Recherche des désintégrations violant la saveur leptonique Bs -> tμ et Bd -> tμ avec l'expérience LHCb

Joan Arnau Romeu

AIX-MARSEILLE UNIVERSITÉ
ECOLE DOCTORALE PHYSIQUE ET SCIENCES DE LA MATIÉRE

Centre de Physique des Particules de Marseille (CPPM) UMR 7346

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline: Physique et Sciences de la Matière
Spécialité: Physique des Particules et Astroparticules

Joan ARNAU ROMEU

Search for the Lepton Flavor Violating decays $B_s^0 \to \tau^{\pm}\mu^{\mp}$ and $B^0 \to \tau^{\pm}\mu^{\mp}$ with the LHCb experiment

Soutenue le 10/09/2018 devant le jury composé de :

Marta CALVI	Università di Milano Bicocca	Rapporteur
Marie-Hélène SCHUNE	LAL / Université Paris-Sud	Rapporteur
Marcella BONA	Queen Mary University of London	Examinateur
Cristinel DIACONU	CPPM / Aix-Marseille Université	Examinateur
Giampiero MANCINELLI	CPPM / Aix-Marseille Université	Directeur de thése
Julien COGAN	CPPM / Aix-Marseille Université	Co-directeur de thése

Numéro national de thése/suffixe local : 2018AIXM0242/021ED352

Suffixe laboratoire : CPPM-T-2018-01

Abstract

A search for $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$ decays is performed using data corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$ of proton–proton collisions recorded by the LHCb experiment during the years 2011 and 2012. For this search, the τ lepton is reconstructed in the 3-prong $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$ channel and the decay with the same topology $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ is used as normalization mode. The full event selection uses Boosted Decision Trees (BDT) based on kinematical and isolation observables. A simultaneous fit to a custom B meson reconstructed mass in bins of a BDT output is performed on the selected data to extract the signal yields. No excess of events is observed. Assuming no contribution from $B^0 \to \tau^\pm\mu^\mp$ decays, an upper limit is set on the $B_s^0 \to \tau^\pm\mu^\mp$ branching fraction of $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp) < 2.5 \cdot 10^{-5}$ at 90% confidence level. If instead no contribution from $B_s^0 \to \tau^\pm\mu^\mp$ decays is assumed, the limit is $\mathcal{B}(B^0 \to \tau^\pm\mu^\mp) < 1.0 \cdot 10^{-5}$ at 90% confidence level. These result correspond the first limit on $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp)$ and the world's best limit on $\mathcal{B}(B^0 \to \tau^\pm\mu^\mp)$.

Acknowledgements

I would like to thank the CPPM LHCb group for this opportunity to design and execute such a pioneer analysis. I want to give special thanks to my supervisors, Julien Cogan and Giampiero Mancinelli, for their guidance and help during the whole process and, above all, for everything they taught to me, from the personal to the professional point of view.

Also, I want to thank my family and friends for their support, even when being far away from them. In particular, I am also grateful to my fellow PhD students for the nice time we have spent together either at work or during our free time.

Contents

Résumé en français **1**

1 Introduction **13**
1.1 The Standard Model of particle physics . . . 13
1.2 Probing the Standard Model with rare b-hadron decays . . . 23
1.2.1 Heavy quark expansion . . . 25
1.2.2 Rare dileptonic B decays . . . 28
1.2.3 Hints of LFU violation in FCNC semileptonic decays . . . 30
1.2.4 Other hints of LFU violation in B decays . . . 33
1.3 Lepton Flavor Violation in $B_{(s)} \to l_i^+ l_j^-$ decays . . . 33
1.3.1 Leptoquark framework . . . 35
1.3.2 Z' models . . . 37
1.4 Lepton Flavor Violation experimental searches . . . 37
1.5 Conclusions . . . 38

2 The LHCb experiment **41**
2.1 Basics on B physics experiments . . . 41
2.2 The Large Hadron Collider . . . 43
2.3 The LHCb detector . . . 47
2.3.1 Tracking system . . . 50
2.3.2 Particle identification systems . . . 53
2.3.3 Trigger systems and the stripping process . . . 57
2.3.4 Simulation . . . 59

