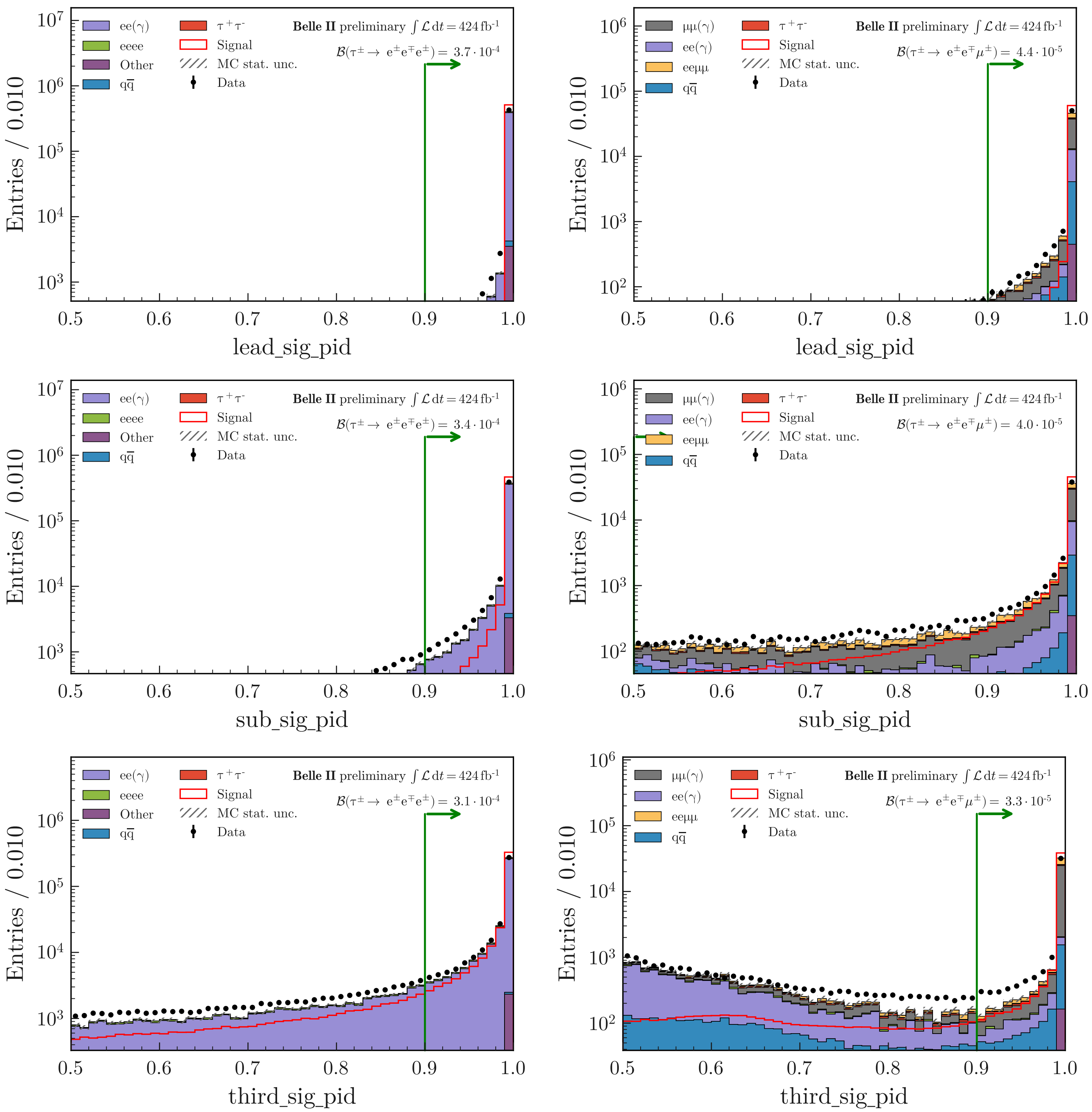


Figure 4.5.: Distributions of the three ranked leptonID values in the signal side, for MC signal and background sources as well as data (top: leading LID, middle: subleading LID, bottom: third LID) for the $e^-e^+e^-$ (left) and $e^-e^+\mu^-$ (right) modes.

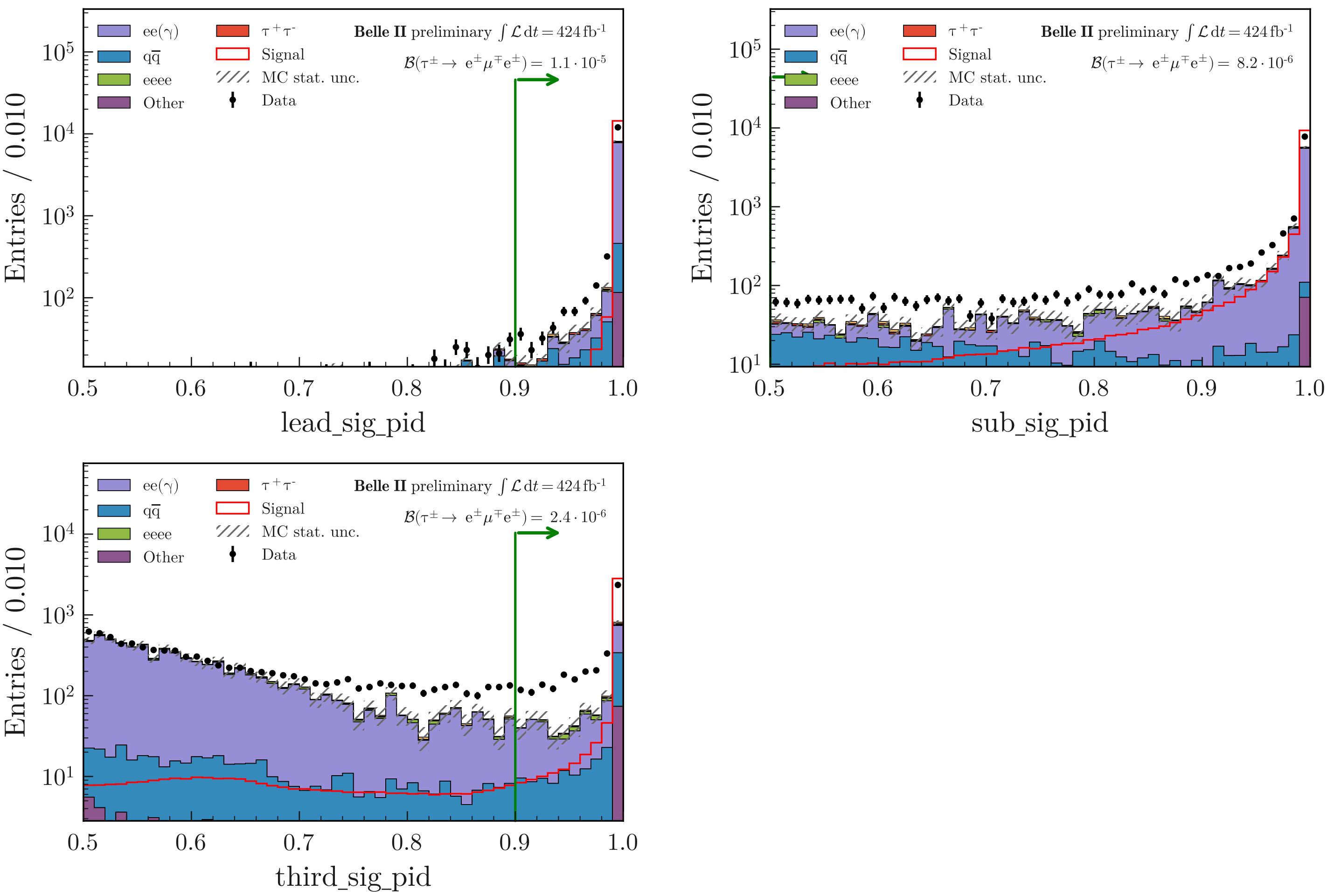


Figure 4.6.: Distributions of the three ranked leptonID values in the signal side, for MC signal and background sources as well as data (upper left: leading LID, upper right: subleading LID, lower left: third LID) for the $e^-\mu^+e^-$ mode.

Table 4.4.: Requirement on the LID for the different modes. e$\mu\ell$ stands for the four modes other than eee, lead and sub refer to the same flavoured leptons, while third to the opposite flavoured lepton (or the third electron by electronID in the $e^-e^+e^-$ mode).

	LID_{lead}	LID_{sub}	LID_{third}
$e^-e^+e^-$	0.9	0.9	0.9
e$\mu\ell$	0.9	0.5	0.9

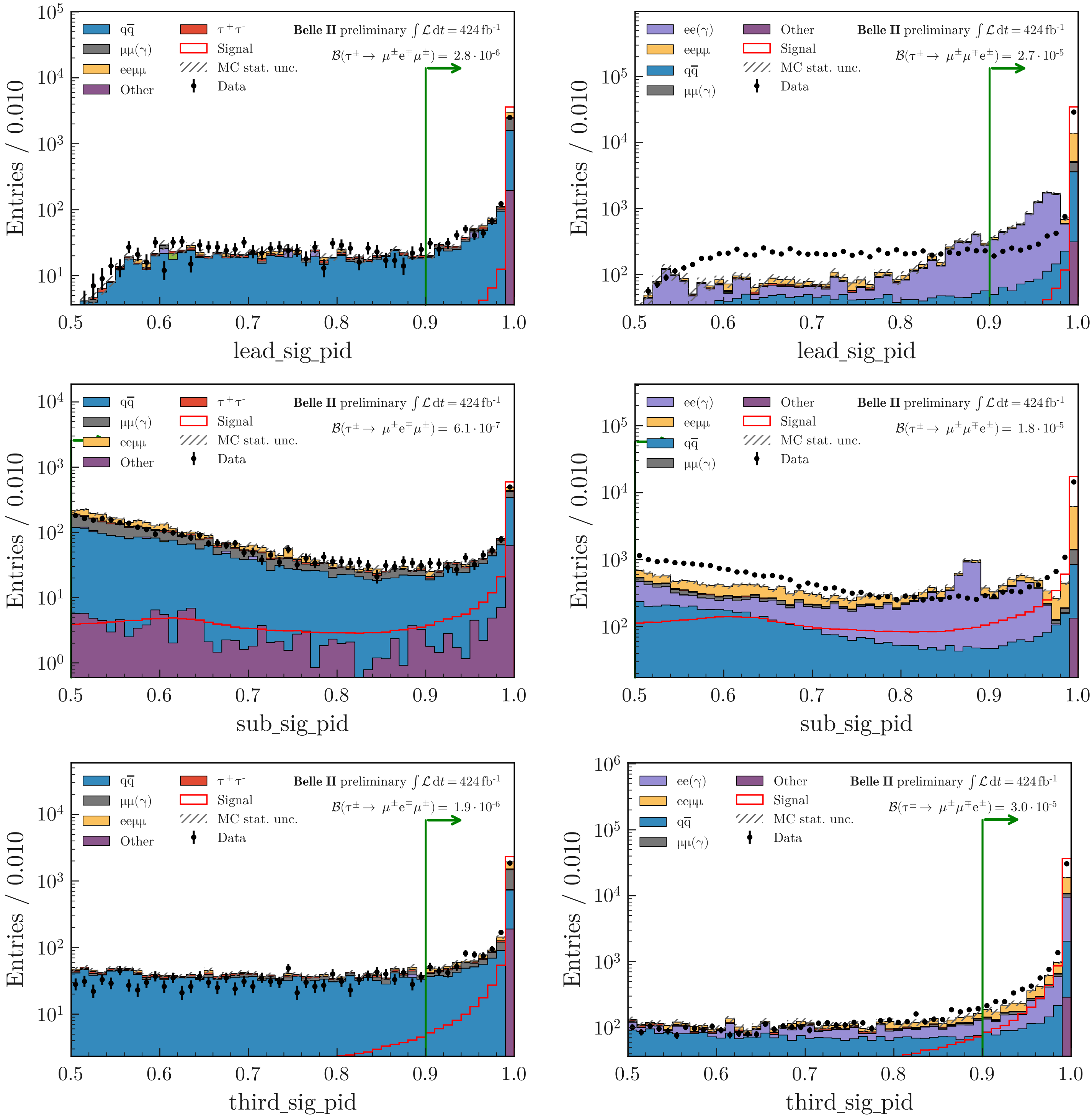


Figure 4.7.: Distributions of the three ranked leptonID values in the signal side, for MC signal and background sources as well as data (top: leading LID, middle: subleading LID, bottom: third LID) for the $\mu^-e^+\mu^-$ (left) and $\mu^-\mu^+e^-$ (right) modes.

Additional criteria are used to reduce the background contributions by removing peaks in different distributions. QED backgrounds are identified using two characteristics : firstly, 2ℓ are accompanied by ISR and a subsequent pair production : thus these events can be rejected by cutting on the invariant mass of opposite charge lepton pairs, under the assumption that it's a converted photon. It is therefore required that these dilepton invariant masses, (computed by assigning the electron mass to the leptons, disregarding their assigned mass) to be greater than 25 MeV/c^2 for all the modes but $\tau^- \to e^-e^+e^-$, as can be seen on Figures 4.8 and 4.9. For $\tau^- \to e^-e^+e^-$, the selection is tightened to 50 MeV/c^2. This cut is justified by the higher background contribution at 0MeV/c^2 for this mode (due to Bhabha scattering). This is seen on Figure 4.10, showing dilepton invariant masses before and after tightening the threshold from 25 MeV/c^2.
Secondly, most of these events are neutrino-less and therefore should have a total visible energy really close to the beam energy, as well as a thrust close to 1. The distributions of these two variables are in Figure 4.11 and Figure 4.12. These two variables are correlated, but it was found more efficient to cut (loosely) on both of them.

Di-photon events with energy along the beam axis (and which are not properly simulated) can be seen in data as peaks at the edges of the polar angle of the missing momentum $\theta^{\rm c.m}_{miss}$. Such events are eliminated by requiring $0.3 < \theta^{\rm c.m}_{miss} < 2.7$. The distribution of $\theta^{\rm c.m}_{miss}$ for each mode is given in Figure 4.13.

Lastly, for the $e^-e^+e^-$ and $\mu^-e^+\mu^-$ modes, events with a total cluster energy of the rest of event above 3.5 GeV are removed, since this variable exhibits a clear separation between background and signal. This is shown in Figure 4.14.

Table 4.5.: Value of the cuts on the various discriminating variables, for all the modes.

	θ^{CM}_{miss} [deg]	$M^{\gamma}_{\ell_1\ell_2}$[GeV/c^2]	$M^{\gamma}_{\ell_2\ell_3}$[GeV/c^2]	E^{CMS}_{vis} [GeV]	thrust	E^{ROE}_{ECL}[GeV]
$e^-e^+e^-$	$\in$ [17.2,154.7]	> 0.05	> 0.05	< 10.4	< 0.97	< 3.5
$e^-e^+\mu^-$	$\in$ [17.2,154.7]	> 0.025	> 0.025	< 10.4	< 0.97	_
$e^-\mu^+e^-$	$\in$ [17.2,154.7]	> 0.025	> 0.025	< 10.4	< 0.97	_
$\mu^-e^+\mu^-$	$\in$ [17.2,154.7]	> 0.025	> 0.025	< 10.4	< 0.97	_
$\mu^-\mu^+e^-$	$\in$ [17.2,154.7]	> 0.025	> 0.025	< 10.4	< 0.97	< 3.5

All the cut values are summarised in Table 4.5, and the resulting signal efficiencies and background yield for simulation and data are shown in Table 4.6.

At this stage, simulations and data suggest that most of the surviving background contributions are QED processes and $q\bar{q}$ events. The $q\bar{q}$ contribution lies at low values of $\Delta E_{e\ell\ell'}$ as can be seen on Figure 4.15 and Figure 4.16, as well as from simulation only inside the blind region on figure 4.19.
These figures also show that the background that needs to be removed (i.e. close to the signal peak in the ($M_{e\ell\ell'}$,$\Delta E_{e\ell\ell'}$) plane) is the contribution corresponding to the data excess, showing a peaking structure around 100 MeV in $\Delta E_{e\ell\ell'}$. The

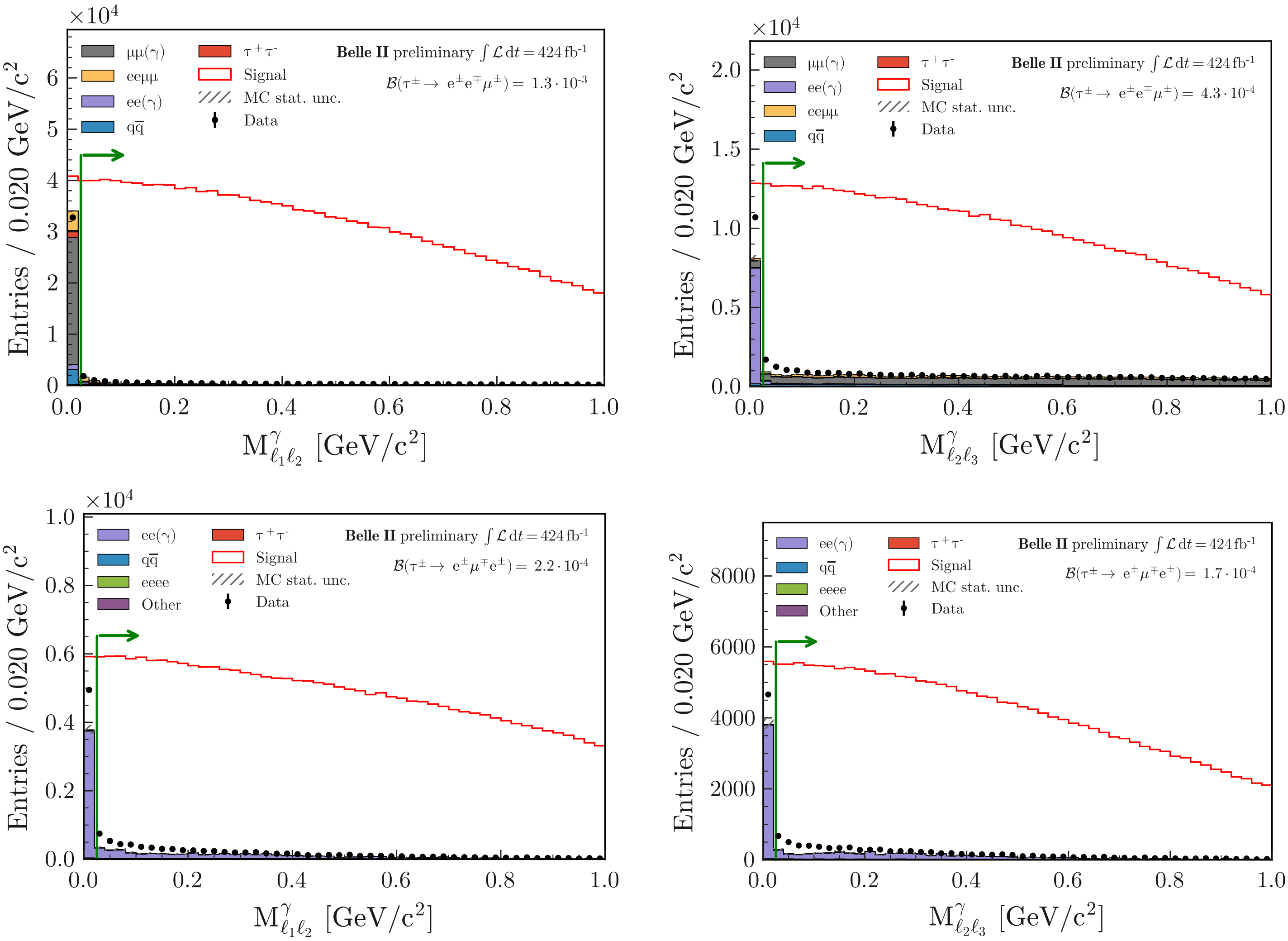


Figure 4.8.: Invariant mass of dilepton systems system under the photon conversion assumption. ℓ_1, ℓ_3 refer to the same sign leptons, while ℓ_2 is the opposite sign lepton. This ordering follows the ordering in the decay chains.

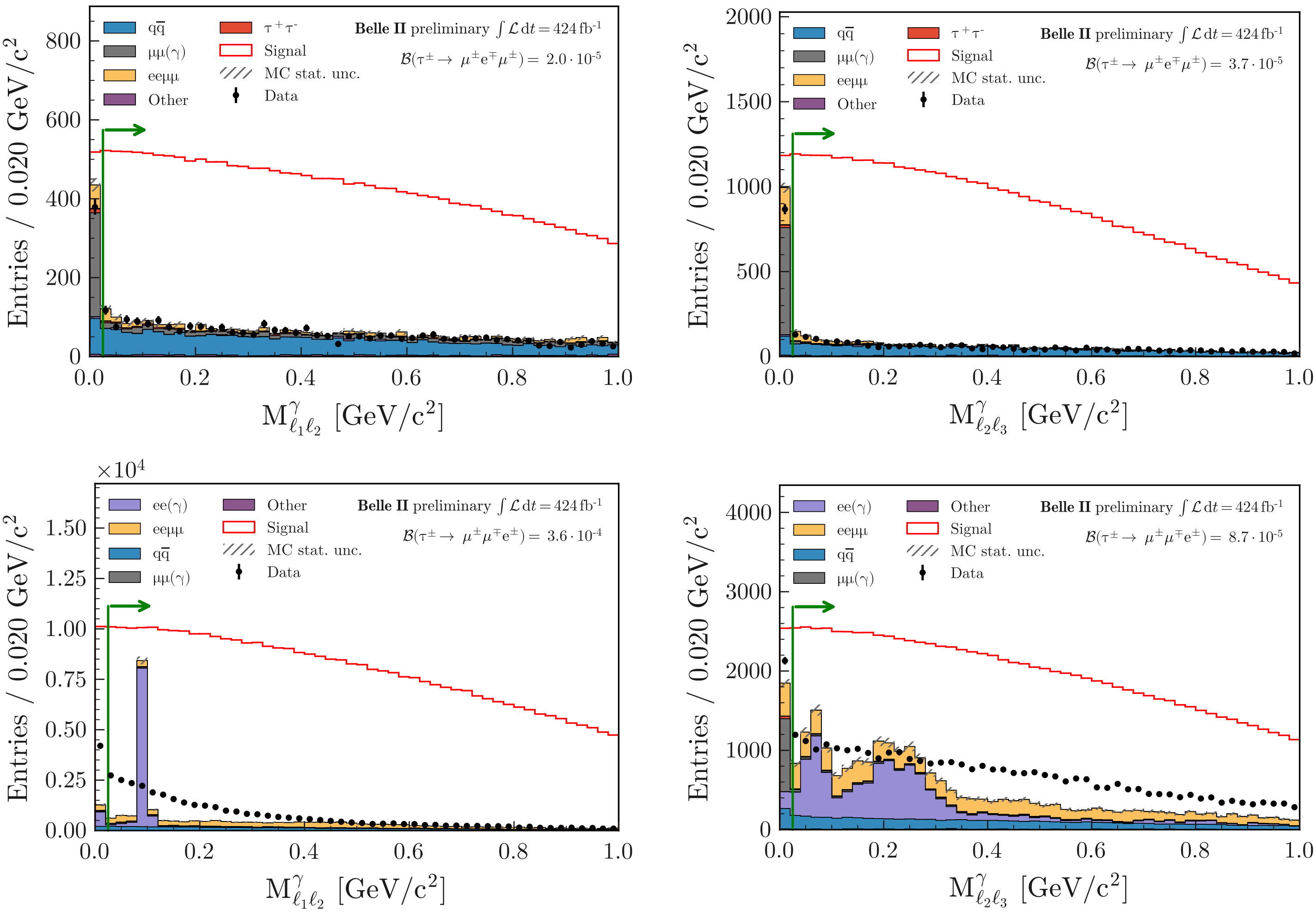


Figure 4.9.: Invariant mass of dilepton systems system under the photon conversion assumption. ℓ_1, ℓ_3 refer to the same sign leptons, while ℓ_2 is the opposite sign lepton. This ordering follows the ordering in the decay chains.

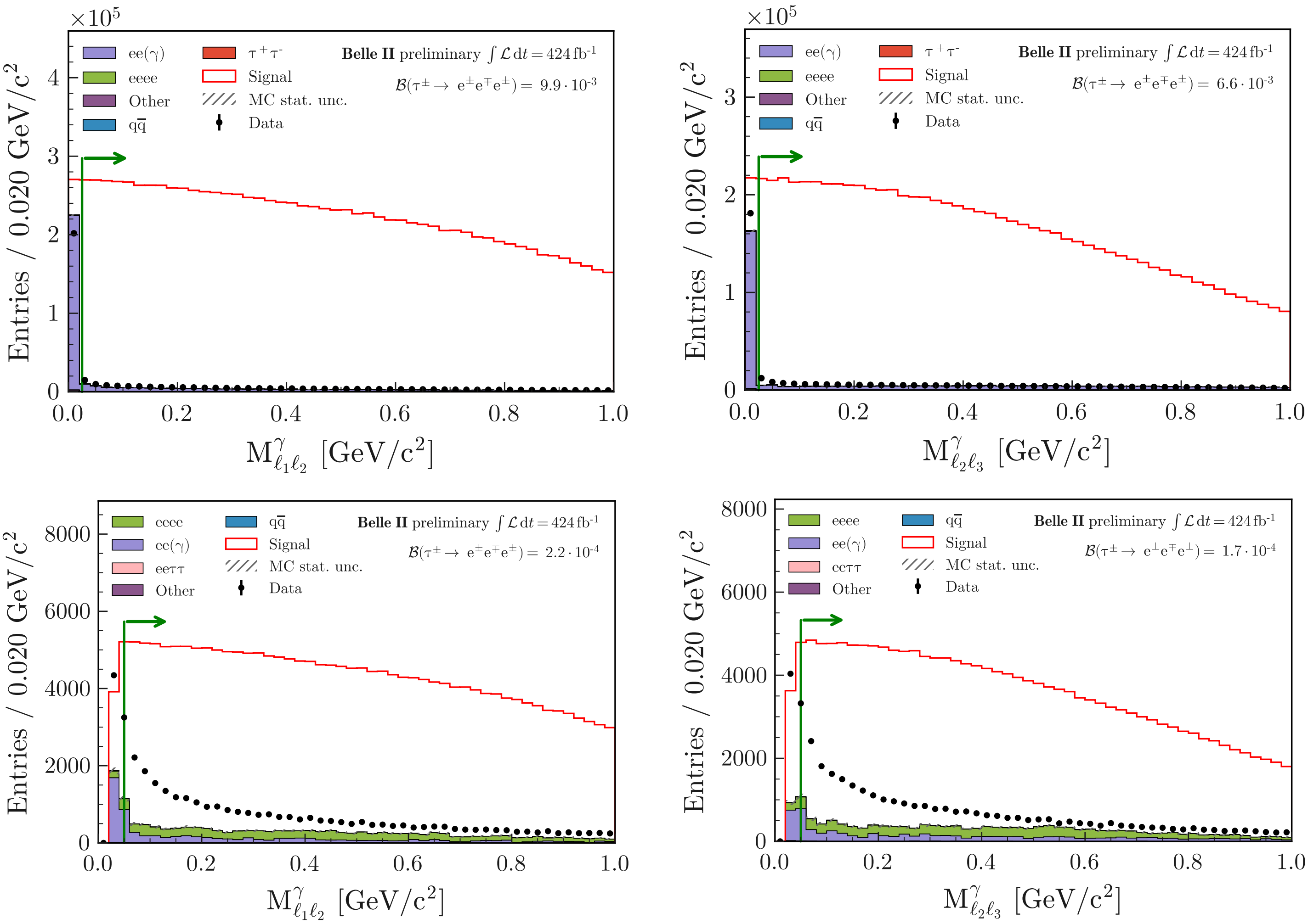


Figure 4.10.: Invariant mass of dilepton systems under the photon conversion assumption, for the three electrons mode. The two upper figures show the initial distribution, and a cut at 25 MeV, while the lower figures show the resulting distribution and the tightened cut at 50 MeV.

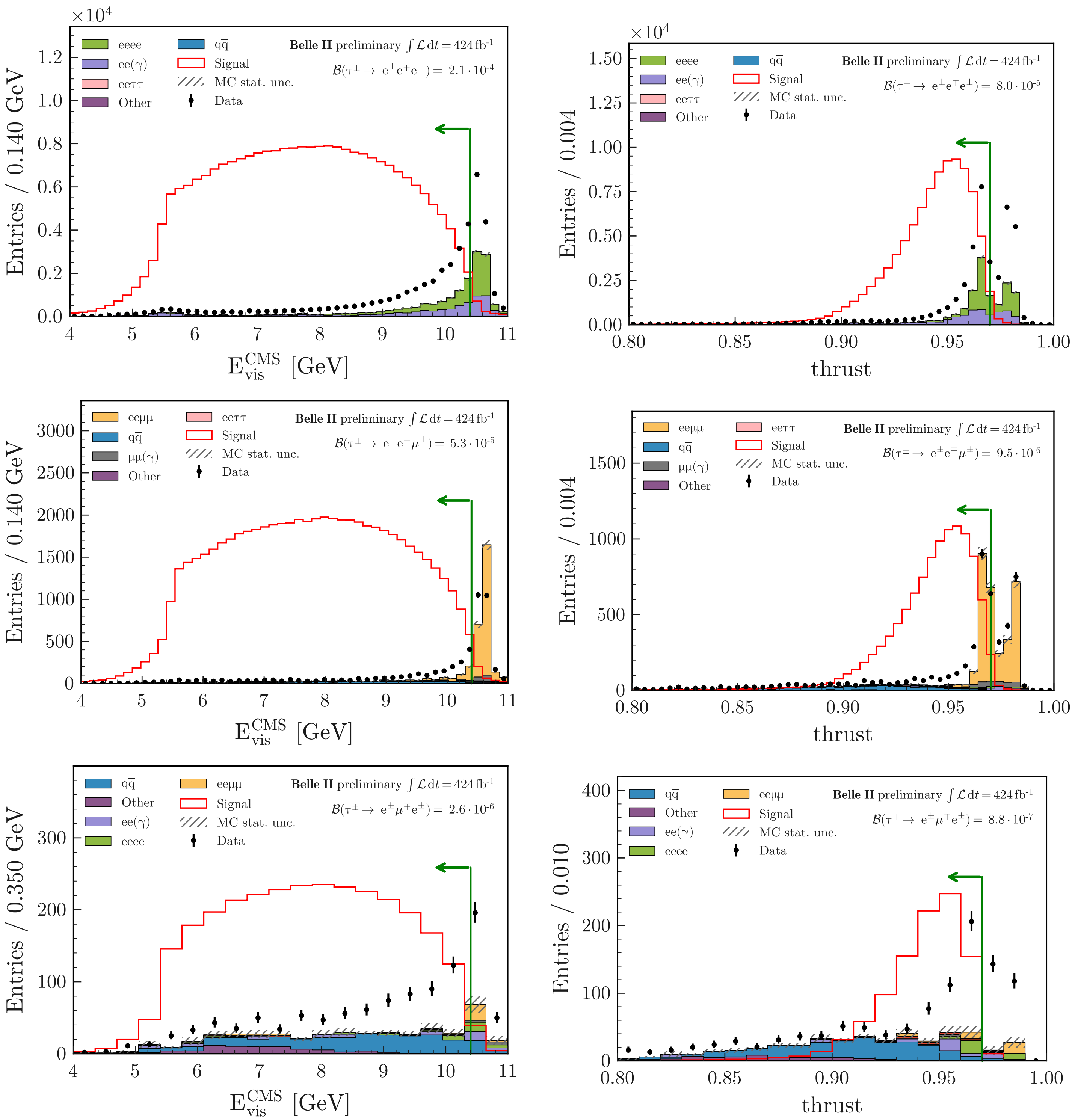


Figure 4.11.: Distributions of thrust (right) and visible energy (left), after the cuts on $\theta_{miss}^{c.m}$ and the dilepton invariant masses, for the 2 and 3 electrons modes.

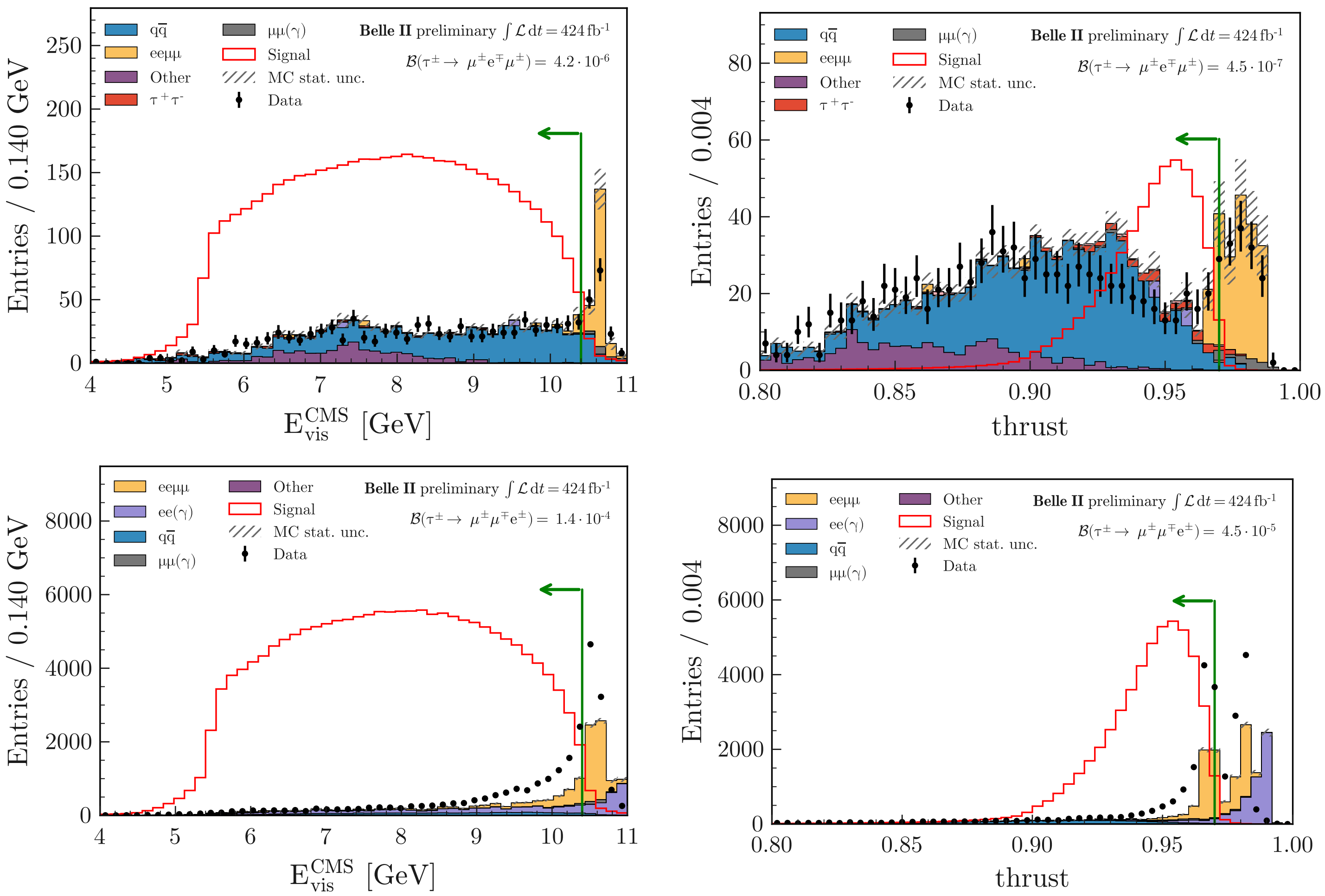


Figure 4.12.: Distributions of thrust (right) and visible energy (left), after the cuts on $\theta_{miss}^{c.m}$ and the dilepton invariant masses, for the 2 and 3 electrons modes.

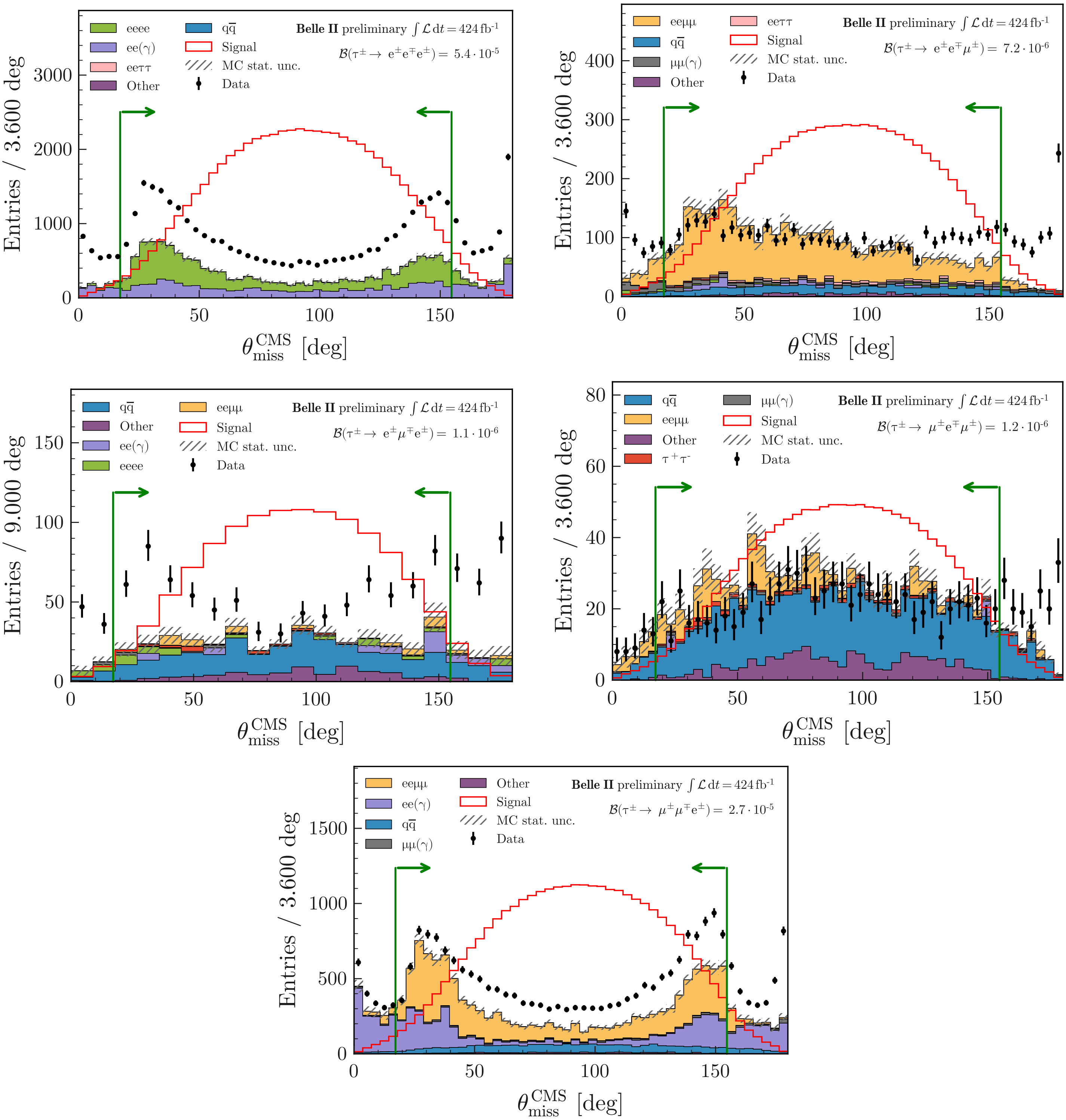


Figure 4.13.: $\theta^{c.m}_{miss}$ distribution for each modes, after the selection on the leptonID and the invariant mass of dilepton pairs.

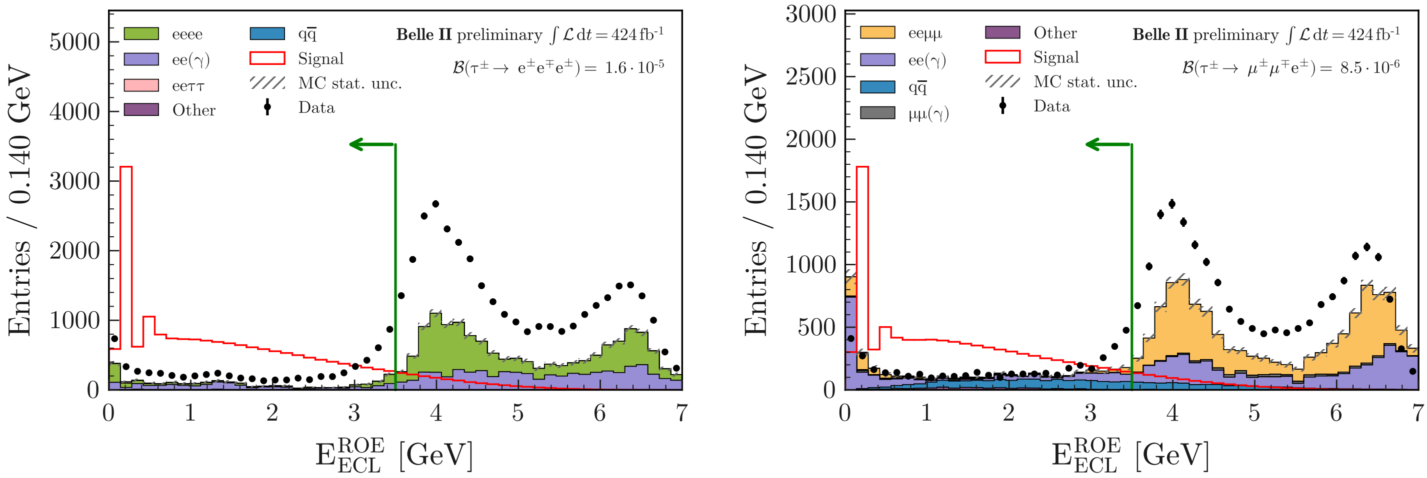


Figure 4.14.: Distribution for the total cluster energy of the rest of event for the $e^-e^+e^-$ (left) and $\mu^-\mu^+e^-$ (right) modes. The other selections are applied.

Table 4.6.: Signal efficiency, background simulation and data yields after the selections. For each mode, the second line correspond to the yields in simulation including the blind region. The MC yields are scaled to 424 fb^{-1}.

	ϵ_{sig}^{tot}	$N_{q\bar{q}}$	$N_{\tau\tau}$	$N_{\ell^+\ell^-}$	$N_{\ell\ell\ell\ell}$	N_{other}	N_{MC}^{tot}	N_{data}
$e^-e^+e^-$	26.3%	21	2	113	281	42	459	1,053
	—	24	3	156	421	44	648	—
$e^-e^+\mu^-$	29.7%	438	38	39	269	134	917	1,229
	—	539	48	41	338	137	1,102	—
$e^-\mu^+e^-$	29.6%	209	13	17	18	67	325	498
	—	262	16	17	22	68	386	—
$\mu^-\mu^+e^-$	27.8%	1,035	51	51	149	244	1,530	1,582
	—	1,198	59	61	201	248	1,767	—
$\mu^-e^+\mu^-$	30.7%	560	26	10	8	175	778	635
	—	678	32	10	8	180	907	—

reason for the peak to be in the positive $\Delta E_{e\ell\ell'}$ region is explained by the fact that it corresponds to two leptons pairs events: each pair carry half of the beam energy, so for a signal candidate to be reconstructed an additional track must be added to the pair. The reason for this contribution to be peaking (instead of a continuum for instance) is however not properly understood, nor is the value of the position of the peak.

It is noteworthy to realise that for most of the backgrounds (for the $e^-e^+e^-$, $e^-\mu^+e^-$ and $\mu^-\mu^+e^-$ modes), the excess is due to events with four tracks in the

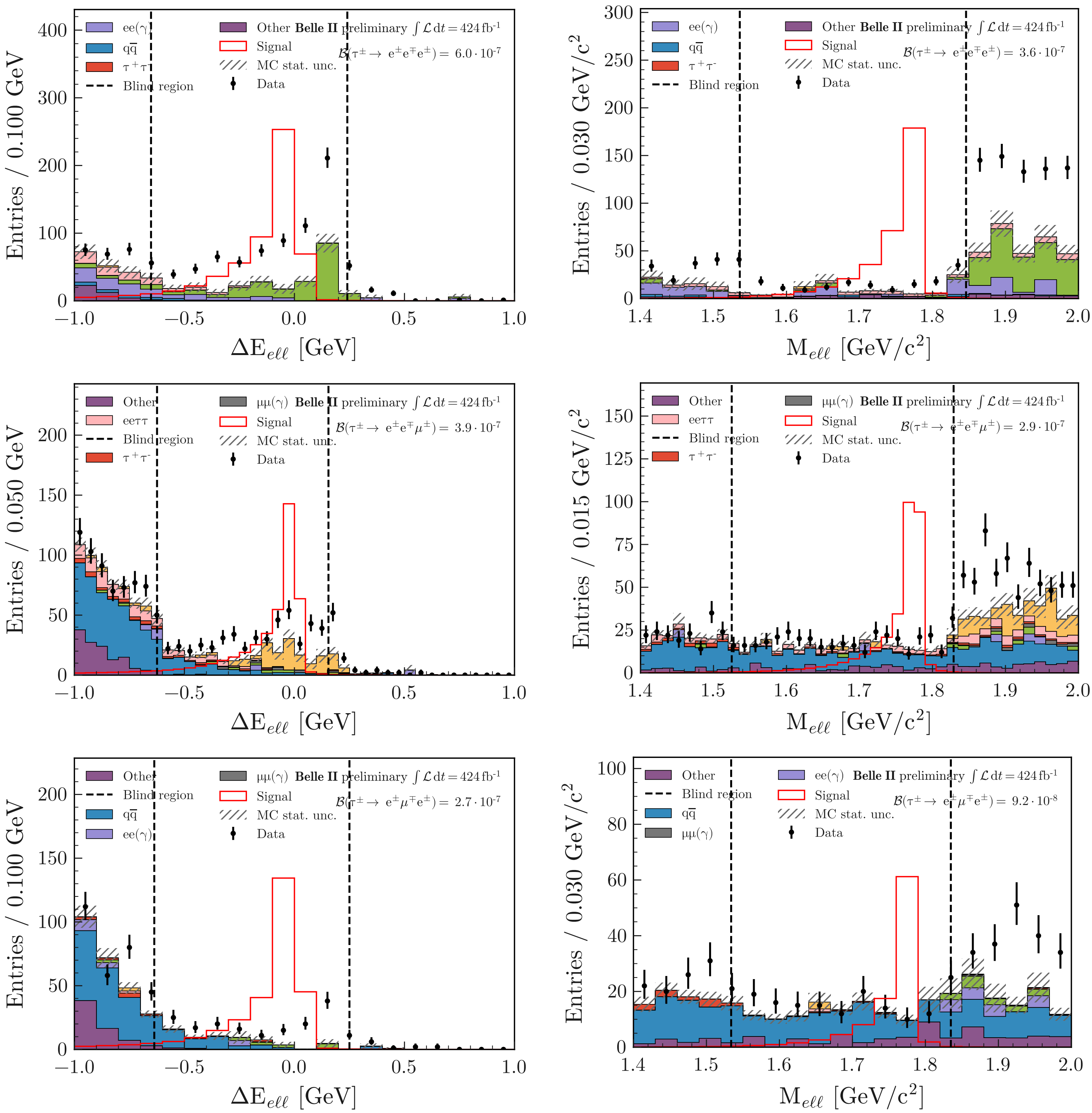


Figure 4.15.: $\Delta E_{e\ell\ell'}$ (left) and $M_{e\ell\ell'}$ (right) distributions after all selections, for the 2 and 3 electrons modes. The dashed black lines show the limit of the blind region.