3 Experimental analysis: Search for the LFV decays $B_s^0 \to \tau^\pm \mu^\mp$ and $B^0 \to \tau^\pm \mu^\mp$ **61**
3.1 Introduction . . . 61
3.1.1 Description of the BaBar analysis . . . 62
3.1.2 LHCb Run I analysis: workflow and analysis strategy . . . 63
3.2 Event samples . . . 67
3.2.1 Data . . . 67
3.2.2 Simulation . . . 67
3.2.3 Stripping selection . . . 68
3.2.4 Trigger requirements . . . 68
3.2.5 Summary . . . 69
3.3 Normalization channel . . . 74
3.3.1 Event selection . . . 74

3.3.2 Event yield . . . 75
3.4 B invariant mass computation . . . 78
3.4.1 Analytic reconstruction of the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ decay . 79
3.4.2 Choice of the B invariant mass variable . . . 80
3.4.3 Decay Tree Fitter . . . 81
3.4.4 Signal blinding . . . 83
3.5 Preselection . . . 83
3.5.1 Cut-based preselection . . . 83
3.5.2 BDT-based preselection . . . 84
3.6 Specific background rejection . . . 91
3.6.1 Combinatorial background . . . 92
3.6.2 Reverse topology partially reconstructed background . . . 94
3.6.3 Remaining background components . . . 100
3.7 Selection efficiency . . . 101
3.7.1 Tracking reconstruction efficiency correction . . . 103
3.7.2 PID efficiency correction . . . 104
3.7.3 Trigger efficiency . . . 108
3.7.4 Offline selection efficiency . . . 114
3.7.5 MC-truth matching inefficiency correction . . . 116
3.7.6 Selection efficiency summary . . . 117
3.7.7 $B^0_s \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^0\nu)\mu^{\mp}$ efficiency contribution . . . 125
3.7.8 Normalization factor . . . 128
3.8 Simultaneous fit . . . 129
3.8.1 Final BDT . . . 129
3.8.2 Fit strategy . . . 135
3.8.3 Fit validation and optimization . . . 141
3.8.4 CLs method and expected upper limit . . . 146
3.9 Post-unblinding strategy . . . 151
3.9.1 Accounting for a potential bias in the fit . . . 151
3.9.2 Potential improvements . . . 152
3.9.3 Post-unblinding strategy . . . 153
3.10 Unblinded data . . . 154
3.10.1 Using the model based on the Same Sign data shape . . . 154
3.10.2 Fit model generalization . . . 156
3.10.3 Branching ratio limits extraction . . . 157
3.11 Prospects . . . 162
3.12 Summary and conclusion . . . 162

Appendix A Analytic reconstruction of the decay kinematics 165

Appendix B Variables data-MC agreement 169

Appendix C 2D simultaneous fit exploration 175
C.1 Absence of signal . . . 176
C.2 Presence of signal . . . 176

Appendix D Mass resolution evolution **179**
D.1 Limits on the branching ratio of the heavy and light eigenstates 179

Appendix E Run 1 and Run 2 variables distribution comparison **183**

References **188**

Résumé en français

Cadre théorique

Les courants neutres avec changement de saveur (FCNC), tels que les transitions $b \to sl^+l^-$, sont supprimés au 1$^{\text{er}}$ ordre (niveau de l'arbre) dans le modèle standard (SM), mais autorisés aux ordres supérieurs de l'expansion perturbative. Ces processus dits *rares* sont excellentes sondes pour les recherches de physique au-delà du modèle standard (BSM), car les contributions potentielles de BSM peuvent apparaître au même niveau que la contribution du SM.

L'objectif principal des recherches $b \to sl^+l^-$ au cours des dernières décennies a été la recherche de la désintégration $B^0_s \to \mu^+\mu^-$, ce qui a permis d'obtenir une mesure du rapport d'embranchement ($\mathcal{B}$) compatible avec la prédiction SM [1].

Des tests sur l'universalité de la saveur leptonique (LFU) ont été effectués dans les transitions $b \to sl^+l^-$. Le SM prédit des couplages universels dans les 3 familles de leptons, alors que dans certains scénarios BSM, les couplages leptoniques pourraient différer les uns des autres. Les tests expérimentaux de LFU montrent des tensions par rapport aux prédictions SM :

- La mesure $R_{K^*} = \frac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}e^+e^-)}$ effectuée par LHCb [2] dévie par environ 2.5σ des prédictions du SM.

- La mesure $R_K = \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to K^+e^+e^-)}$ effectuée par LHCb [3] dévie par environ 2.6σ de la prédiction du SM.

D'autres tests du LFU, dans ce cas des courants chargés (FCCC), montrent également des écarts par rapport aux prédictions SM.

- La combinaison de $R_D = \frac{\mathcal{B}(B^0 \to D\tau^+\nu_\tau)}{\mathcal{B}(B^0 \to Dl^+\nu)}$ et $R_{D^*} = \frac{\mathcal{B}(B^0 \to D^*\tau^+\nu_\tau)}{\mathcal{B}(B^0 \to D^*l^+\nu)}$ effectué par les collaborations BaBar, Belle et LHCb sont differentes des predictions du SM jusqu'a 4σ [4].

Les déviations possibles de LFU motivent les recherches sur la violation de la saveur leptonique (LFV) car les modèles comprenant la non-universalité leptonique impliquent généralement aussi LFV. Les désintégrations LFV, telles le mode $B^0_{(s)} \to \tau^\pm\mu^\mp$, sont interdites dans le SM en absence des masses des neutrinos, mais peuvent se produire via des diagrammes à une boucle si des oscillations des masses des neutrinos sont incluses, comme montre le schéma de la Figure 1 où le neutrino virtuel doit osciller pour que la désintégration puisse avoir lieu. Le taux des processus LFV étant très supprimé, $(m_\nu/m_W)^4 \sim \mathcal{O}(10^{-48})$, se situent au-delà des sensibilités expérimentales actuelles et à venir. Cependant, une

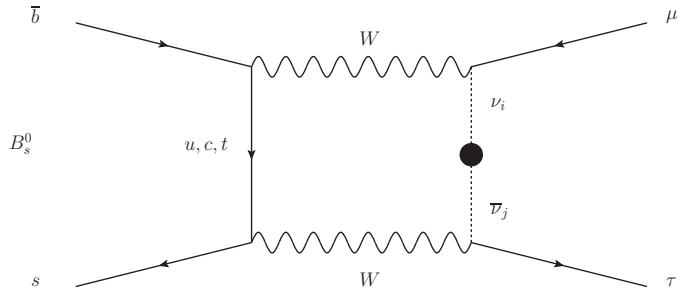

Figure 1: Diagramme de la contribution du modèle standard au processus LFV $B_s^0 \to \tau^{\pm}\mu^{\mp}$ lorsque des oscillations des masse des neutrinos sont incluses.

grande variété de scénarios BSM prédisent des taux considérablement plus élevés pour ces processus.

Pour un scénario général de physique BSM, dans le cadre du Operator Product Expansion (OPE), les nouvelles contributions physiques proviennent généralement de la modification des coefficients de Wilson du SM C_9 et C_{10} (y compris les contributions QCD pingouin) en ajoutant aussi une partie avec la chiralité droite C_9' et C_{10}'. D'autres coefficients d'opérateurs de Wilson qui sont négligés dans le SM comme $C_V^{(\prime)}$, $C_A^{(\prime)}$, $C_S^{(\prime)}$ et $C_P^{(\prime)}$, peuvent acquérir de l'importance dans certains modèles BSM.

Pour un modèle général, le rapport d'embranchement de la désintégration LFV $B_s^0 \to l_i^+ l_j^-$, en termes des coefficients de Wilson, peut être exprimé comme suit: [5]:

$$\mathcal{B}\left(B_s^0 \to l_i^+ l_j^-\right) = \frac{\tau_{B_s^0}}{64\pi^2} f_{B_s^0}^2 G_F^2 m_{B_s^0} \alpha_{em}^2 |V_{tb}V_{ts}^*|^2 \sqrt{\left[1-\left(\frac{m_{l_i}+m_{l_j}}{m_{B_s^0}}\right)^2\right]\left[1-\left(\frac{m_{l_i}-m_{l_j}}{m_{B_s^0}}\right)^2\right]}$$
$$\cdot \left\{ \left[1-\left(\frac{m_{l_i}+m_{l_j}}{m_{B_s^0}}\right)^2\right] \left| \left(C_9 + C_V - C_V'\right)\left(m_{l_i}-m_{l_j}\right) + \frac{m_{B_s^0}^2}{m_b+m_s}\left(C_S - C_S'\right)\right|^2 \right.$$
$$\left. + \left[1-\left(\frac{m_{l_i}-m_{l_j}}{m_{B_s^0}}\right)^2\right] \left| \left(C_{10} + C_A - C_A'\right)\left(m_{l_i}-m_{l_j}\right) + \frac{m_{B_s^0}^2}{m_b+m_s}\left(C_P - C_P'\right)\right|^2 \right\}, \tag{1}$$

où $\tau_{B_s^0}$ est la vie B_s^0, $f_{B_s^0}$ est la fraction de fragmentation du méson B_s^0, G_F la constante de Fermi, α_{em} la constante de couplage électromagnétique et V_{tb}, V_{ts} les éléments de matrice CKM correspondants.