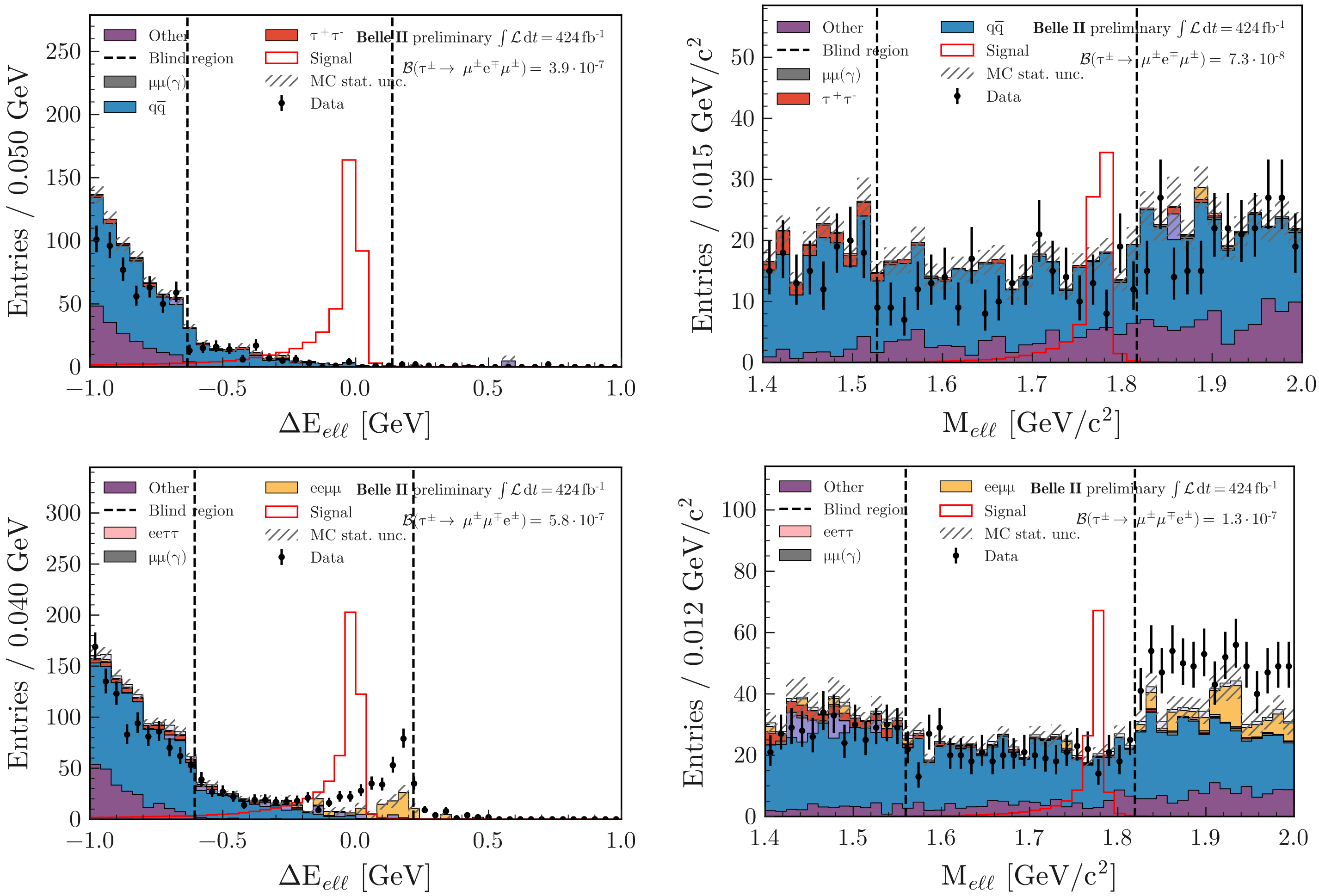


Figure 4.16.: $\Delta E_{\ell\ell'}$ (left) and $M_{\ell\ell'}$ (right) distributions after all selections, for the 2-muons modes. The dashed black lines show the limit of the blind region.

event[1], as can be seen on Figure 4.17. Figure 4.18 shows that the excess corresponds to particle with a high electron ID, adding more support to the idea that the data excess is indeed coming from four leptons processes.

The $\mu^- e^+ \mu^-$ final state is the only one with no data excess, as explained earlier, and is actually showing the opposite behaviour with a deficit in data with respect to simulations. The main background for this mode comes from $q\bar{q}$ events, for which the cross-section is underestimated in the simulation.

One could try to reject all these remaining background events using more cuts, by leveraging the simple topology of these events and for instance apply electron veto on the ROE. However, as will be explained in the following section, this final background rejection step is performed by using a BDT.

Due to the large discrepancy between data and simulation and the low generated luminosity for event types such as $e^+e^- \to e^+e^-(\gamma)$ (100 fb^{-1}), $e^+e^- \to ee\mu\mu$ or $e^+e^- \to e^+e^-e^+e^-$ (200 fb^{-1}), the BDT are trained directly on data. This training must exclude events that will be used after the final selection to obtain the $\mathcal{B}(\tau^- \to e^\mp \ell^\pm \ell'^-)$ measurements. I thus define at this stage the fit region, which is the range in $M_{e\ell\ell'}$ and $\Delta E_{e\ell\ell'}$ in which the fit will be performed.

4.3.3. Fit region definition

The first step is to construct the fit region from which the branching fraction will be inferred. While it would be more optimal to do this at the very end of the background rejection, this region must be kept completely independent of the BDT training. This region is therefore constructed as a slice of $\Delta E_{e\ell\ell'}$, whose width is chosen following the same procedure as for the blind box at the beginning of this section :

$$N_{sig}^{\delta_{\Delta E}^{low}} = 0.9 \cdot N_{sig}^{\Delta E, low} \tag{4.10}$$
$$N_{sig}^{\delta_{\Delta E}^{high}} = 0.9 \cdot N_{sig}^{\Delta E, high} \tag{4.11}$$
$$\tag{4.12}$$

in order to obtain a band covering 90% of the signal events passing through all selections presented until this point. The ISR mask from equation 4.9 is also applied for this region definition. This region is defined smaller than the blind box : this choice of 90% coverage ensures that the peak of the background at around 100 MeV is excluded while still maintaining a high signal efficiency.

Note that this definition is arbitrary, but it would be complicated to optimise (against some figure of merit, such as the Punzi[2]) the width of the band. If the MC

[1] i.e. one track in the ROE

[2] The Punzi figure of merit [111] is defined as $S/(\alpha/2 + \sqrt{B})$, where S is the signal yield (or the efficiency), B the background yield in the signal region and $\alpha = 3$.

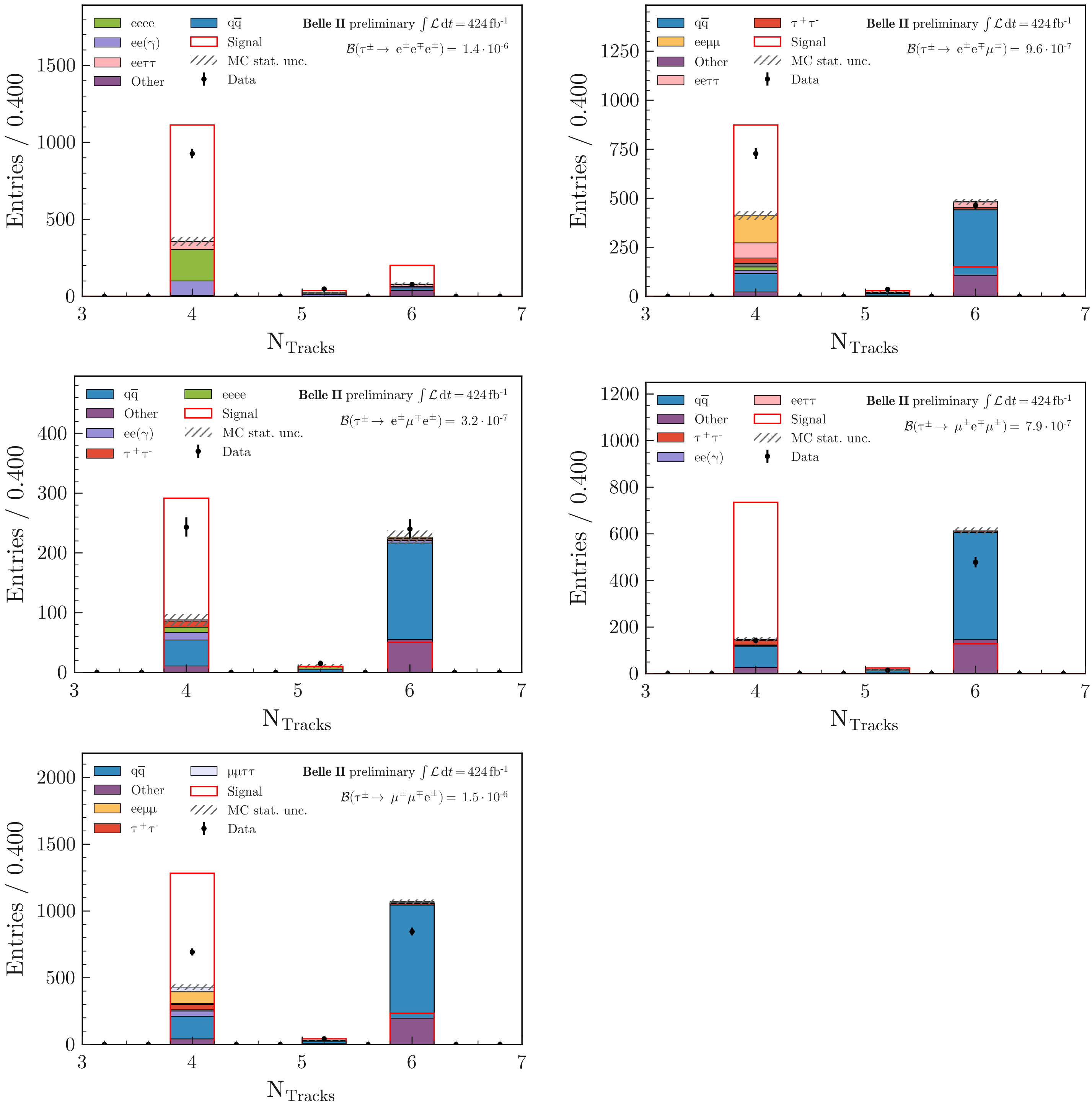


Figure 4.17.: Distribution of the number of tracks in the events after all selections.

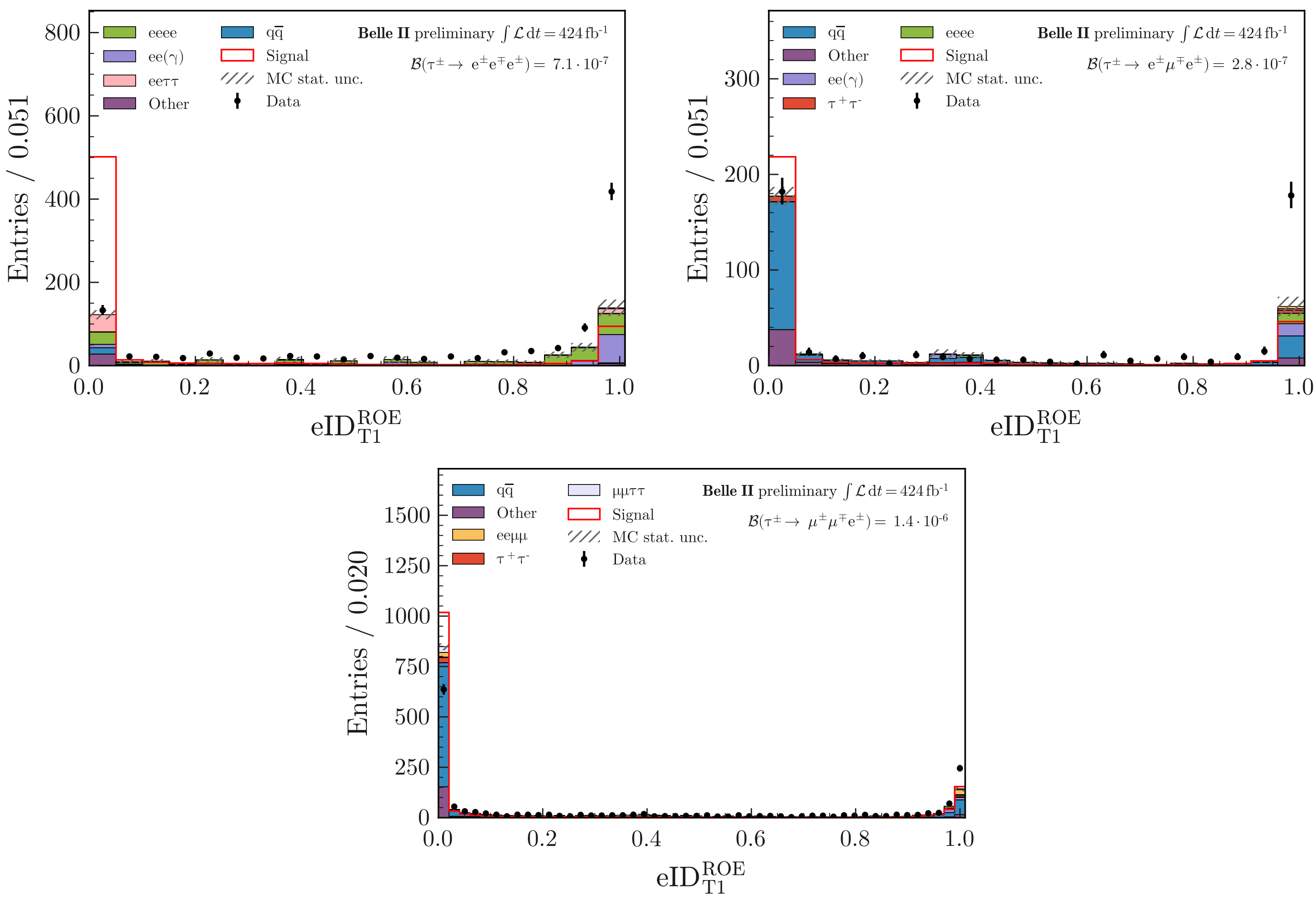


Figure 4.18.: Distribution of the electronID of the ROE side track for the $e^-e^+e^-$, $e^-\mu^+e^-$ and $\mu^-\mu^+e^-$ modes.

were reliable, the size of the fit region (or more generally, the region from which the measurement is made) would be optimised once the background suppression is completed.

Table 4.7.: Definition of the $\Delta E_{e\ell\ell'}$ interval for the fit band, with its relative (w.r.t the total 2D plane) and absolute signal efficiencies.

	$[\Delta E_{e\ell\ell'}^{fit,low}, \Delta E_{e\ell\ell'}^{fit,high}]$ [GeV]	ϵ_{sig}^{rel}	ϵ_{sig}^{abs}
$e^-e^+e^-$	[−0.34,0.04]	79.1%	22.8%
$e^-e^+\mu^-$	[−0.29,0.05]	78.6%	24.5%
$e^-\mu^+e^-$	[−0.30,0.03]	78.2%	24.2%
$\mu^-\mu^+e^-$	[−0.21,0.03]	77.0%	24.3%
$\mu^-e^+\mu^-$	[−0.21,0.03]	76.2%	24.4%

This entire region is called the fit region, while its sub-range outside the blind region is referred to as sidebands. The signal region $\Delta E_{e\ell\ell'}$ interval is given in Table 4.7, with its relative and total signal efficiency. The band can be seen in Figure 4.20, on the ($M_{e\ell\ell'}$,$\Delta E_{e\ell\ell'}$) plane.

4.3.4. Boosted Decision Tree classifier

The BDT selection aims at removing the remaining background events. It is trained on data and MC signal samples. The BDT used in this analysis is based on the `XGBoost` library [112], and the hyper-parameters optimisation is done using the `Optuna` library [113].

Data training sample The training sample contains 15000[1] signal events from simulation (and are not used anymore after the training for signal efficiency estimation), and background from data outside the blind box to avoid any potential bias, and outside the entire fit region.

Table 4.8.: Number of the data training events.

	N_{bkg}^{train}
$e^-e^+e^-$	768
$e^-e^+\mu^-$	991
$e^-\mu^+e^-$	452
$\mu^-\mu^+e^-$	1,471
$\mu^-e^+\mu^-$	625

The numbers of background events used for training are shown in Table 4.8 and range from a few hundred to a thousand. For the $e^-\mu^+e^-$ and $\mu^-e^+\mu^-$ modes the

[1]Much more signal events are available, but it would not improve the quality of the training to pour endlessly millions of signal events in the BDT.

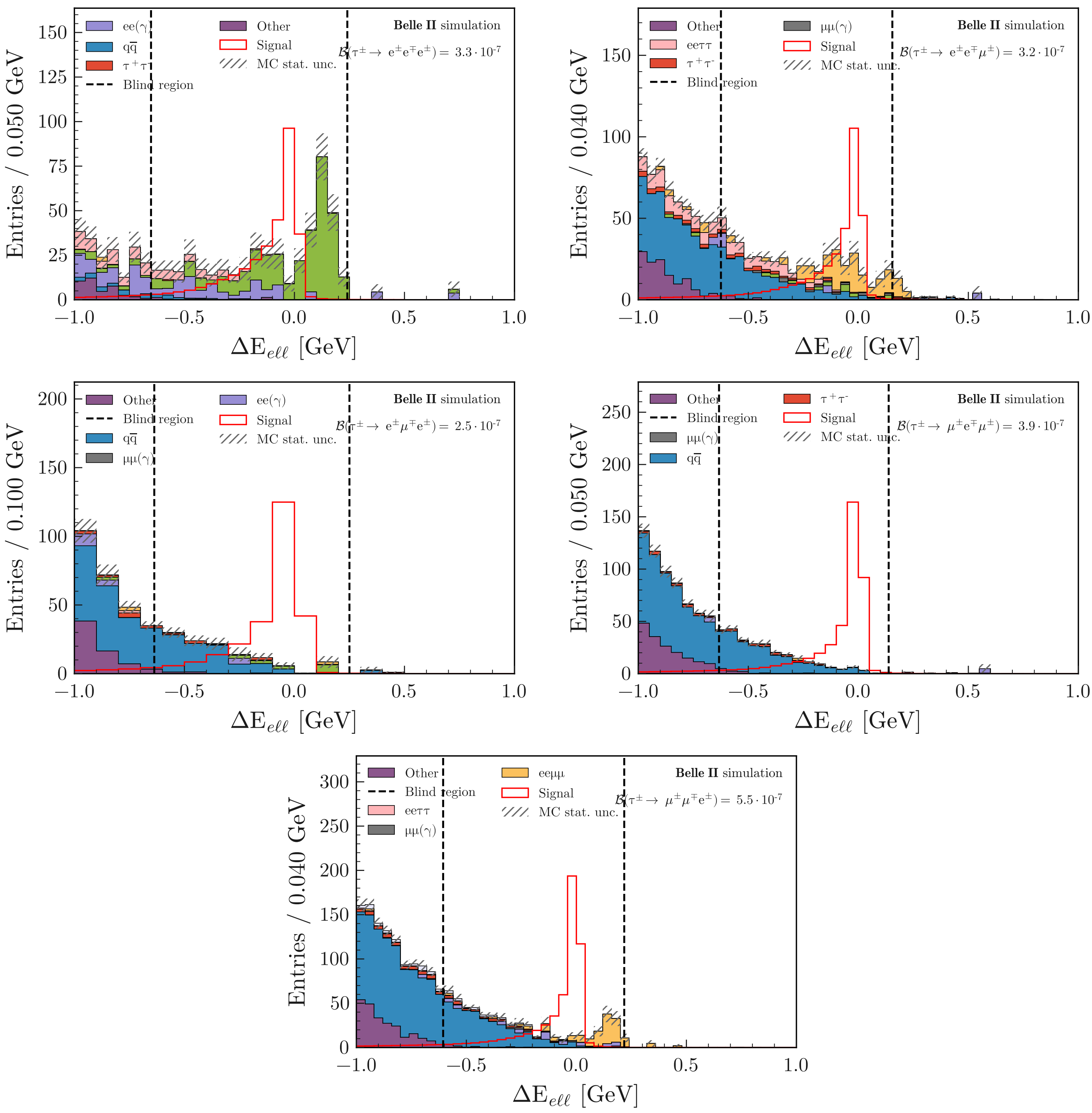


Figure 4.19.: Distributions of $\Delta E_{e\ell\ell'}$ in simulations (not blinded). The dashed lines correspond to the limits of the blind region.

training samples are quite small, with only a few hundred events (as these modes are the one with the smallest background); it proves to be enough for the BDT training as discussed in section 4.4 on the validation of the BDT.

This entire sample (background and signal) is then further split, where 80% is used for training, and 20% for validation, control over-training and optimisation of the hyper-parameters. For these samples, weights are applied so that the total weight of the backgrounds events is equal to the total weight of the signal events (considering the asymetry in the number of signal and background events, not doing this reweighting could lead to an important bias).

Discriminating variables for the BDT classifier The choice of variables used as the BDT inputs is made among a set of roughly 40 variables, removing the ones that show too much correlations between them (or to $M_{\ell\ell'}$) or that have low discriminating power. In case the correlation between two variables was significantly different for signal and background, both variables were kept. The correlation threshold was set at 85%, while the relative difference in correlation for signal and background to keep pairs was set at 7.5%, these two values being arbitrary. The correlation matrices for the $e^-e^+e^-$ mode are shown in Figure 4.21 and Figure 4.22. The correlation matrices for the other modes can be found in Appendix A.3.3. The variables used can be split into three categories:

- The first category consists of variables associated with the signal side, such as the lepton's polar angles and ordered energies, the invariant mass of the three lepton-pair combinations, the flight time of the τ divided by its uncertainty, the τ azimutal angle and the cosine of its polar angle.
- The second category involves variables related to the ROE properties such as its mass, its energy, the total energy of the clusters in the ROE and the number of neutral clusters in the ROE. The ROE mode, defined as n_e^{ROE} $+2\cdot n_\mu^{ROE}$ $+3\cdot n_\pi^{ROE}$ where $n_{e,\mu,\pi}^{ROE}$ are the number of electrons, muons and pions in the ROE, is also included.
- The third category of input variables comprises event based variables : the thrust value, the cosine of the polar angle between the τ momentum and the thrust axis, the total visible energy in the event, the numbers of photons and the total photon energy in the event, and variables related to the missing momentum of the event: its magnitude and polar angle in the c.m. frame, the angle between the missing momentum and each of the three leptons, and the invariant missing mass squared. Lastly, variables related to the event shape as the reduced Fox Wolfram moment R2 and the first three CLEO cones [114] are used.

The precise list of variables for each modes is given in Table 4.9. The BDT importance of each input variable is shown in Appendix A.3.1. Comparisons of

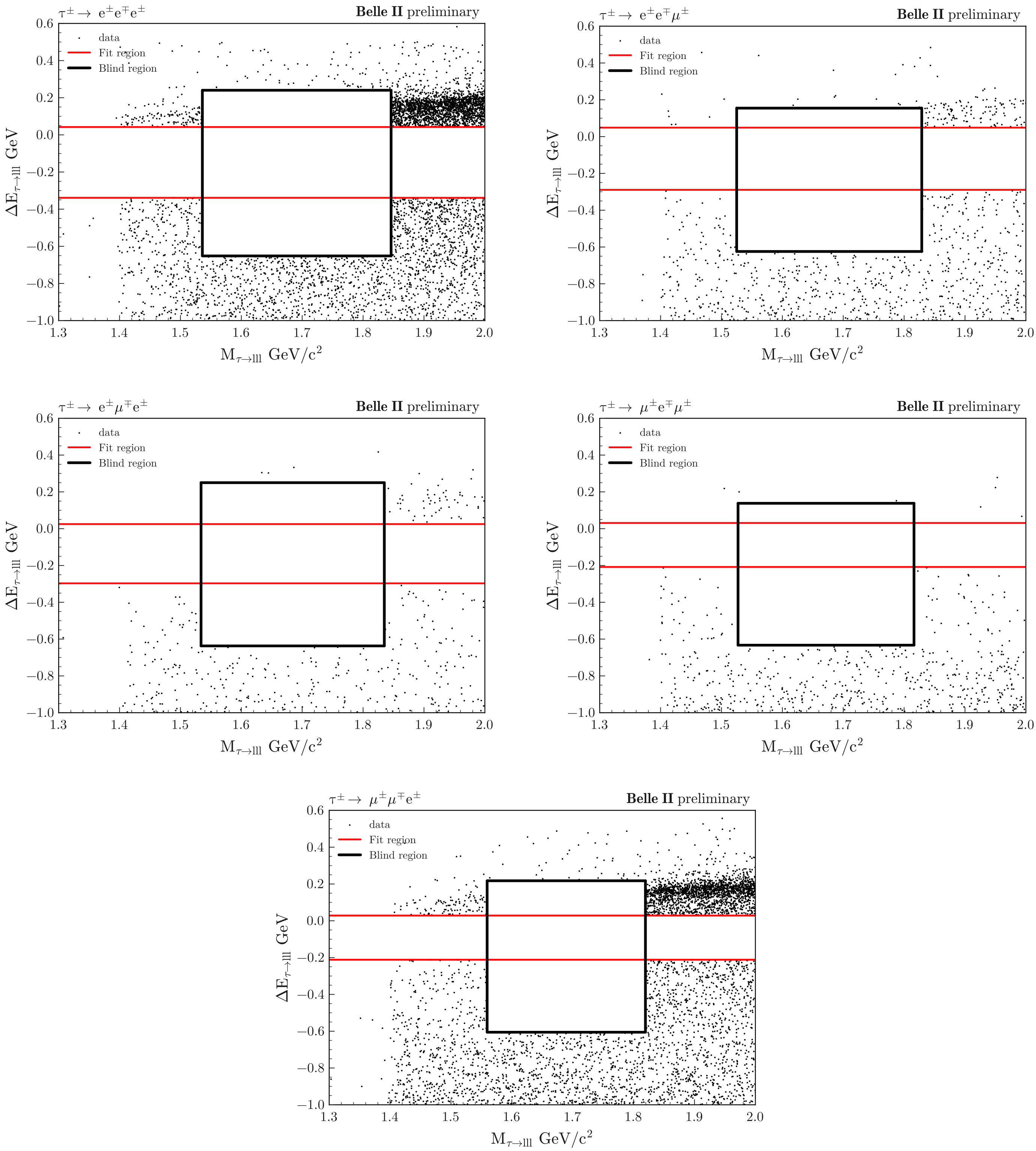


Figure 4.20.: Distribution of the entire training samples (before validation splitting) in the ($M_{e\ell\ell'}$,$\Delta E_{e\ell\ell'}$) plane. The red lines delimit the fit region, and the black lines delimit the blinded region. The sub-range of the fit region outside the blinded region corresponds to the sidebands.

signal and background distribution of the training sample for each input variable can be found in Appendix A.3.2.

Table 4.9.: List of all the signal/background discriminating variables used as input for BDT classifier algorithm.

Description	Variables	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
Angle between the leptons and the missing momentum	θ_{l1}^{miss}	✓	✓	—	✓	✓
	θ_{l2}^{miss}	✓	✓	—	—	✓
	θ_{l3}^{miss}	✓	✓	✓	✓	—
Polar angle of the leptons	θ_{l1}	—	✓	—	✓	—
	θ_{l2}	—	✓	✓	✓	—
	θ_{l3}	—	✓	✓	✓	—
Ranked energy of the leptons	E_{lead}	✓	✓	✓	✓	✓
	E_{sub}	✓	—	—	—	✓
	E_{third}	✓	✓	✓	✓	✓
Invariant mass of pair of leptons	$\mathrm{M}_{inv}^{1,2}$	✓	✓	✓	✓	✓
	$\mathrm{M}_{inv}^{1,3}$	✓	—	—	—	✓
	$\mathrm{M}_{inv}^{2,3}$	✓	✓	—	—	✓
Mass and energy of the ROE	M_{ROE}	✓	✓	✓	✓	✓
	E_{ROE}	✓	✓	✓	✓	✓
Total energy of the clusters in the ROE	E_{ROE}^{extra}	✓	✓	✓	✓	—
Total energy of the neutral clusters in the ROE	N_{ROE}^{cls}	✓	—	—	—	✓
"Tag" side mode	tag_mode	✓	✓	✓	✓	✓

Table 4.9.: List of all the signal/background discriminating variables used as input for BDT classifier algorithm.

Description	Variables	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
τ flight time over its error	$d^{\tau-flight}/d^{\tau-flight}_{err}$	✓	✓	✓	✓	✓
τ azimuthal angle	dϕ_{tau}	✓	✓	_	✓	✓
τ cosine of polar angle	dcosθ_τ	✓	✓	✓	✓	✓
cosine of polar angle of τ to thrust	cosθ^{thrust}_τ	✓	✓	✓	✓	✓
Thrust of the event	Thrust	✓	✓	✓	✓	✓
Reduced Fox-Wolfram R2	R2	_	✓	_	✓	_
Total photon energy of the event	E^{tot}_γ	_	_	✓	✓	_
Total number of photons in the event	Nγ	_	✓	_	✓	✓
First three CLEO cones	cleoConeThrust 0	_	✓	_	✓	_
	cleoConeThrust 1	✓	✓	✓	_	✓
	cleoConeThrust 2	✓	✓	_	_	✓
Polar angle of the missing momentum in the c.m	θ^{CMS}_{miss}	✓	✓	✓	✓	✓
Visible energy of the event	E_{vis}	✓	_	_	_	✓
Momentum and mass squared of the missing energy	P_{miss}	✓	✓	✓	✓	✓
	M^2_{miss}	✓	✓	✓	✓	✓

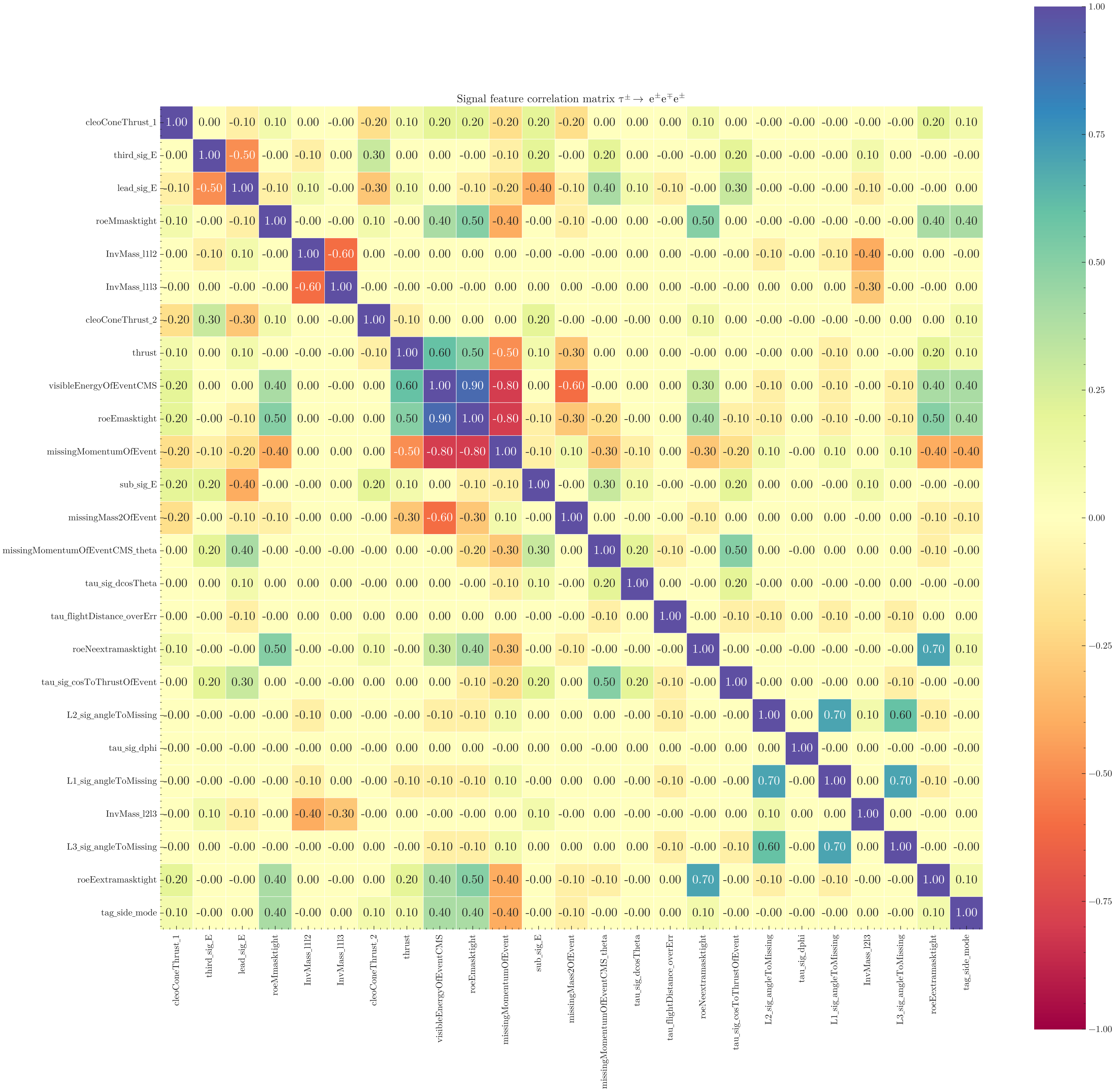


Figure 4.21.: Correlation matrix for the signal, $e^-e^+e^-$ mode

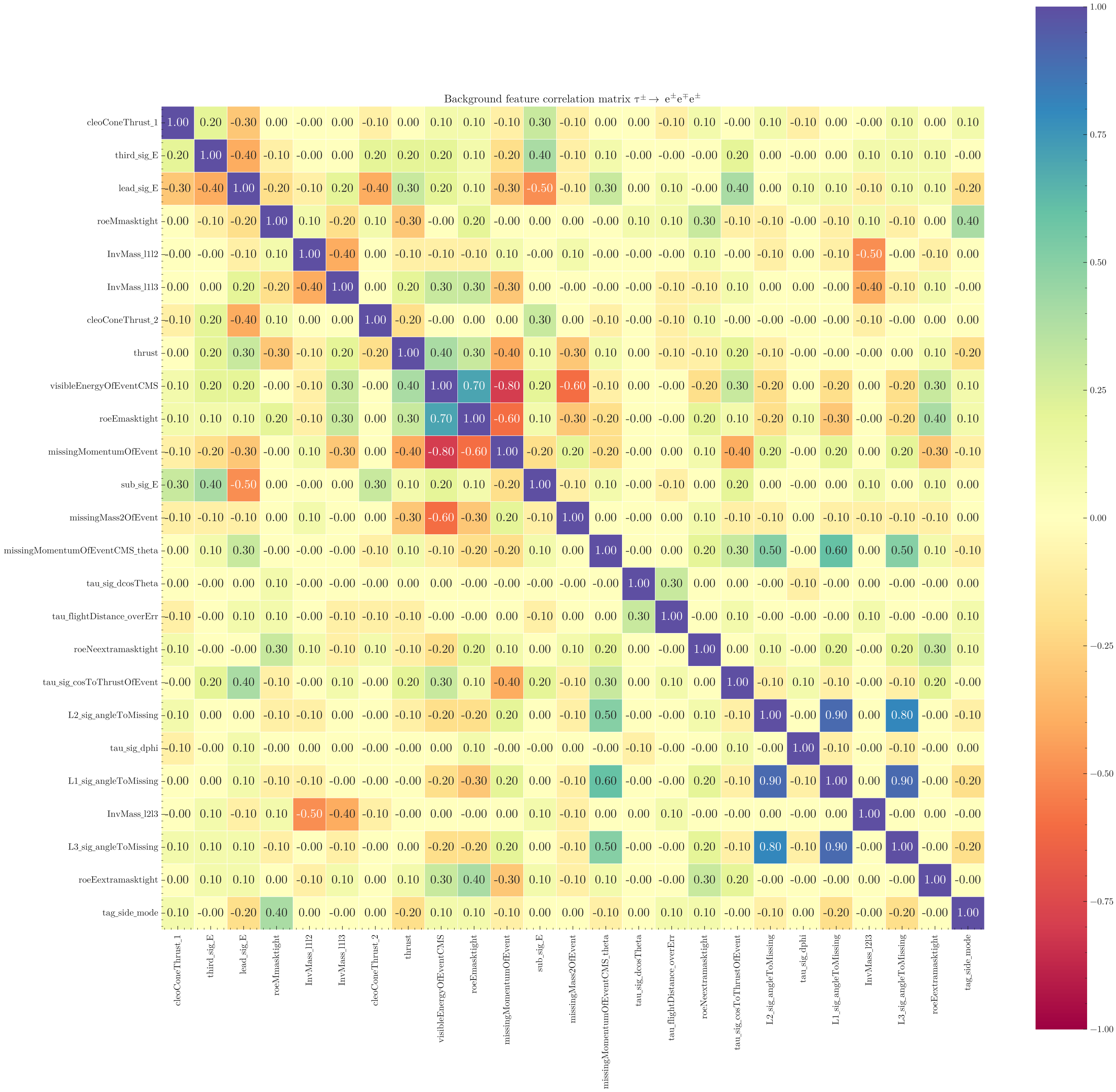


Figure 4.22.: Correlation matrix for the background, $e^{-}e^{+}e^{-}$ mode

4.3.4.1. BDT performances

Figure 4.23 shows the score distribution in the sidebands for data and simulation for the best iteration of the BDT, after training and optimisation (described in section 4.4). A cut is applied on the score ; its value is chosen in order to minimize the expected upper limit, as described in section 4.7. Table 4.10 shows the nominal cut on the BDT score, the absolute signal efficiency, as well as the number of remaining events in the data sidebands.

Table 4.10.: Nominal cut on the BDT output for each modes, with the corresponding total signal efficiencies and number of remaining background events in the sidebands.

	s^{BDT}	N_{SB}	ϵ^{tot}_{sig}
$e^-e^+e^-$	0.95	3	15.0%
$e^-e^+\mu^-$	0.8	6	20.4%
$e^-\mu^+e^-$	0.5	6	23.5%
$\mu^-\mu^+e^-$	0.8	12	20.1%
$\mu^-e^+\mu^-$	0.5	4	24.1%

4.4. Selection validation

In this section, the validation procedure for the BDT is presented, through its optimisation and various checks.

BDT parameters optimisation Several hyperparameters (HP) are optimised to provide the best separation and avoid overfitting and overtraining. These parameters and their optimal values are given in Table 4.11. The python library `Optuna` [113] is used to scan and find the best configuration in the hyperparameter space in a much faster way than a regular grid search. The algorithm samples hyperparameters using a Tree-structured Parzen Estimator [115], and try to maximize the so-called objective function, F :

$$F = \frac{AUC_{val}}{L^{val}_{log}} \cdot (\frac{AUC_{val}}{AUC_{train}})^2 \cdot \frac{L^{train}_{log}}{L^{val}_{log}} \tag{4.13}$$

where AUC_{val} and AUC_{train} respectively are the area under the curve (AUC) of the receiver operating characteristic curve (ROC) of the BDT for the validation and training sample, and L_{val} and L_{train} are the logarithmic loss (logloss) functions evaluated on the validation and training samples. The logloss function is computed

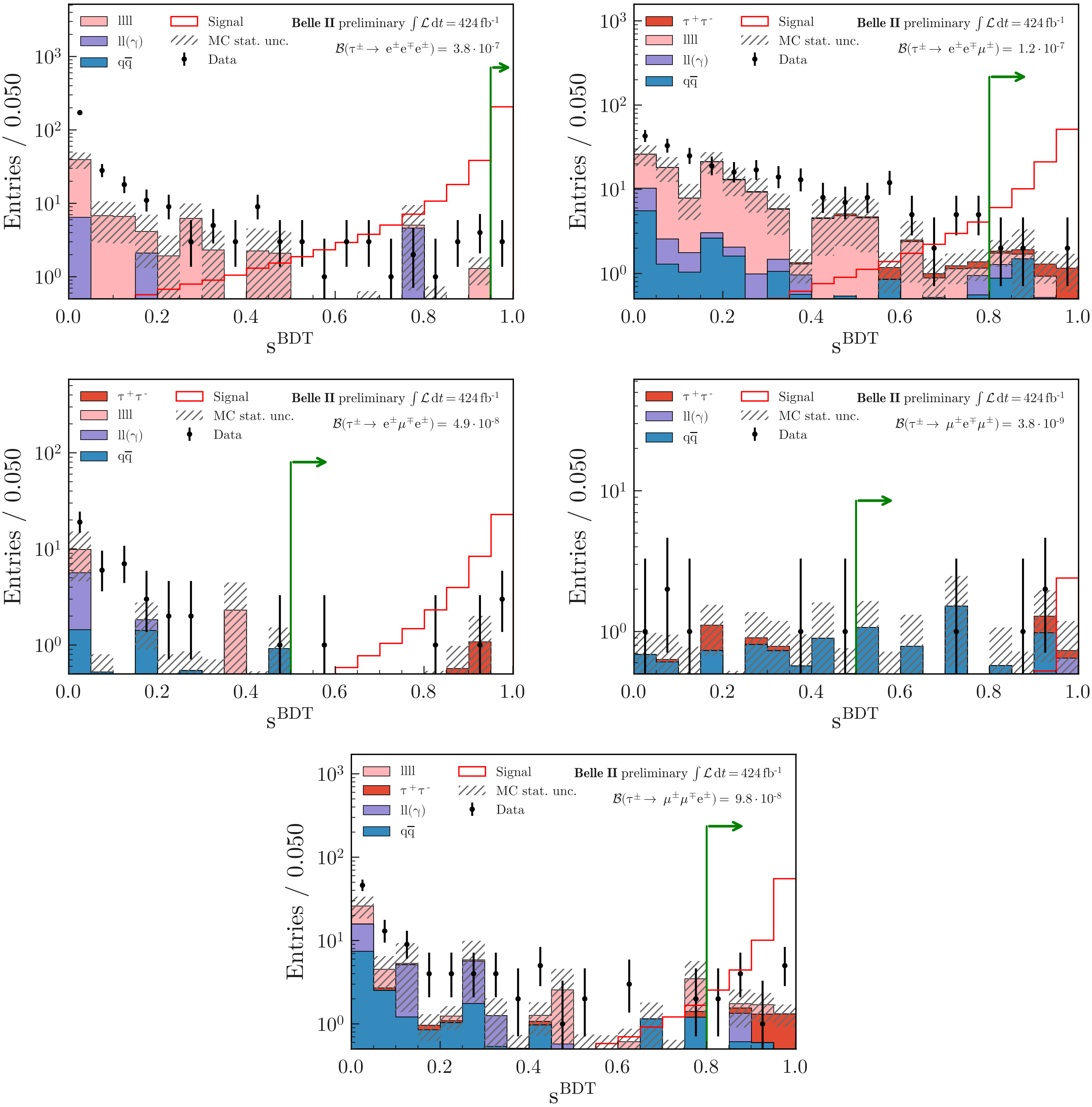


Figure 4.23.: BDT score distribution in the data sidebands. Background simulations are also in the sidebands while the simulated signal is in the entire $(M_{e\ell\ell'}, \Delta E_{e\ell\ell'})$ plane.

as an average over the N_s events of a sample :

$$L_{log}(\vec{y}, \vec{s}^{BDT}) = -\frac{1}{N_s}\sum_{i}^{N_s}(y_i \times log(s_i^{BDT}) + (1 - y_i) \times (log(1 - s_i^{BDT}))), \quad (4.14)$$

where $\vec{y}$ and $\vec{s}^{BDT}$ are respectively the class (0 for background and 1 for signal) and the BDT score (in $[0, 1]$) of each event.

The ratio terms $\frac{AUC_{val}}{AUC_{train}}$ and $\frac{L_{log}^{train}}{L_{log}^{val}}$ in Eq. 4.13 prevent overtraining, while the additional term $\frac{AUC_{val}}{L_{log}^{val}}$ will favour good performances and generalisation power of the model.

Table 4.11.: Definition of the `XGBoost` BDT classifier hyperparameters. The values correspond to the output of the optimisation process.

Name	Definition	Value				
		$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
`n_estimators`	Number of boosting rounds.	190	30	130	140	220
`max_depths`	Maximum depth of a tree, number of intermediate nodes.	2	2	1	1	1
`min_child_weight`	Minimum sum of instance weight needed in a child node.	3	5	1	0	9
`learning_rate`	Step size shrinkage used after each boosting update.	0.313	0.378	0.234	0.454	0.08
`gamma`	Minimum loss reduction required to make a further partition on a leaf node of the tree	9	5	5	10	5
`alpha`	L1 regularization term on weights	3	8	3	6	9
`lambda`	L2 regularization term on weights	6	6	5	10	9

Figure 4.24 shows the logloss for the training and validation samples. There indeed is no relevant discrepancy between the two samples, the overall slightly better performances for the training sample not being of any danger. The validation logloss is used to stop the training when no further progress is made. The ROC curves show the same slight discrepancy and can also be found in figure 4.24; their integral is close to 1, which showcase the strong separation power of the BDT. It is important to note that in particular for the $e^-e^+e^-$, $e^-e^+\mu^-$ and $e^-\mu^+e^-$ modes, the validation metric is poorer than the training one : it is understood as a

consequence of the low training background statistic. A higher number of events of the training would probably improve this, but it is not possible to add more relevant background to the training. This is anyway not considered to be an issue : the training metrics show no worrying signs and the final background suppression is very good.

4.4.0.1. Check of correlation

Since the branching fraction is extracted from a fit to $M_{e\ell\ell'}$, it is important to make sure that the BDT score does not present any dependency as a function of $M_{e\ell\ell'}$, and in particular that there is no peaking structure near the known τ mass value (i.e. around 1.777 GeV/c^2). The profile plots (the average of the score in bins of $M_{e\ell\ell'}$) for data sidebands on Figure 4.25 show that there indeed is no such problematic sculpting from the BDT. Profile plots for MC within the fit region are also included ; while the background present in MC might not be entirely representative of the actual data background, it allows to control the BDT output in the relevant region in $\Delta E_{e\ell\ell'}$ and $M_{e\ell\ell'}$.

4.4.0.2. Partial unblinding

In order to further verify that the BDT is correctly rejecting the background and not sculpting events near the τ mass, a partial unblinding is performed in the upper section of the blind region, above the fit region. This region is chosen because it corresponds to the peak of the four-lepton background events while also having a low signal efficiency. The distribution of these events in the $(M_{e\ell\ell'}, \Delta E_{e\ell\ell'})$ plane is shown in figure 4.26.

Table 4.12 shows the number of events before and after the BDT selection in the partially unblinded region, with all the other selection being applied. The number of surviving events is extremely low, displaying one more time that the BDT efficiently removes background events. The score distribution in this unblinded region is also shown in figure 4.27.

	N_{before}	N_{after}
$e^-e^+e^-$	177	1
$e^-e^+\mu^-$	28	0
$e^-\mu^+e^-$	35	1
$\mu^-e^+\mu^-$	1	1
$\mu^-\mu^+e^-$	79	4

Table 4.12.: Number of events before and after the BDT from the data partially unblinded.

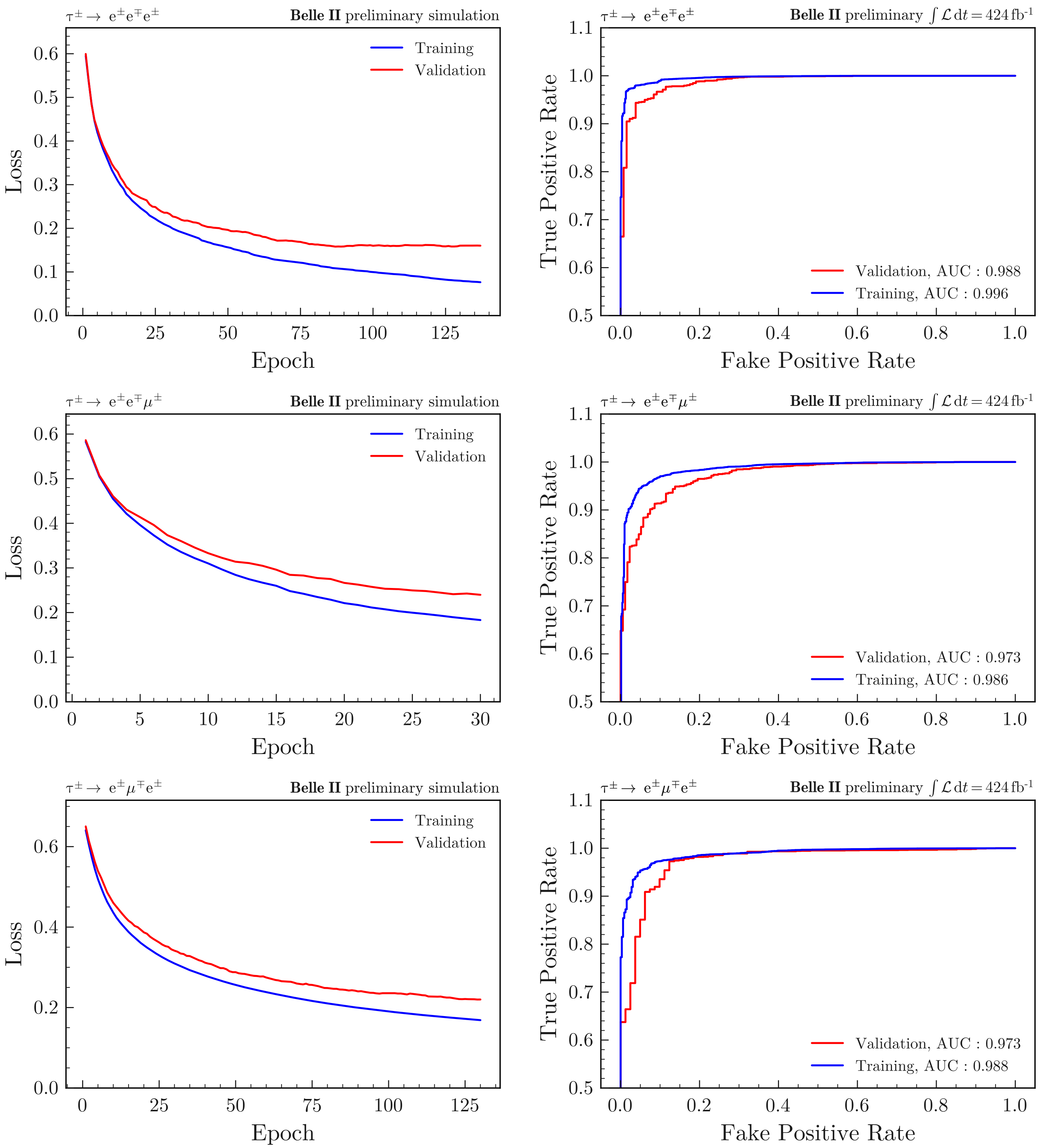


Figure 4.24.: Left : training and validation logloss as a function of the number of estimators. Right : training and validation ROC curves.

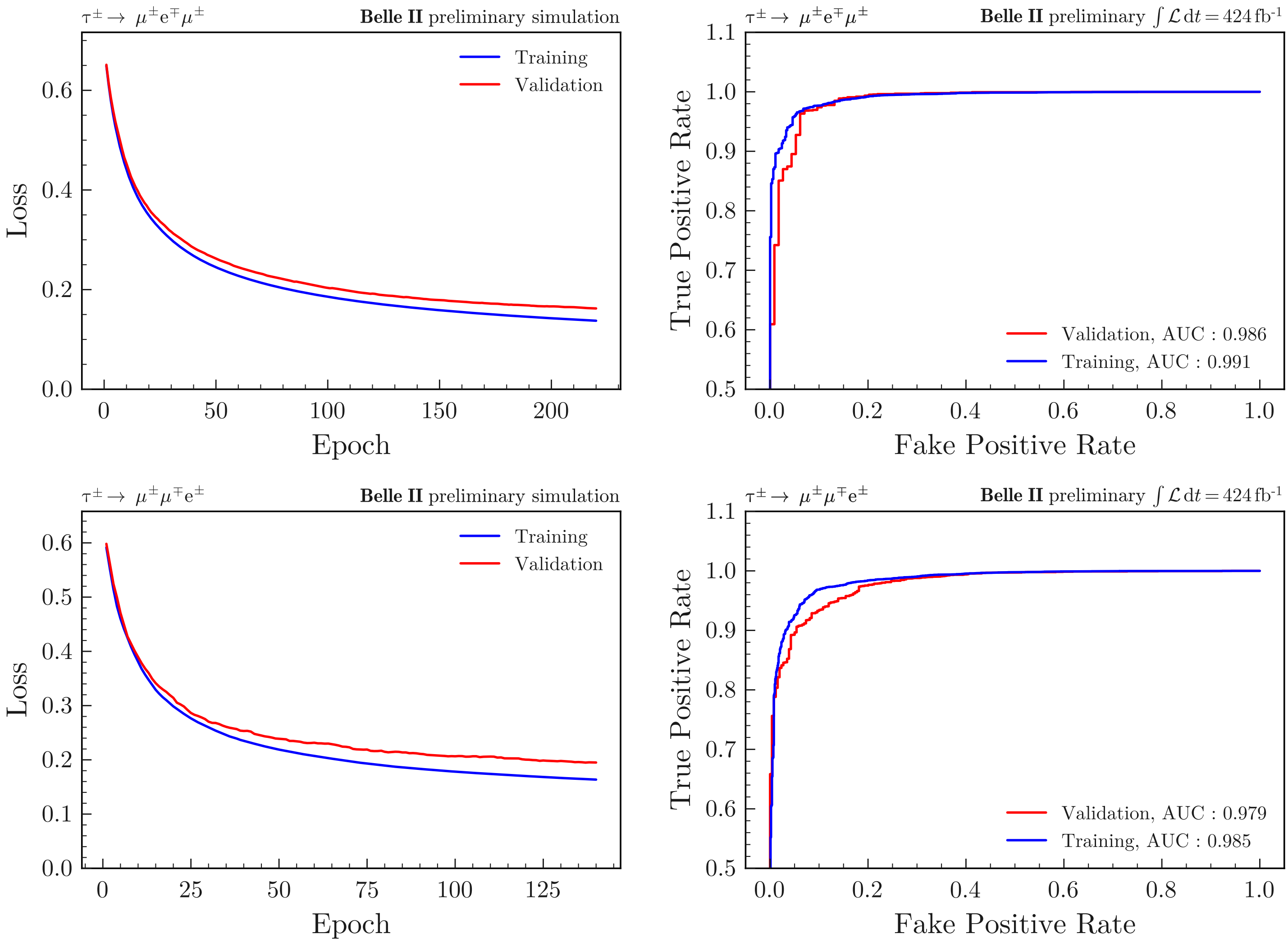


Figure 4.24.: Left : training and validation logloss as a function of the number of estimators. Right : training and validation ROC curves

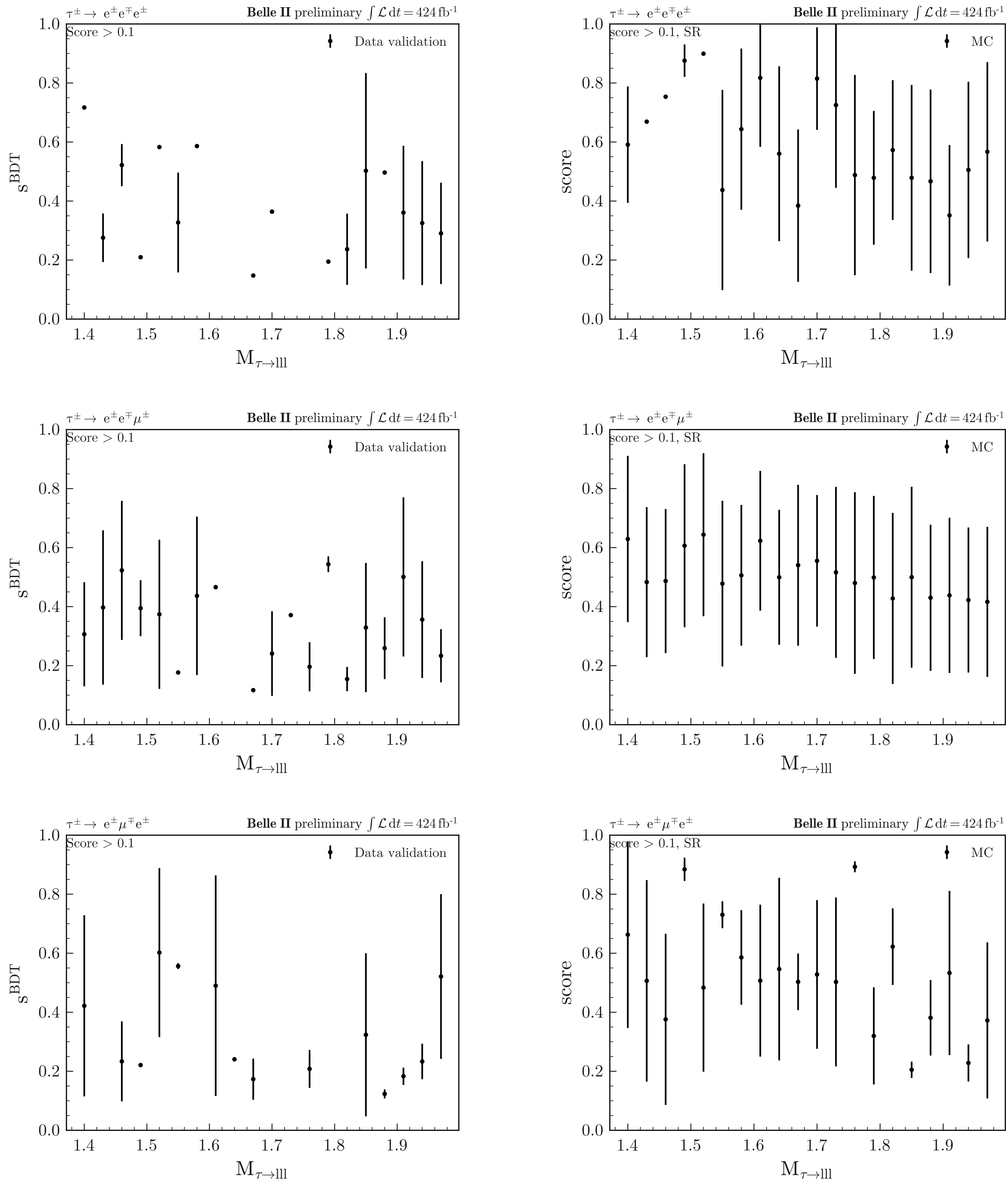


Figure 4.25.: Profile plots with the average of the BDT score in bins of $M_{e\ell\ell'}$. Left : data validation. Right : background simulations in the fit region. In both cases, a cut on the BDT score at 0.1 is applied. The error bars correspond to the RMS of the BDT score distribution. Bins with no errors have only one entry.

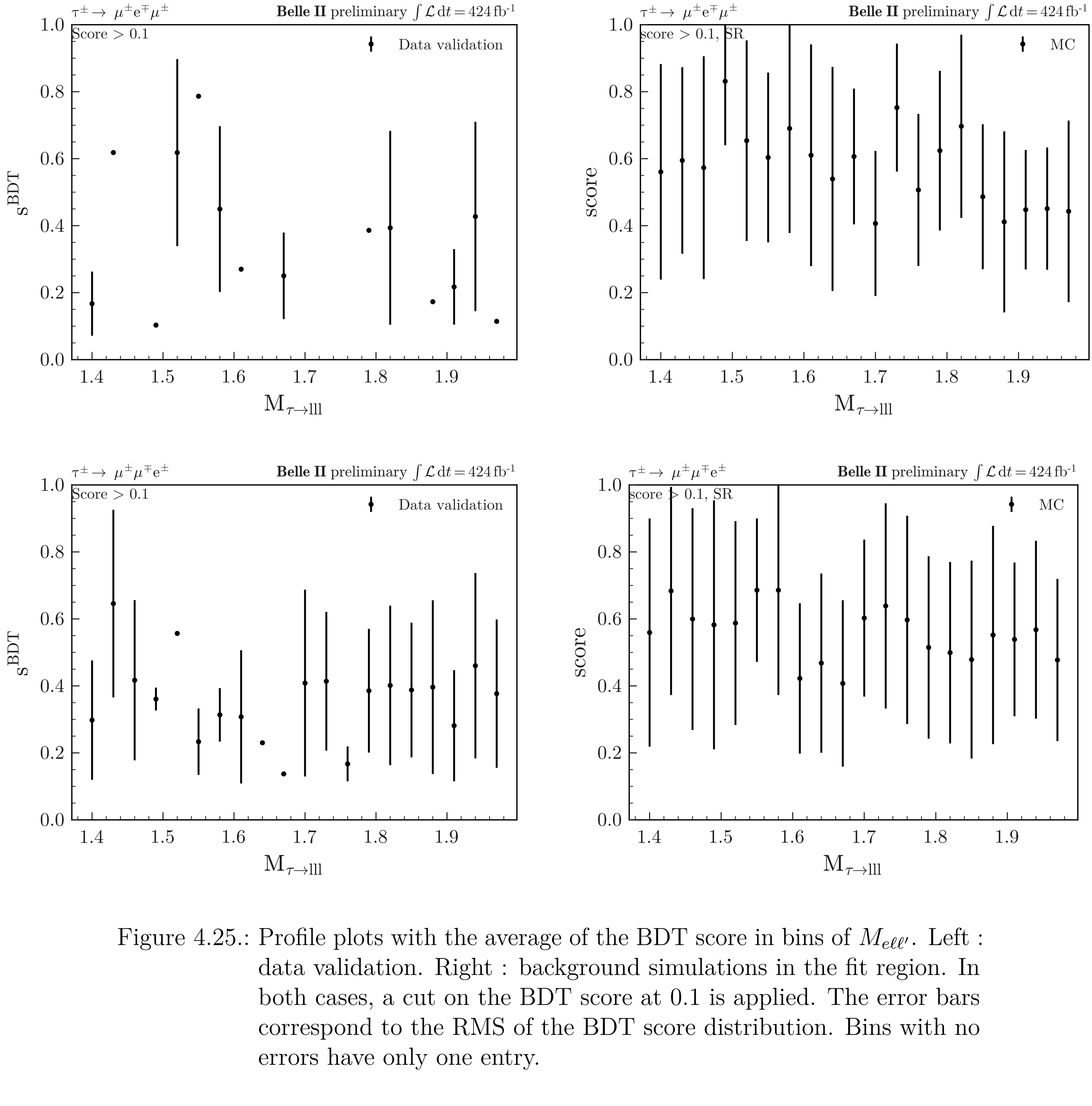


Figure 4.25.: Profile plots with the average of the BDT score in bins of $M_{e\ell\ell'}$. Left : data validation. Right : background simulations in the fit region. In both cases, a cut on the BDT score at 0.1 is applied. The error bars correspond to the RMS of the BDT score distribution. Bins with no errors have only one entry.

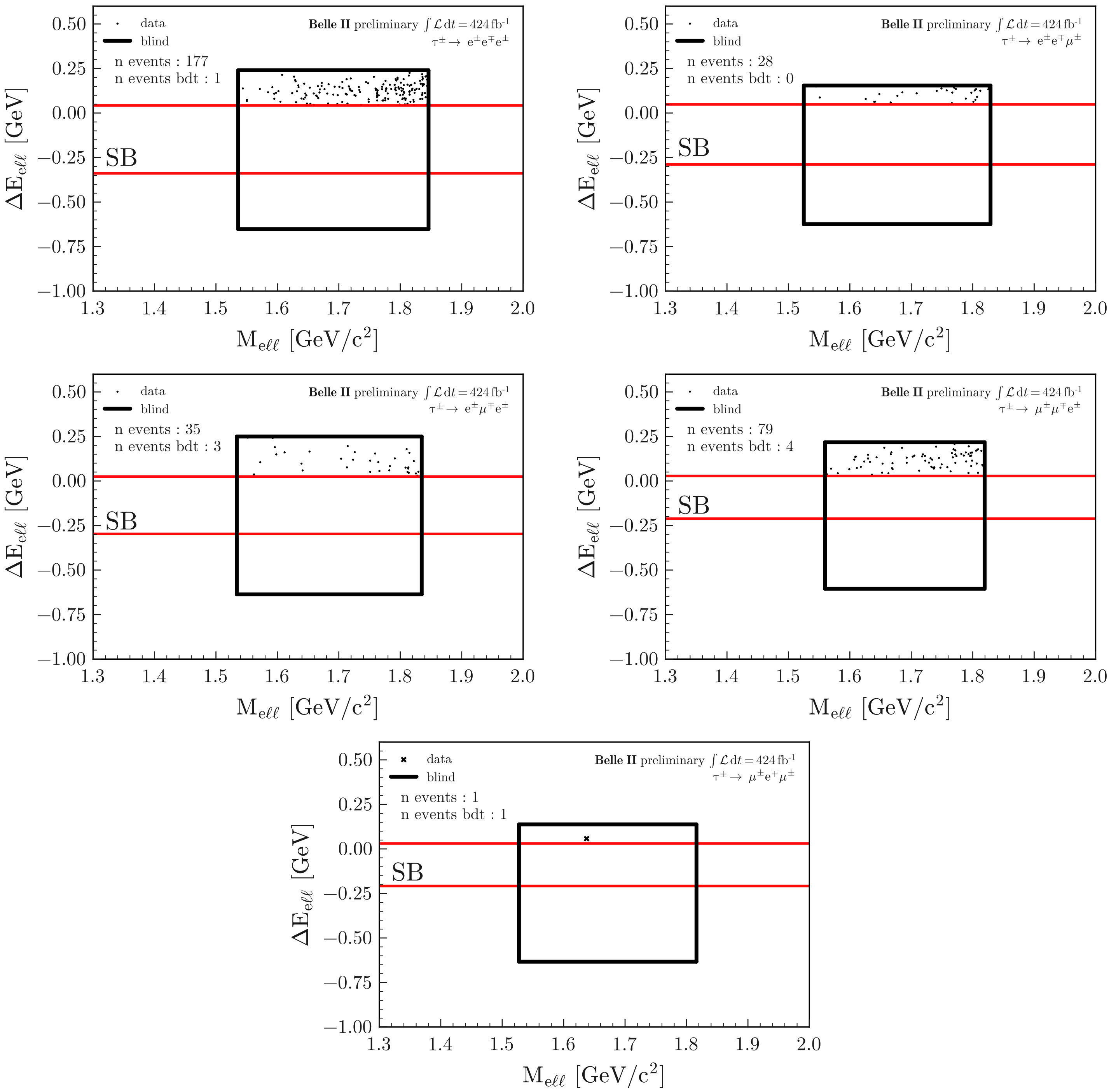


Figure 4.26.: Distribution of the events in the $(M_{e\ell\ell'}, \Delta E_{e\ell\ell'})$ plane before the BDT selection, in the partially unblinded dataset.

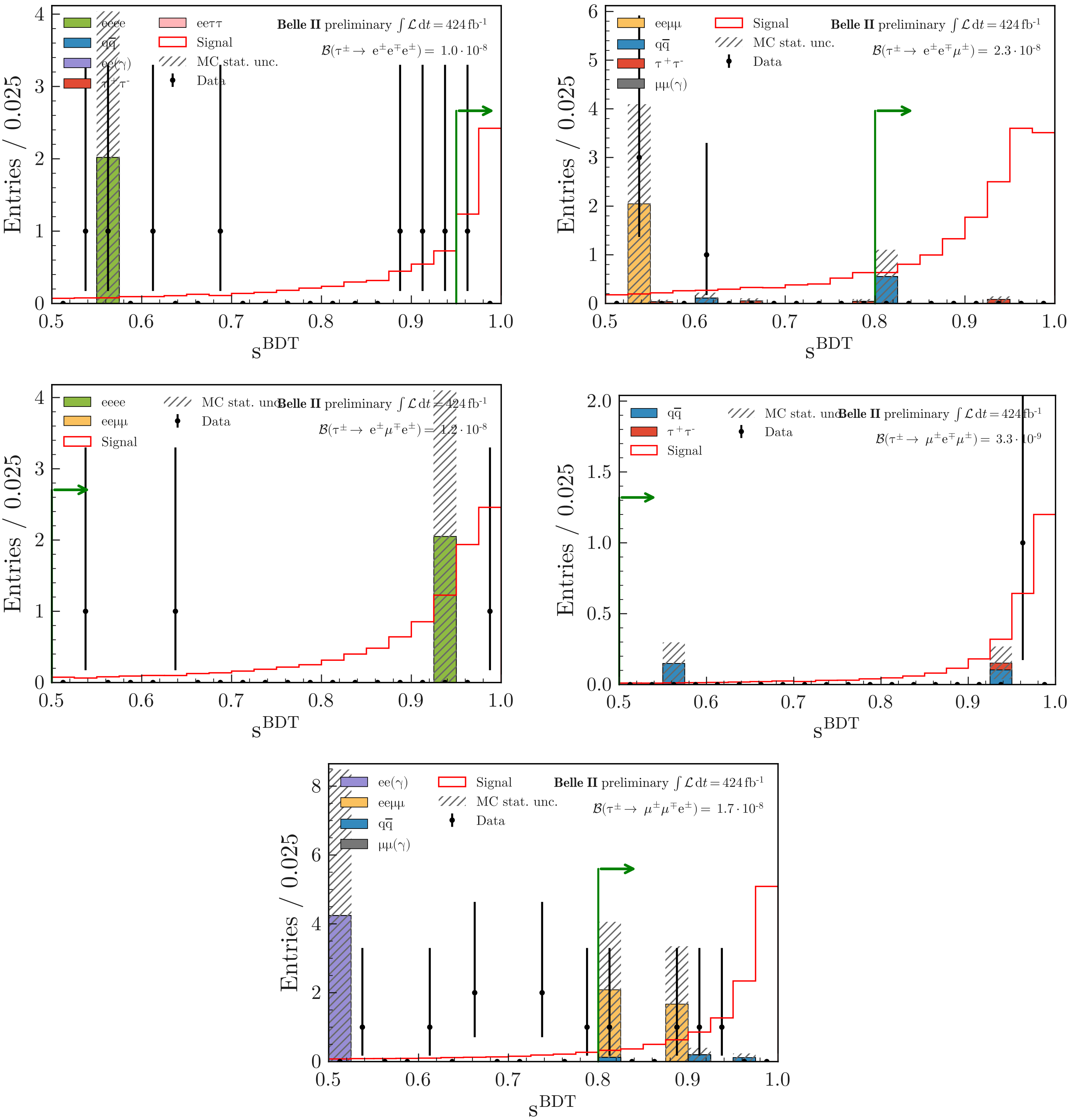


Figure 4.27.: BDT score in the unblinded sub-region.

To add another layer of verification, profile plots in $M_{e\ell\ell'}$ with these events are also shown in figure 4.28 : no sign of peak near the τ mass (or any other worrying structure) is observed.

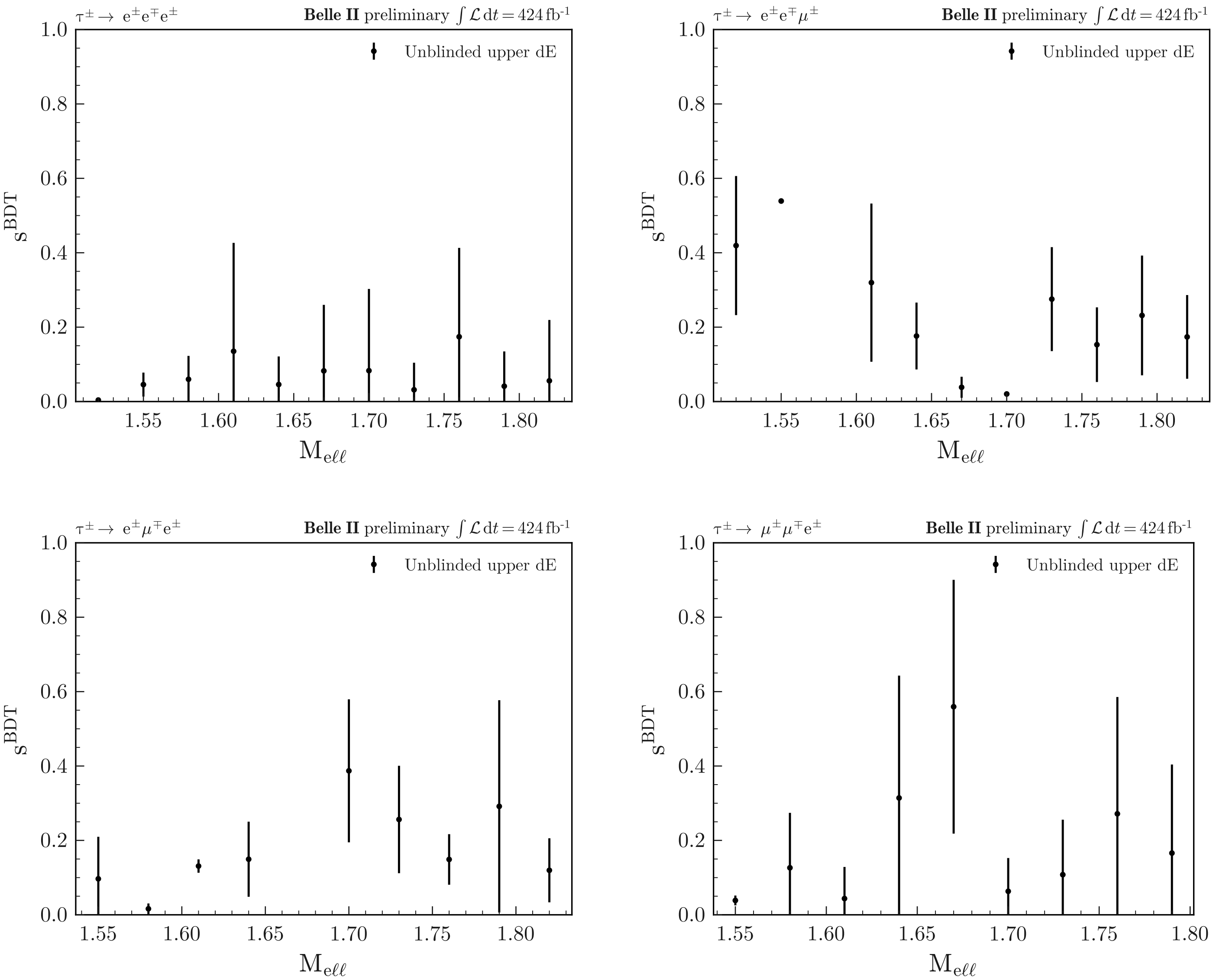


Figure 4.28.: Profile plots in the unblinded region. The $\mu^{-}e^{+}\mu^{-}$ mode is not included as only one event was observed before the BDT selection.

Nb of photons from $\pi^0 < 7$
Nb of photons in ROE < 8
$0.89 < thrust < 0.97$
$17.2° < \theta_{miss}^{CMS} < 154.7°$
0.5 GeV/c$^2 < M_{3\pi} < 1.7$ GeV/c^2
-1GeV $< \Delta E_{3\pi} < 0$ GeV
$\Delta E_{ROE} <$ -0.5 GeV
$E_{vis}^{CMS} <8$
$M_{ROE} <3$ GeV/c^2
$E_{ROE} < 5$ GeV

Table 4.13.: Additional requirements for background suppression in the $\tau \to 3\pi\nu_\tau$ control channel

4.4.0.3. Control channel

The $\tau \to 3\pi\nu_\tau$ decay is used as a control channel[1], to validate the agreement between data and simulation. This control channel is reconstructed using 42.25 fb^{-1} of data and 1 ab^{-1} of simulation samples. The control channel is reconstructed similarly to $\tau^- \to e^\mp\ell^\pm\ell'^-$, as described in Section 4.2. Three charged pions from a τ^- replace the combination of three leptons. To ensure orthogonal reconstruction and avoid accidental unblinding, a veto on the 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. Further selections are applied to increase the purity of the sample and are listed in Table 4.13. After the selection, there are 114535 events in data and 112568^{+258}_{-258} in simulation, showing a good agreement. The uncertainties are statistical only and the simulation yields is rescaled to match the data luminosity.

The comparison of the BDT score (trained on each mode) between data and simulation of the control sample is shown on Figures 4.29 and 4.30. The agreement is not perfect, although the slight discrepancy will be accounted for as a systematic uncertainty (see section 4.6). The comparison of the BDT input variables for $\tau \to \pi\pi\pi\nu$ between data and MC shows that all variables are properly modelled (and therefore are "safe" variables to use in the BDT). These distributions are found in Appendix A.1.

4.5. Fitting procedure

In this section, I present the fitting procedure, used to extract the branching fraction. The branching fraction is directly fitted from data in the signal region

[1] Obviously a three hadrons final state is not ideal to control a three leptons final state, but this is the dominant decay of a τ into three charged particles

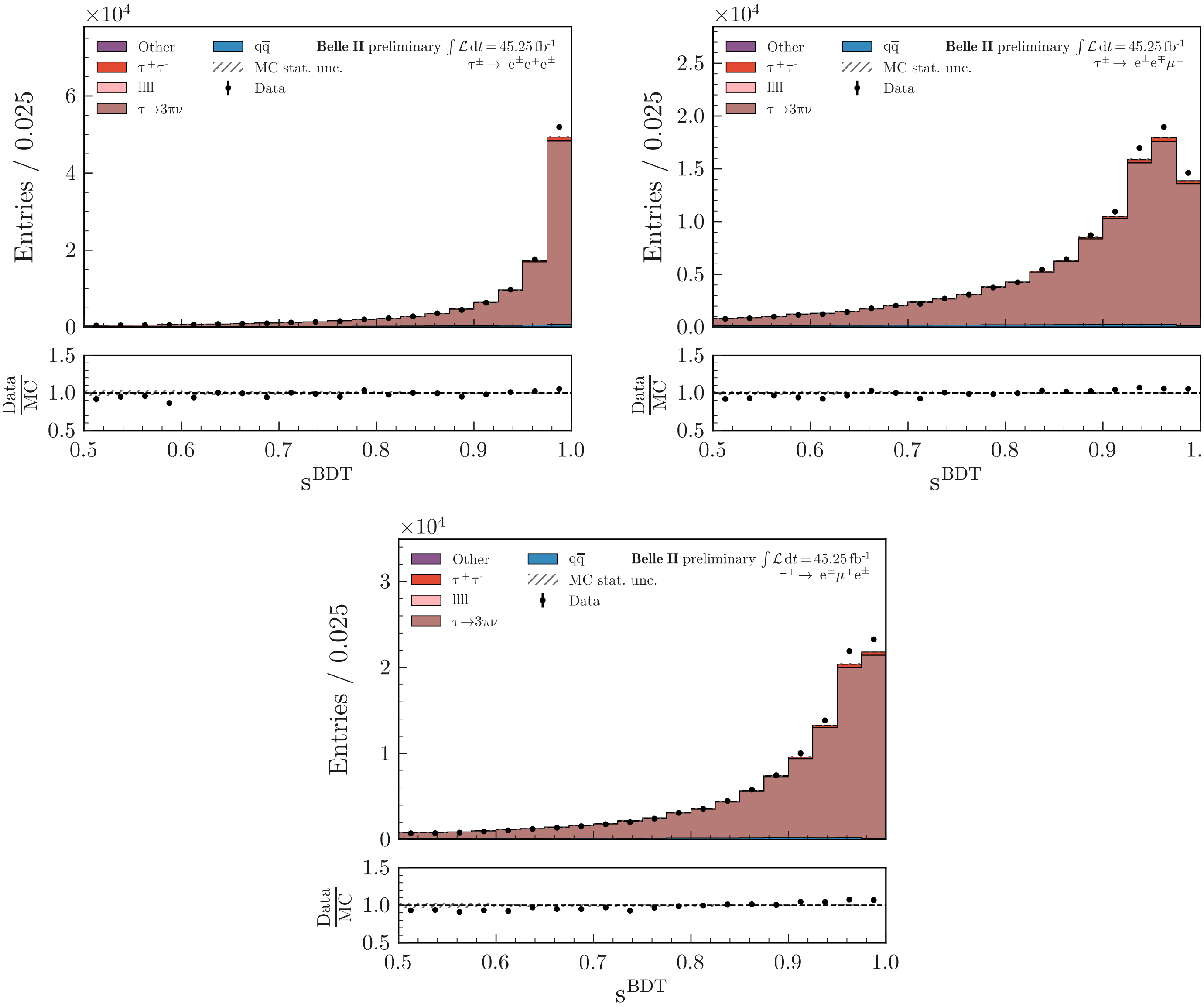


Figure 4.29.: Score of the BDT trained on the different $\tau \to e\ell\ell$ signal decays for the $\tau \to 3\pi\nu_\tau$ control sample.

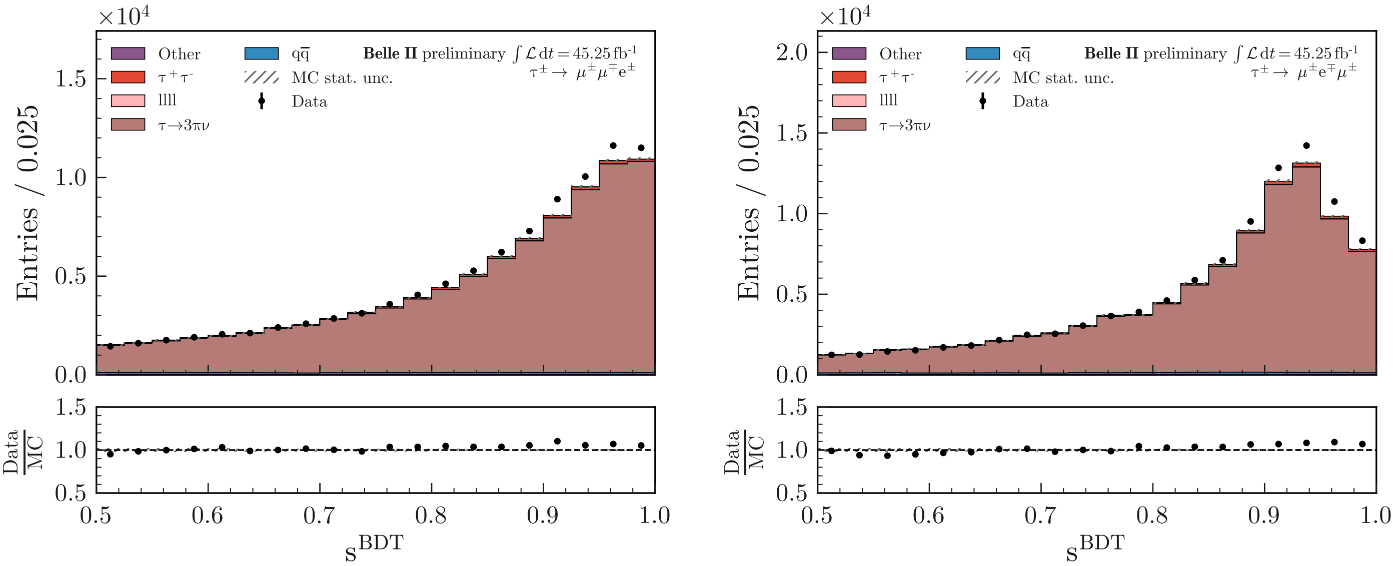


Figure 4.30.: Score of the BDT trained on the different $\tau \to e\ell\ell$ signal decays for the $\tau \to 3\pi\nu_\tau$ control sample.

using an unbinned maximum likelihood fit with the following probability density function (p.d.f) :

$$PDF_{tot}(M_{e\ell\ell'}) = \mathcal{B}(\tau^- \to e^{\mp}\ell^{\pm}\ell'^{-}) \cdot \mu \cdot PDF_{sig}(M_{e\ell\ell'}) + N_{bg} \cdot PDF_{bg}(M_{e\ell\ell'}) \quad (4.15)$$

with $\mu = 2 \cdot \epsilon_{sig} \cdot \sigma_{\tau\tau} \cdot \mathcal{L}$, where ϵ_{sig} is the signal reconstruction efficiency, $\mathcal{L}$ the luminosity of the Run 1 dataset, and $\sigma_{\tau\tau}$ the tau-pair production cross section. The signal and background are described below.

4.5.0.1. Signal p.d.f

The signal p.d.f is an asymmetrical double-sided Crystal Ball (CB) function; such a shape corresponds to the roughly Gaussian behaviour of the signal close to the τ mass, while also accounting for the long ISR tail at low mass, and sharp cut-off at higher masses.

$$PDF_{sig}(M_{e\ell\ell'}) = \begin{cases} A_L \cdot (B_L - \frac{M_{e\ell\ell'} - m_0}{\sigma_L})^{-n_L}, & \text{for } \frac{M_{e\ell\ell'} - m_0}{\sigma_L} < -\alpha_L \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{M_{e\ell\ell'} - m_0}{\sigma_L}\right]^2\right), & \text{for } \frac{M_{e\ell\ell'} - m_0}{\sigma_L} \leq 0 \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{M_{e\ell\ell'} - m_0}{\sigma_R}\right]^2\right), & \text{for } \frac{M_{e\ell\ell'} - m_0}{\sigma_R} \leq \alpha_R \\ A_R \cdot (B_R + \frac{M_{e\ell\ell'} - m_0}{\sigma_R})^{-n_R}, & \text{otherwise,} \end{cases} \quad (4.16)$$

where

$$A_i = \left(\frac{n_i}{|\alpha_i|}\right)^{n_i} \cdot \exp\left(-\frac{|\alpha_i|^2}{2}\right), \tag{4.17}$$

$$B_i = \frac{n_i}{|\alpha_i|} - |\alpha_i|, \tag{4.18}$$

$$\tag{4.19}$$

and m_0 is the peak of the Gaussian components, σ_i are the widths of the two sides of the Gaussian, n_i the exponents of the power-law tails and α_i the locations of the transition from the Gaussian to power-law tails. All the parameters are fixed to their fitted value, obtained on 10 000 simulated signal events. The signal events and the fitted p.d.f for the $\tau \to e^-\mu^+e^-$ mode can be seen on figure 4.31, and the other modes are shown in Appendix A.4.

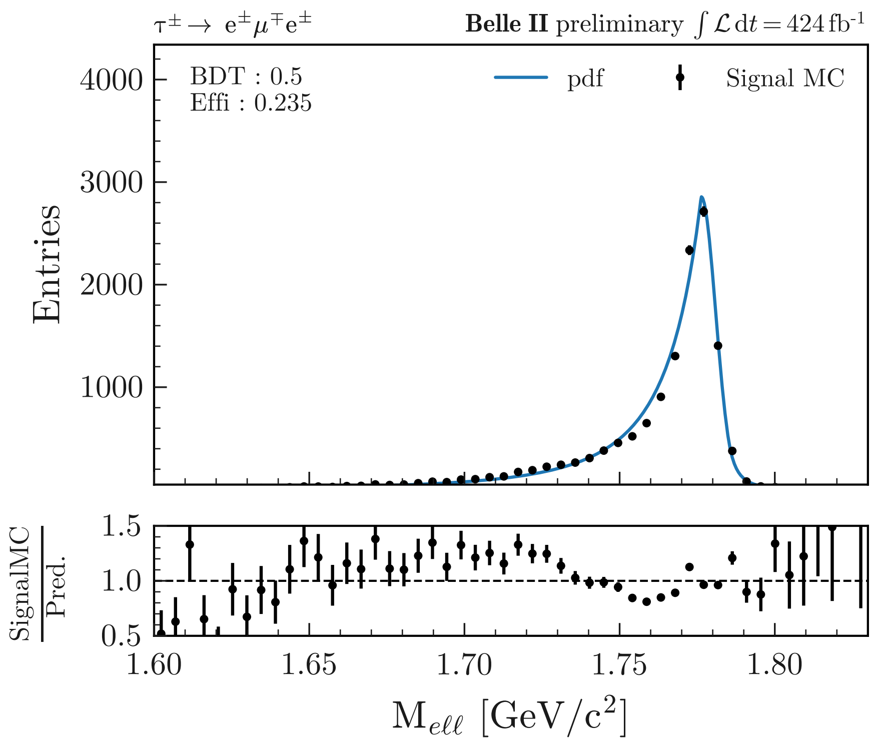


Figure 4.31.: Fitted signal p.d.f for $\tau^- \to e^-\mu^+e^-$.

4.5.0.2. Background PDF

The low number of background events suggests avoiding complex functions. It also implies the need to ensure that the p.d.f does not become negative at any point, which could occur with a polynomial function. Consequently, an exponential function is assumed:

$$PDF_{bg}(M_{e\ell\ell'}) = A \cdot \exp(C_{bg} \cdot M_{e\ell\ell'}) \tag{4.20}$$

with A the normalisation constant :

$$A = \frac{C_{bg}}{e^{2 \cdot C_{bg}} - e^{1.4 \cdot C_{bg}}} \tag{4.21}$$

where C_{bg} gives the overall shape of the distribution and is a free parameter in the fit.

4.5.0.3. Fitting procedure

The three free parameters in the total p.d.f from Eq 4.15 are the background shape C_{bg}, the background yield N_{bg} and the branching fraction $\mathcal{B}$ $(\tau^- \to e^{\mp}\ell^{\pm}\ell'^-)$. The p.d.f is fitted to data by performing an unbinned maximum likelihood fit while constraining $\mathcal{B}$ to be positive : considering the really low background yields, almost any negative value of $\mathcal{B}$ produces negative values for the p.d.f. In the case where no signal is found (the branching fraction is compatible with 0), an upper limit is set on the branching fraction.

4.5.0.4. Upper limit

The upper limit calculation is done using the CLs method [116, 117], implementing the `FrequentistCalculator` and `HypoTestInverter` from `RooStat` which make use of toys.

10000 pseudo experiments are used for each of the 10 scanned points in the $[0, 3] \cdot 10^{-8}$ interval for the branching fraction. The toys are generated using the total p.d.f from equation 4.15, with the fitted value of C_{bg}, the observed number of events N_{obs} and the signal p.d.f from equation 4.16 (whose parameters have been fitted from simulation, as detailed above). Systematic uncertainties on C_{bg}, ϵ_{sig}, $\mathcal{L}$ and $\sigma_{\tau\tau}$ are accounted for by applying Gaussian constraints using the values described in Section 4.6. Confidence level (CL) or p-values are computed for the background only hypothesis (CL_b) and signal+background hypothesis (CL_{s+b}), for the scanned values of $\mathcal{B}$. The upper limit on the branching fraction is then

Table 4.14.: Fitted values of the signal p.d.f parameters.

	α_L	α_R	m_0	n_L	n_R	σ_L	σ_R
$e^-e^+e^-$	0.02	1.93	1.78	4.54	2.90	0.0003	0.006
$e^-e^+\mu^-$	0.06	1.50	1.78	3.03	4.25	0.0007	0.005
$e^-\mu^+e^-$	0.03	1.86	1.78	3.13	3.05	0.0003	0.004
$\mu^-\mu^+e^-$	0.71	2.00	1.78	2.72	3.15	0.004	0.004
$\mu^-e^+\mu^-$	0.33	2.00	1.78	2.79	3.20	0.002	0.005

constructed as the value for which the confidence level for signal $CL_s = 10\%$, with

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{4.22}$$

4.5.1. Fit validation

In order to validate the fitting procedure, the fitted value of the branching fraction are controlled with toys, for different values of the generated branching fraction. At each point, 10000 toys dataset are generated and then used to fit the p.d.f from equation 4.15. The toys are produced using the background p.d.f obtained from fitting the data sidebands, and the signal p.d.f from equation 4.16.

Figure 4.32 show the pulls (the pull value p is defined as $p = (v_f - v_t)/v_e$, with v_f the fitted value, v_e its error and v_t the true value) for two different values of the generated branching fraction (0.33 and 1.33 $\cdot 10^{-8}$). The pulls are expected to showcase a Gaussian profile, so they are fitted with a Gaussian function where the mean and width should be 0 and 1 respectively. However there is a peaking contribution at negative values that deviates from the Gaussian profile. This is due to toys where the fitted branching fraction is 0, but the fit would move to negative values if it was not constrained. In such cases the error on the fitted branching fraction is underestimated, leading to this peaking contribution. A Gaussian with mean and width at (0,1) is also overlaid on the plots, showing that excluding this problematic contribution, the pulls are indeed well described by a Gaussian distribution. The more signal is injected in a toy, the less likely the fit is to converge to a negative value, and the less the peaking contribution is visible. Consequently, in the case of a fitted value of $\mathcal{B} = 0$ in the data, the error would be estimated by using the coverage from toys. For other fitted values, the error from the fit is used.

The values of the two fitted parameters ($\mathcal{B}$ and C_{bg}) in toys, as a function of the generated $\mathcal{B}$ are found in Figures 4.33 and 4.34. The fitted values correspond to their generated values, showing that no bias is introduced. The slight underestimation of C_{bg} for the $e^-\mu^+e^-$ and $\mu^-\mu^+e^-$ modes comes from the extremely small value of C_{bg} for theses modes, that leads to the generation of biased toys with no events in the upper mass range.