Pour les différents modèles BSM, non exclus par les contraintes expérimentales actuelles, les rapports d'embranchement attendus pour le processus LFV $B_s^0 \to \tau^{\pm}\mu^{\mp}$ sont listés ci dessous. La Figure 2 contient les diagrammes des différents processus BSM.

- **Modèles des leptoquarks scalaires $(3,2)_{1/6}$ [6] :** Un Leptoquark (LQ) est un état coloré qui peut médier des interactions entre des quarks et des leptons portant à la fois des nombres baryoniques et leptoniques. Dans $(SU(3), SU(2)_L)_Y$, il est indiqué si le champ est un -singlet, -doublet ou -triplet de $SU(3)$ et $SU(2)_L$ respectivement, et Y marque l'hypercharge. Pour le LQ scalaire, la contribution au processus $B_s^0 \to \tau^{\pm}\mu^{\mp}$ est au niveau de la boucle et les masses des LQ sont

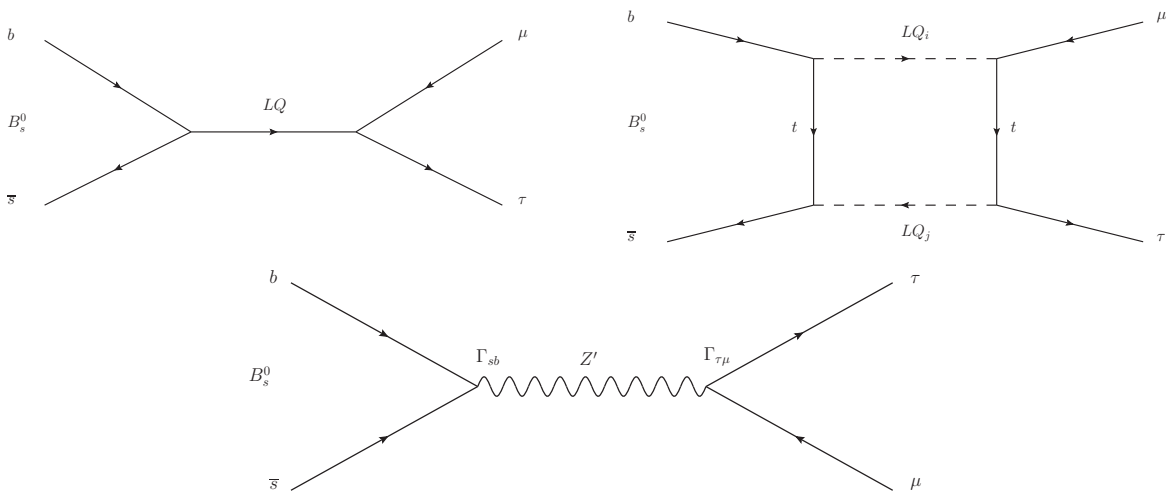

Figure 2: Exemple des diagrammes de Feynman pour les modèles BSM contribuant au processus LFV $B_s^0 \to \tau^\pm\mu^\mp$. 1ère ligne : Leptoquark vecteur (à gauche) et Leptoquark scalaire (à droite). 2ème ligne : Z' avec des couplages aux quarks et leptons.

supposées être $m_{LQ} = m_{LQ^{2/3}} = m_{LQ^{-1/3}}$. En utilisant les mesures expérimentales actuelles comme contraintes, et une certaine supposition pour les couplages, le $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp)$ peut être aussi grand que 10^{-5}, 10^{-6} ou 10^{-7} pour $m_{LQ} = 1$, 5 et 10 TeV respectivement.

- **Modèles des leptoquarks vecteurs $(3,1)_{2/3}$ et $(3,3)_{2/3}$ [6] [7] :** La principale caractéristique ici, en plus des nouvelles interactions, est que les courants LQ induisent un mélange de fermions. Dans modèles $(3,1)_{2/3}$ ($(3,3)_{2/3}$) le $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp)$ peut être aussi grand que 10^{-9} (10^{-6}).