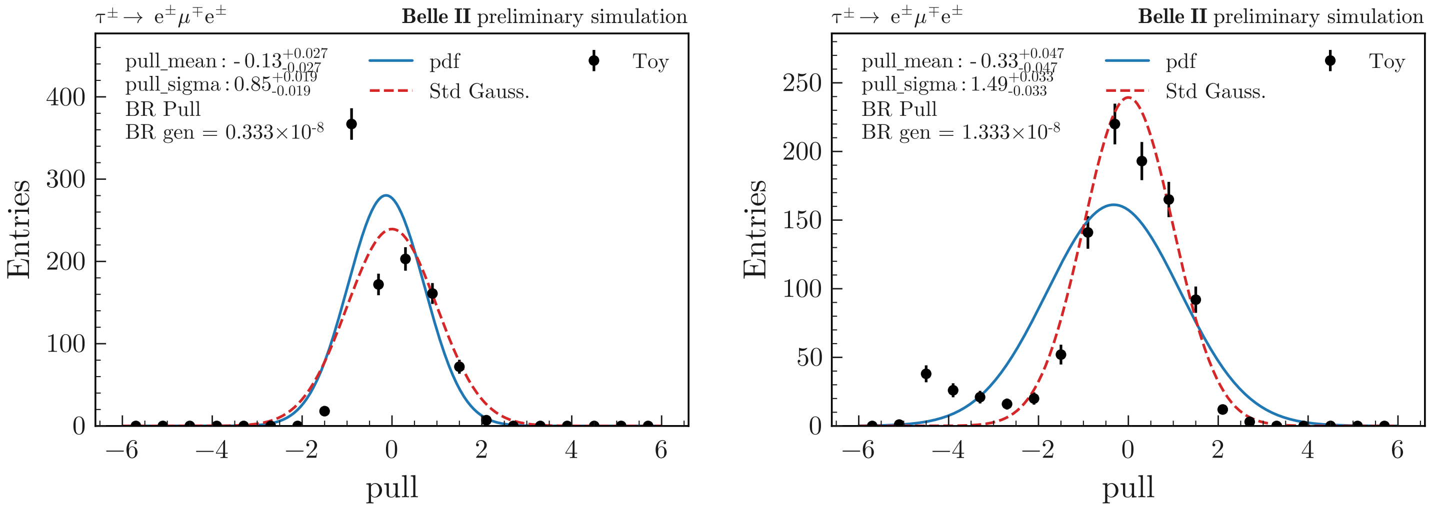


Figure 4.32.: The pulls for the fitted branching fraction for toys generated with a branching fraction of $0.33 \cdot 10^{-8}$ (left) and $1.33 \cdot 10^{-8}$ (right). The blue line corresponds to a Gaussian fit of the distribution, while the red dashed line is a Gaussian with fixed mean and width in (0,1). The fitted mean and width deviate from (0,1) because of the presence of toys with a fitted $\mathcal{B}$ of 0. A similar behaviour is observed for the other modes.

4.6. Study of the systematics uncertainties

In this section, I study the sources of systematic uncertainties that can affect various quantities involved in the measurement of branching fraction of the $\tau^- \to e^{\mp}\ell^{\pm}\ell'^{-}$ decays, given in Eq. 4.15. It is necessary to consider these uncertainties when calculating the upper limit accurately. Systematic uncertainties arise from the imperfect representation of data in simulations, the methods employed, or external inputs. They 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 the uncertainty is provided in Table 4.18. One should nonetheless keep in mind that the main source of uncertainty in this work is statistical.

4.6.1. Uncertainty on signal efficiency

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

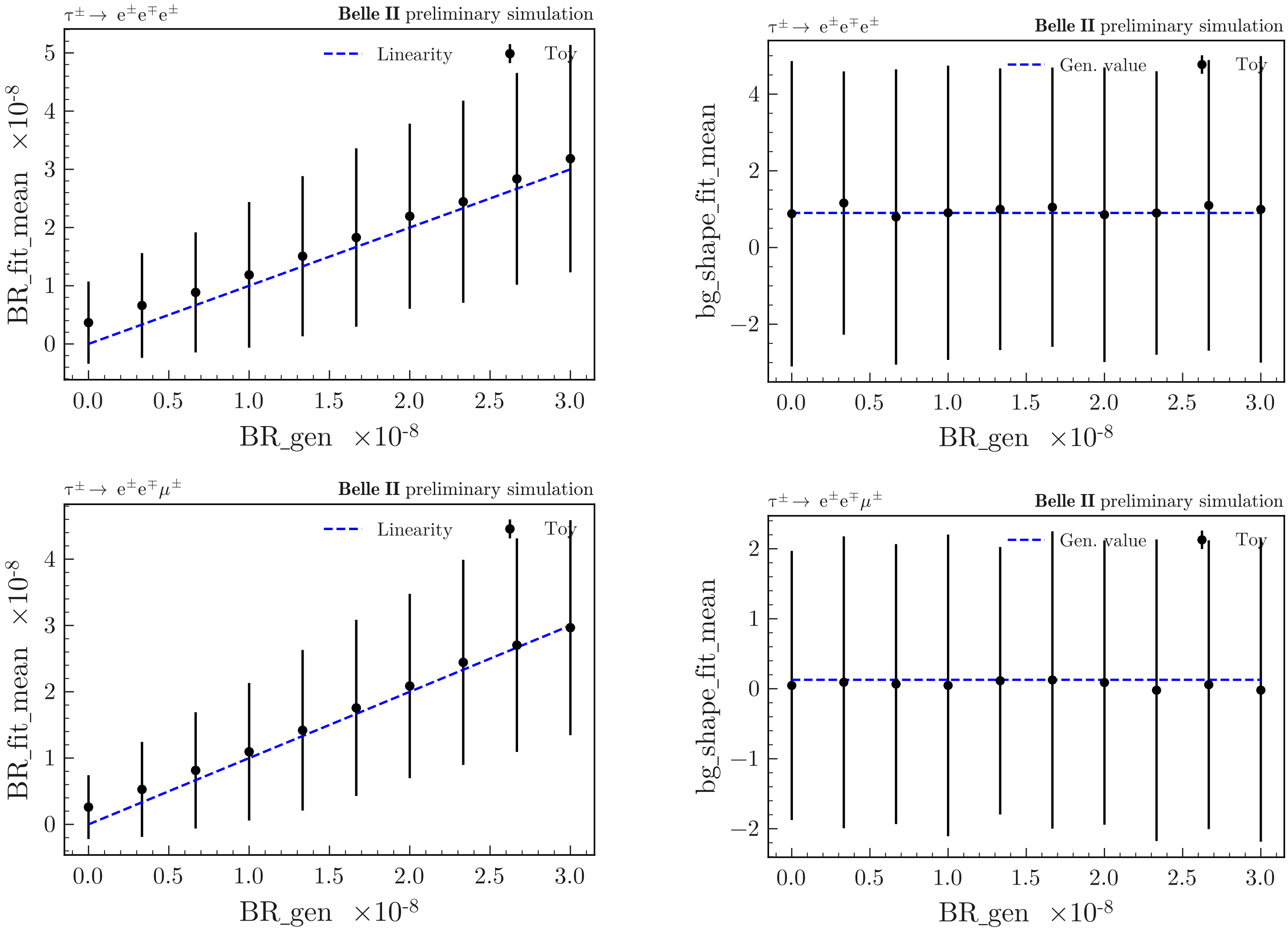


Figure 4.33.: Mean of the fitted values for $\mathcal{B}$ (left) and C_{bg} (right) for toys corresponding to differents generated values of $\mathcal{B}$. The dashed lines correspond to the generated values. The errors correspond to the RMS of the distributions.

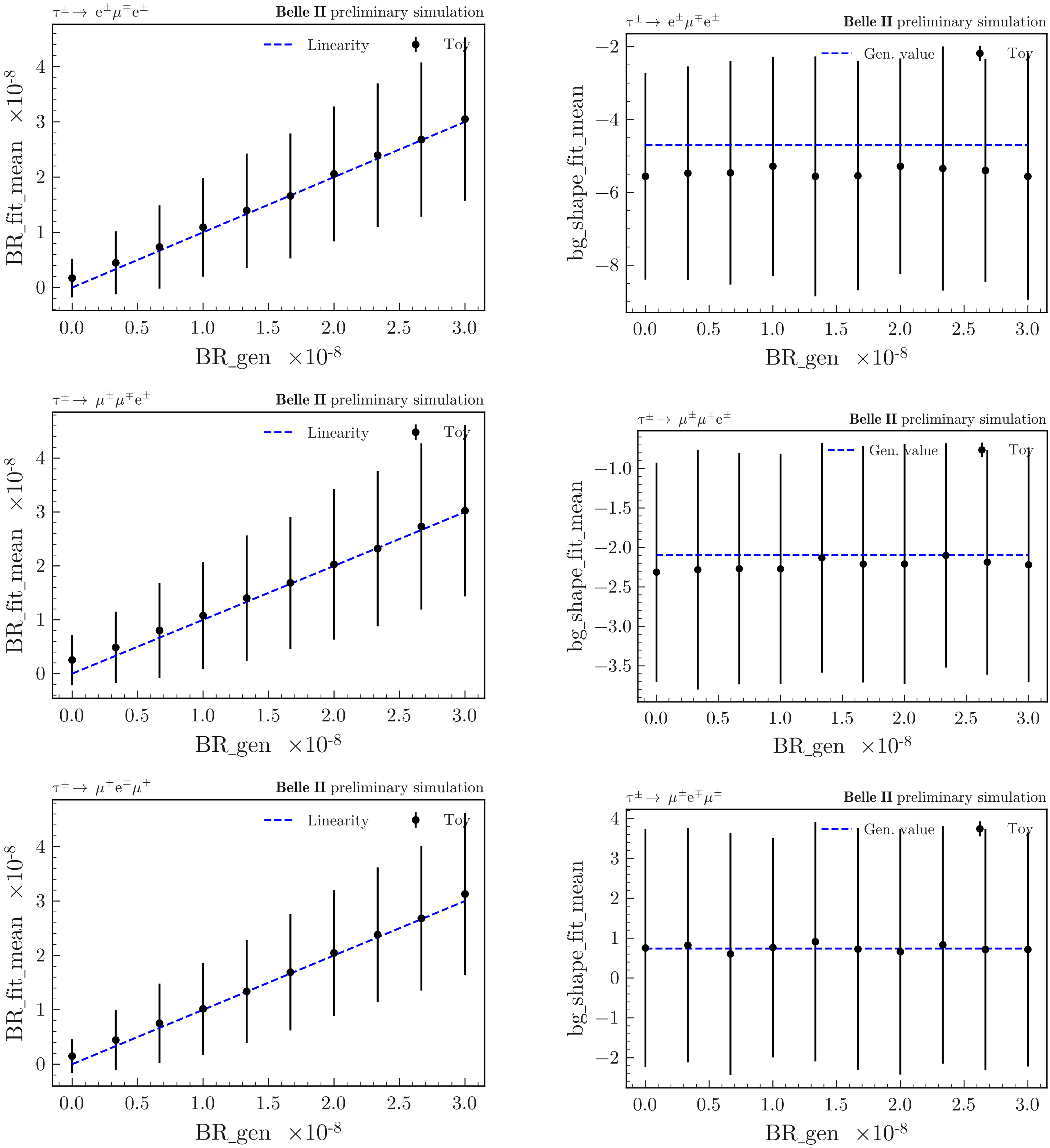


Figure 4.34.: Mean of the fitted values for $\mathcal{B}$ (left) and C_{bg} (right) for toys corresponding to different generated values of $\mathcal{B}$. The error bars correspond to the standard deviations, while the dashed lines represent the generated values.

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

$$w_{LID}^{\substack{sys \\ stat},\uparrow\downarrow} = \prod_{i=1}^{3} w_{LID,\ell i} \times w_{LID,\ell i}^{\substack{sys \\ stat},\uparrow\downarrow}. \quad (4.23)$$

Then, the signal efficiency is recomputed using the statistical and systematics LID variation, and a relative error is computed as:

$$\sigma_{LID}^{\substack{sys \\ stat},\uparrow\downarrow} = \sqrt{\left(\frac{n_{LID} - n_{LID}^{\substack{sys \\ stat},\uparrow\downarrow}}{n_{produced}}\right)^2} \times \frac{1}{\varepsilon_{sig}^{abs}}, \quad (4.24)$$

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

The detailed errors are in Table 4.16. The overall relative systematic uncertainty is then the sum in quadrature of the maximal error for the statistical and systematic contribution.

Table 4.15.: Detailed signal efficiencies and deviation with the different LID correction variations. After applying the whole background rejection selections, the numbers are obtained in the fit region.

	σ_{LID}^{statup}	$\sigma_{LID}^{statdown}$	σ_{LID}^{systup}	$\sigma_{LID}^{systdown}$	σ_{LID}^{tot}
$e^-e^+e^-$	0.3%	0.1%	1.0%	0.7%	1.0%
$e^-e^+\mu^-$	1.0%	0.9%	0.7%	0.6%	1.2%
$e^-\mu^+e^-$	1.0%	0.9%	0.74%	0.6%	1.3%
$\mu^-\mu^+e^-$	1.6%	1.5%	0.7%	0.5%	1.7%
$\mu^-e^+\mu^-$	1.7%	1.5%	0.7%	0.5%	1.8%

Trigger efficiency As explained in Section 4.2, the selected events are required to fire the ECL-based trigger lines. In order to assess the overall difference in trigger efficiency between data and simulation, the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample described in section 4.3, is used. The trigger systematic uncertainty is then evaluated as the relative variation in trigger efficiency between data and MC, using the CDC trigger as independent selection. Figure 4.35 shows the trigger efficiency as a function of the transverse momentum of the leading pion; the uncertainty is estimated to be 0.5%. The possible bias coming from the use of the CDC trigger as a reference is estimated on simulations: a difference of 0.5% is found with respect to the absolute efficiency. It is added in quadrature to the previous uncertainty, leading to a total trigger uncertainty of 0.7%.

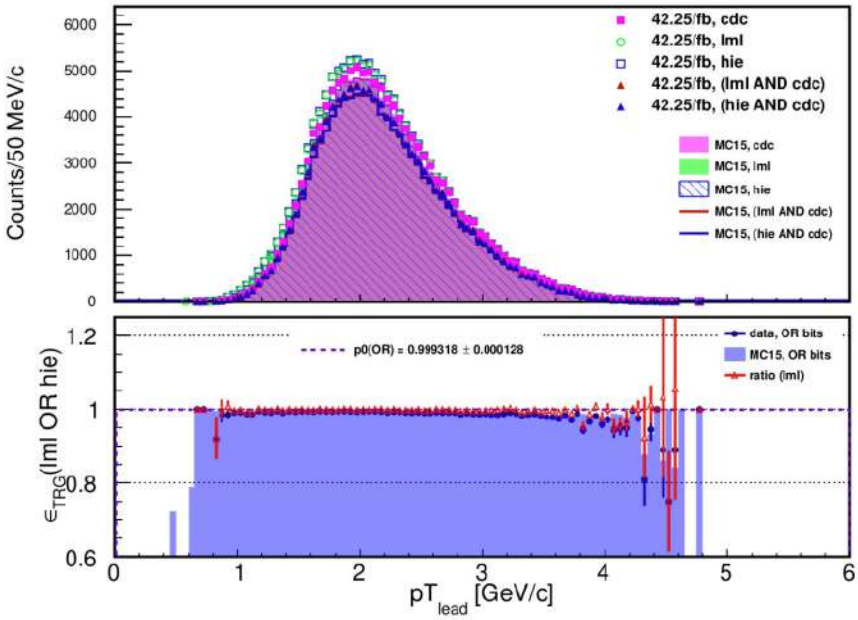


Figure 4.35.: Up : Distribution of the leading transverse momentum in $\tau \to 3\pi\nu$ in data and MC, for the CDC, lml and hie triggers. Down : efficiency of the trigger efficiency (lml or hie) using the CDC trigger as a reference, in data and MC, as well as the data/MC ratio.

Tracking efficiency Tracking efficiency was measured by the tracking group using tag-and-probe techniques in $e^+e^- \to \tau^+\tau^-$ events, targeting 1-prong ($\tau^- \to e^-\bar{\nu}_e\nu_\tau$) and 3-prong ($\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$) decays. 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 since the total charge of the event must be equal to zero. The per-track reconstruction efficiency $\varepsilon_{tracking}$ was computed:

$$\varepsilon_{track} = \frac{N_4}{N_3 + N_4}, \tag{4.25}$$

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. The tracking uncertainty is then obtained by averaging over the 4 and 6 tracks events.

BDT The systematic uncertainty introduced by the BDT are estimated by using the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ control sample described in Section 4.3. The idea is to estimate the discrepancy in efficiency between the data and MC samples while applying a cut on the BDT output.

The first step is to apply the BDT trained for $\tau^- \to e^\mp\ell^\pm\ell'^-$ samples on the control sample, and find the cut value that produces the same relative signal efficiency on $\tau \to 3\pi\nu_\tau$ as the various $\tau^- \to e^\mp\ell^\pm\ell'^-$ modes as shown on Figure 4.36.

The discrepancy between data and MC efficiencies is taken as BDT systematic uncertainty, the values being compiled in Table 4.16.

$$\sigma_{BDT} = \left|1 - \frac{\varepsilon^{rel;3\pi}_{data}}{\varepsilon^{rel;3\pi}_{MC}}\right| \tag{4.26}$$

Table 4.16.: BDT systematic uncertainties.

	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
σ_{BDT}	0.7%	2.5%	0.7%	1.5%	0.4%

4.6.2. Uncertainty on the background modeling

The invariant-mass shape of the background is modelled with an exponential p.d.f (see equation 4.20) and therefore can be parametrised with two numbers : C_{bg} and N_{bg}. They are fitted together and entirely correlated, so a systematic error is assigned to C_{bg} only.

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 [118]. 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 $9998.7^{+3.8}_{-5.7} \times 10^{-4}$ (i.e. a small deviation from 1), with two additional data samples created by applying variations in both directions. The systematic uncertainty is calculated by fitting C_{bg} in the background p.d.f from the sidebands varying the scale factor as :

$$\sigma^{rel}_{MS} = max(\left|\frac{C_{True} - C_{Low}}{C_{True}}\right|, \left|\frac{C_{True} - C_{High}}{C_{True}}\right|) \tag{4.27}$$

To reduce sensitivity to statistical fluctuations, the fit is performed on sidebands with a loose cut of 0.5 on the BDT output. The $M_{e\ell\ell'}$ distribution for the three momentum scale values for all the modes are shown on Figure 4.37.

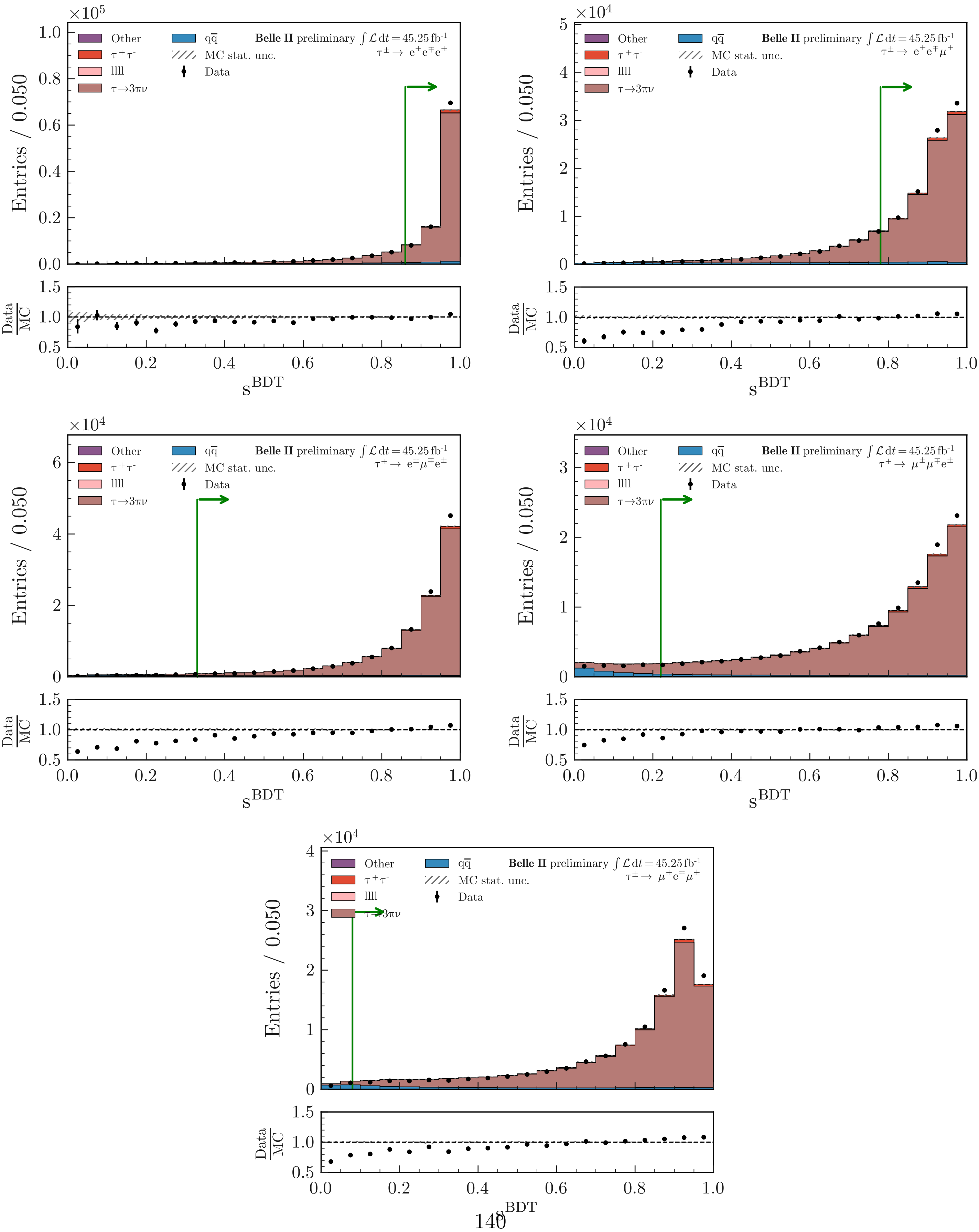


Figure 4.36.: Distribution of the BDT score for $\tau \to 3\pi\nu_\tau$. The cuts displayed are the one giving the same relative signal efficiencies as for the signal $\tau^- \to e^\mp \ell^\pm \ell'^-$.

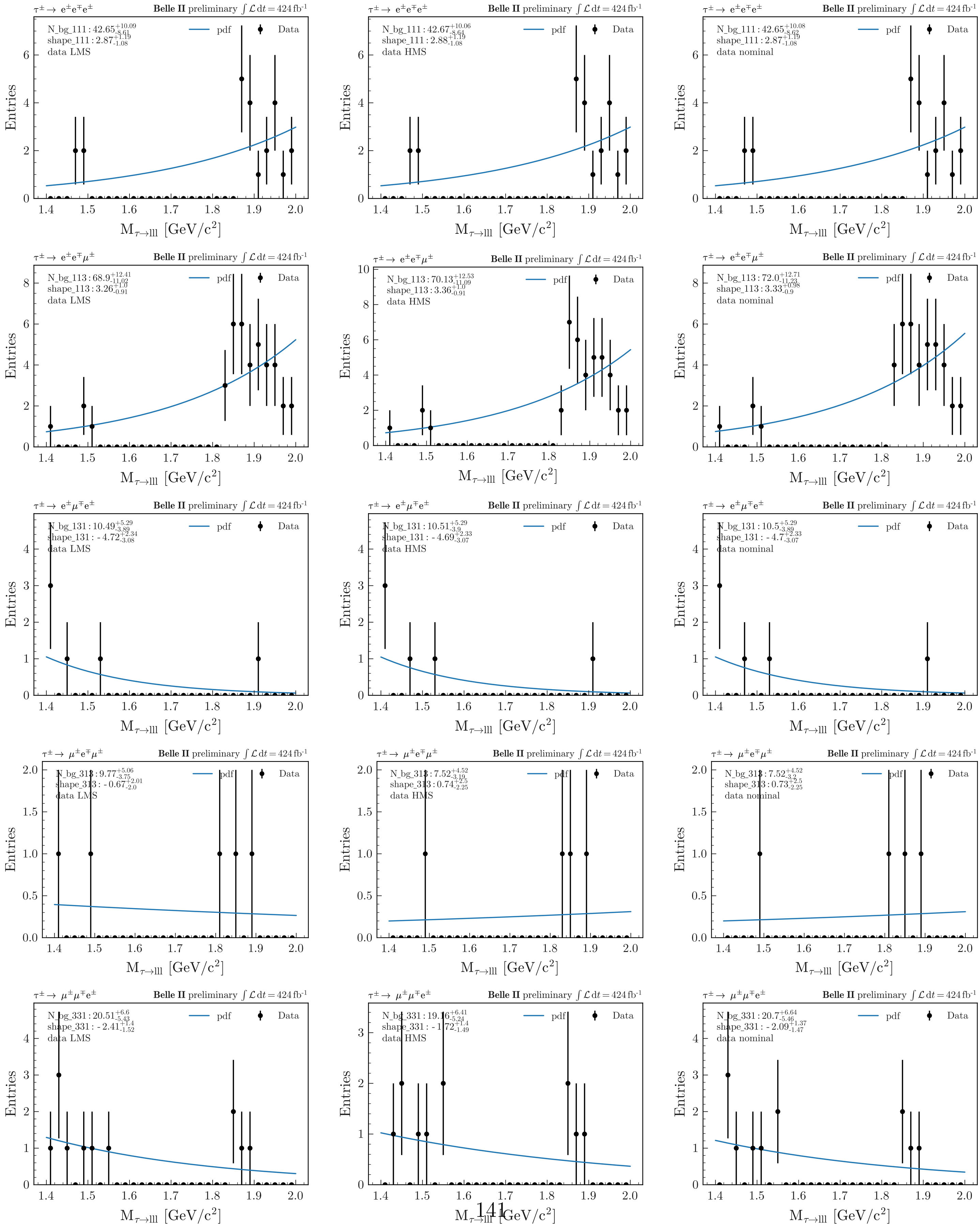


Figure 4.37.: $M_{e\ell\ell'}$ distribution in the case of low (left), high (center) and nominal (right) momentum scale applied.

The systematic uncertainties values for the track momentum scale can be found in Table 4.17. The $\mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$ modes have an uncertainty of 17.9% and 190% respectively, way larger than the other modes. In the case of the $\mu^-e^+\mu^-$ final state, this important relative variation is a consequence of the low number of events and the nominal value being close to 0, so a small absolute variation leads to an important relative variation. Since the nominal fitted value of C_{bg} is small, this important relative variation correspond to a small absolute variation and therefore this uncertainty is not dominant (compared to the statistical one).

Table 4.17.: Momentum scale systematic uncertainties.

	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
σ_{MS}	0.2%	2.1%	0.6%	17.9%	190.7%

4.6.3. Other sources

- Luminosity: the systematic uncertainty on the integrated luminosity has been measured in [119], 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 [120] as 0.003 nb.

4.6.4. Systematics uncertainties summary

Table 4.18 displays the summary of the systematic uncertainties considered for computing the upper limit of exclusion at 90% CL. The total uncertainty on signal efficiency σ_ϵ^{tot} is computed as the sum in quadrature of the LID uncertainty σ_{LID}, the tracking uncertainty $\sigma_{tracking}$ and the BDT uncertainty σ_{BDT}.

Quantity	Uncertainty	$e^-e^+e^-$	$e^-e^+\mu^-$	$e^-\mu^+e^-$	$\mu^-\mu^+e^-$	$\mu^-e^+\mu^-$
$\mathcal{L}$	$\sigma_{\mathcal{L}}$	0.6%	0.6%	0.6%	0.6%	0.6%
$\sigma_{\tau\tau}$	$\sigma_{\tau\tau}$	0.3%	0.3%	0.3%	0.3%	0.3%
ϵ_{sig}	$\sigma_{tracking}$	1.0%	1.0%	1.0%	1.0%	1.0%
	σ_{BDT}	0.7%	2.5%	0.7%	1.5%	0.4%
	σ_{LID}	1.0%	1.2%	1.3%	1.7%	1.8%
	$\sigma_{trigger}$	0.7%	0.7%	0.7%	0.7%	0.7%
	σ_{ϵ}^{tot}	1.8%	2.9%	1.8%	2.6%	2.2%
C_{bg}	$\sigma_{C_{bg}}$	0.2%	2.1%	0.6%	17.9%	190.7%

Table 4.18.: Summary of the systematic uncertainty, where σ_{ϵ}^{tot} is the total uncertainty on the signal efficiencies and $\sigma_{C_{bg}}$ the one on the background modeling (from the track momentum scale only).

4.7. Upper limits

In this section I compute the expected upper-limit to evaluate the sensitivity of the analysis. The calculation presented below is repeated for every cut value on the BDT score between 0.5 and 0.95, with a step of 0.05 ; the cut on the score that gives the best expected upper limit is kept (and is the one that can be found in Table 4.10). For each BDT cut value, the expected background is estimated on sidebands and then used for the expected upper limit estimation.

Background estimation Following the same procedure that would be used for unblinding, the first step is to fit the background component (p.d.f from equation 4.20) of the total p.d.f 4.15 to data sidebands.
Only the background component is fitted as the signal is expected to be close to 0 in the sidebands, and would be difficult to constrain from a fit to sidebands only.
From this fit, we obtain an expected value for the exponential factor C_{bg}^{exp}, and an expected number of events N_{bg}^{exp}, by normalising the background p.d.f to the sidebands. These two expected values for the background are presented in Table 4.19, while the fits are shown in Figure 4.38.

Table 4.19.: Fitted value of the background p.d.f term (C_{bg}^{exp}) and expected total number of events in the fit band.

	C_{bg}^{exp}	N_{bg}^{exp}
$e^-e^+e^-$	$0.92^{+2.76}_{-2.54}$	$6.08^{+4.29}_{-2.87}$
$e^-e^+\mu^-$	$0.13^{+1.86}_{-1.79}$	$12.11^{+5.70}_{-4.31}$
$e^-\mu^+e^-$	$-4.70^{+2.33}_{-3.07}$	$10.51^{+5.28}_{-3.90}$
$\mu^-\mu^+e^-$	$-2.09^{+1.37}_{-1.47}$	$20.70^{+6.64}_{-5.45}$
$\mu^-e^+\mu^-$	$0.74^{+2.49}_{-2.25}$	$7.53^{+4.51}_{-3.20}$

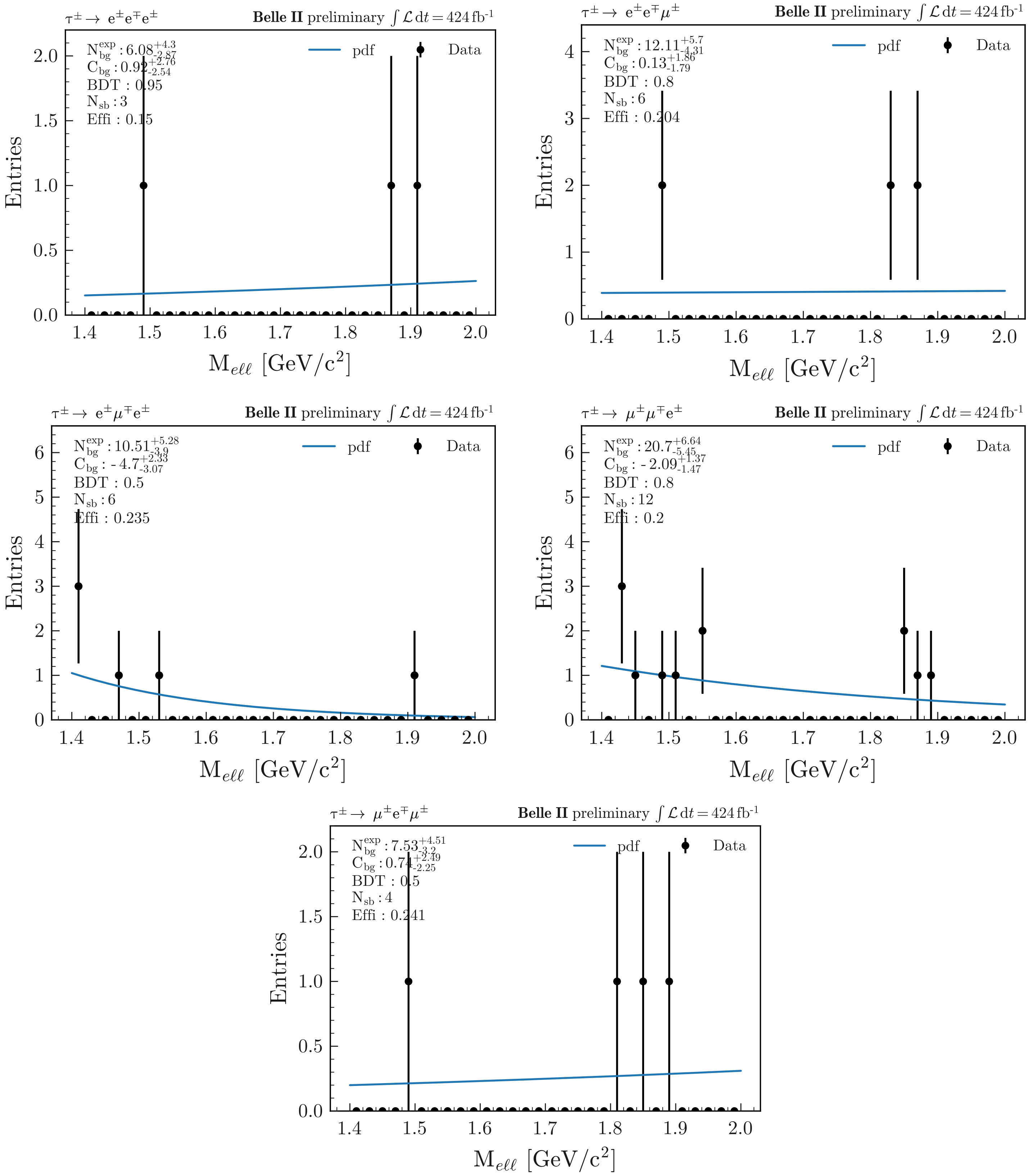


Figure 4.38.: Fits to the sidebands to obtain the expected background yields.

Expected upper limit The expected upper limit is computed using the CL_S method described in section 4.5, with the inputs being $N_{bg}^{exp}$and C_{bg}^{exp}. All the uncertainties are taken into account in the procedure. The total uncertainty on C_{bg} is the sum in quadrature of the statistical uncertainty (from the fit) and the systematic uncertainty. The values of the expected upper limits at 90% CL are presented in Table 4.20, and the CL_S scans are shown in Figure 4.39. The procedure will be repeated with the unblinded data to obtain the observed upper limits.

In addition to the upper limits, the signal distributions in the plane defined by the invariant masses squared of the two opposite charge lepton pairs are provided in Appendix A.1. The signal distribution is impacted by the reconstruction and selection criteria on the invariant masses of the lepton pairs, and might differ between the different BSM models.

Table 4.20.: Expected values of the upper limit at 90% CL on the branching fractions of the five $\tau^- \to e^{\mp}\ell^{\pm}\ell'^-$ decays.

	$\mathcal{B}_{UL}^{90,exp} \cdot 10^{-8}$
$e^-e^+e^-$	2.50
$e^-e^+\mu^-$	2.00
$e^-\mu^+e^-$	1.54
$\mu^-\mu^+e^-$	1.80
$\mu^-e^+\mu^-$	1.54

The expected upper limits are better than the observed limits by Belle[15] for the $e^-e^+e^-$, $\mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$ modes (with the biggest improvement for the $\mu^-\mu^+e^-$ mode), similar for the $e^-\mu^+e^-$ mode and slightly worse for the $e^-e^+\mu^-$ mode.

4.8. Results

This analysis was approved for unblinding by the Belle II collaboration, although after the initial submission of this document (but fortunately before the defense !). This section presents the results obtained upon unblinding.

The $M_{e\ell\ell'}$ distributions for the five modes are shown in Figure 4.40. The fits are also shown in the same figure, with the background and signal components separated. No overall excess with respect to what was expected is observed and no signal is found. For all the modes but $\mu^-\mu^+e^-$, no event is observed near the peak of the τ mass. As a consequence, the observed upper limits will be better than the expected ones.

The upper limits are computed using the same procedure as for the expected limits, but with the observed numbers. The observed upper limits are presented in Table 4.21, together with the number of events as well as the expected numbers. The obtained bounds, between 1.4 and 2.4 $\times 10^{-8}$, are the most stringent to date.

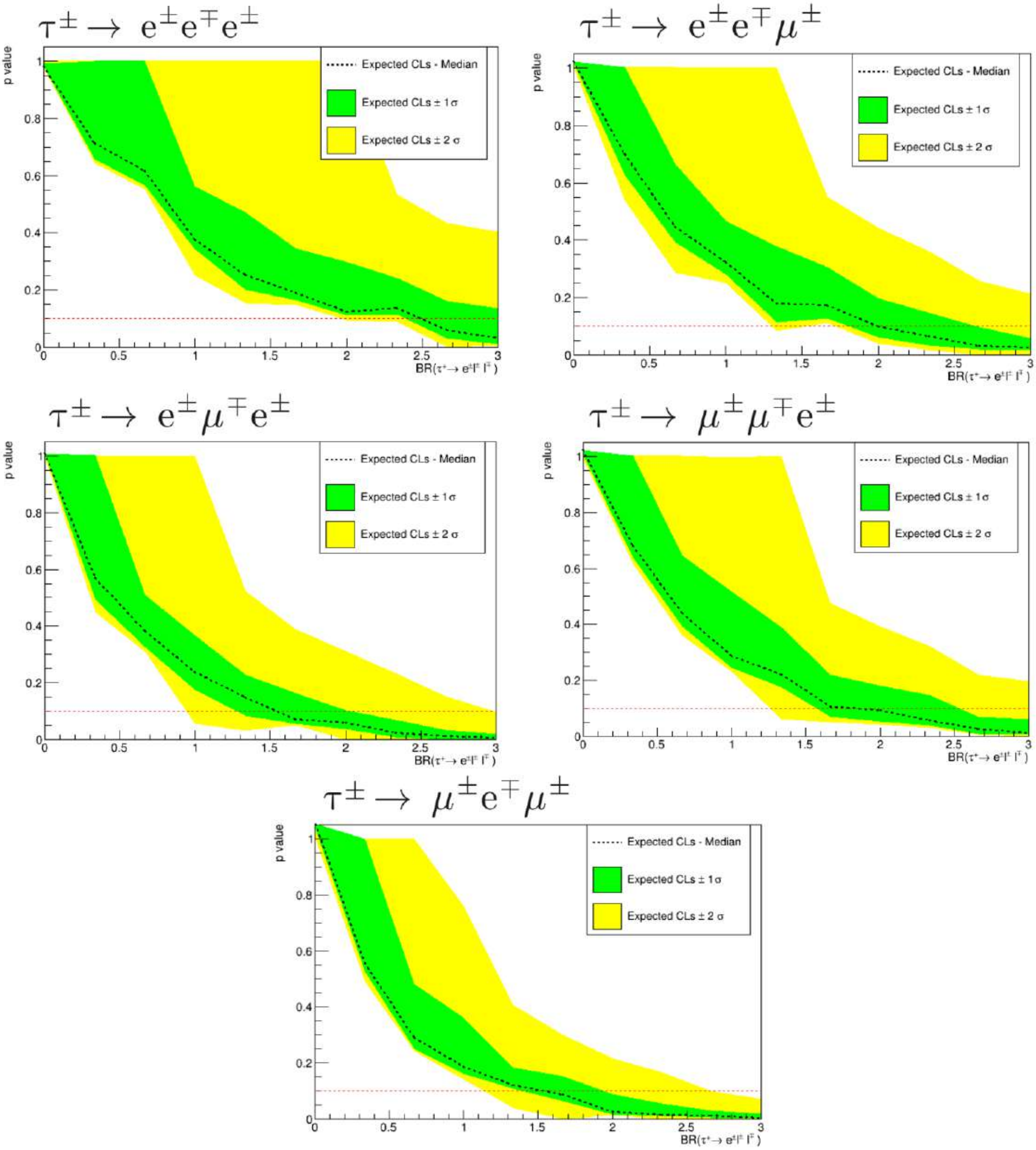


Figure 4.39.: CL_S scans and expected upper limits on $\mathcal{B}$, for each of the five considered decay modes. The red dashed lines correspond to the 90% confidence level. The yellow and green bands corresponds to the +1 and +2 σ intervals.

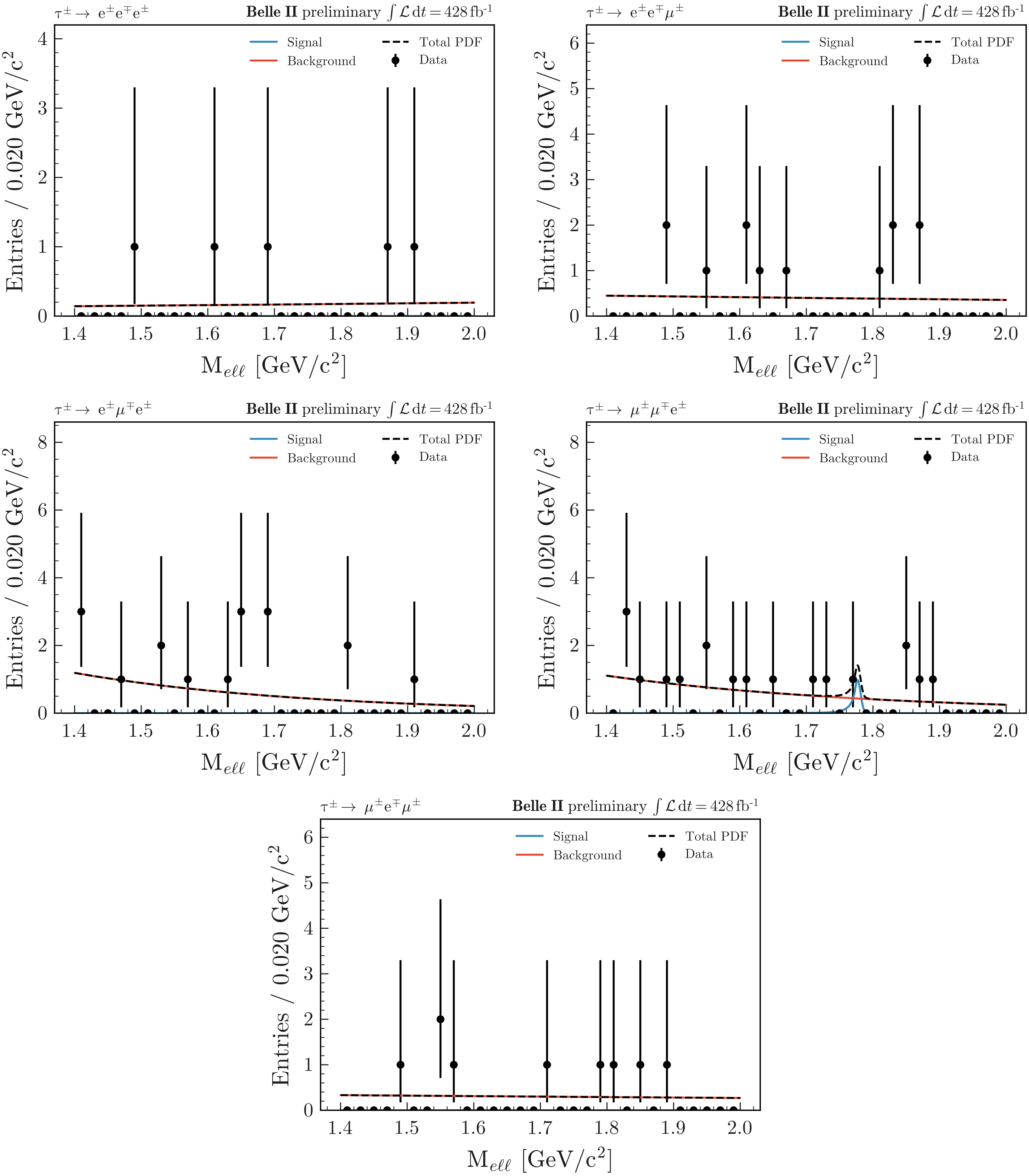


Figure 4.40.: Distribution of data events and fits to the $M_{e\ell\ell'}$ variable. The background and signal components of the PDF are shown in red and blue, respectively, while the black dashed line is the total fitted PDF.

Table 4.21.: Expected and observed number of events, fitted value of C_{bg} and branching fractions, and expected and observed upper limits at 90% C.L.

	N_{exp}	N_{obs}	C_{bg}	$\mathcal{B} \cdot 10^{-8}$	$\mathcal{B}^{UL}_{exp} \cdot 10^{-8}$	$\mathcal{B}^{UL}_{obs} \cdot 10^{-8}$
$e^-e^+e^-$	$6.1^{+4.3}_{-2.9}$	5	$0.52^{+2.64}_{-2.60}$	0	2.50	2.48
$e^-e^+\mu^-$	$12.1^{+5.7}_{-4.3}$	12	$-0.40^{+1.67}_{-1.68}$	0	2.00	1.57
$e^-\mu^+e^-$	$10.5^{+5.3}_{-4.3}$	17	$-2.90^{+1.48}_{-1.54}$	0	1.54	1.55
$\mu^-\mu^+e^-$	$20.7^{+6.6}_{-5.5}$	18	$-2.50^{+1.45}_{-1.52}$	$0.48^{+0.90}_{-0.48}$	1.80	2.43
$\mu^-e^+\mu^-$	$7.5^{+4.5}_{-3.2}$	9	$-0.34^{+1.93}_{-1.94}$	0	1.54	1.33

Conclusion

In this manuscript, I presented the research conducted over the course of last three years in the Belle II collaboration.
I outlined three different tracking studies to estimate and reduce the rate of fake tracks in the vertex detector with simulated $e^+e^- \to e^+e^-$ events.
The first one was a study of the fake tracks rate in the proposed upgrade of the vertex detector, the VTX, which will have to operate in a high luminosity environment. I explored various means implemented in the tracking software to reduce the fake tracks rate and highlighted flaws in the tracking chains in high luminosity scenarios.
Additionally, I presented two ways of taking advantage of the time information from the SVD, a sub-system of the current vertex detector:

- with the implementation of the track time variable in the tracking software, which allows to reject half of the fake tracks and can be used at the analysis level, and
- with the rejection of detector hits using a classifier to cope with the degraded time resolution due to deteriorated data acquisition conditions.

The main subject of the thesis was the search for the five lepton flavour violating decays $\tau \to e\ell\ell'$ using the Run 1 dataset of 424 fb^{-1}.
The analysis relies on an untagged reconstruction of the signal candidates in $e^+e^- \to \tau^+\tau^-$ events, where only the signal τ is reconstructed while the other is left unconstrained. This search is affected by background contributions from QED processes ($e^+e^- \to \ell\ell'\ell'$), that are mis-modeled in the simulations; thus the entire background selection is performed on data sidebands using a combination of selection criteria and a BDT classifier.
The branching fraction is extracted from an unbinned maximum likelihood fit to the $M_{e\ell\ell'}$ distribution, and the upper limits are computed using the CL_S method. In four of the modes ($e^-e^+e^-$, $e^-e^+\mu^-$, $\mu^-\mu^+e^-$ and $\mu^-e^+\mu^-$), the analysis set more stringent limit compared to the previous Belle's results, while the fifth mode ($e^-\mu^+e^-$) is comparable to Belle's value.
The improvement in sensitivity, despite a lower statistic (424 fb^{-1} for this work, compared to the 782 fb^{-1} used by Belle) is due several factors. The overall performances of the Belle II detector, tracking and reconstruction have been improved with respect to the Belle era. To complement this, the analysis techniques used push the sensitivity at its best with the use of an untagged event candidate

	$\mathcal{B}_{UL}^{90,exp} \times 10^{-8}$	$\mathcal{B}_{UL,25}^{90,exp} \times 10^{-8}$	$\mathcal{B}_{UL,50}^{90,exp} \times 10^{-8}$
$e^-e^+e^-$	2.5	$11.3 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$
$e^-e^+\mu^-$	2.0	$12.4 \cdot 10^{-2}$	$10.1 \cdot 10^{-2}$
$e^-\mu^+e^-$	1.5	$8.7 \cdot 10^{-2}$	$5.3 \cdot 10^{-2}$
$\mu^-\mu^+e^-$	1.8	$14.4 \cdot 10^{-2}$	$8.5 \cdot 10^{-2}$
$\mu^-e^+\mu^-$	1.5	$9.1 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$

Table 4.22.: Projection of the expected values of the upper limit at 90% CL on the branching fractions at the integrated luminosities of 25 and 50 ab^{-1}. The expected upper limits are computed by rescaling the expected background yield to the two luminosities. The first column corresponds to the expected upper limits at 424 fb^{-1} from this work.

selection, a BDT for background rejection and extraction of the signal yield with an unbinned maximum likelihood fit to $M_{e\ell\ell'}$.
In the final unblinded distributions, no signal was observed and the most stringent limits to date were set, between 1.4 and 2.4 10^{-8}.

In the coming years, Belle II will likely be the only experiment capable of improving the upper limits on these decays. While LHC experiments can also probe certain τ LFV decays, as can be seen from the recent results from CMS [62] and LHCb [63] on $\tau \to 3\mu$ (although with a lesser sensitivity than Belle II'), the presence of electrons in the final state of the $\tau^- \to e^\mp\ell^\pm\ell'^-$ channels severely limits their ability to search these modes. Additionally, it's not clear yet whether the (hypothetical) Future Circular Collider (FCC-ee) [121] will be able to perform better.

It is therefore interesting to consider the future prospects for this analysis at Belle II. By simply rescaling the background yield to 25 and 50 ab^{-1}, one obtain expected upper limits in the 5-14 $\times 10^{-10}$. The values are presented in Table 4.22.

This projection is conservative, as the underlying assumption is that no improvement in the analysis will be made. A potential lead to explore is the use of the WHIZARD generator [105, 106] for simulating four-lepton final states, with accurate modelling of ISR and FSR. This approach could enable the use of MC for BDT training or for extracting the background p.d.f.

A. Appendix

A.1. Dalitz plots

The background rejection uses selection criteria on the invariant masses of the lepton pairs. This can affect the phase space distribution of signal events, and reduce the sensitivity to some BSM models. To account for this, the signal distributions in the plane defined by the invariant mass squared of the two pairs of opposite sign leptons are shown in the following figures.

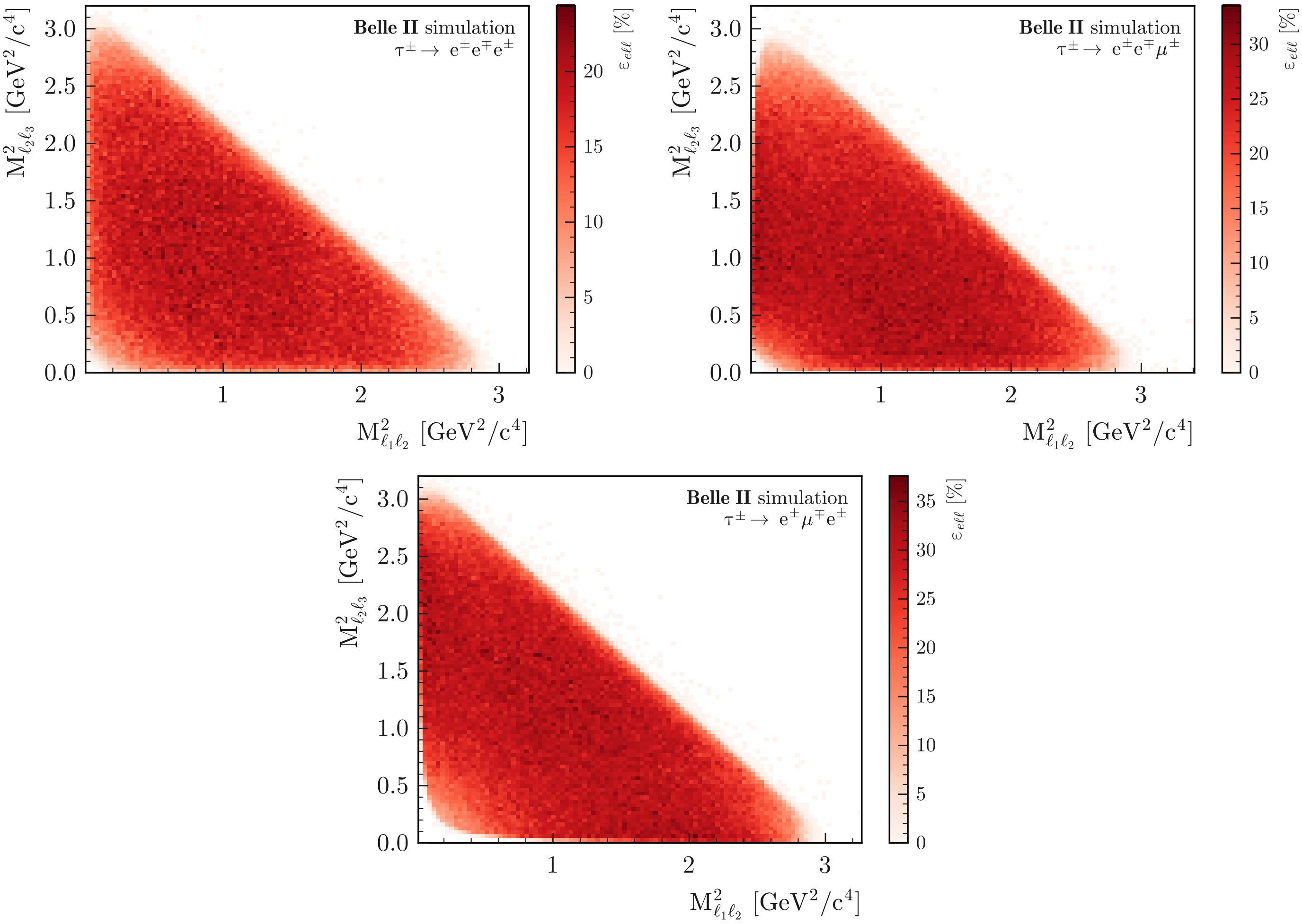


Figure A.1.: Signal distribution in the plane defined by the invariant mass squared of the two pairs of opposite sign leptons for the $e^-e^+e^-$ (upper left), $e^-e^+\mu^-$ (upper right) and $e^-\mu^+e^-$ (bottom) channels. The indices 1,2 and 3 follow the ordering of the decay chains.

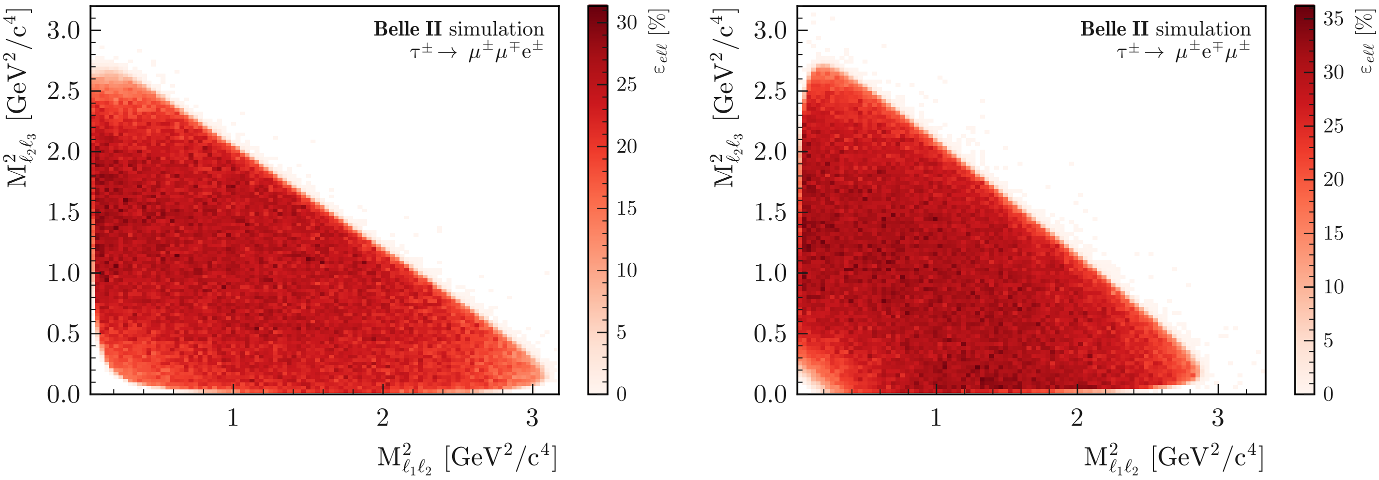


Figure A.2.: Signal distribution in the plane defined by the invariant mass squared of the two pairs of opposite sign leptons for the $\mu^-\mu^+e^-$ (left) and $\mu^-e^+\mu^-$ (right) channels. The indices 1,2 and 3 follow the ordering of the decay chains.

A.2. Distributions of physics observables for $\tau \to 3\pi\nu$

The following plots are the comparison between data and simulation of all the variables used as input in the BDTs, for the $\tau \to 3\pi\nu$ control sample.

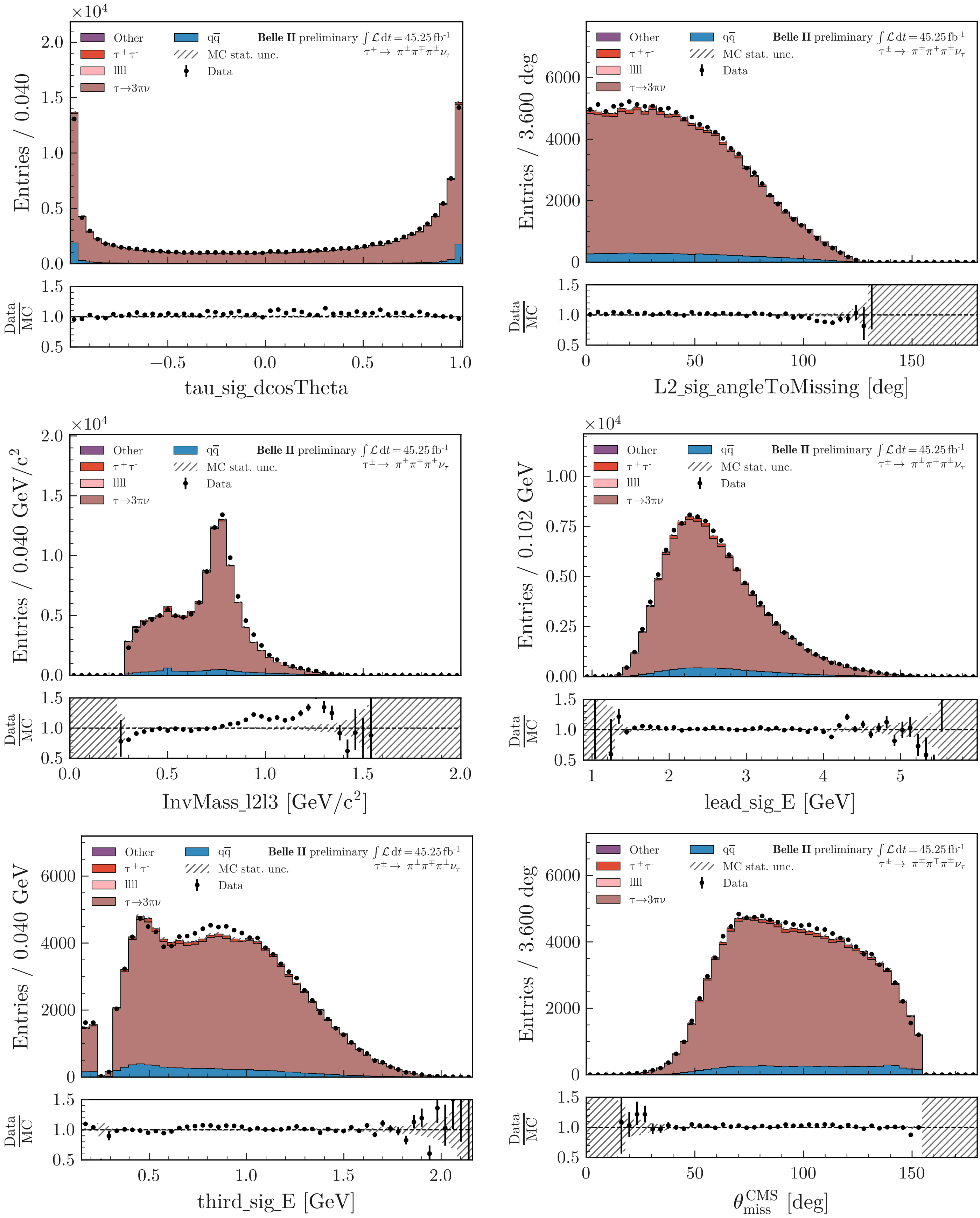


Figure A.3.: Data-MC comparison of the BDT input variables for the control channel

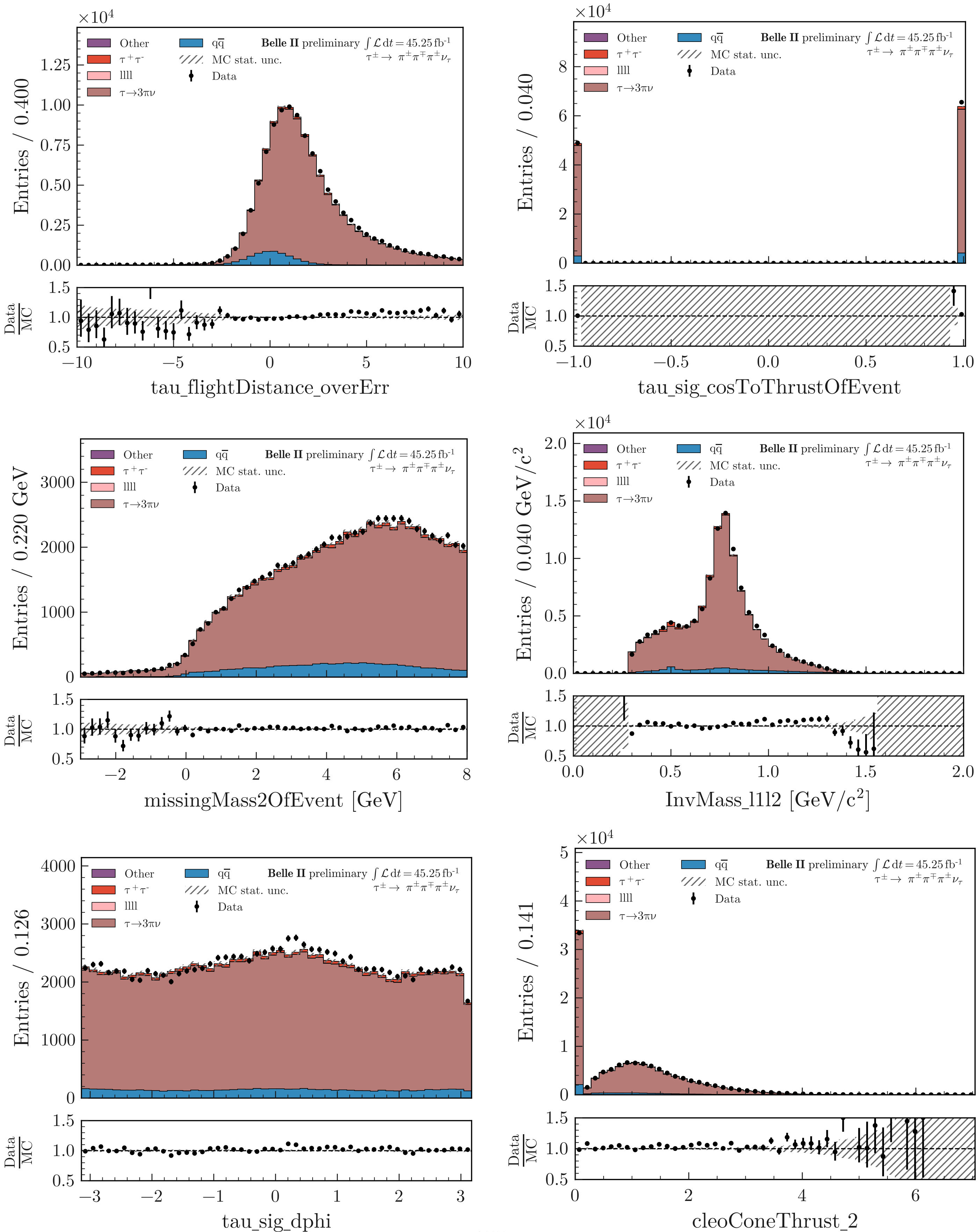


Figure A.4.: Data-MC comparison of the BDT input variables for the control channel

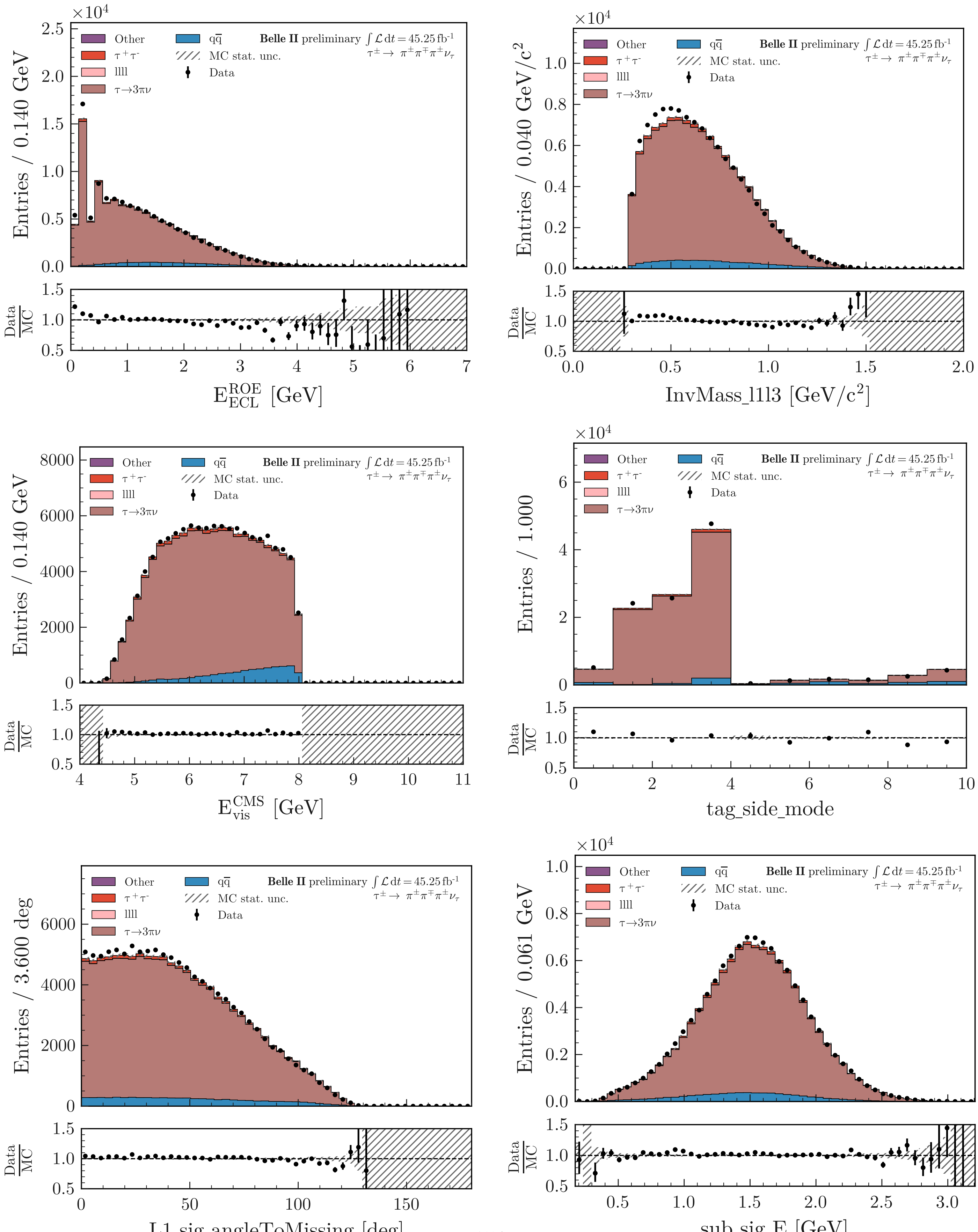


Figure A.5.: Data-MC comparison of the BDT input variables for the control channel

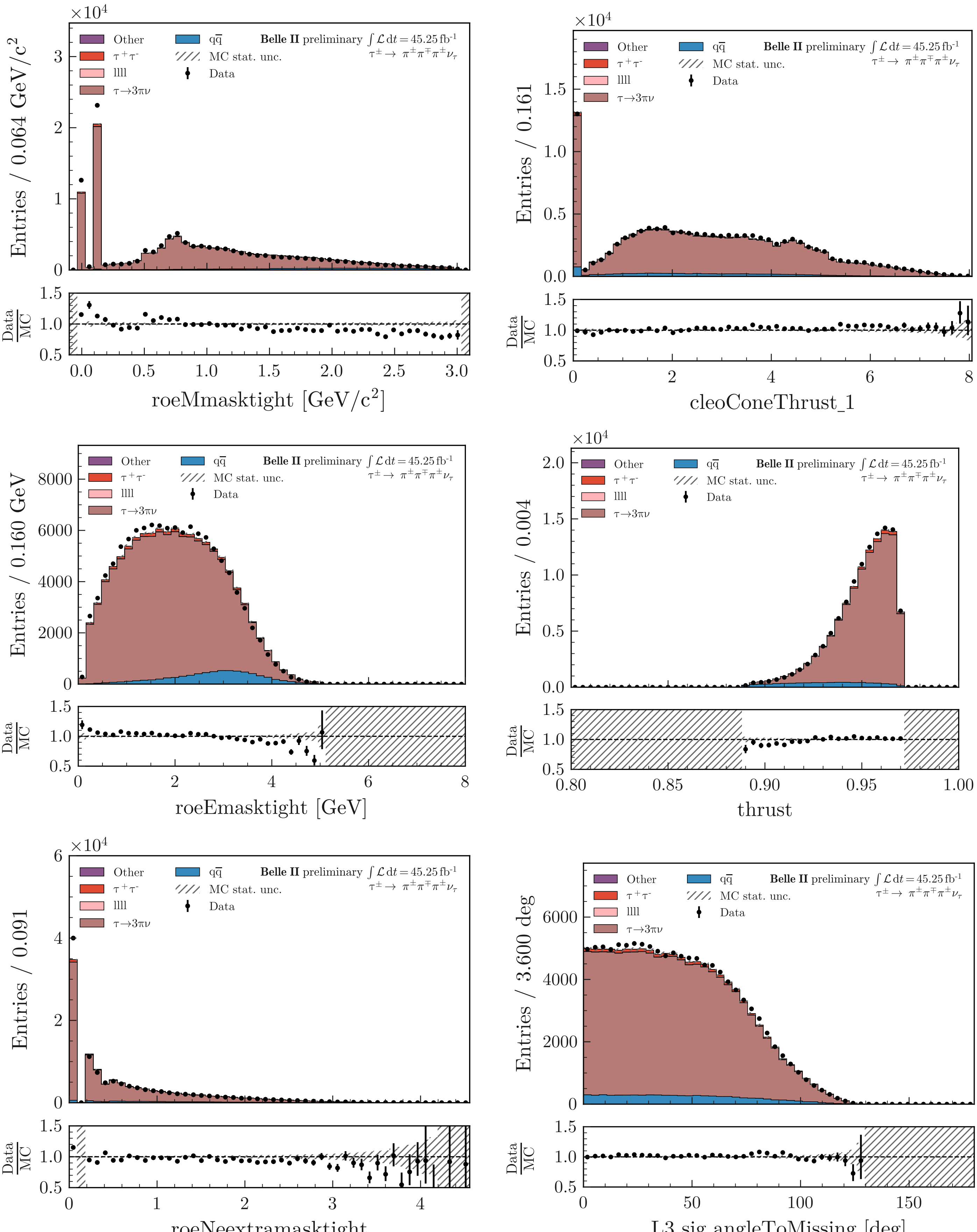


Figure A.6.: Data-MC comparison of the BDT input variables for the control channel

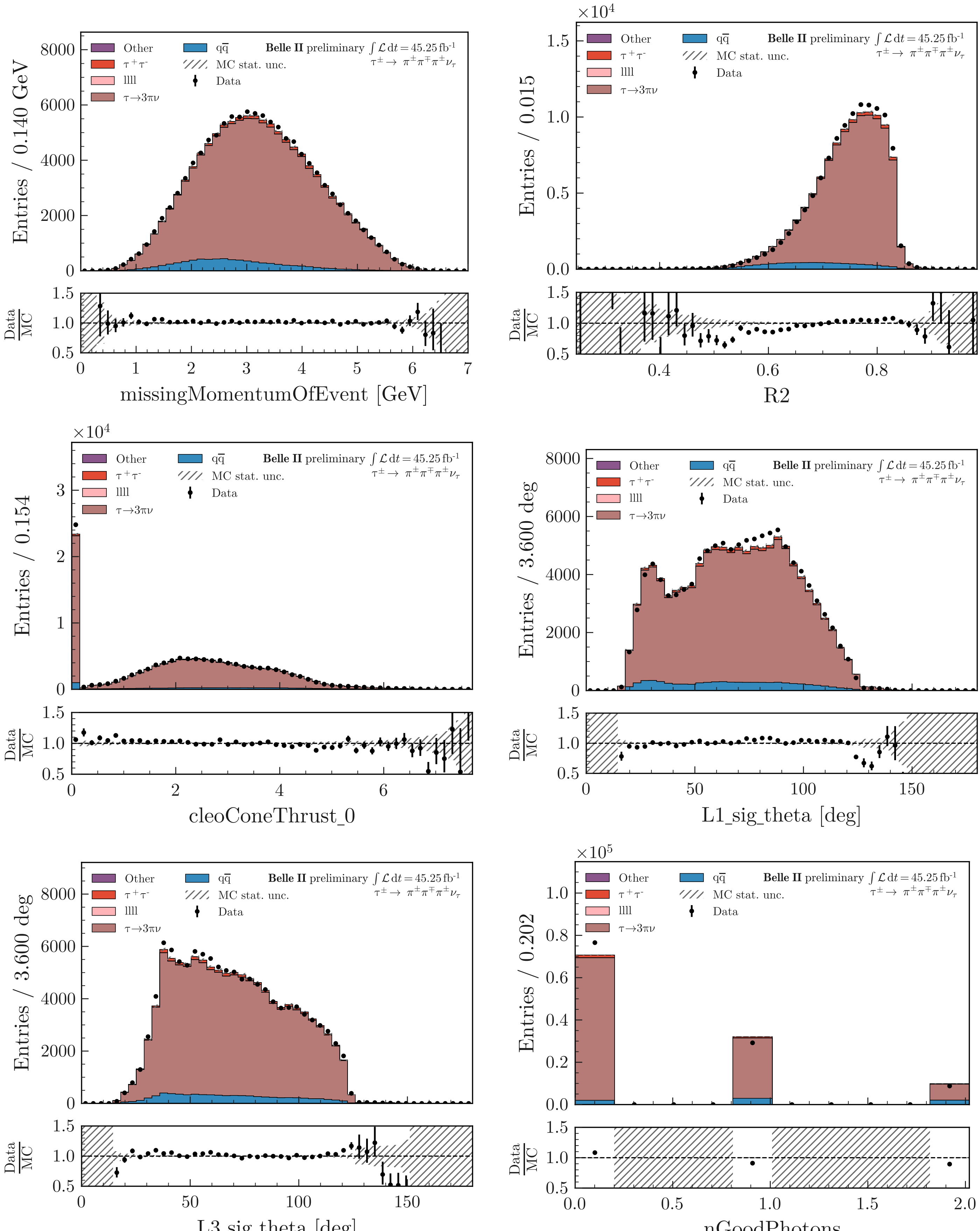


Figure A.7.: Data-MC comparison of the BDT input variables for the control channel

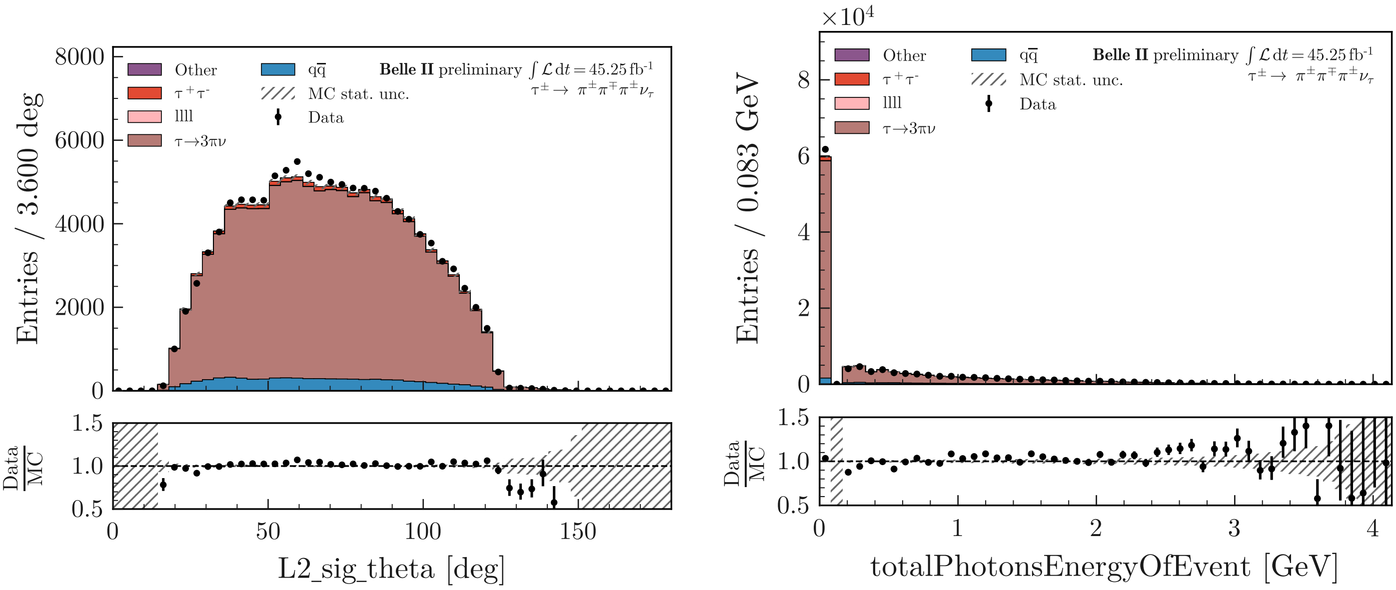


Figure A.8.: Data-MC comparison of the BDT input variables for the control channel

A.3. BDT appendix

A.3.1. Variable importance

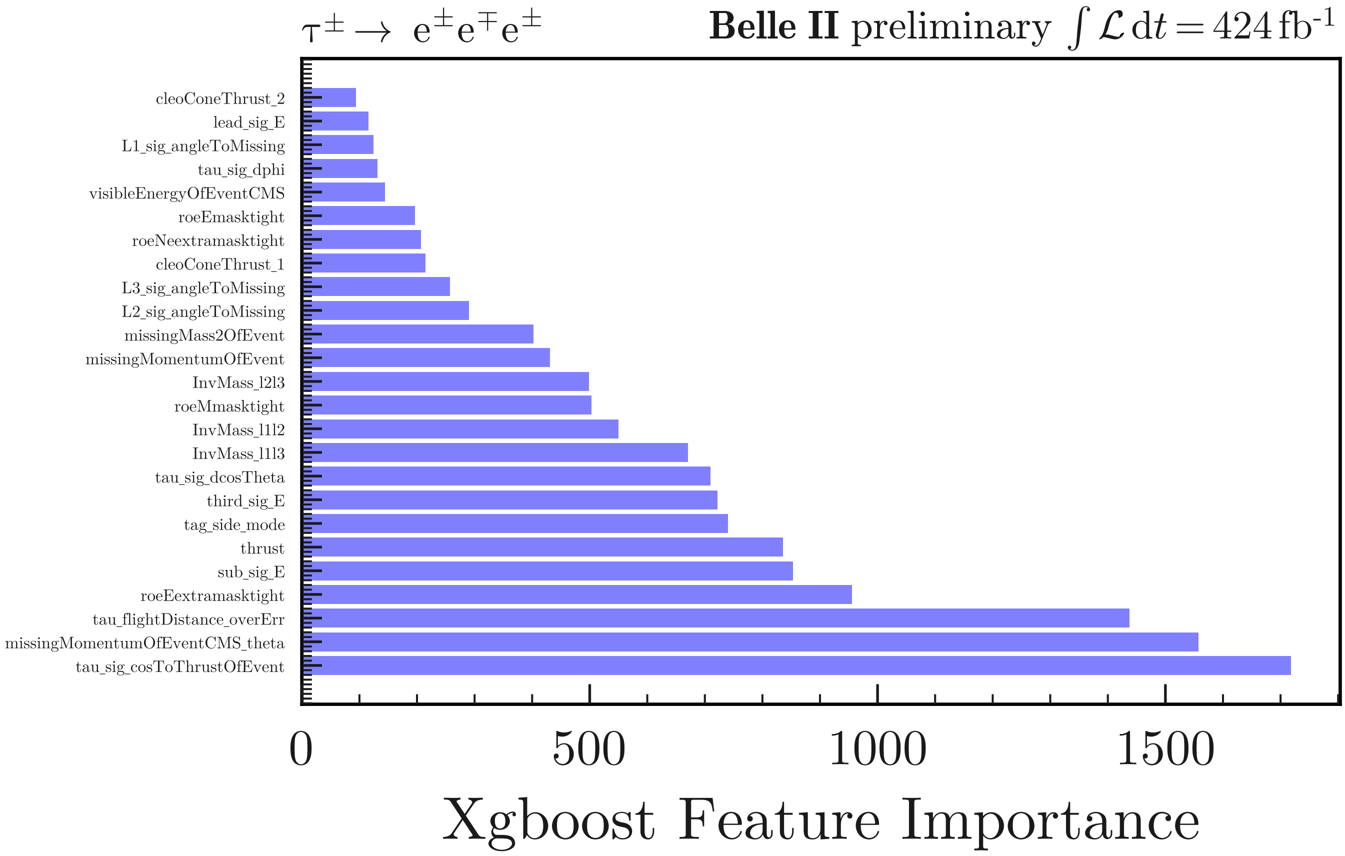


Figure A.9.: Importance of each variable in the BDT

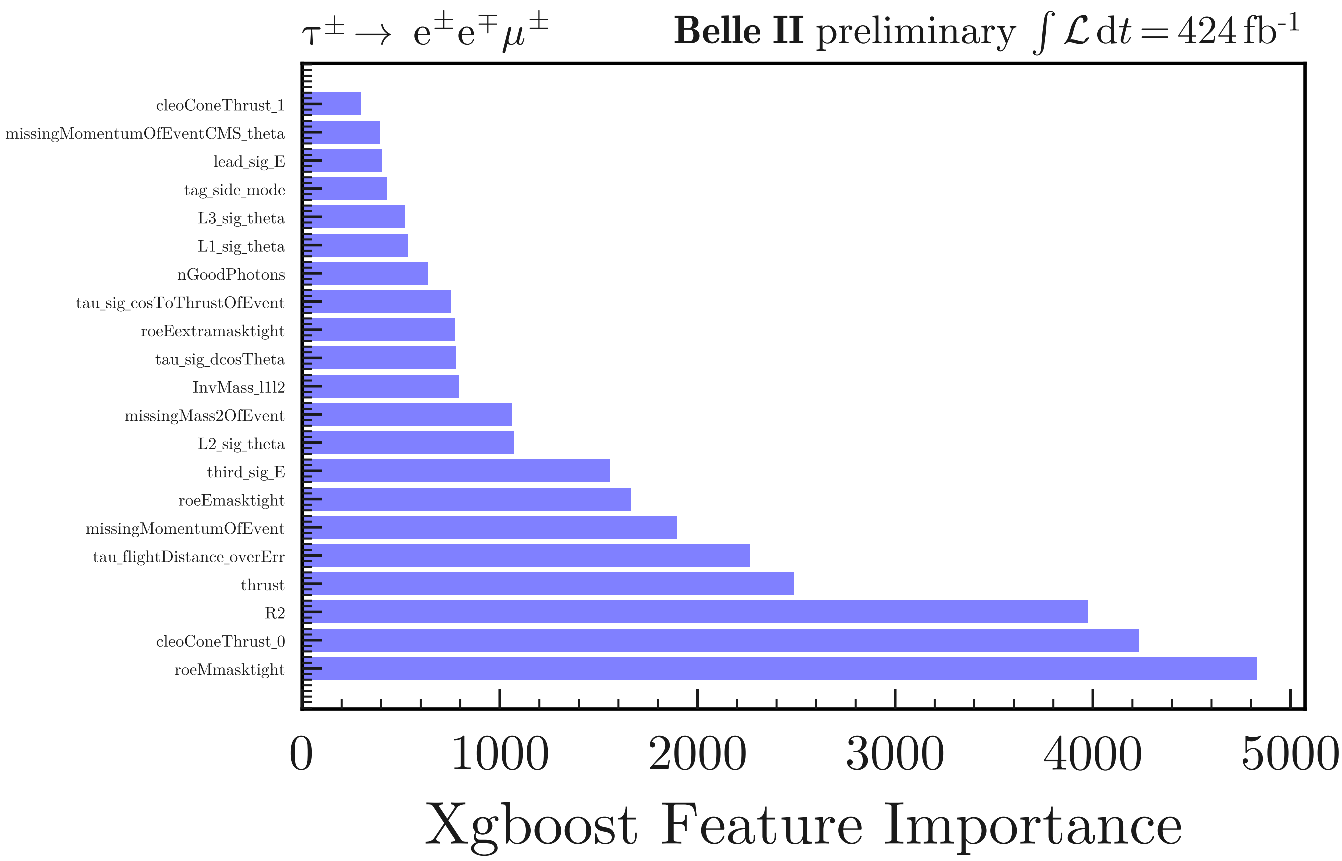


Figure A.10.: Importance of each variable in the BDT

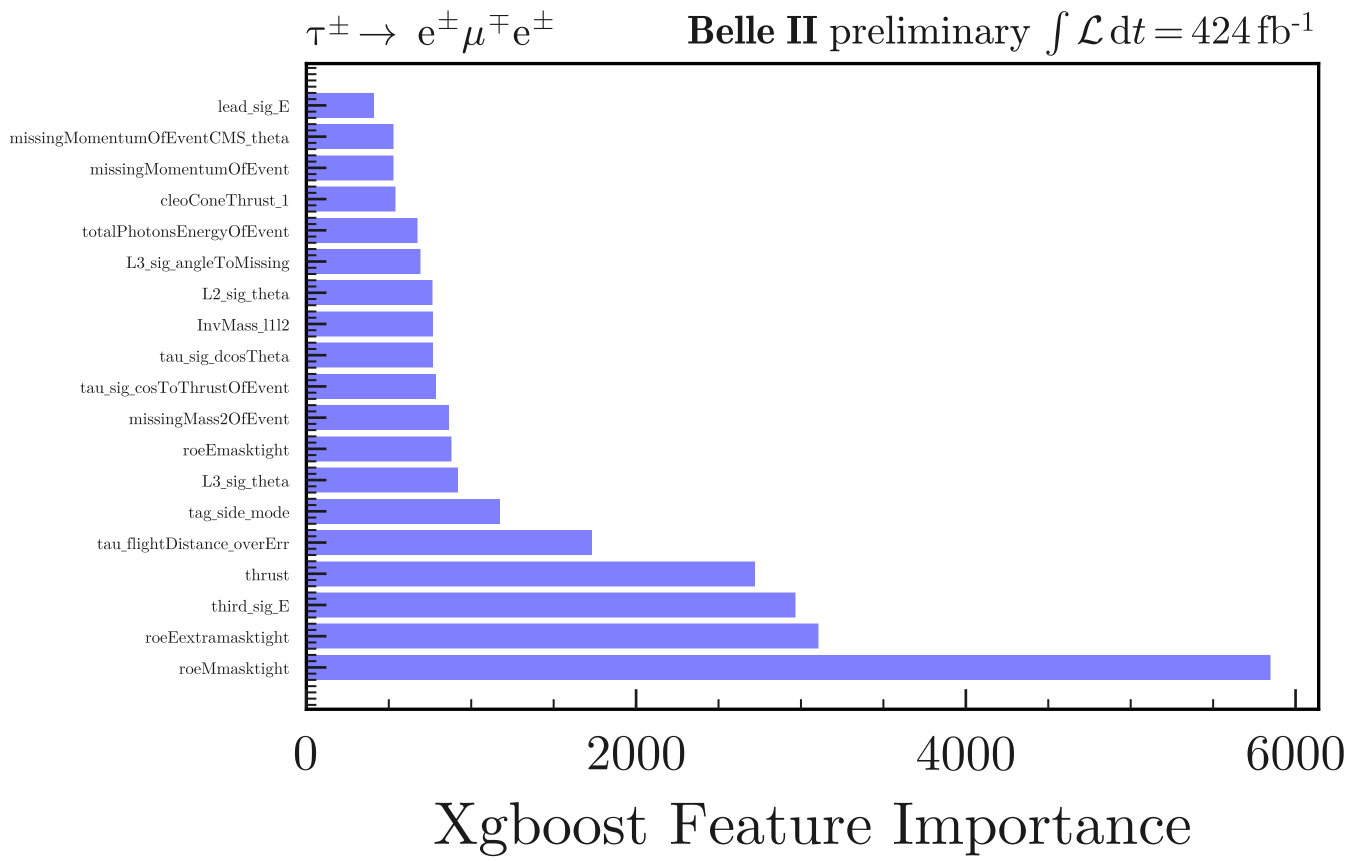


Figure A.11.: Importance of each variable in the BDT

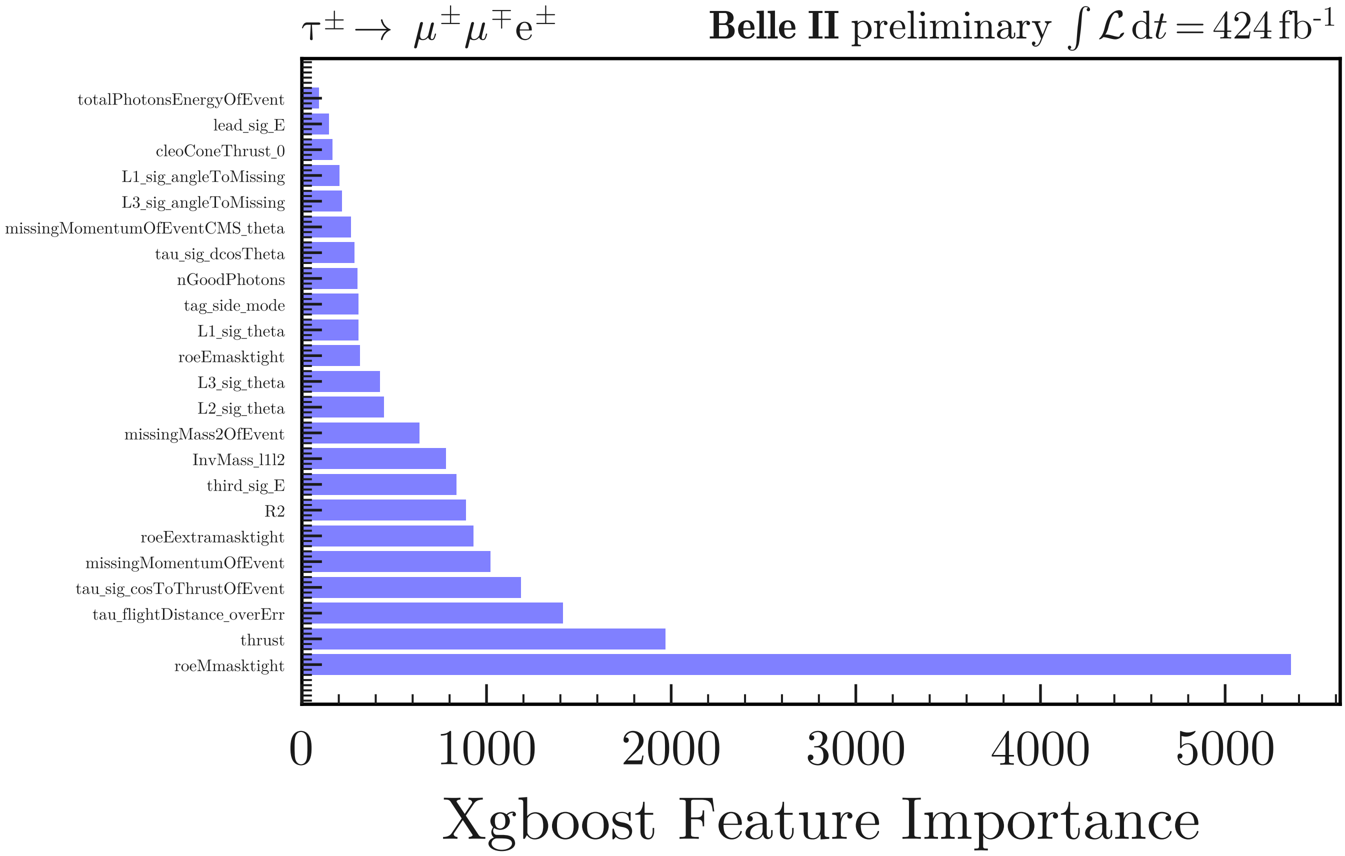


Figure A.12.: Importance of each variable in the BDT

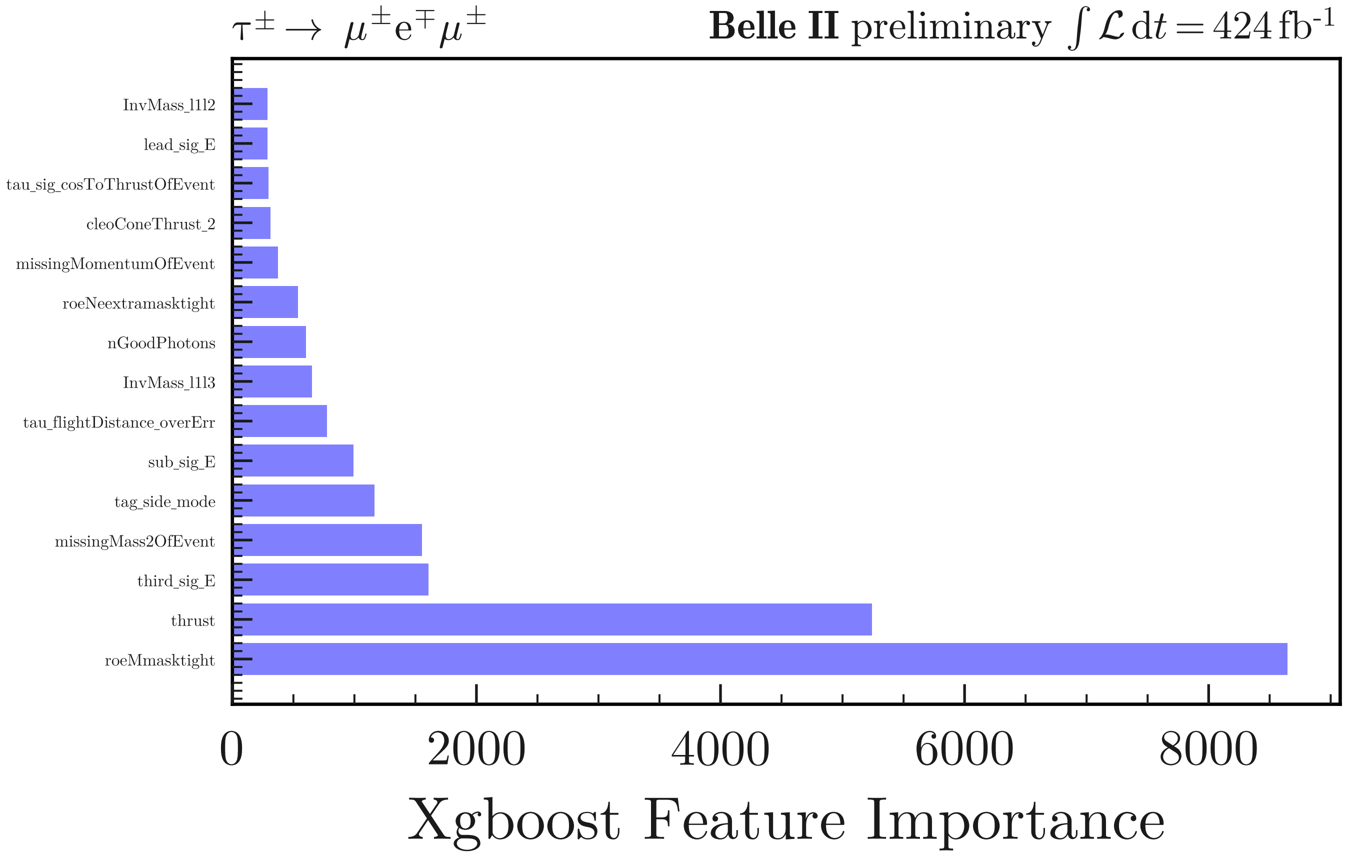


Figure A.13.: Importance of each variable in the BDT

A.3.2. Comparison of background and signal distributions

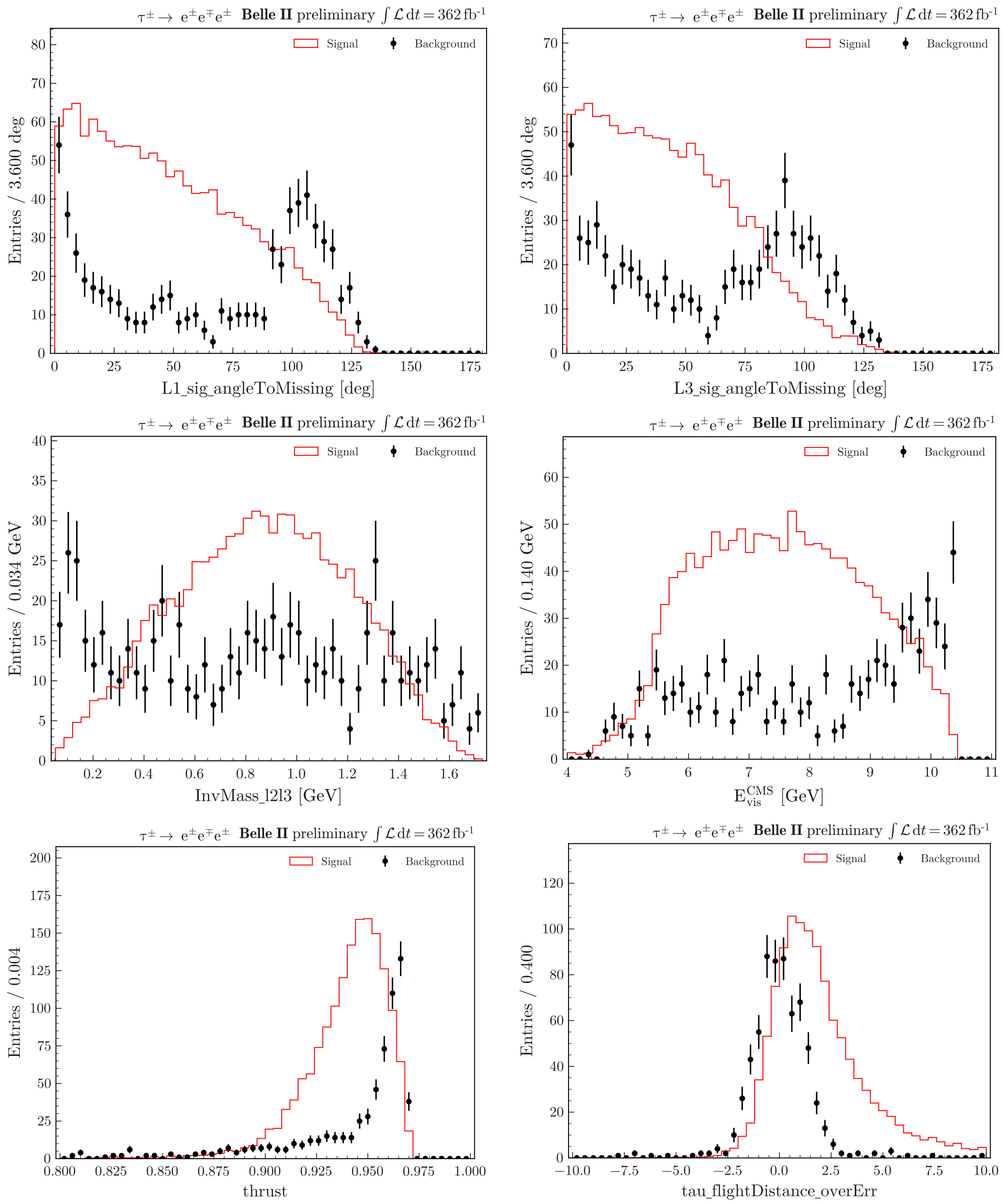


Figure A.14.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}e^{-}$ mode.