- **Modèles Z' [8] [9] [10] :** De nombreux modèles proposés pour expliquer les données $b \to sl^+l^-$ introduisent un boson vectoriel lourd et neutre (Z') qui génère une contribution au niveau de l'arbre à $C_9^{(\prime)}$ et $C_{10}^{(\prime)}$. Le Z' est supposé être plus lourd que le boson du SM Z^0 et avoir des couplages au niveau des arbres de magnitude différente pour les leptons e, μ et τ et parmi les quarks. Avec les contraintes expérimentales actuelles, les différents modèles prédisent un $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp)$ entre 10^{-8} et 10^{-9}. Cependant, un document publié cette année [31] indique que $B_s^0 \to \tau^\pm\mu^\mp$ est incroyablement sensible à certains modèles Z' (UV complets) utilisés pour décrire les anomalies $R_{K^{(*)}}$ et $R_{D^{(*)}}$ anomalies simultanément. Le rapport de branchement de $B_s^0 \to \tau^\pm\mu^\mp$ est $\sim 10^{-4}$ dans ces types de modèles.

Du côté expérimental, une limite supérieure sur le canal $B^0 \to \tau^\pm\mu^\mp$ a été mesurée par la collaboration BaBar: $\mathcal{B}\left(B^0 \to \tau^\pm\mu^\mp\right) < 2.2 \cdot 10^{-5}$ à 90% CL [11]. La première recherche du canal $B_s^0 \to \tau^\pm\mu^\mp$ est effectuée dans le cadre de l'analyse décrite dans le présent document. Pour cette raison, et parce que les prédictions des rapports d'embranchement parmi les différents modèles sont généralement plus élevées dans le canal B_s^0, l'analyse est optimisée pour $B_s^0 \to \tau^\pm\mu^\mp$.

Le LHC et le detecteur LHCb

Le grand collisionneur de hadrons (LHC) est situé au Centre Européen de Recherche Nucléaire (CERN) à Genève (Suisse). Le LHC est un collisionneur circulaire proton-proton (pp) d'une circonférence de ~ 27 km et situé à une profondeur moyenne de 100m souterrain. Il accélère les protons circulant dans des directions opposées à une énergie de 3.5, 4 et 6.5 TeV correspondant à $\sqrt{s}$ de 7 TeV en 2011, 8 TeV en 2012 (Run I) et 13 TeV de 2015 à la fin de 2017 (Run II). L'analyse décrite dans ce document exploite les données de collision de Run I.

Le LHC fournit 4 points d'interaction, dans lesquels les deux faisceaux de protons se croisent et les collisions ont lieu. C'est là que les détecteurs sont installés. Les quatre détecteurs sont : ATLAS, CMS, LHCb et ALICE.

LHCb [12] est une expérience de précision dédiée à la physique b et c au LHC qui recherche la physique BSM à travers l'étude des désintégrations très rares des hadrons de charme et de beauté et des mesures de précision des observables violant le CP.

La conception du détecteur est celle d'un spectromètre à bras unique dans la région avant, avec une couverture angulaire d'environ 15 mrad 300(250) mrad dans le plan de flexion de l'aimant du détecteur (non-flexion). La disposition du détecteur LHCb dans le plan de flexion est montrée dans la Figure 3. La géométrie du détecteur est pilotée par la production angulaire de $b\bar{b}$ dans les collisions proton-proton. La pseudo-rapidité d'acceptation (η) de LHCb, c'est-à-dire sa couverture angulaire, est unique par rapport aux autres détecteurs du LHC.

Le détecteur LHCb est divisé en différents sous-systèmes en fonction de leur fonctionnalité:

- **Système trajectographique :** en charge de mesurer les observables dynamiques des traces chargées.
- **Système d'identification de particules (PID) :** il distingue les différentes particules finales.
- **Trigger et acquisition des données (DAQ) :** il enregistre uniquement les événements intéressants et traite les données pour les analyses physiques ultérieures.

Analyse expérimentale

L'analyse est effectuée sur les données enregistrées par le détecteur LHCb au cours des années 2011 et 2012. Le but est d'effectuer la mesure du rapport d'embranchement des processus $B^0_{(s)} \to \tau^{\pm}\mu^{\mp}$ en utilisant le mode de désintégration $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ comme canal de normalisation.

$$\mathcal{B}\left(B^0_{(s)} \to \tau^{\pm}\mu^{\mp}\right) \propto \frac{N^{obs}_{B^0_{(s)}\to\tau^{\pm}\mu^{\mp}}}{\epsilon^{sig}_{B^0_{(s)}\to\tau^{\pm}\mu^{\mp}}} \mathcal{B}_{norm} \frac{\epsilon^{sig}_{norm}}{N^{obs}_{norm}} \tag{2}$$

En l'absence de signal, comme prévu dans le SM, une limite supérieure pour le rapport d'embranchement sera définie.

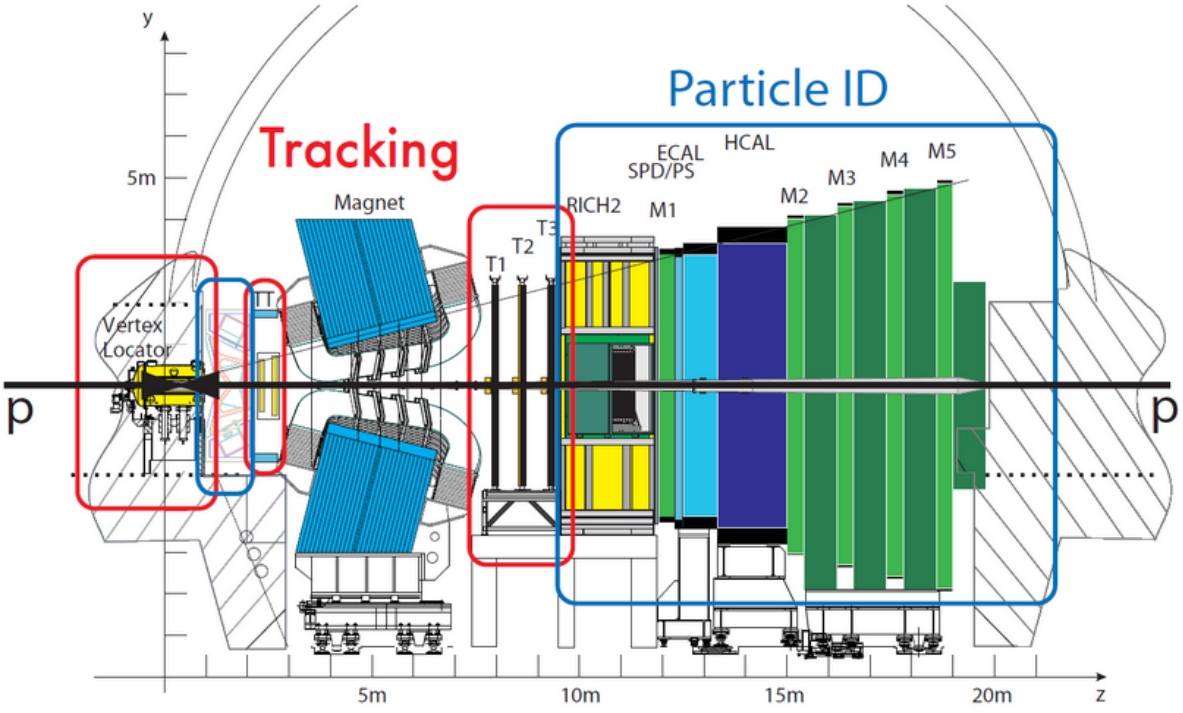

Figure 3: Disposition du détecteur LHCb dans le plan de flexion de l'aimant.

Le mode de désintégration τ choisi est $\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu$, où $\mathcal{B}\left(\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu\right) = (9.02 \pm 0.05)\%$ [13]. Le τ lepton se désintègre via deux résonances intermédiaires qui sont utiles dans la sélection du signal :

$$\tau^{+} \to a^{+}(1260)\nu \to \rho(770)\pi^{+}\nu \to \pi^{+}\pi^{-}\pi^{+}\nu.$$

Le mode de désintégration $\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^{0}\nu$ avec $\mathcal{B}\left(\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^{0}\nu\right) = (4.49 \pm 0.05)\%$ [13] contribue au signal, bien que de manière moins significative.