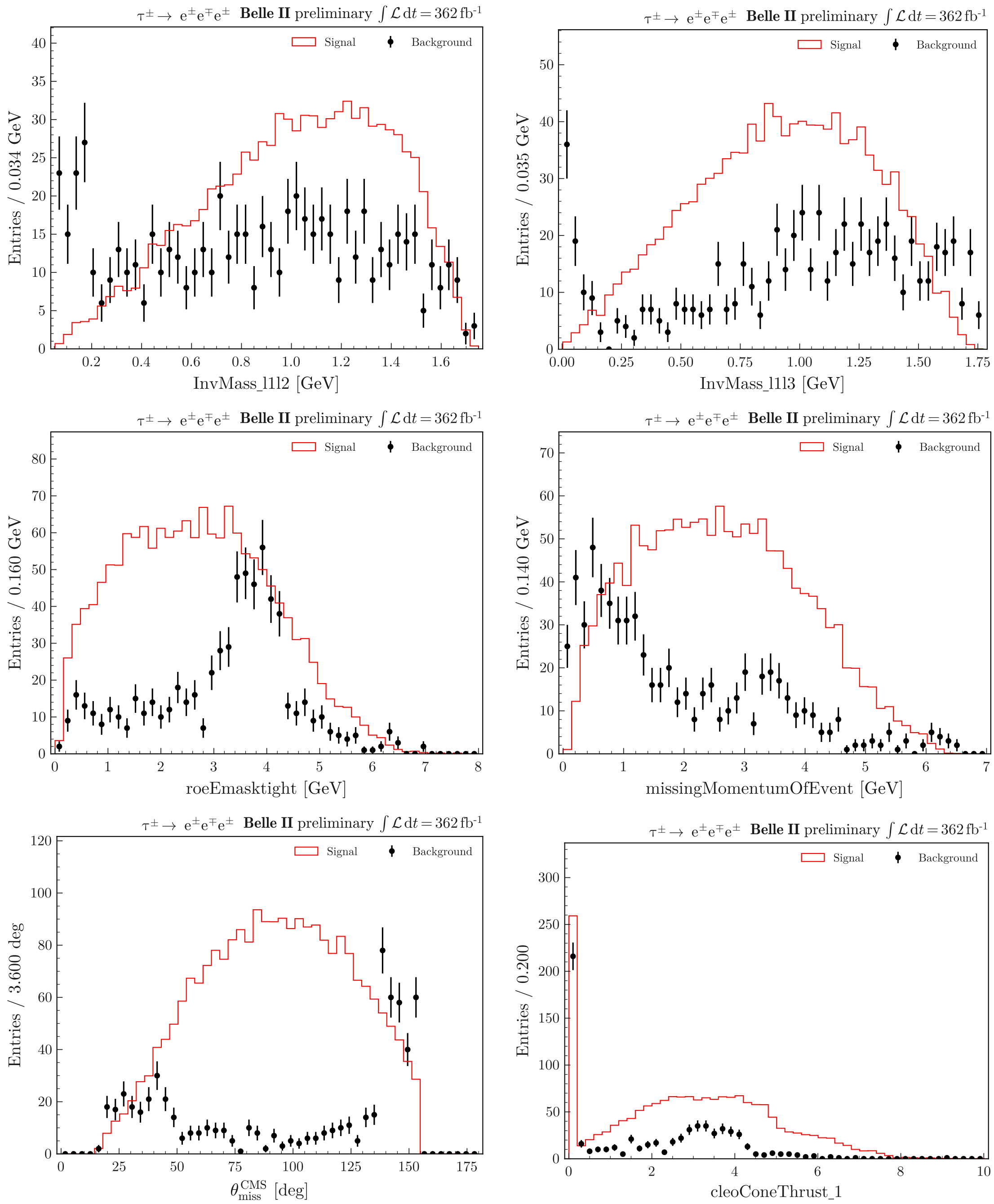


Figure A.15.: Comparison of background and signal BDT input variables for the $e^-e^+e^-$ mode.

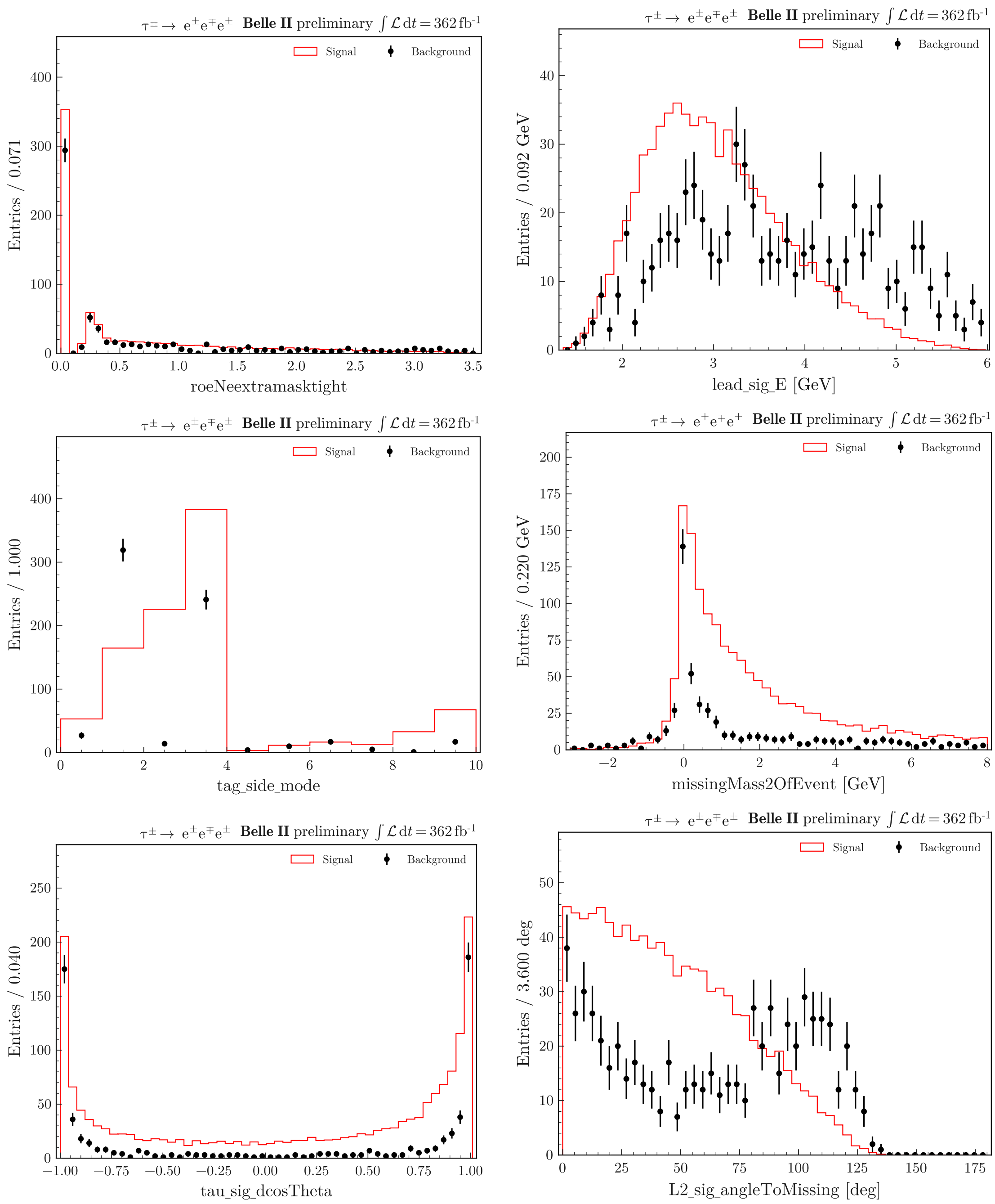


Figure A.16.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}e^{-}$ mode.

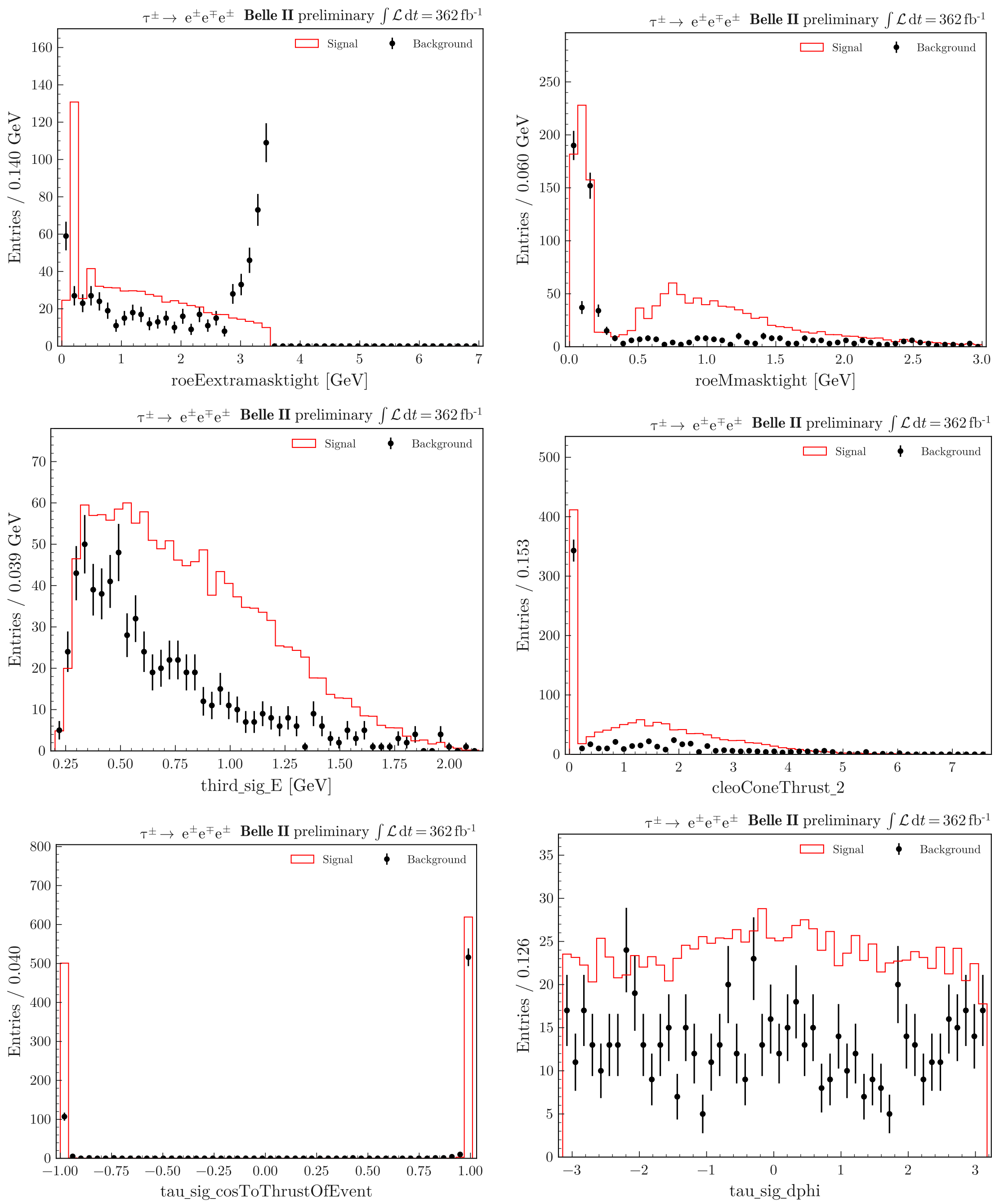


Figure A.17.: Comparison of background and signal BDT input variables for the $e^-e^+e^-$ mode.

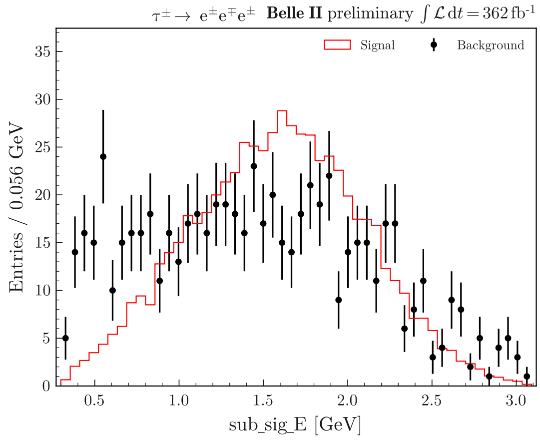


Figure A.18.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}e^{-}$ mode.

2267 ### A.3.3. Correlation matrices

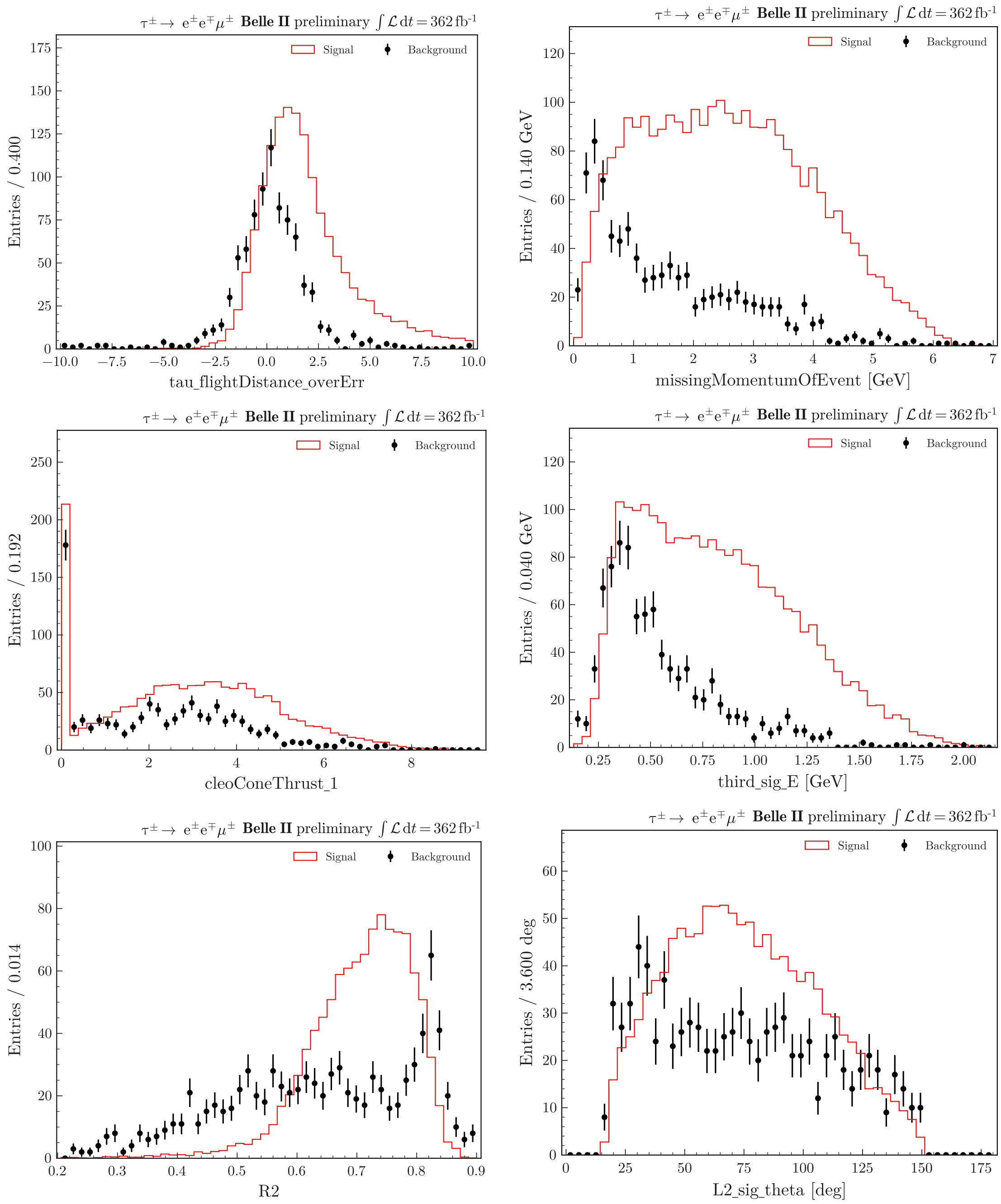


Figure A.19.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}\mu^{-}$ mode.

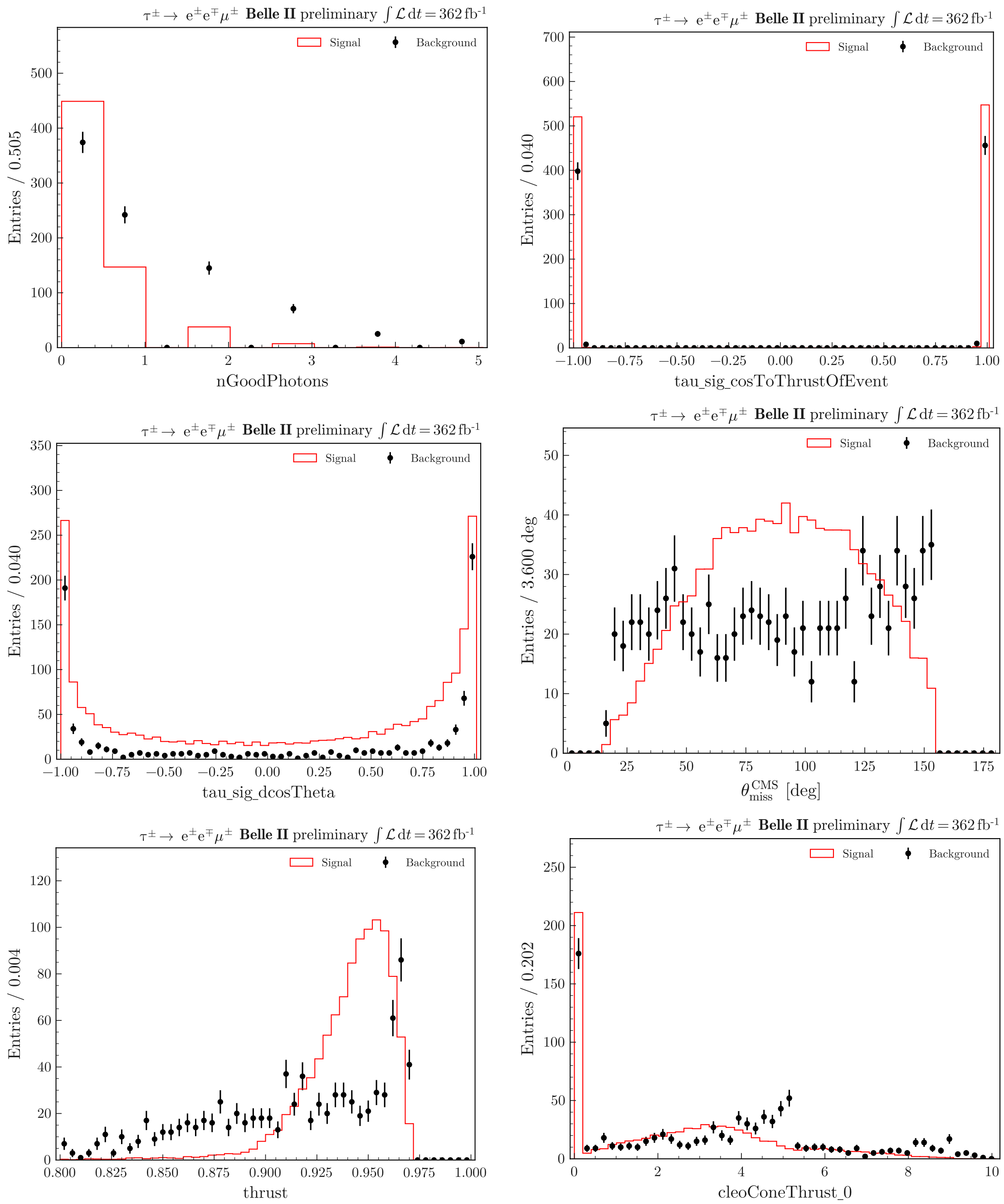


Figure A.20.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}\mu^{-}$ mode.

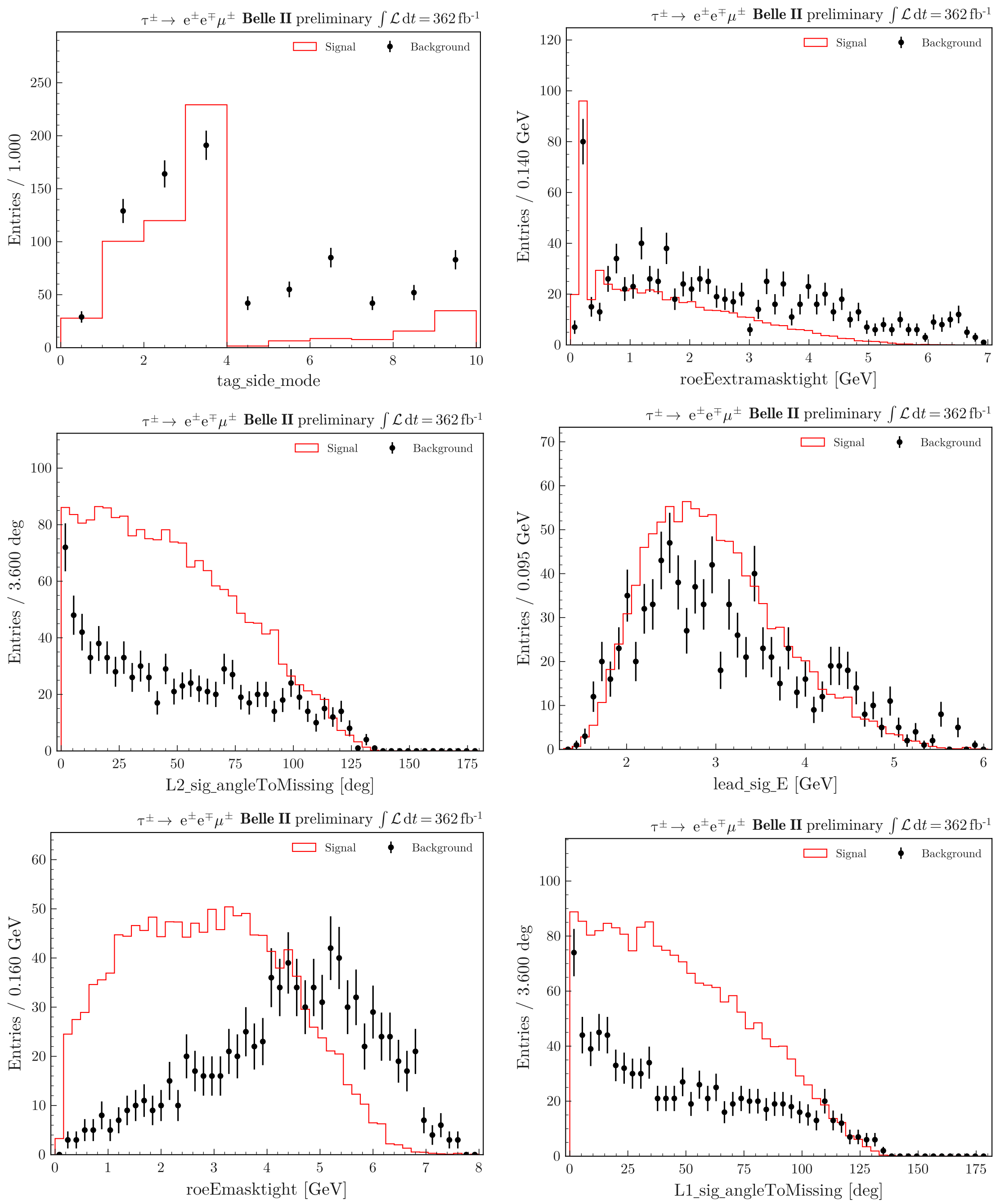


Figure A.21.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}\mu^{-}$ mode.

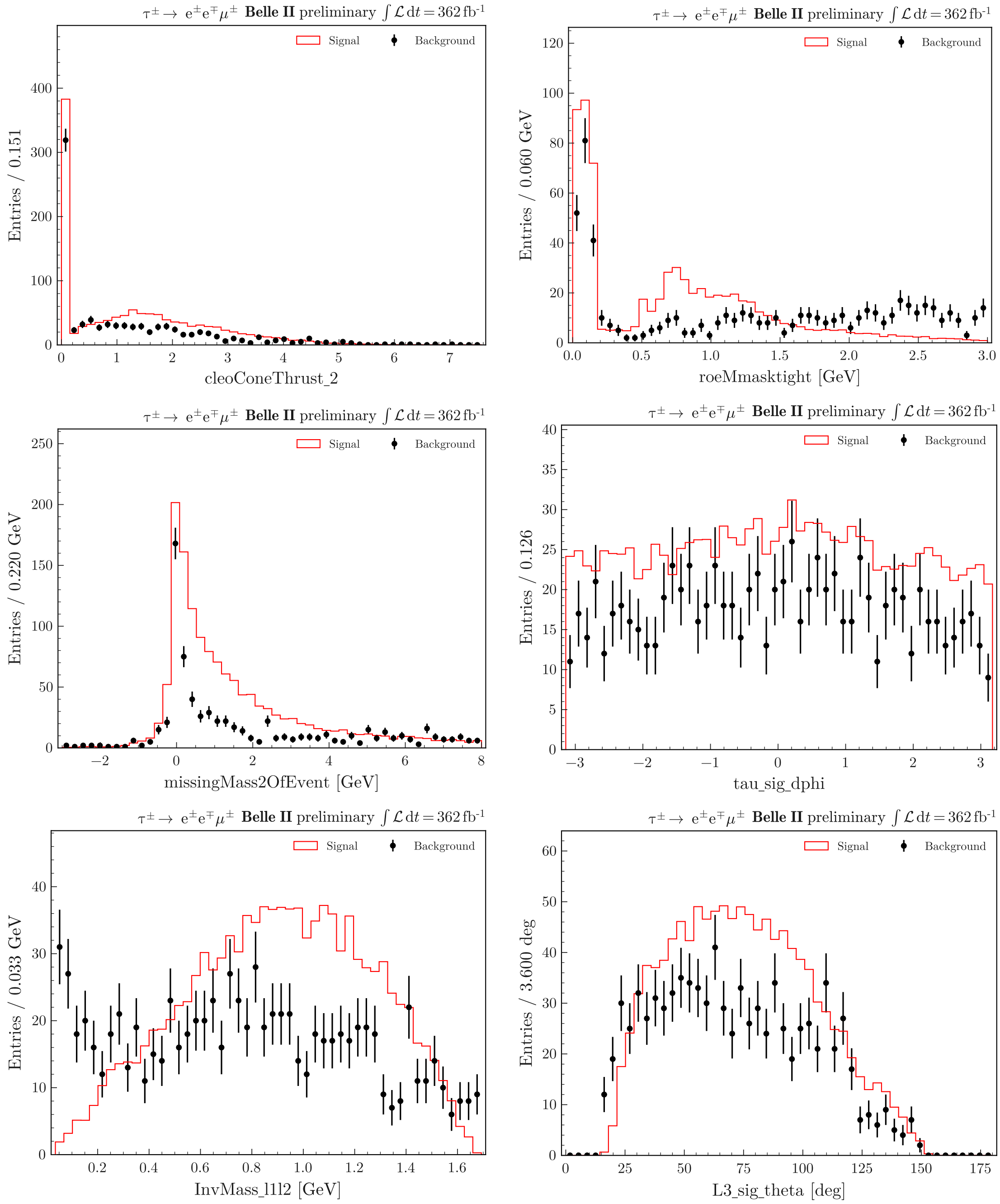


Figure A.22.: Comparison of background and signal BDT input variables for the $e^-e^+\mu^-$ mode.

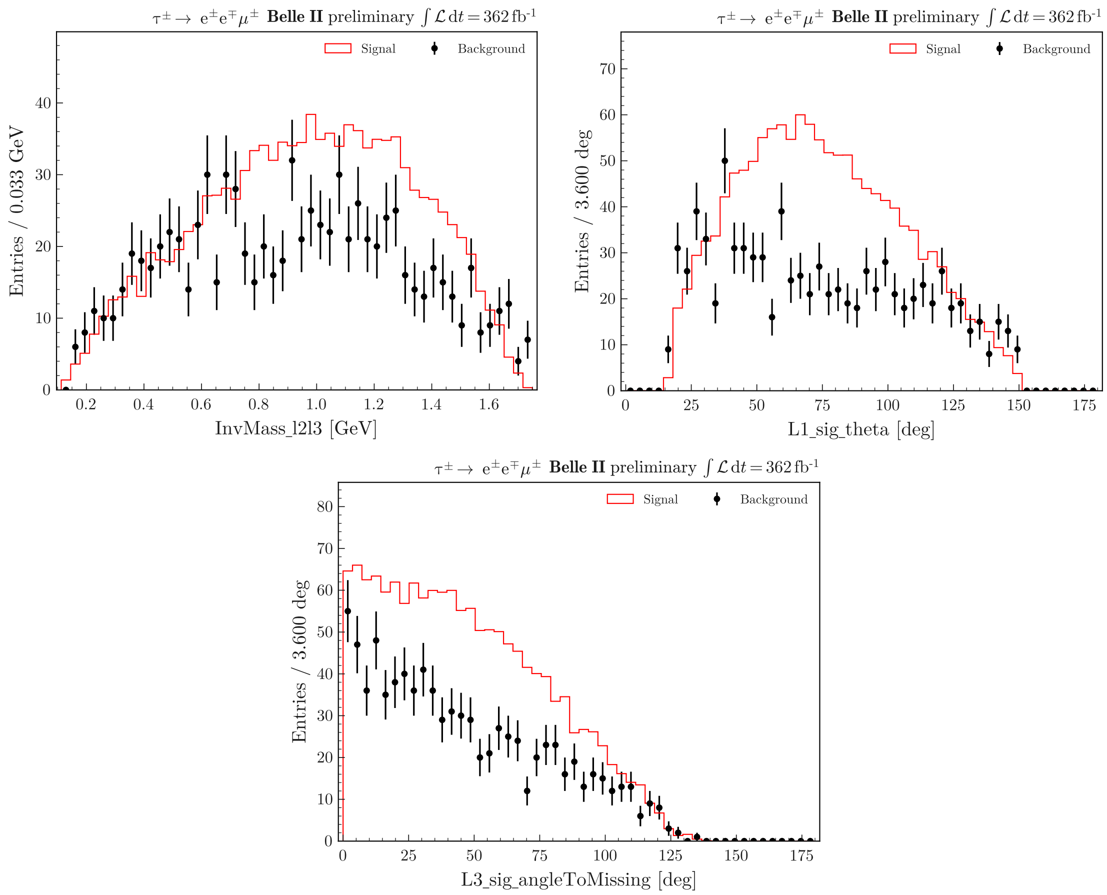


Figure A.23.: Comparison of background and signal BDT input variables for the $e^{-}e^{+}\mu^{-}$ mode.

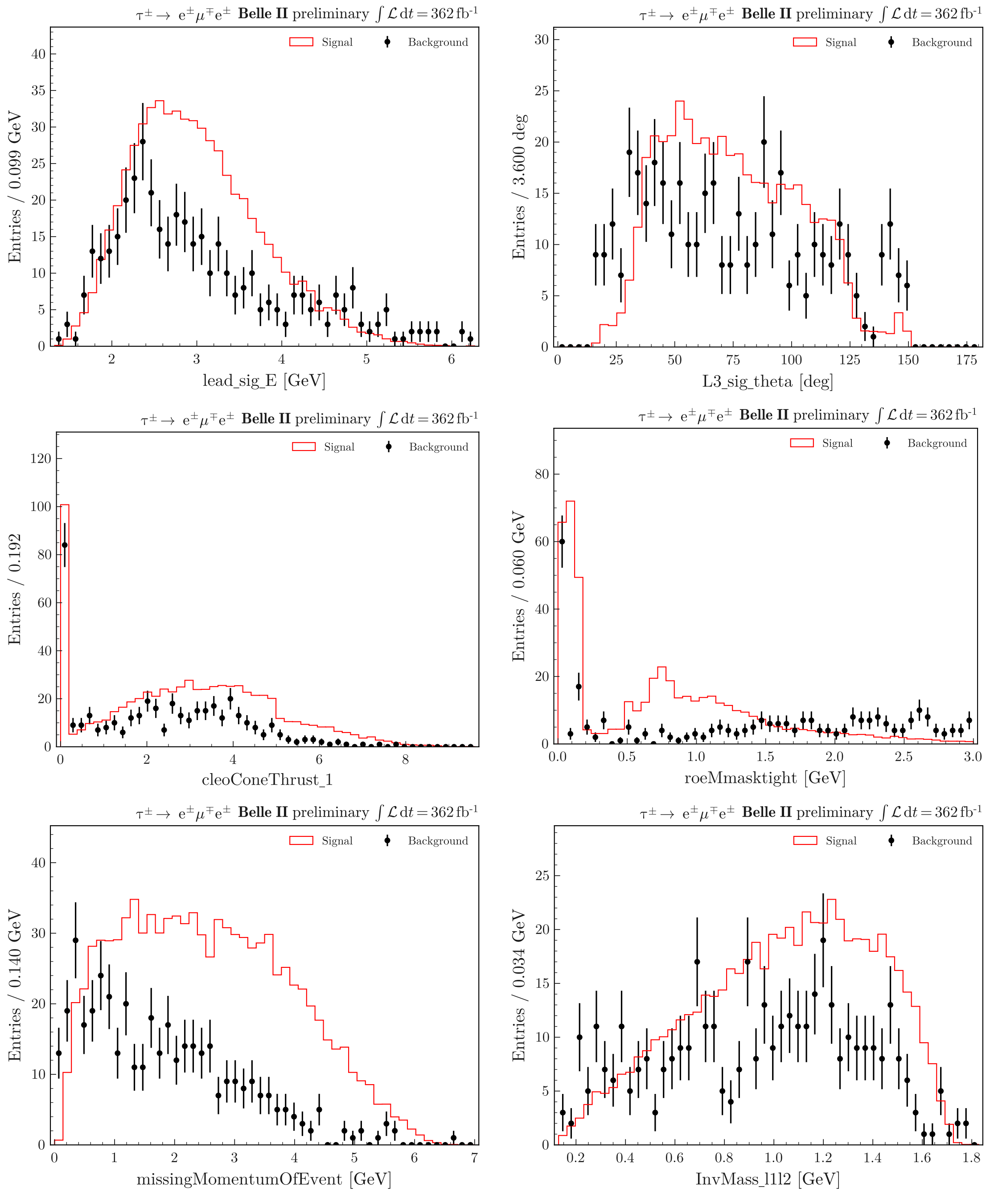


Figure A.24.: Comparison of background and signal BDT input variables for the $e^{-}\mu^{+}e^{-}$ mode.

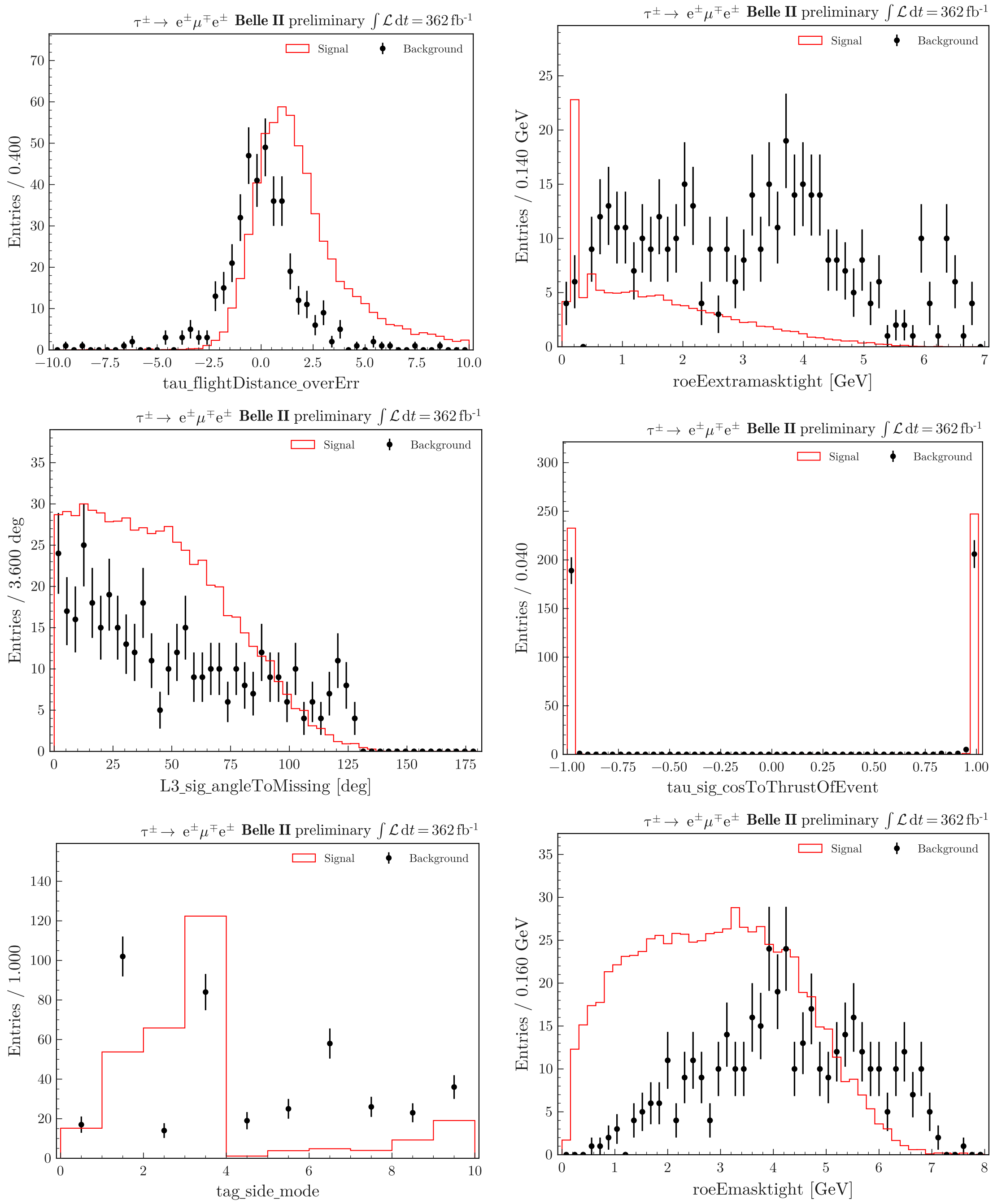


Figure A.25.: Comparison of background and signal BDT input variables for the $e^-\mu^+e^-$ mode.

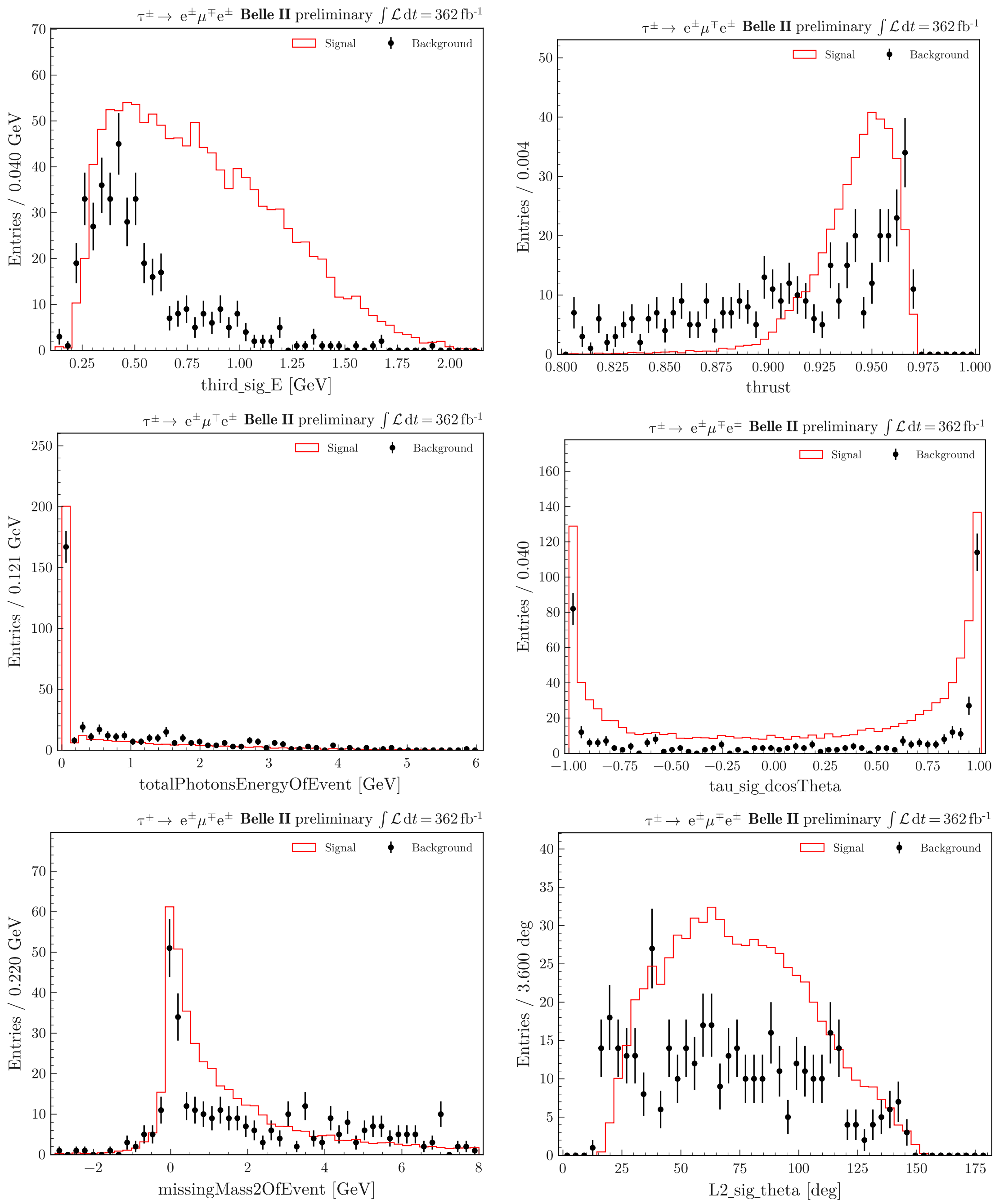


Figure A.26.: Comparison of background and signal BDT input variables for the $e^{-}\mu^{+}e^{-}$ mode.