La Figure 4 montre la topologie de la désintégration. Une paire $b\bar{b}$ est produite par les collisions des parton au vertex primaire (PV) suivi de l'hadronisation de l'un des quarks dans un méson $B^{0}_{(s)}$. Le méson $B^{0}_{(s)}$ vole jusqu'à ce qu'il se désintègre en un μ, qui traverse généralement tout le détecteur, et un τ, qui se désintègre presque immédiatement en 3π et un neutrino. La signature du signal présente les propriétés suivantes :

- 3 traces de pion provenant d'un vertex commun déplacé du PV,
- Une trajectoire de muons ne pointant pas vers le PV.

Le mode $B^{0} \to D^{-}(\to K^{+}\pi^{-}\pi^{-})\pi^{+}$ a une topologie très proche de celle du signal, c'est-à-dire le même nombre de traces à l'état final et trois hadrons légers provenant d'un vertex déplacé. Un ensemble de coupures sur des variables fournies par la reconstruction hors-ligne des données est appliqué pour selectionner les candidats $B^{0} \to D^{-}(\to K^{+}\pi^{-}\pi^{-})\pi^{+}$. La distribution de la masse invariante des candidates sélectionnés est presentée sur la Figure 5.

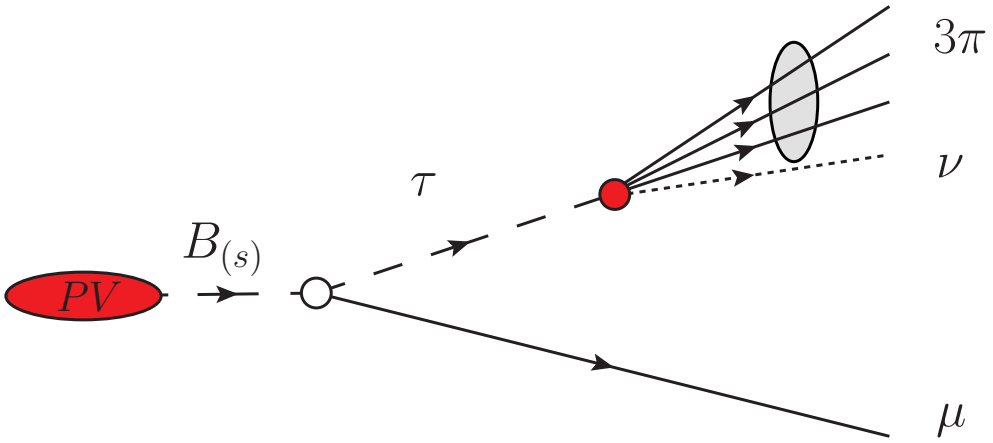

Figure 4: Topologie du processus $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$.