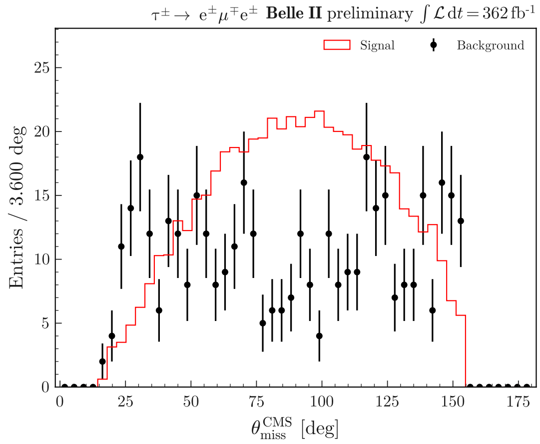


Figure A.27.: Comparison of background and signal BDT input variables for the $e^{-}\mu^{+}e^{-}$ mode.

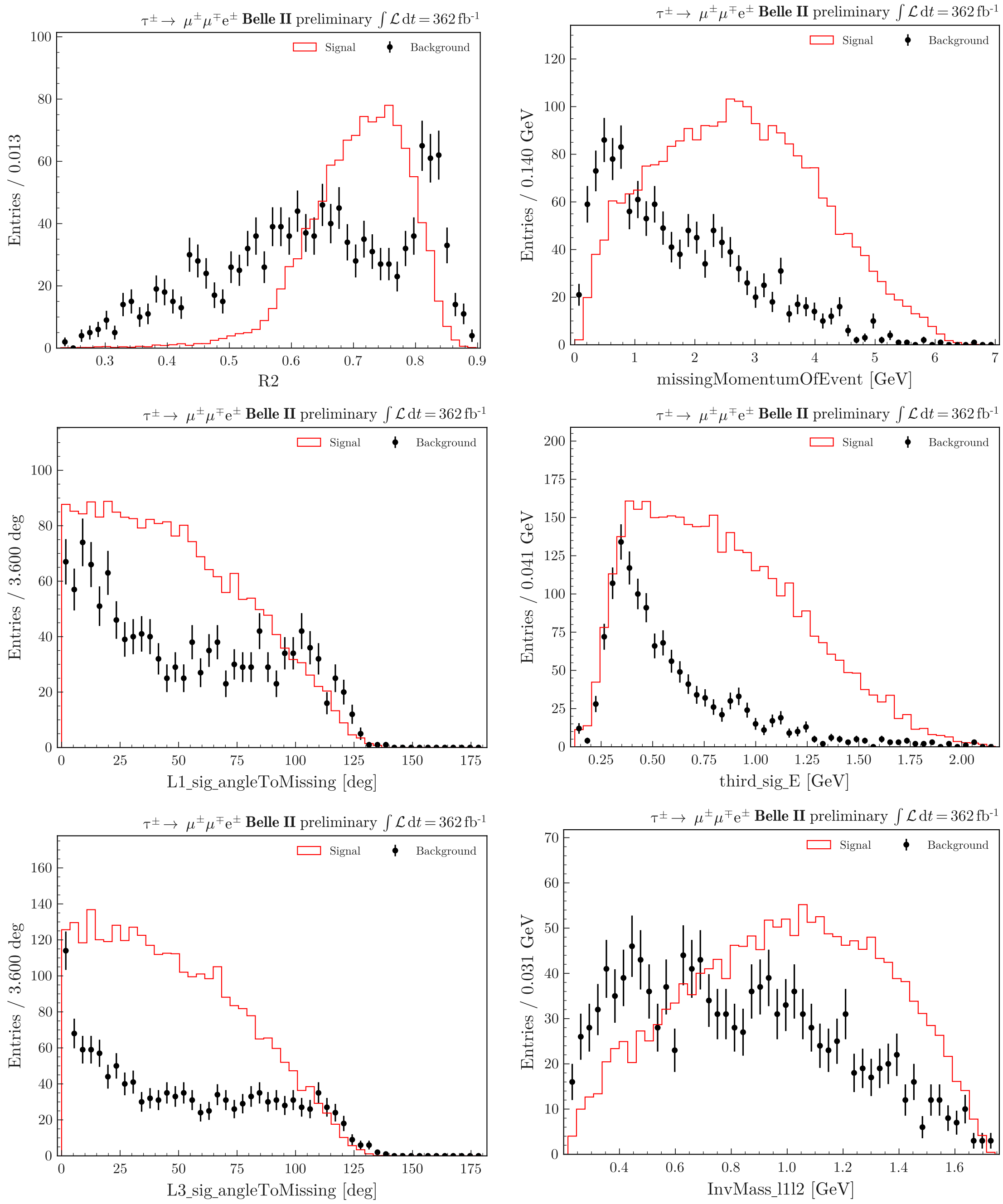


Figure A.28.: Comparison of background and signal BDT input variables for the $\mu^-\mu^+e^-$ mode.

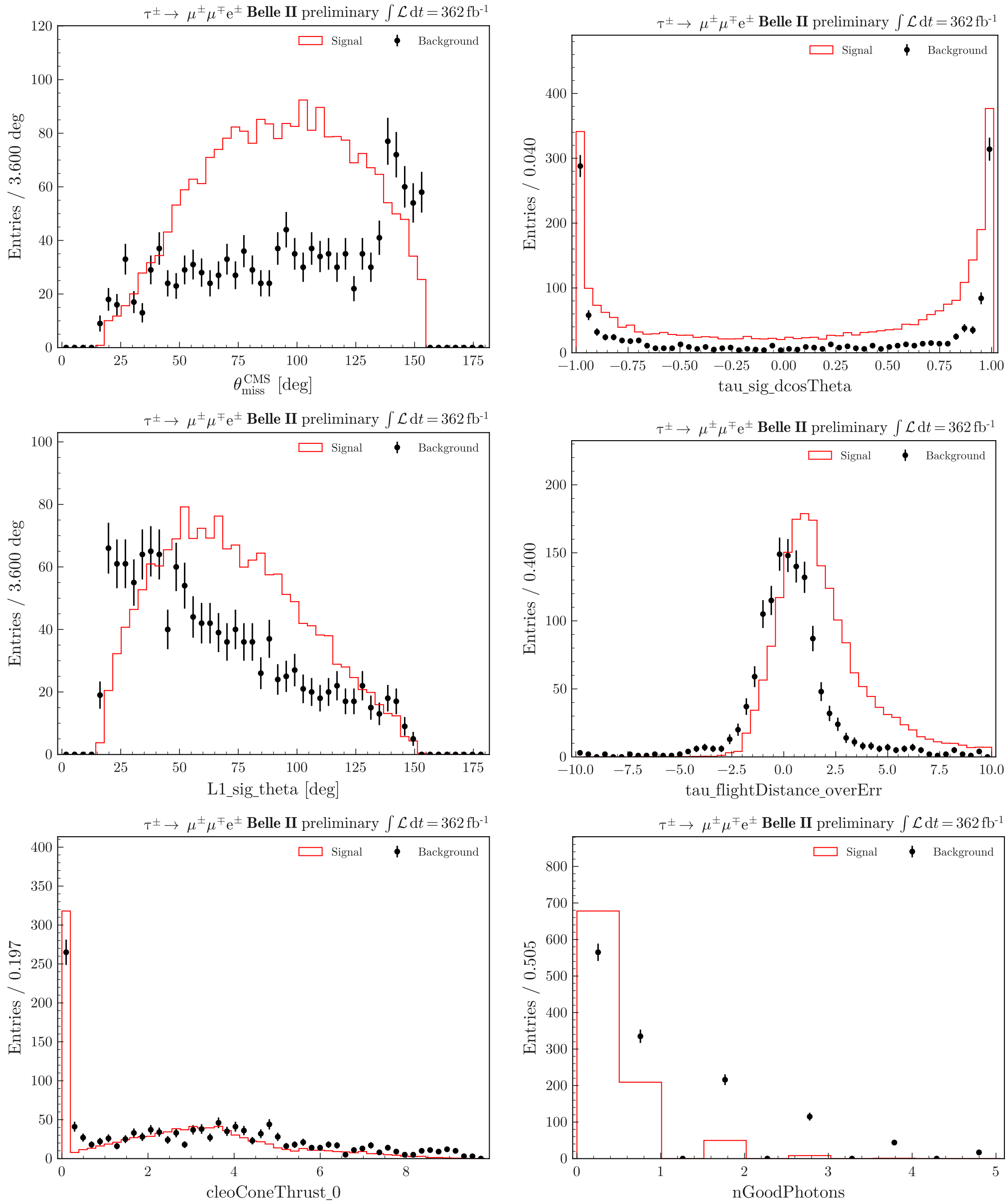


Figure A.29.: Comparison of background and signal BDT input variables for the $\mu^-\mu^+e^-$ mode.

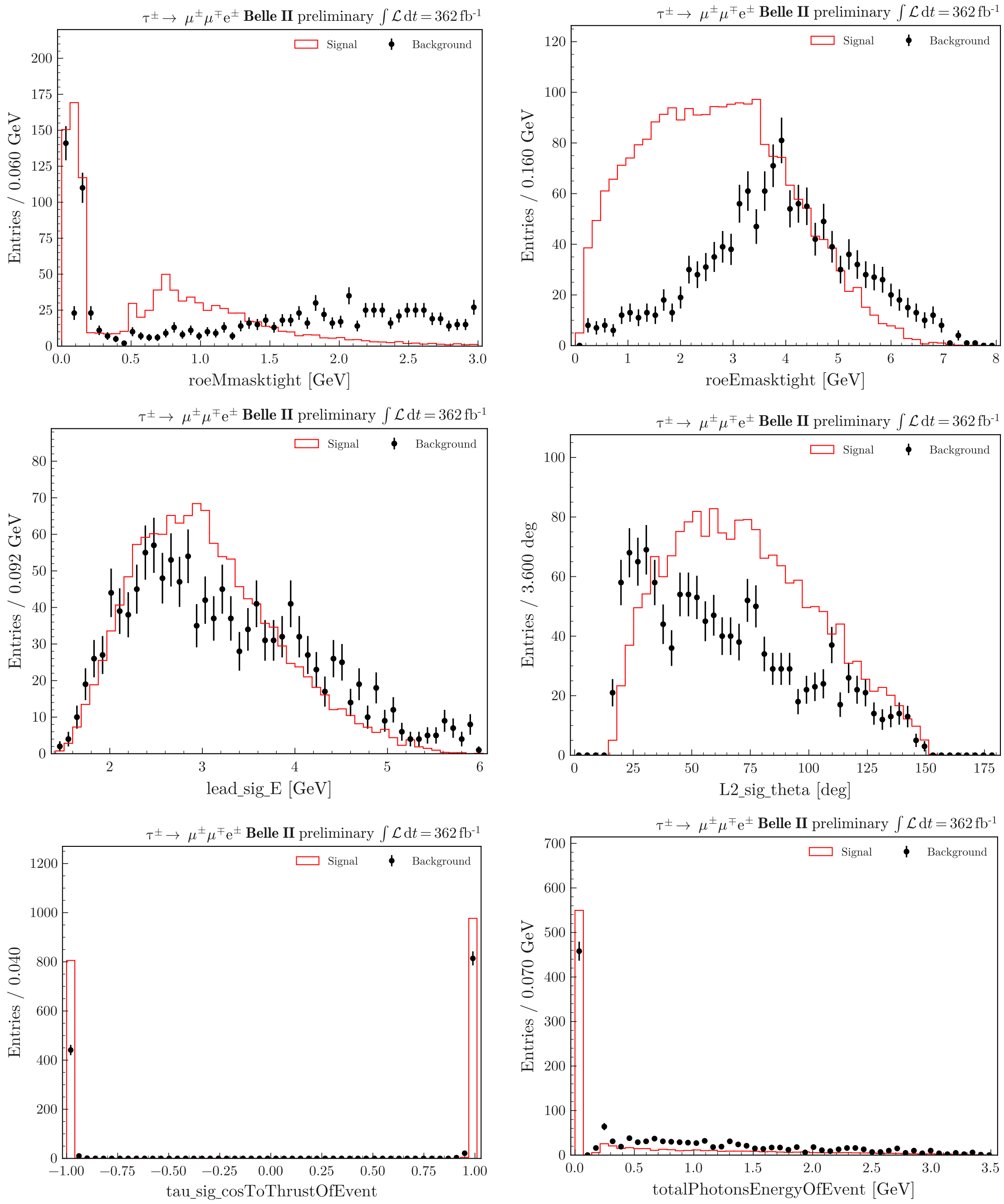


Figure A.30.: Comparison of background and signal BDT input variables for the $\mu^-\mu^+e^-$ mode.

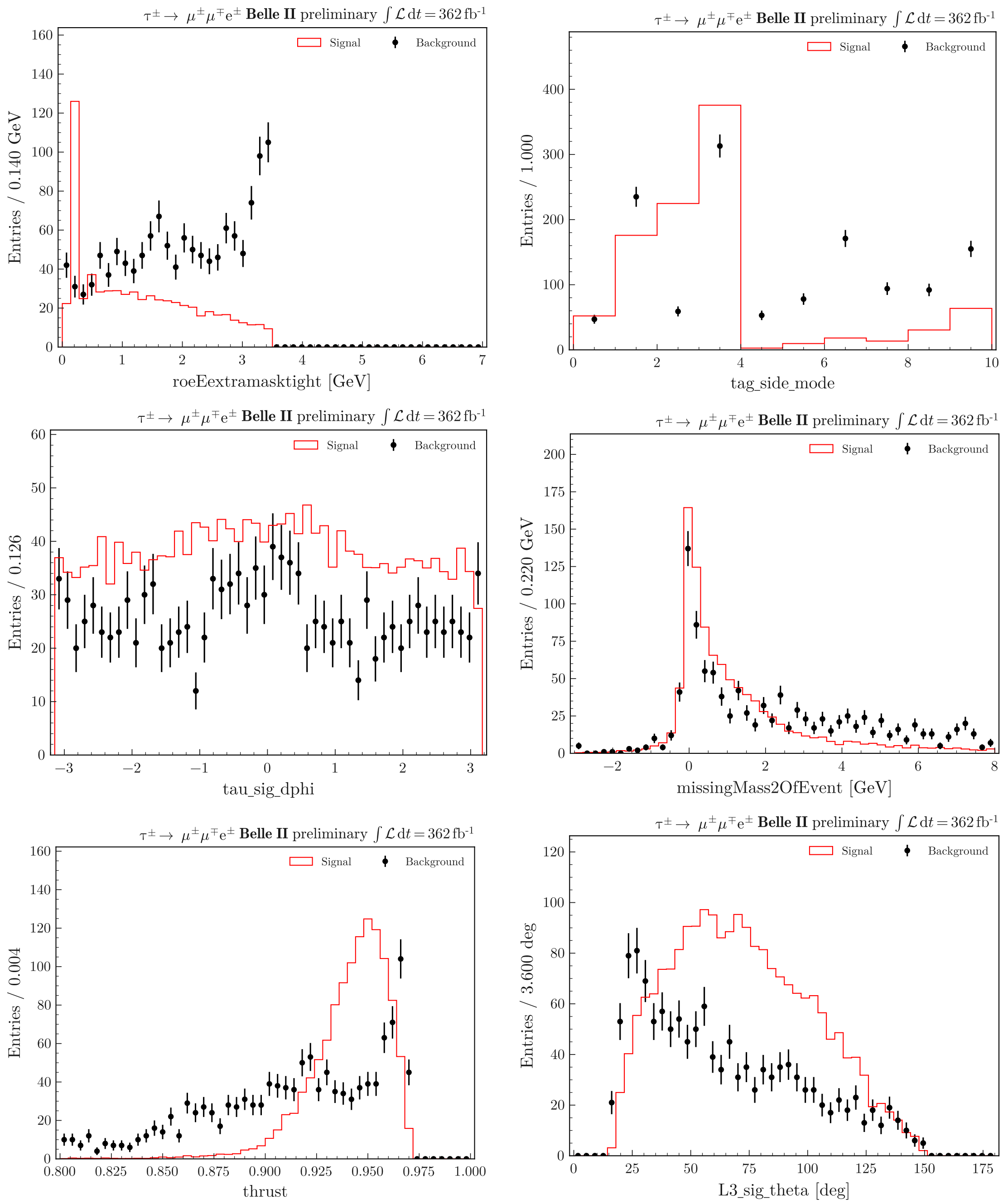


Figure A.31.: Comparison of background and signal BDT input variables for the $\mu^-\mu^+e^-$ mode.

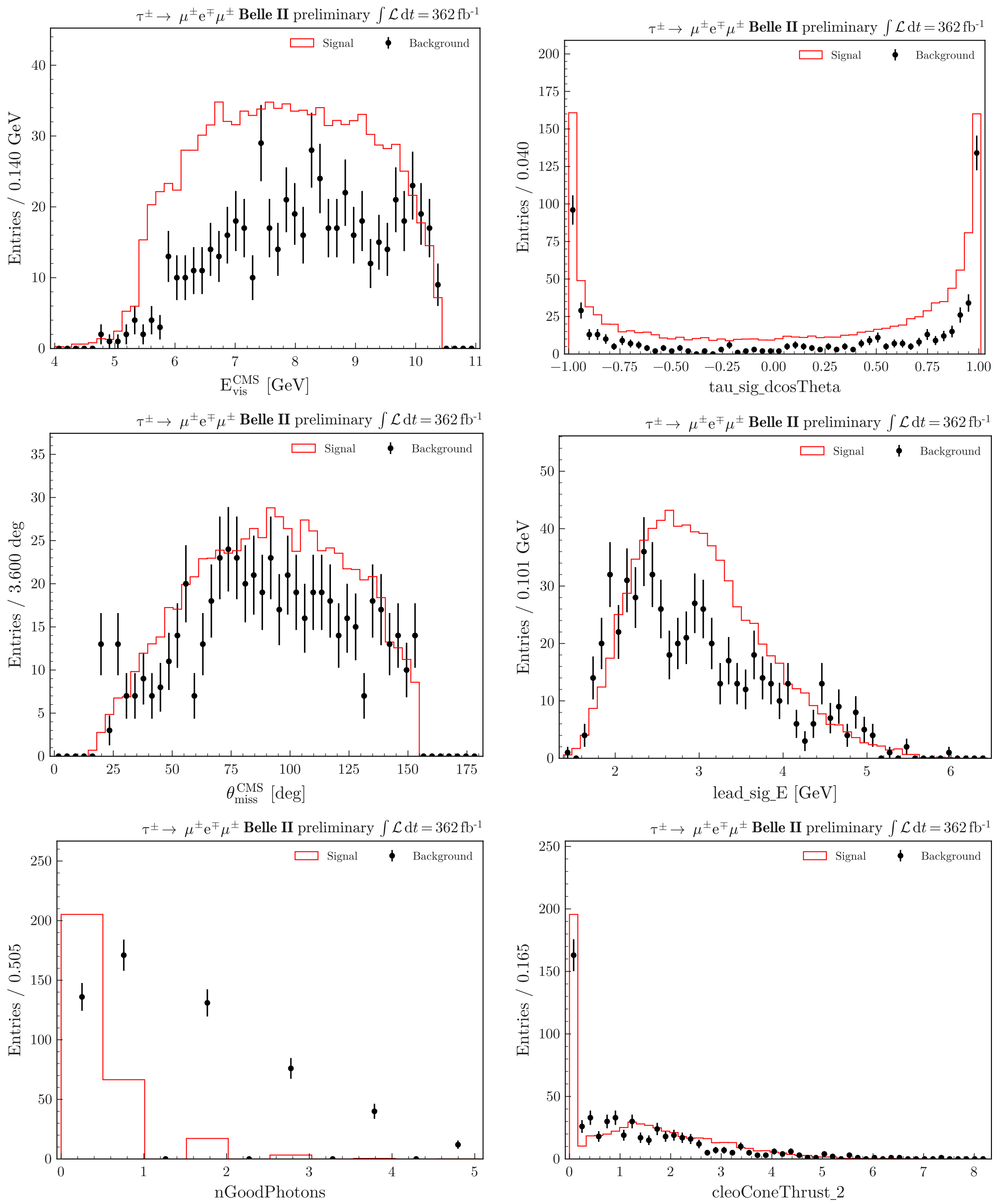


Figure A.32.: Comparison of background and signal BDT input variables for the $\mu^-e^+\mu^-$ mode.

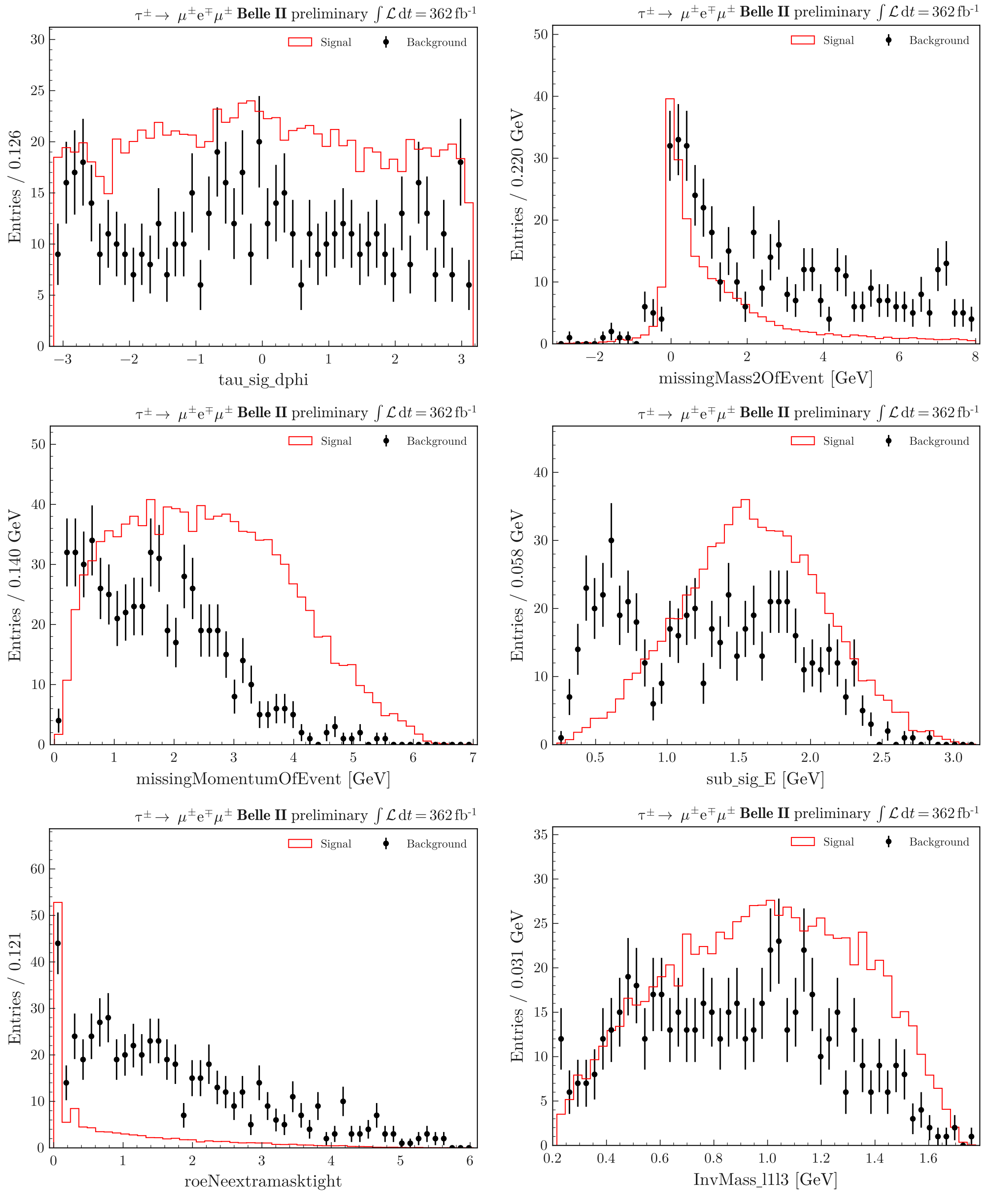


Figure A.33.: Comparison of background and signal BDT input variables for the $\mu^{-}e^{+}\mu^{-}$ mode.

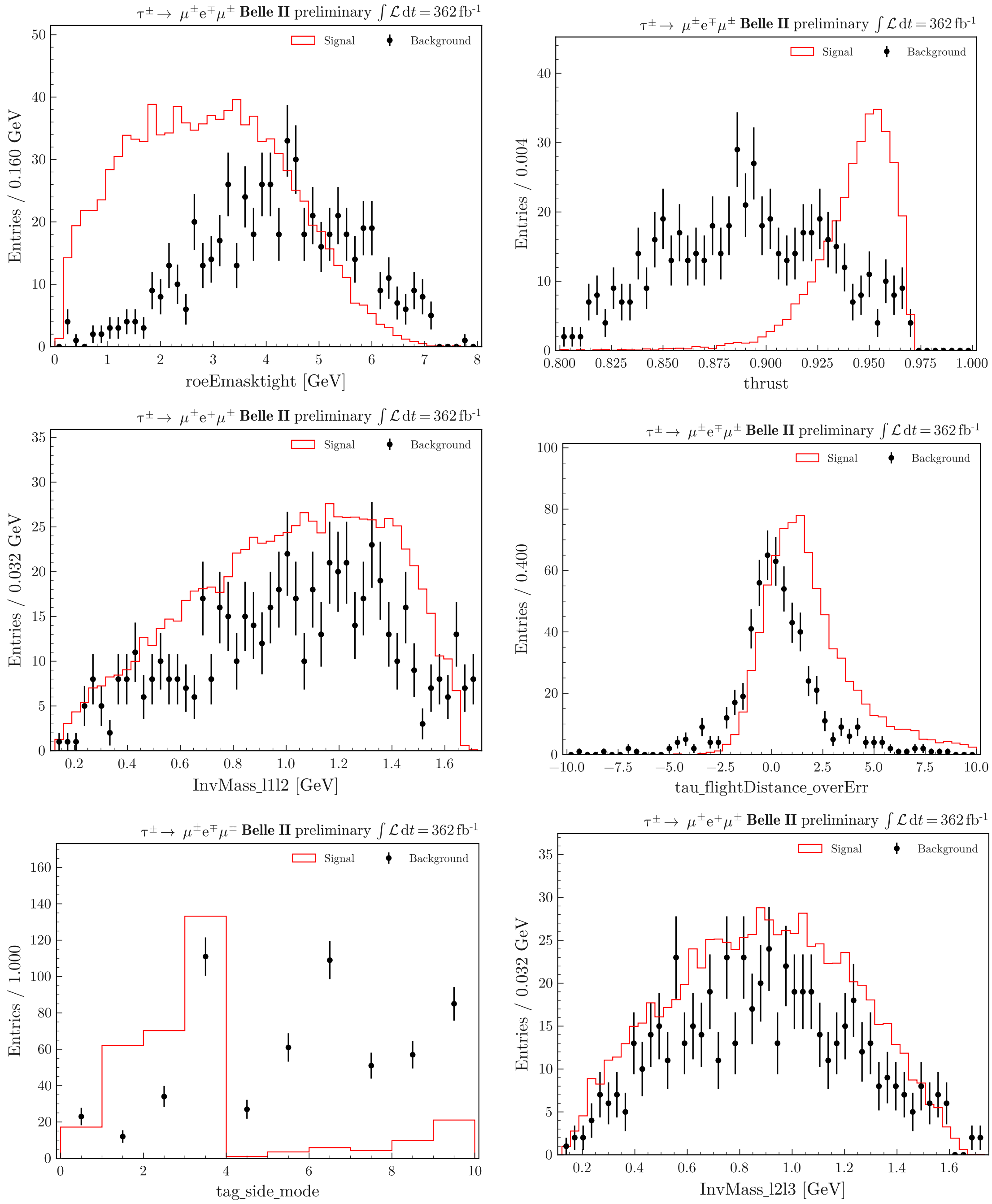


Figure A.34.: Comparison of background and signal BDT input variables for the $\mu^- e^+ \mu^-$ mode.

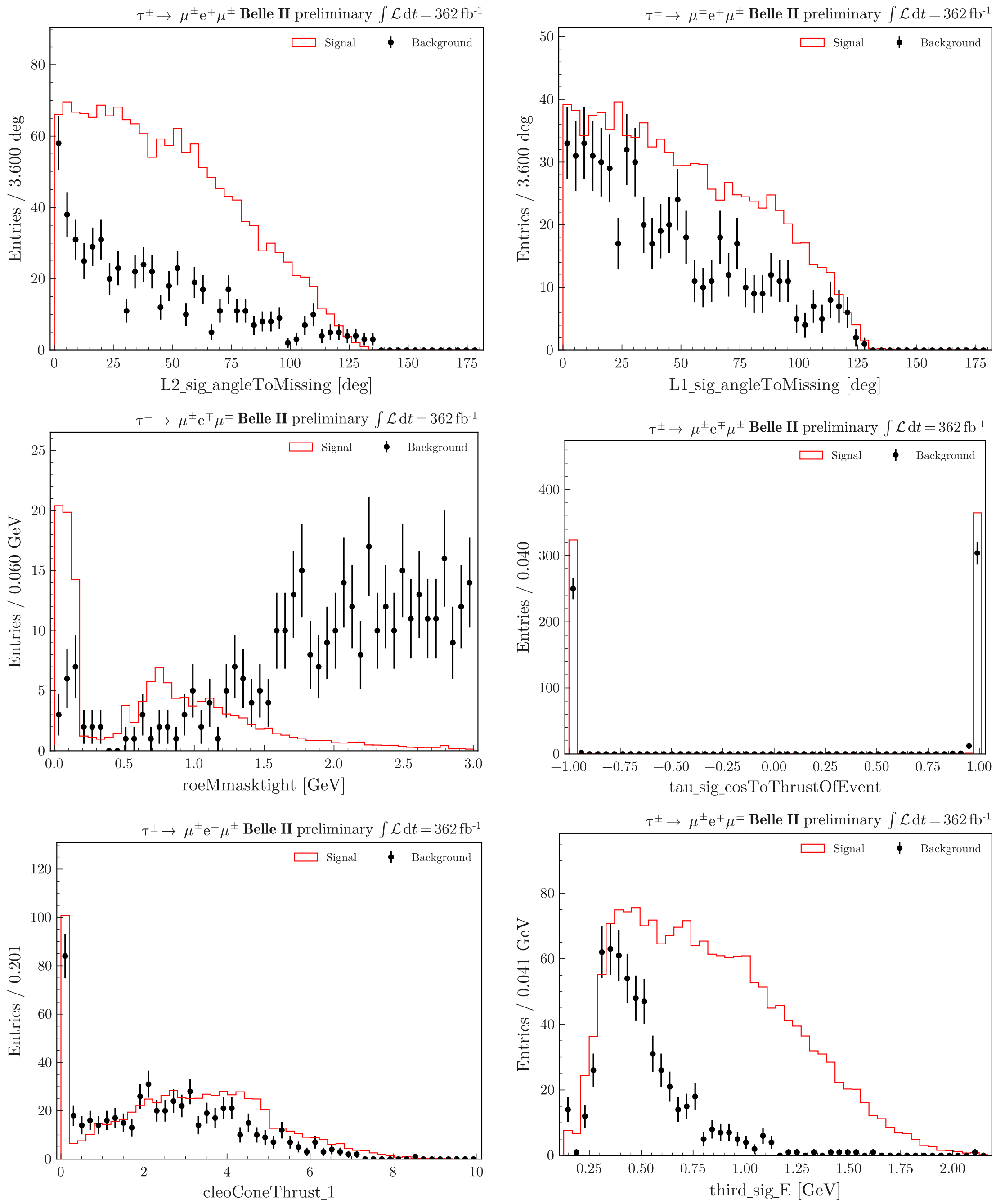


Figure A.35.: Comparison of background and signal BDT input variables for the $\mu^-e^+\mu^-$ mode.

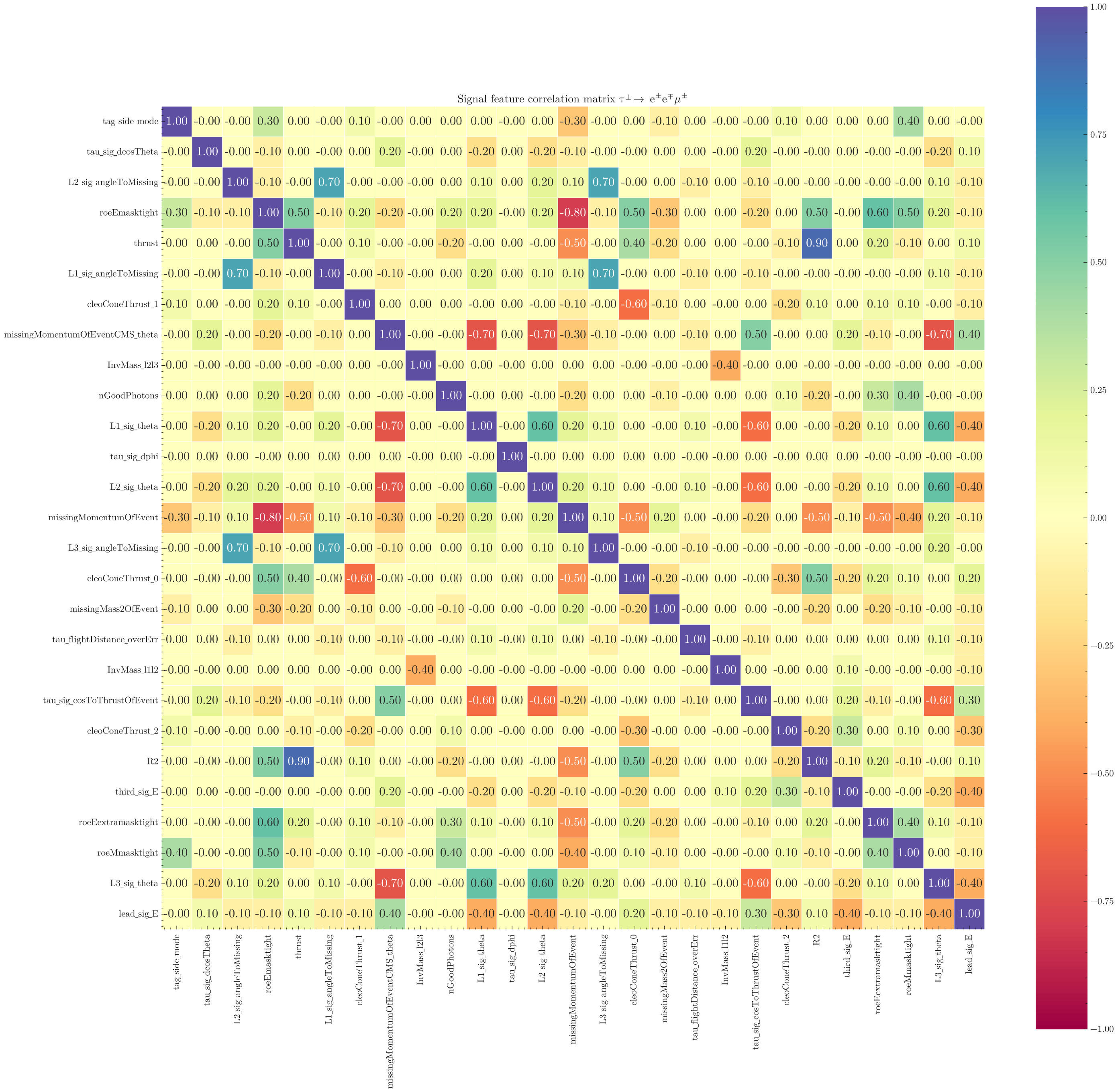


Figure A.36.: Correlation matrix for the signal, $e^{-}e^{+}\mu^{-}$ mode

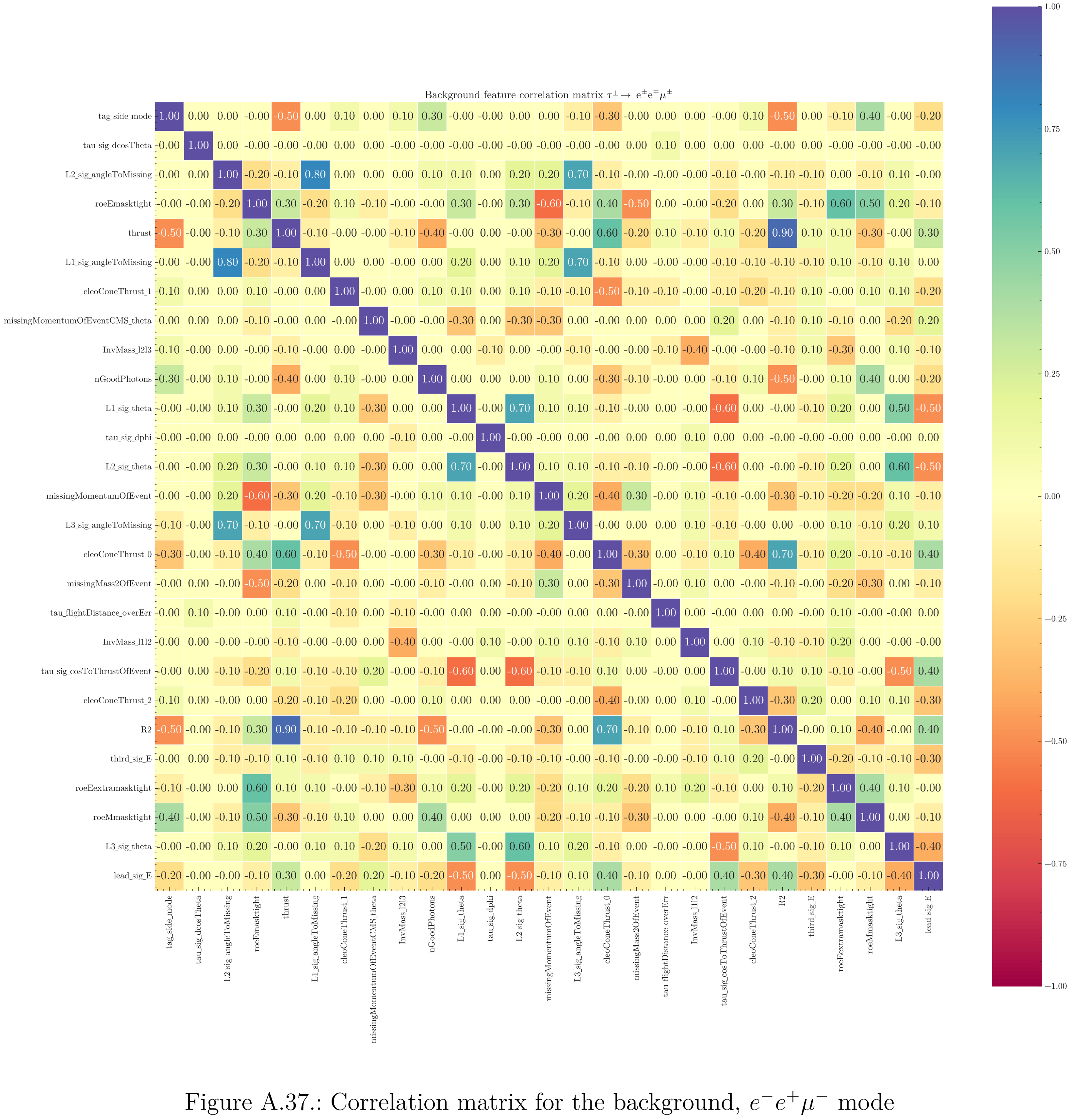


Figure A.37.: Correlation matrix for the background, $e^{-}e^{+}\mu^{-}$ mode

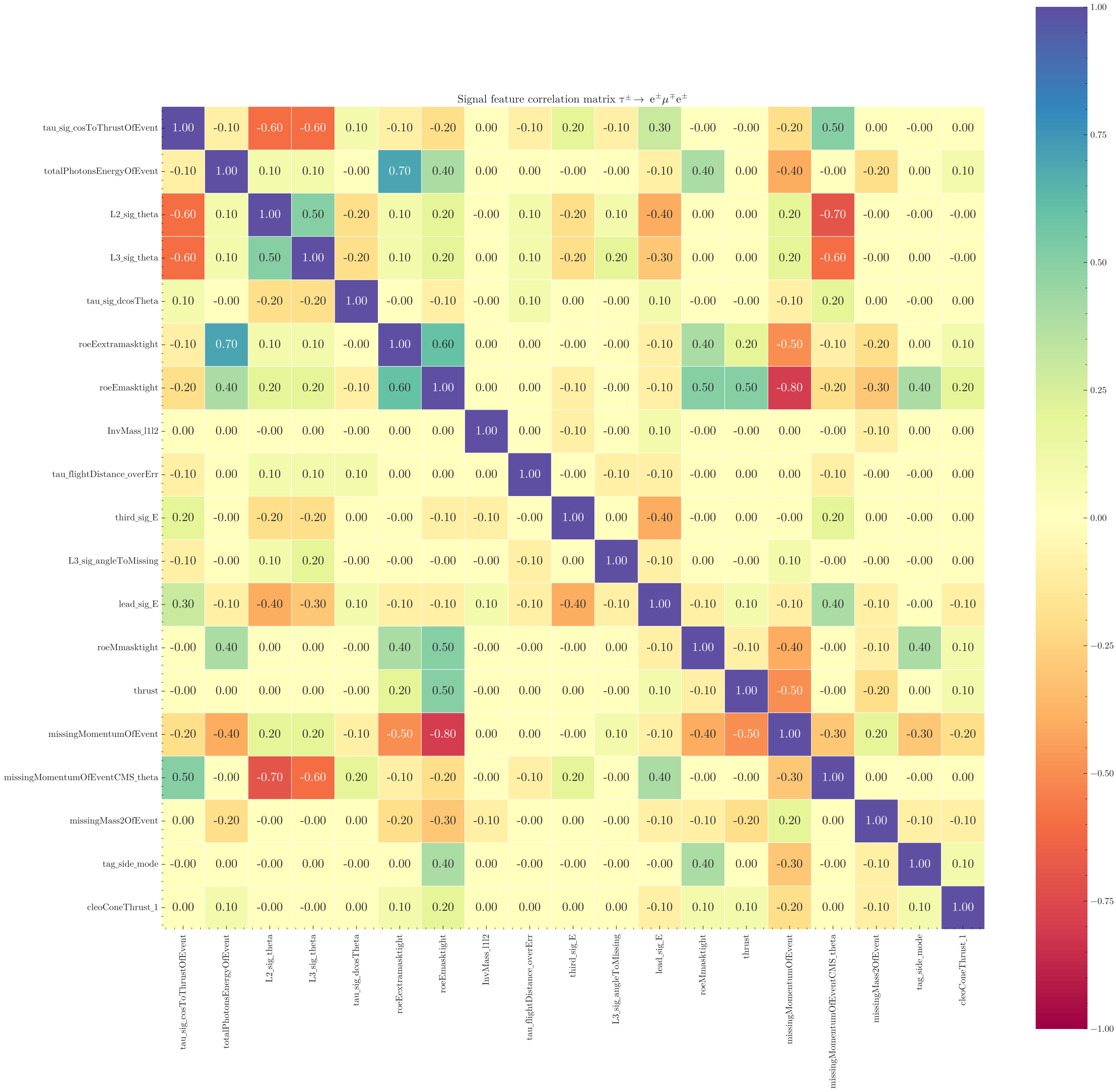


Figure A.38.: Correlation matrix for the signal, $e^{-}\mu^{+}e^{-}$ mode

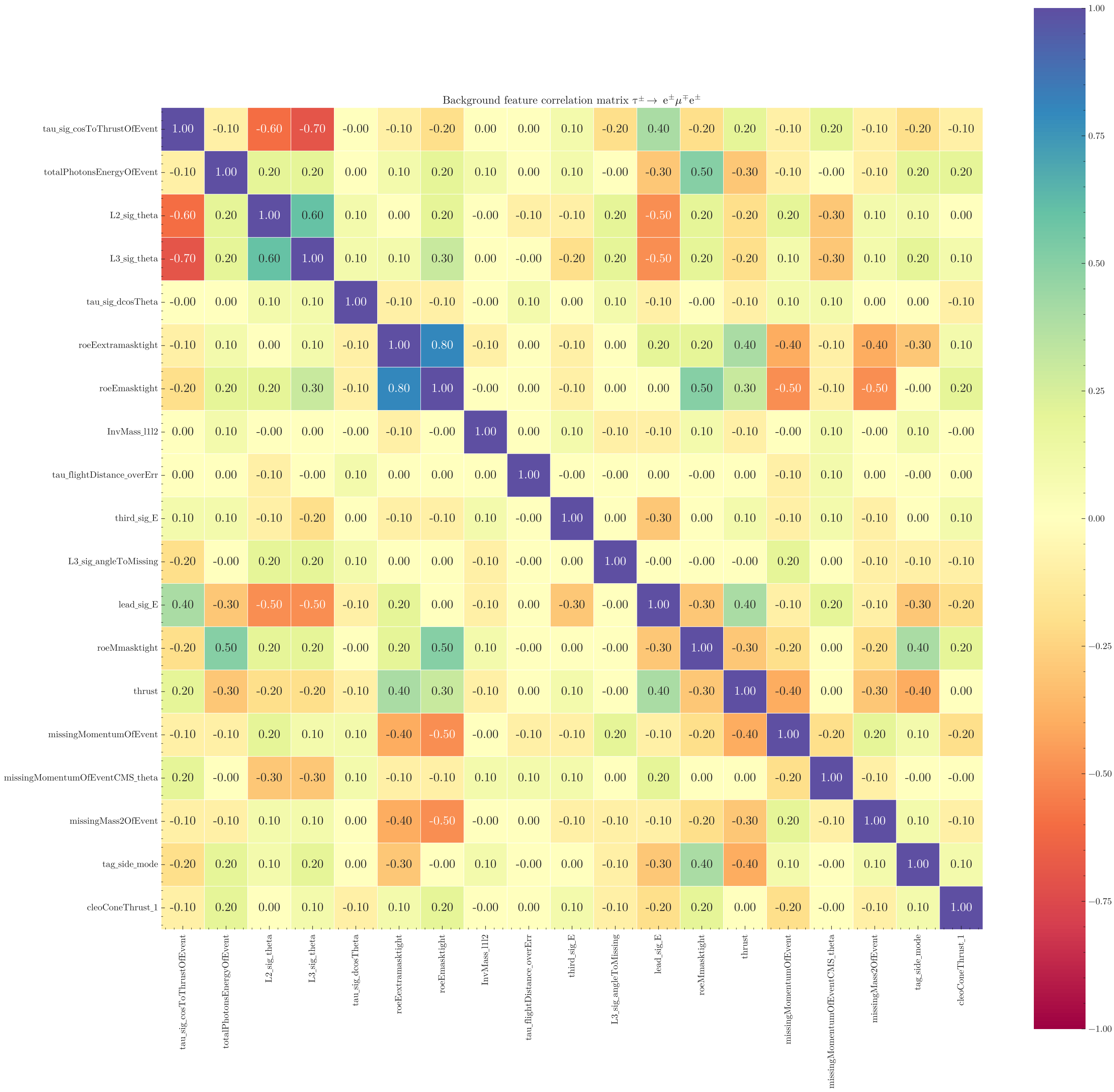


Figure A.39.: Correlation matrix for the background, $e^{-}\mu^{+}e^{-}$ mode

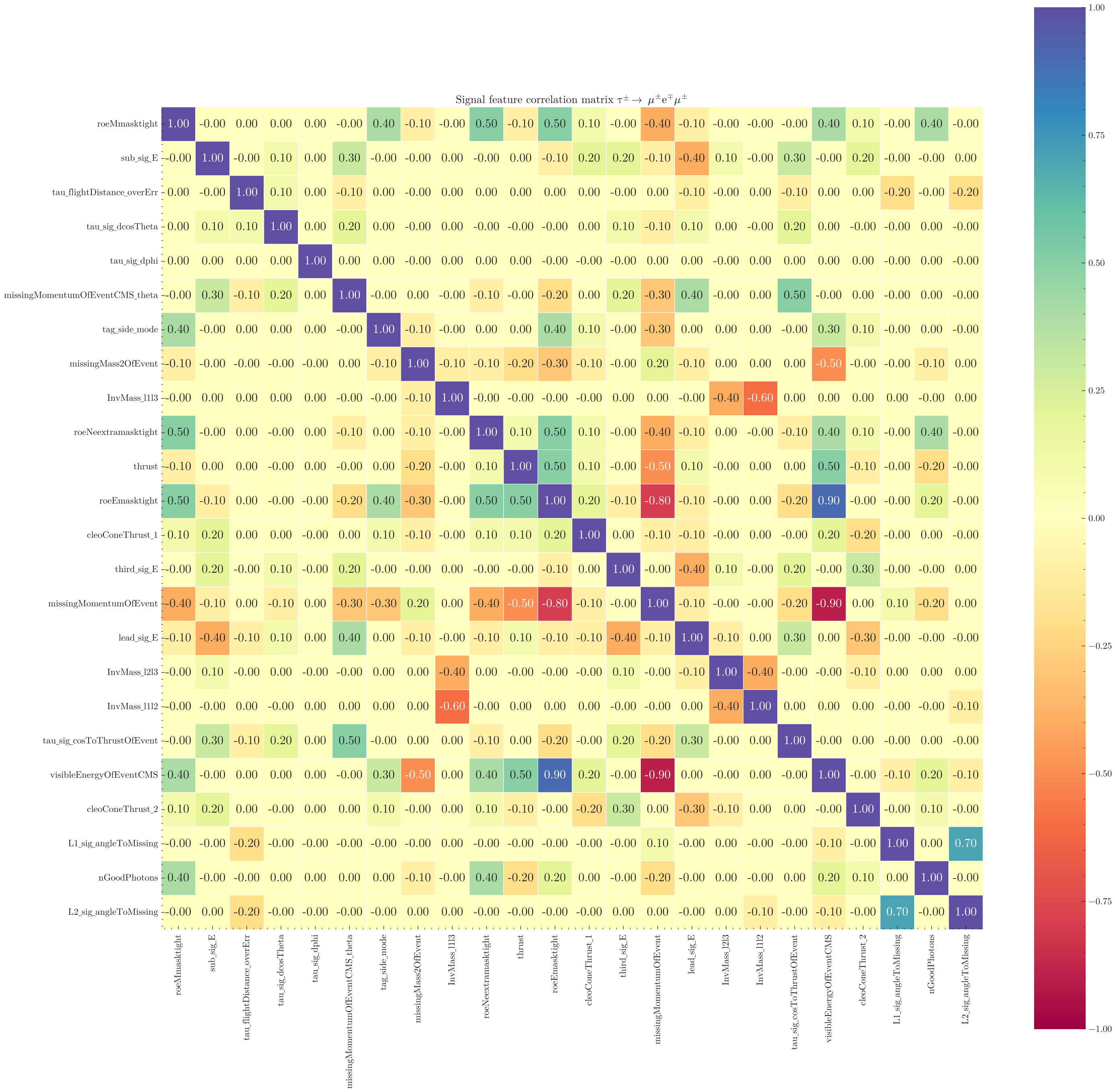


Figure A.40.: Correlation matrix for the signal, $\mu^{-}e^{+}\mu^{-}$ mode

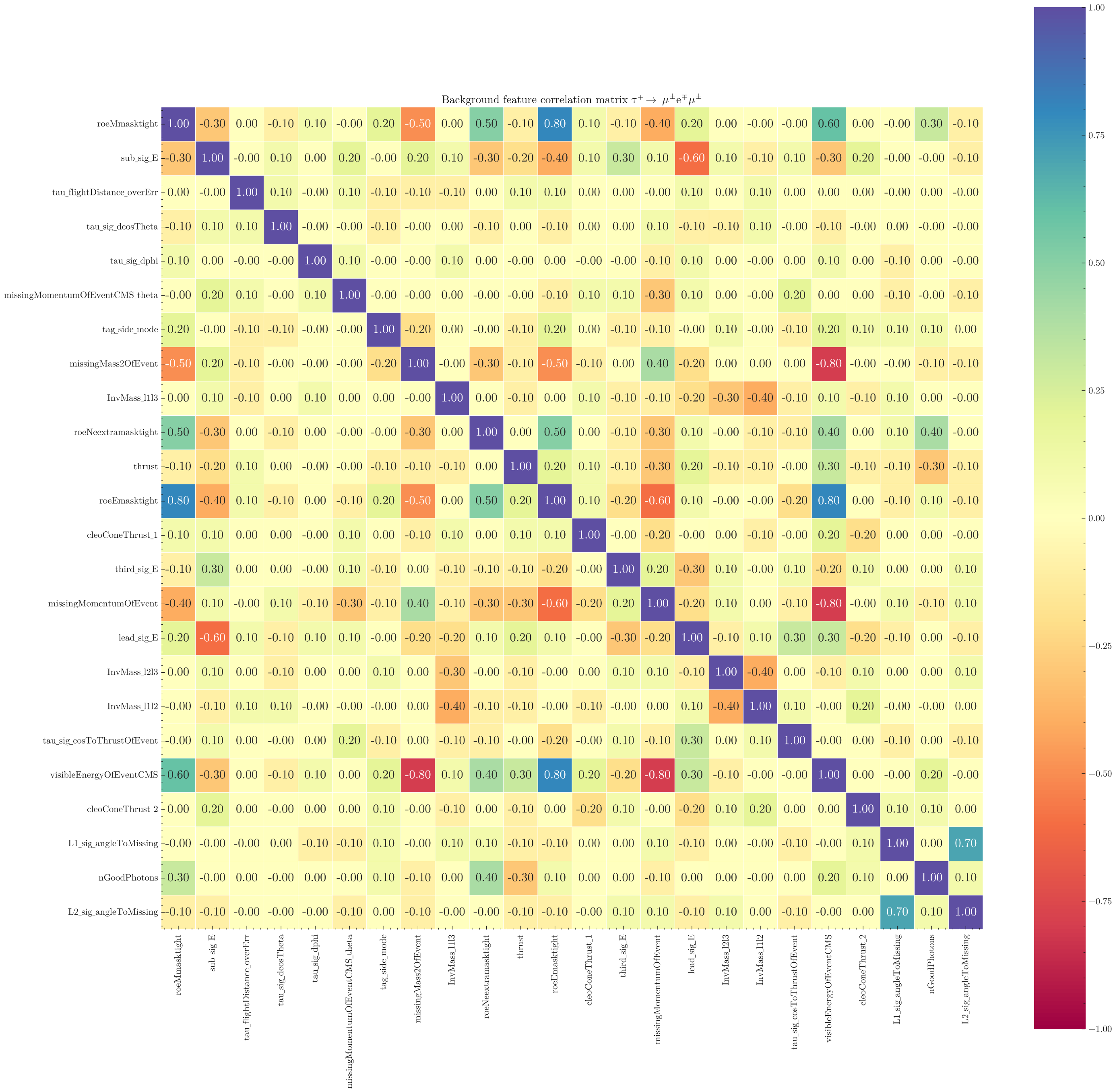


Figure A.41.: Correlation matrix for the background, $\mu^{-}e^{+}\mu^{-}$ mode

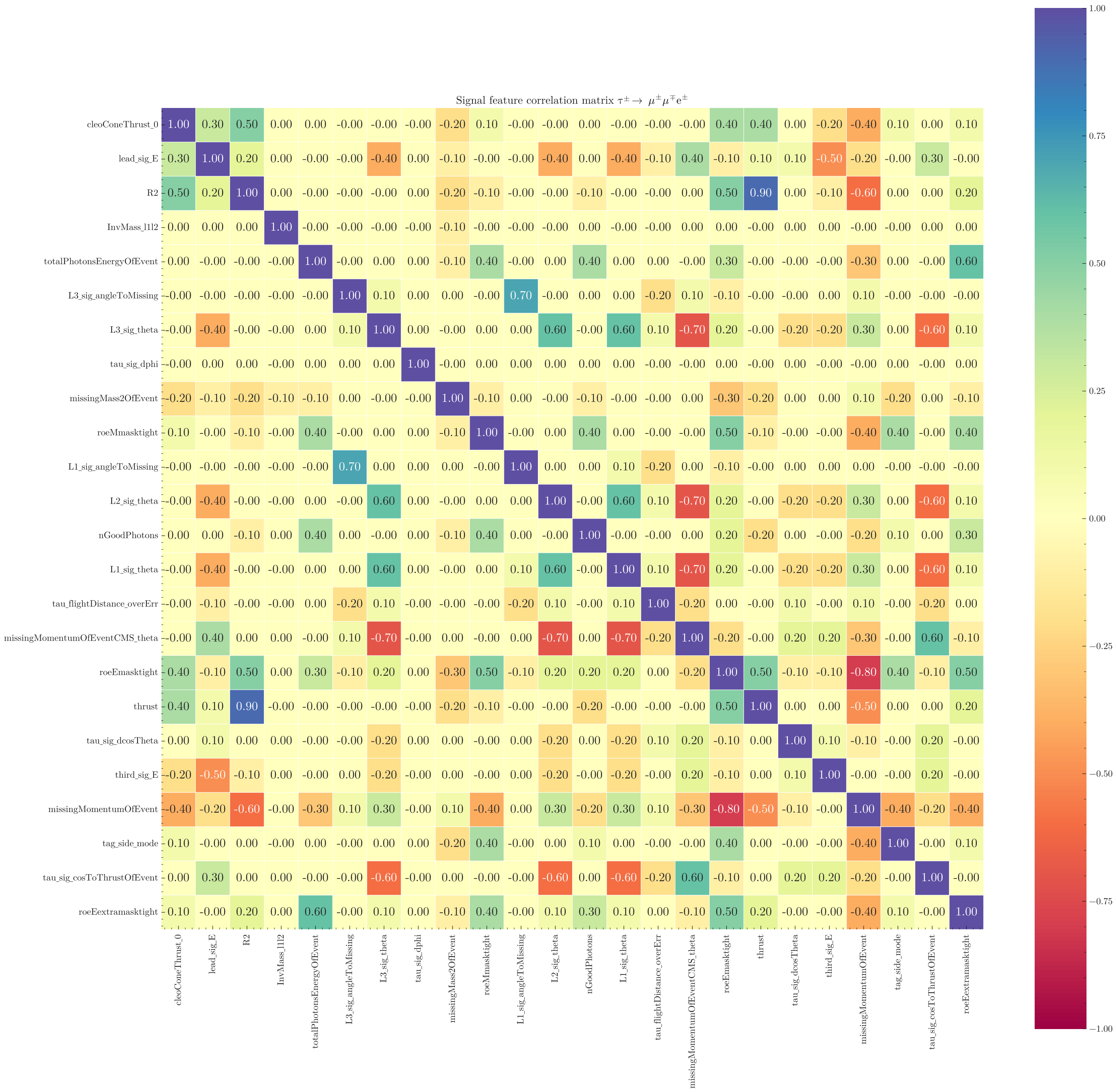


Figure A.42.: Correlation matrix for the signal, $\mu^{-}\mu^{+}e^{-}$ mode

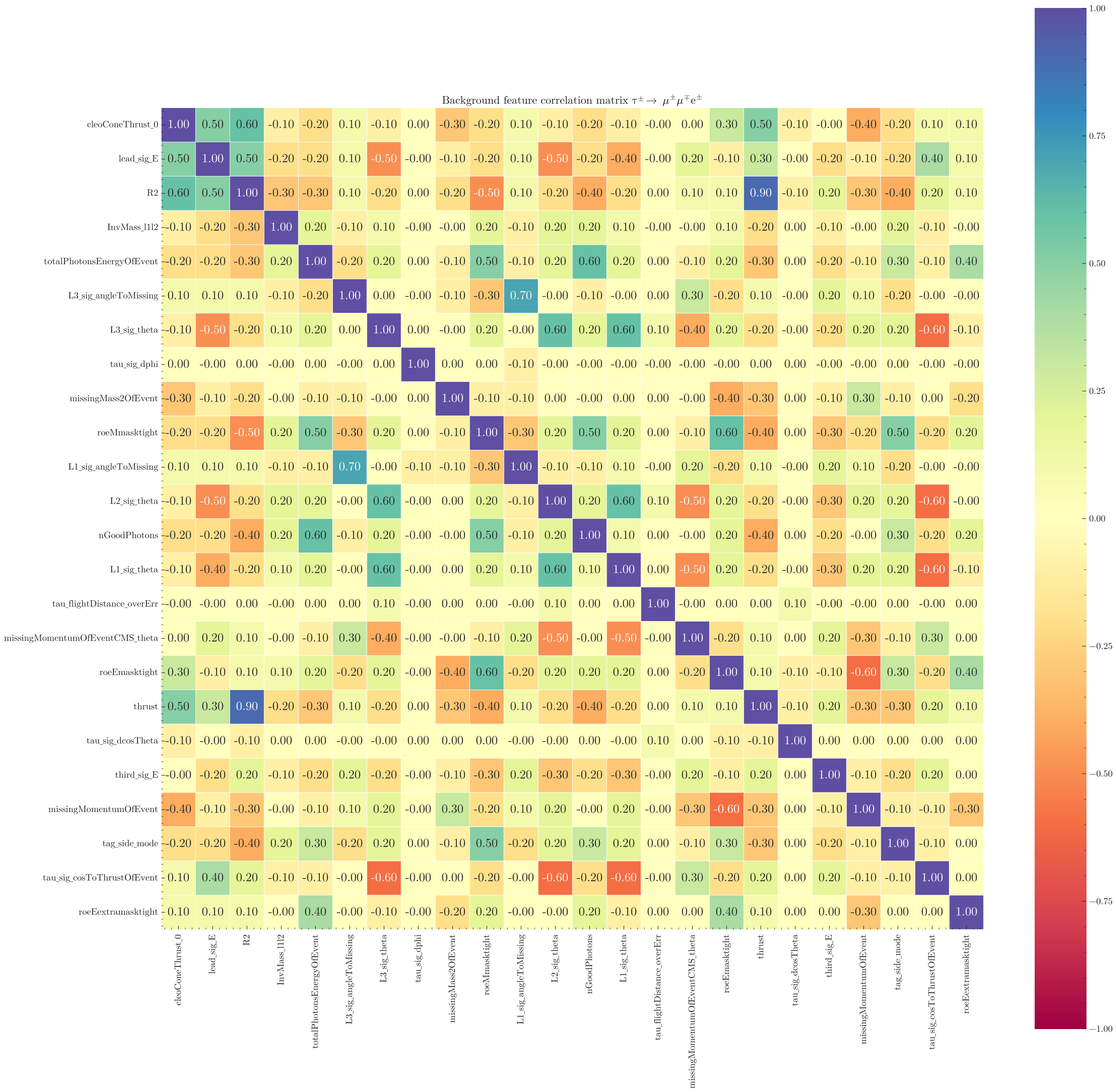


Figure A.43.: Correlation matrix for the background, $\mu^{-}\mu^{+}e^{-}$ mode

A.4. Fits to the signal $M_{\ell\ell'}$ distributions

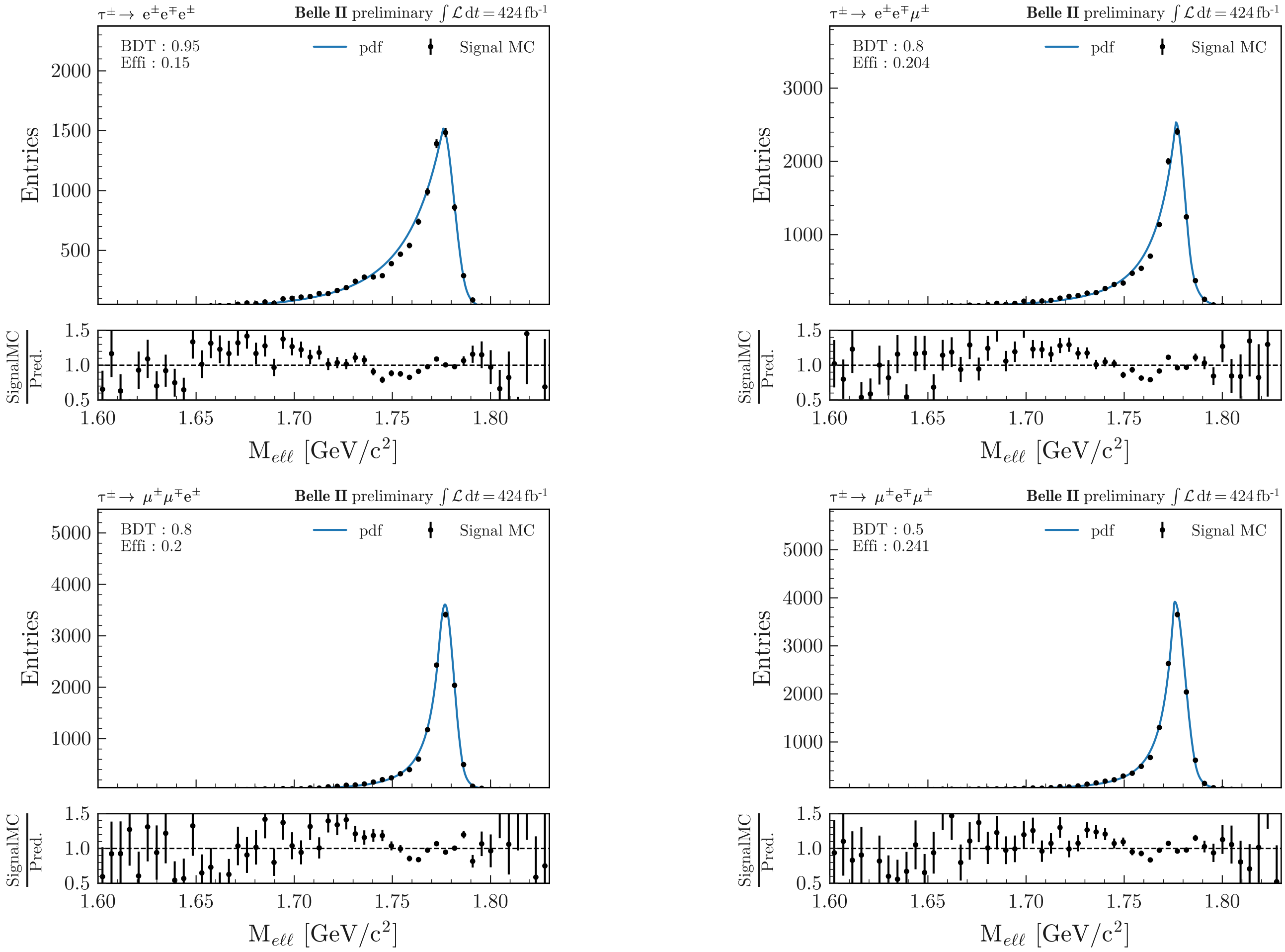


Figure A.44.: Fitted signal p.d.f for the $e^-e^+e^-$ (upper left), $e^-e^+\mu^-$ (upper right), $\mu^-\mu^+e^-$ (lower left) and $\mu^-e^+\mu^-$ (lower right) modes.

List of Figures

List of Tables

Bibliography

[1] Claude Itzykson and Jean-Bernard Zuber. *Quantum Field Theory.* New edition. New York Hamburg: McGraw-Hill Inc.,US, May 1986. ISBN: 978-0-07-066353-4.

[2] Steven Weinberg. *The Quantum Theory of Fields: Volume 1: Foundations.* Vol. 1. Cambridge: Cambridge University Press, 1995. ISBN: 978-0-521-67053-1. DOI: 10.1017/CBO9781139644167.

[3] Michio Kaku. *Quantum Field Theory: A Modern Introduction.* 1st edition. New York: Oxford University Press, Mar. 1993. ISBN: 978-0-19-507652-3.

[4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction To Quantum Field Theory.* Reading, Mass: Westview Press, Oct. 1995. ISBN: 978-0-201-50397-5.

[5] E. Noether. "Invariante Variationsprobleme". In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257.

[6] Atlas Collaboration. "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020.

[7] The CMS Collaboration. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 03702693. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[8] Super-Kamiokande Collaboration, Y. Fukuda, T. Hayakawa, et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81.8 (Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[9] SNO Collaboration, Q. R. Ahmad, R. C. Allen, et al. "Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Solar Neutrinos at the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 87.7 (July 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301.

[10] KamLAND Collaboration, K. Eguchi, S. Enomoto, et al. "First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance". In: *Phys. Rev. Lett.* 90.2 (Jan. 2003), p. 021802. DOI: 10.1103/PhysRevLett.90.021802.

[11] 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 1962), pp. 36–44. DOI: 10.1103/PhysRevLett.9.36.

[12] Benjamin W. Lee and Robert E. Shrock. "Natural suppression of symmetry violation in gauge theories: Muon- and electron-lepton-number nonconservation". In: *Phys. Rev. D* 16.5 (Sept. 1977), pp. 1444–1473. DOI: 10.1103/PhysRevD.16.1444.

[13] Ta-Pei Cheng and Ling-Fong Li. "Muon-number-nonconservation effects in a gauge theory with $V + A$ currents and heavy neutral leptons". In: *Phys. Rev. D* 16.5 (Sept. 1977), pp. 1425–1443. DOI: 10.1103/PhysRevD.16.1425.

[14] BaBar Collaboration. "Limits on τ Lepton-Flavor Violating Decays in three charged leptons". In: *Phys. Rev. D* 81.11 (June 2010), p. 111101. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.81.111101. arXiv: 1002.4550 [hep-ex].

[15] Belle Collaboration, K. Hayasaka, K. Inami, 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. arXiv: 1001.3221 [hep-ex].

[16] T. Abe, I. Adachi, K. Adamczyk, et al. *Belle II Technical Design Report.* Nov. 2010. DOI: 10.48550/arXiv.1011.0352. arXiv: 1011.0352 [hep-ex, physics:physics].

[17] E. Kou, P. Urquijo, W. Altmannshofer, et al. "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019), p. 123C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex, physics:hep-lat, physics:hep-ph].

[18] Yukiyoshi Ohnishi, Tetsuo Abe, Toshikazu Adachi, et al. "Accelerator Design at SuperKEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 2013), 03A011. ISSN: 2050-3911. DOI: 10.1093/ptep/pts083.

[19] H. Ye et al. "Commissioning and performance of the Belle II pixel detector". In: *Nucl. Instrum. Meth. A* 987 (2021), p. 164875. DOI: 10.1016/j.nima.2020.164875.

[20] 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. 2022), P11042. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/11/P11042.

[21] *Particle identification performance of the prototype aerogel RICH counter for the Belle II experiment | Progress of Theoretical and Experimental Physics | Oxford Academic.* https://academic.oup.com/ptep/article/2016/3/033H01/2461082.

[22] H. Aihara, A. Aloisio, D. P. Auguste, et al. *The Belle II Detector Upgrades Framework Conceptual Design Report.* July 2024. DOI: 10.48550/arXiv.2406.19421. arXiv: 2406.19421 [hep-ex, physics:physics].

[23] Belle II Collaboration, F. Abudinén, et al. "Search for lepton-flavor-violating $\tau^- \to \ell^- \phi$ decays in 2019-2021 Belle II data". In: (May 2023).

[24] Belle II Collaboration, I. Adachi, L. Aggarwal, et al. *Search for lepton-flavor-violating $\tau^- \to \mu^- \mu^+ \mu^-$ decays at Belle II.* May 2024. DOI: 10.48550/arXiv.2405.07386. arXiv: 2405.07386 [hep-ex].

[25] Wikipedia contributors. "Elementary_particle". In: *Wikipedia, The Free Encyclopedia* (July 2024).

[26] C. S. Wu, E. Ambler, R. W. Hayward, et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105.4 (Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[27] B. Pontecorvo. "Mesonium and Anti-Mesonium". In: *Sov. Phys. JETP* 6 (1957), p. 429.

[28] B. Pontecorvo. "Inverse Beta Processes and Nonconservation of Lepton Charge". In: *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247.

[29] Léonard Polat and Justine Serrano. "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". PhD thesis. Marseille: CPPM, 2022.

[30] Robin Leboucher and Justine Serrano. "Search for $\tau^- \to \mu^-\mu^+\mu^-$ lepton flavour violating decays and measurement of the SVD cluster position resolution at Belle II experiment". PhD thesis. Marseille: CPPM, 2023.

[31] The Muon $g - 2$ Collaboration, D. P. Aguillard, T. Albahri, et al. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm". In: *Phys. Rev. Lett.* 131.16 (Oct. 2023), p. 161802. DOI: 10.1103/PhysRevLett.131.161802.

[32] BaBar Collaboration, J. P. Lees, V. Poireau, et al. "Evidence for an Excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ Decays". In: *Phys. Rev. Lett.* 109.10 (Sept. 2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802.

[33] LHCb collaboration, R. Aaij, B. Adeva, et al. "Measurement of the ratio of branching fractions $\mathcal{B}(\overline{B}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$". In: *Phys. Rev. Lett.* 115.11 (Sept. 2015), p. 111803. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.115.111803. arXiv: 1506.08614 [hep-ex].

[34] Swagato Banerjee. "Searches for Lepton Flavor Violation in Tau Decays at Belle II". In: *Universe* 8.9 (Sept. 2022), p. 480. ISSN: 2218-1997. DOI: 10.3390/universe8090480.

[35] Aditya Batra, Praveen Bharadwaj, Sanjoy Mandal, et al. "Phenomenology of the simplest linear seesaw mechanism". In: *J. High Energ. Phys.* 2023.7 (July 2023), p. 221. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2023)221. arXiv: 2305.00994 [hep-ex, physics:hep-ph].

[36] Jernej F. Kamenik and Miha Nemevsek. "Lepton flavor violation in type I + III seesaw". In: *J. High Energy Phys.* 2009.11 (Nov. 2009), pp. 023–023. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2009/11/023. arXiv: 0908.3451 [hep-ph].

[37] E. Akhmedov, M. Lindner, E. Schnapka, et al. "Dynamical Left-Right Symmetry Breaking". In: *Phys. Rev. D* 53.5 (Mar. 1996), pp. 2752–2780. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.53.2752. arXiv: hep-ph/9509255.

[38] Marco Ardu, Sacha Davidson, and Stéphane Lavignac. "Constraining new physics models from $\mu \to e$ observables in bottom-up EFT". In: *Eur. Phys. J. C* 84.5 (May 2024), p. 458. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-12782-x.

[39] J. Schechter and J. W. F. Valle. "Neutrino masses in SU(2) $\otimes$ U(1) theories". In: *Phys. Rev. D* 22.9 (Nov. 1980), pp. 2227–2235. DOI: 10.1103/PhysRevD.22.2227.

[40] Rabindra N. Mohapatra and Goran Senjanovic. "Neutrino Masses and Mixings in Gauge Models with Spontaneous Parity Violation". In: *Phys. Rev. D* 23 (1981), p. 165. DOI: 10.1103/PhysRevD.23.165.

[41] Andrea Brignole and Anna Rossi. "Anatomy and Phenomenology of μ-τ Lepton Flavour Violation in the MSSM". In: *Nuclear Physics B* 701.1-2 (Nov. 2004), pp. 3–53. ISSN: 05503213. DOI: 10.1016/j.nuclphysb.2004.08.037. arXiv: hep-ph/0404211.

[42] John Ellis, Junji Hisano, Martti Raidal, et al. "A New Parametrization of the Seesaw Mechanism and Applications in Supersymmetric Models". In: *Phys. Rev. D* 66.11 (Dec. 2002), p. 115013. ISSN: 0556-2821, 1089-4918. DOI: 10.1103/PhysRevD.66.115013. arXiv: hep-ph/0206110.

[43] 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. 2003), p. 033012. DOI: 10.1103/PhysRevD.68.033012.

[44] Monika Blanke, Andrzej J. Buras, Bjoern Duling, et al. *FCNC Processes in the Littlest Higgs Model with T-Parity: an Update.* Mar. 2010. DOI: 10.48550/arXiv.0906.5454. arXiv: 0906.5454 [hep-ex, physics:hep-ph].

[45] Monika Blanke, Andrzej J. Buras, Anton Poschenrieder, et al. "Rare and CP-Violating K and B Decays in the Littlest Higgs Model with T-Parity". In: *J. High Energy Phys.* 2007.01 (Jan. 2007), pp. 066–066. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/01/066. arXiv: hep-ph/0610298.

[46] Monika Blanke, Andrzej J. Buras, Björn Duling, et al. "Charged lepton flavour violation and $(g\text{-}2)_\mu$ in the Littlest Higgs model with T-Parity: a clear distinction from Supersymmetry". In: *J. High Energy Phys.* 2007.05 (May 2007), p. 013. ISSN: 1126-6708. DOI: 10.1088/1126-6708/2007/05/013.

[47] Kingman Cheung, Abner Soffer, Zeren Simon Wang, et al. "Probing Charged Lepton Flavor Violation with Axion-like Particles at Belle II". In: *Journal of High Energy Physics* 2021.11 (Nov. 2021), p. 218. ISSN: 1126-6708. DOI: 10.1007/JHEP11(2021)218.

[48] Fernando Arias-Aragón, Jérémie Quevillon, and Christopher Smith. "Axion-like ALPs". In: *J. High Energ. Phys.* 2023.3 (Mar. 2023), p. 134. ISSN: 1029-8479. DOI: 10.1007/JHEP03(2023)134. arXiv: 2211.04489 [hep-ph].

[49] Roberto D. Peccei. “The Strong CP Problem and Axions”. In: *Axions: Theory, Cosmology, and Experimental Searches*. Ed. by Markus Kuster, Georg Raffelt, and Berta Beltrán. Berlin, Heidelberg: Springer, 2008, pp. 3–17. ISBN: 978-3-540-73518-2. DOI: 10.1007/978-3-540-73518-2_1.

[50] 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: 1434-6052. DOI: 10.1140/epjc/s10052-021-09824-z.

[51] The MEG Collaboration. “Search for the Lepton Flavour Violating Decay $\mu^+ \to e^+\gamma$ with the Full Dataset of the MEG Experiment”. In: arXiv:1605.05081 (July 2016). DOI: 10.48550/arXiv.1605.05081. arXiv: 1605.05081 [hep-ex, physics:physics].

[52] SINDRUM Collaboration, U. Bellgardt, and G. Otter. *Search for the decay $\mu^+ \to e^+e^-e^+$.* Mar. 1988. DOI: 10.1016/0550-3213(88)90462-2.

[53] The COMET Collaboration, R. Abramishvili, and G. Adamov. *COMET Phase-I Technical Design Report*. https://arxiv.org/abs/1812.09018v3. Dec. 2018. DOI: 10.1093/ptep/ptz125.

[54] N. Teshima. *Status of the DeeMe Experiment, an Experimental Search for μ-e Conversion at J-PARC MLF*. Nov. 2019. DOI: 10.48550/arXiv.1911.07143. arXiv: 1911.07143 [hep-ex, physics:physics].

[55] L. Bartoszek, E. Barnes, J. P. Miller, et al. *Mu2e Technical Design Report*. Mar. 2015. DOI: 10.48550/arXiv.1501.05241. arXiv: 1501.05241 [hep-ex, physics:physics].

[56] ATLAS Collaboration, G. Aad, B. Abbott, et al. “Search for the Charged-Lepton-Flavor-Violating Decay $Z\to\ E\mu$ in p p Collisions at s = 13 TeV with the ATLAS Detector”. In: *Phys. Rev. D* 108.3 (Aug. 2023), p. 032015. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.108.032015.

[57] ATLAS Collaboration, G. Aad, B. Abbott, et al. “Search for Charged-Lepton-Flavour Violation in Z-boson Decays with the ATLAS Detector”. In: *Nat. Phys.* 17.7 (July 2021), pp. 819–825. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01225-z.

[58] ATLAS Collaboration. "Searches for lepton-flavour-violating decays of the Higgs boson into $e\tau$ and $\mu\tau$ in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". In: *J. High Energ. Phys.* 2023.7 (July 2023), p. 166. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2023)166. arXiv: 2302.05225 [hep-ex].

[59] CMS Collaboration. "Search for lepton-flavor violating decays of the Higgs boson in the $\mu\tau$ and $e\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 104.3 (Aug. 2021), p. 032013. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.104.032013. arXiv: 2105.03007 [hep-ex].

[60] LHCb Collaboration, R. Aaij, C. Abellán Beteta, et al. "Search for the Lepton-Flavor-Violating Decays $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$". In: *Phys. Rev. Lett.* 123.21 (Nov. 2019), p. 211801. DOI: 10.1103/PhysRevLett.123.211801.

[61] Particle Data Group Collaboration, S. Navas, C. Amsler, et al. "Review of Particle Physics". In: *Phys. Rev. D* 110.3 (Aug. 2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[62] CMS Collaboration. *Search for the Lepton Flavor Violating $\tau \to 3\mu$ Decay in Proton-Proton Collisions at $\sqrt{(s)} = 13$ TeV*. Apr. 2024. DOI: 10.48550/arXiv.2312.02371. arXiv: 2312.02371 [hep-ex].

[63] LHCb collaboration. "Search for the Lepton Flavour Violating Decay $\tau \to 3\mu$". In: *J. High Energ. Phys.* 2015.2 (Feb. 2015), p. 121. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2015)121.

[64] Swagato Banerjee, Vincenzo Cirigliano, Mogens Dam, et al. *Snowmass 2021 White Paper: Charged lepton flavor violation in the tau sector*. May 2022. DOI: 10.48550/arXiv.2203.14919. arXiv: 2203.14919 [hep-ex, physics:hep-ph].

[65] SuperB Collaboration. *SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report*. Sept. 2007. DOI: 10.48550/arXiv.0709.0451. arXiv: 0709.0451 [hep-ex].

[66] Mikhail Zobov. "New generation electron-positron factories". In: *Phys. Part. Nuclei* 42.5 (Sept. 2011), pp. 782–799. ISSN: 1531-8559. DOI: 10.1134/S1063779611050066.

[67] M. J. French, L. L. Jones, Q. Morrissey, et al. "Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 4th Int. Symp. on Development and Application of Semiconductor Tracking

Detectors 466.2 (July 2001), pp. 359–365. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(01)00589-7.

[68] 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. arXiv: 2209.03006 [physics].

[69] Valerio Bertacchi, Tadeas Bilka, Nils Braun, et al. “Track finding at Belle II”. In: *Computer Physics Communications* 259 (Feb. 2021), p. 107610. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2020.107610.

[70] Konstantinos Moustakas. “Design and Development of Depleted Monolithic Active Pixel Sensors with Small Collection Electrode for High-Radiation Applications”. PhD thesis. Bonn U., 2021.

[71] T. Kuhr, C. Pulvermacher, M. Ritter, et al. “The Belle II Core Software”. In: *Comput Softw Big Sci* 3.1 (Nov. 2018), p. 1. ISSN: 2510-2044. DOI: 10.1007/s41781-018-0017-9.

[72] *Documentation of the Belle II software*. https://software.belle2.org/.

[73] The Belle II Collaboration. *Belle II Analysis Software Framework (basf2)*. Zenodo. Aug. 2022. DOI: 10.5281/zenodo.6949513.

[74] 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*. New Computing Techniques in Physics Research V 389.1 (Apr. 1997), pp. 81–86. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00048-X.

[75] Nils Lennart Braun. “Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment”. PhD thesis. KIT, Karlsruhe, 2018.

[76] T. Alexopoulos, M. Bachtis, E. Gazis, et al. “Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 592.3 (July 2008), pp. 456–462. ISSN: 0168-9002. DOI: 10.1016/j.nima.2008.04.038.

[77] I. Abt, D. Emeliyanov, I. Kisel, et al. “CATS: A Cellular Automaton for Tracking in Silicon for the HERA-B Vertex Detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 489.1 (Aug. 2002), pp. 389–405. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(02)00790-8.

[78] A. Glazov, I. Kisel, E. Konotopskaya, et al. “Filtering Tracks in Discrete Detectors Using a Cellular Automaton”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 329.1 (May 1993), pp. 262–268. ISSN: 0168-9002. DOI: 10.1016/0168-9002(93)90945-E.

[79] Martin Gardner. *Mathematical Games*. Oct. 1970.

[80] D. Funke, T. Hauth, V. Innocente, et al. “Parallel Track Reconstruction in CMS Using the Cellular Automaton Approach”. In: *J. Phys.: Conf. Ser.* 513.5 (June 2014), p. 052010. ISSN: 1742-6596. DOI: 10.1088/1742-6596/513/5/052010.

[81] Tadeas Bilka, Nils Braun, Thomas Hauth, et al. *Implementation of GENFIT2 as an experiment independent track-fitting framework*. Feb. 2019. DOI: 10.48550/arXiv.1902.04405. arXiv: 1902.04405 [physics].

[82] The CMS Collaboration. “Description and Performance of Track and Primary-Vertex Reconstruction with the CMS Tracker”. In: *J. Inst.* 9.10 (Oct. 2014), P10009. ISSN: 1748-0221. DOI: 10.1088/1748-0221/9/10/P10009.

[83] R. Mankel. “Pattern Recognition and Event Reconstruction in Particle Physics Experiments”. In: *Rep. Prog. Phys.* 67.4 (Mar. 2004), p. 553. ISSN: 0034-4885. DOI: 10.1088/0034-4885/67/4/R03.

[84] ATLAS Collaboration. “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2”. In: *Eur. Phys. J. C* 77.10 (Oct. 2017), p. 673. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-017-5225-7. arXiv: 1704.07983 [hep-ex].

[85] Rainer Mankel and Alexander Spiridonov. “The Concurrent Track Evolution Algorithm: Extension for Track Finding in the Inhomogeneous Magnetic Field of the HERA-B Spectrometer”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 426.2 (May 1999), pp. 268–282. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(99)00013-3.

[86] Alexander Glazov, Petar Rados, Ami Rostomyan, et al. *Measurement of the track reconstruction efficiency and fake rate with $e^+e^- \to \tau^+\tau^-$ events.* https://docs.belle2.org/record/2035.

[87] Marco Milesi, Justin Tan, and Phillip Urquijo. "Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers". In: *EPJ Web Conf.* 245 (2020), p. 06023. ISSN: 2100-014X. DOI: 10.1051/epjconf/202024506023.

[88] S. Jadach, B. F. L. Ward, and Z. Was. "The precision Monte Carlo event generator $\mathcal{KK}$ for two-fermion final states in e^+e^- collisions". In: *Computer Physics Communications* 130.3 (Aug. 2000), pp. 260–325. ISSN: 0010-4655. DOI: 10.1016/S0010-4655(00)00048-5.

[89] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. "TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons". In: *Computer Physics Communications* 64.2 (May 1991), pp. 275–299. ISSN: 0010-4655. DOI: 10.1016/0010-4655(91)90038-M.

[90] Elisabetta Barberio, Bob van Eijk, and Zbigniew Was. "Photos — a universal Monte Carlo for QED radiative corrections in decays". In: *Computer Physics Communications* 66.1 (July 1991), pp. 115–128. ISSN: 0010-4655. DOI: 10.1016/0010-4655(91)90012-A.

[91] Torbjorn Sjostrand, Stefan Ask, Jesper R. Christiansen, et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2015.01.024.

[92] 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.* BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines 462.1 (Apr. 2001), pp. 152–155. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(01)00089-4.

[93] Giovanni Balossini, Carlo M. Carloni Calame, Guido Montagna, et al. "Matching perturbative and parton shower corrections to Bhabha process at flavour factories". In: *Nuclear Physics B* 758.1 (Dec. 2006), pp. 227–253. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2006.09.022.

[94] 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: 0370-2693. DOI: 10.1016/j.physletb.2008.04.007.

[95] C. M. Carloni Calame, G. Montagna, O. Nicrosini, et al. "The BABAYAGA event generator". In: *Nuclear Physics B - Proceedings Supplements*. SIGHAD03 131 (Apr. 2004), pp. 48–55. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2004.02.008.

[96] Carlo Michel Carloni Calame. "An improved parton shower algorithm in QED". In: *Physics Letters B* 520.1 (Nov. 2001), pp. 16–24. ISSN: 0370-2693. DOI: 10.1016/S0370-2693(01)01108-X.

[97] 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 (Sept. 2000), pp. 459–479. ISSN: 0550-3213. DOI: 10.1016/S0550-3213(00)00356-4.

[98] F. A. Berends, P. H. Daverveldt, and R. Kleiss. "Radiative corrections to the process $e^+e^- \to e^+e^-\mu^+\mu^-$". In: *Nuclear Physics B* 253 (Jan. 1985), pp. 421–440. ISSN: 0550-3213. DOI: 10.1016/0550-3213(85)90540-1.

[99] 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: 0550-3213. DOI: 10.1016/0550-3213(85)90541-3.

[100] F. A Berends, P. H Daverveldt, and R Kleiss. "Monte Carlo simulation of two-photon processes: II: Complete lowest order calculations for four-lepton production processes in electron-positron collisions". In: *Computer Physics Communications* 40.2 (June 1986), pp. 285–307. ISSN: 0010-4655. DOI: 10.1016/0010-4655(86)90115-3.

[101] Sadaharu Uehara. *TREPS: A Monte-Carlo Event Generator for Two-photon Processes at e^+e^- Colliders using an Equivalent Photon Approximation*. Oct. 2013. DOI: 10.48550/arXiv.1310.0157. arXiv: 1310.0157 [hep-ph].

[102] 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: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8.

[103] *Trigger Simulation (TSIM) - Belle II - DESY Confluence*. https://confluence.desy.de/pages/viewpage.action?pageId=47822559.

[104] Cobham Technical Services, Vector Fields Software. *OPERA-3D User Guide Ver.15R3*. Tech. rep. Oxford, England, 2012.

[105] Mauro Moretti, Thorsten Ohl, and Juergen Reuter. *O'Mega: An Optimizing Matrix Element Generator*. Feb. 2001. DOI: 10.48550/arXiv.hep-ph/0102195. arXiv: hep-ph/0102195.

[106] Wolfgang Kilian, Thorsten Ohl, and Jurgen Reuter. "WHIZARD: Simulating Multi-Particle Processes at LHC and ILC". In: *Eur. Phys. J. C* 71.9 (Sept. 2011), p. 1742. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-011-1742-y. arXiv: 0708.4233 [hep-ex, physics:hep-ph].

[107] *SVD status and plans, 44th B2GM at KEK*. https://indico.belle2.org/event/8084/contributions/48584/.

[108] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth. A* 434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006.

[109] Alexander L. Read. "Presentation of search results: The CL(s) technique". In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[110] Robin Leboucher. *Good Track Selection for Tau Events*. 2020.

[111] Giovanni Punzi. *Sensitivity of searches for new signals and its optimization*. Dec. 2003. DOI: 10.48550/arXiv.physics/0308063. arXiv: physics/0308063.

[112] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. arXiv: 1603.02754 [cs.LG].

[113] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". en. In: (July 2019). arXiv:1907.10902. URL: http://arxiv.org/abs/1907.10902.

[114] D. M. Asner, M. Athanas, D. W. Bliss, et al. "Search for Exclusive Charmless Hadronic B Decays". In: *Phys. Rev. D* 53.3 (Feb. 1996), pp. 1039–1050. DOI: 10.1103/PhysRevD.53.1039.

[115] James Bergstra, R. Bardenet, Yoshua Bengio, et al. "Algorithms for Hyper-Parameter Optimization". In: Advances in Neural Information Processing Systems 24 (Dec. 2011). Ed. by J. Shawe-Taylor, R.S. Zemel, P. Bartlett, et al. URL: https://inria.hal.science/hal-00642998.

[116] Thomas Junk. "Confidence Level Computation for Combining Searches with Small Statistics". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 434.2-3 (Sept. 1999), pp. 435–443. ISSN: 01689002. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006.

[117] A. L. Read. "Presentation of search results: the CLs technique". In: *J. Phys. G: Nucl. Part. Phys.* 28.10 (Sept. 2002), p. 2693. ISSN: 0954-3899. DOI: 10.1088/0954-3899/28/10/313.

[118] Qi-Dong Zhou. *Correction for Tracking Momentum Bias Based on Invariant Mass Peak Studies*. 2020.

[119] F. Abudinén et al. "Measurement of the integrated luminosity of the Phase 2 data of the Belle II experiment". In: *Chinese Physics C* 44.2 (2020), p. 021001. DOI: 10.1088/1674-1137/44/2/021001. arXiv: 1910.05365 [hep-ex].

[120] Swagato Banerjee, Bolek Pietrzyk, J. Michael Roney, et al. "Tau and muon pair production cross sections in electron-positron annihilations at $\sqrt{s} = 10.58$ GeV". In: *Physical Review D* 77.5 (2008). DOI: 10.1103/physrevd.77.054012. arXiv: 0706.3235 [hep-ph].

[121] A. Abada, M. Abbrescia, S. S. AbdusSalam, et al. "FCC-ee: The Lepton Collider". In: *Eur. Phys. J. Spec. Top.* 228.2 (June 2019), pp. 261–623. ISSN: 1951-6401. DOI: 10.1140/epjst/e2019-900045-4.