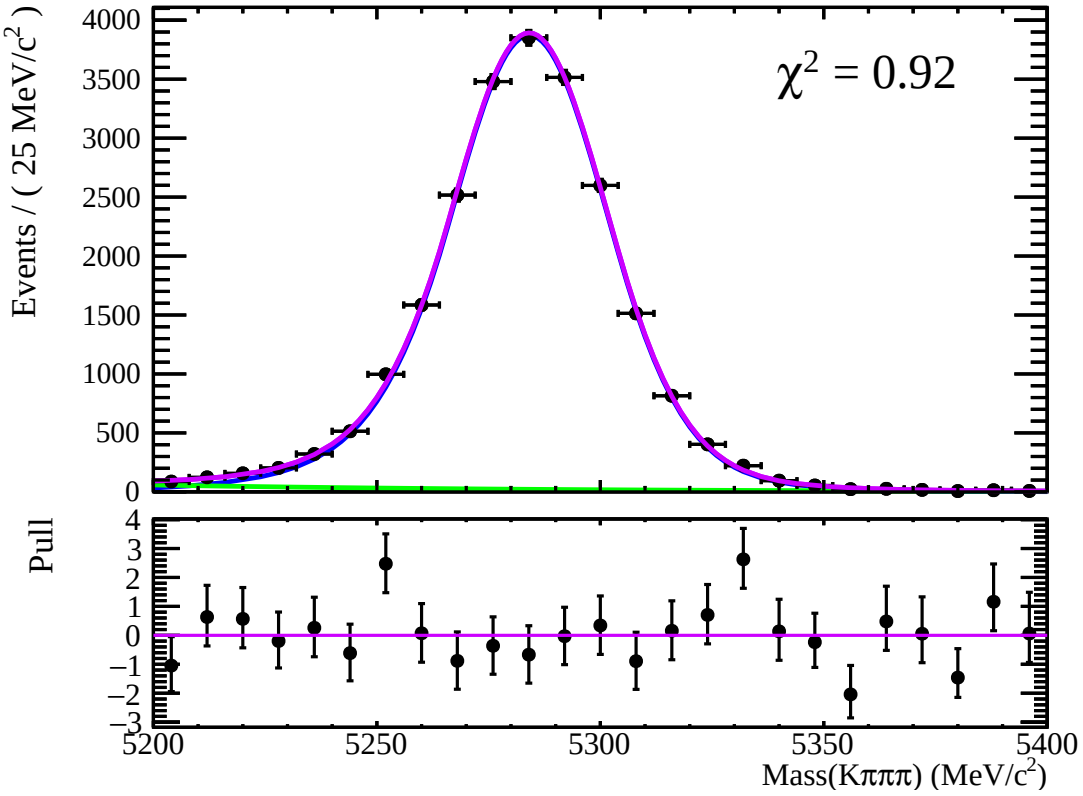

Figure 5: Ajustement des donées du $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ sur la masse invariante B après de la sélection hors ligne. Le PDF total est représenté en violet, le signal en bleu et le fond en vert.

La première étape de l'analyse est la reconstruction du signal étant donné que le neutrino échappe à la détection. Une technique de reconstruction spécifique est utilisée pour déduire l'énergie du ν, en profitant de la position reconstruite du vertex des 3π. De cette façon, la cinématique complète du processus peut être résolue à une double ambiguïté

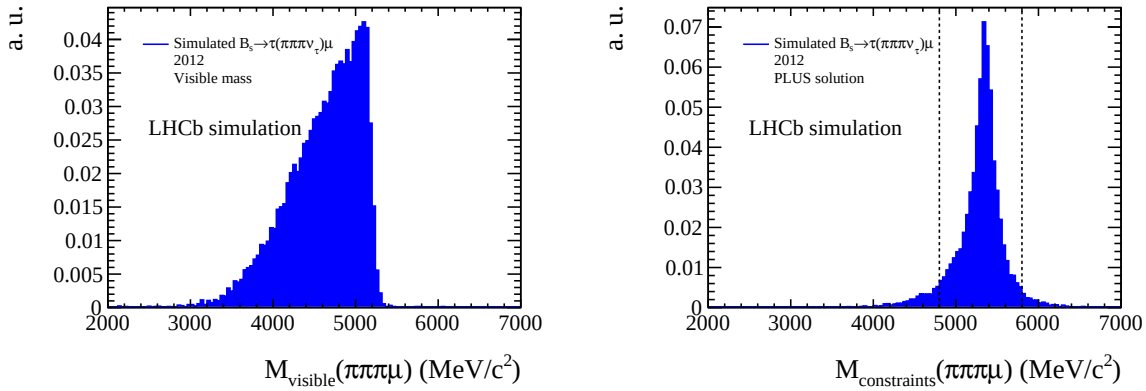

Figure 6: Simulation du signal $B_s^0 \to \tau^\pm \mu^\mp$: les distributions de la masse invatriante visible (sans le neutrino) du meson B_s^0 (gauche) et la masse invariante du meson B_s^0 avec la reconstruction de masse dédiée.

près. Une fois la procédure de reconstruction terminée, la masse invariante du méson $B^0_{(s)}$ peut être calculée. La Figure 6 montre les distributions de la masse invariante du B_s^0 dans l'échantillon de simulation sans tenir compte du neutrino et avec la reconstruction dédiée.

En utilisant la masse invariante B reconstruite à l'aide d'échantillons de simulation de signal, une région de signal est définie et l'échantillon de données est aveuglé dans cette région.

Afin de séparer le signal du bruit de fond, une sélection hors ligne composée de différentes étapes est appliquée. Des techniques d'analyse multivariées, telles comme des Boosed Decision Trees (BDT) [14], sont utilisées pendant le processus de sélection.

D'abord, une présélection est appliquée en deux niveaux étapes:

- Une présélection basée sur des coupures sur les variables fournissant un pouvoir de discrimination élevé entre le signal et du bruit de fond pour éliminer des bruits de fond potentiellement dangereux (i.e. $B^0_{(s)} \to D_s\,(\to \pi\pi\pi)\,X$).

- Une présélection basée en utilisant un BDT basé sur des variables d'isolation, construit pour fournir un rejet de fond élevé et réduire les données à un niveau gérable. Les variables d'isolation examinent la présence de traces non désirées au voisinage des traces et/ou des vertex candidats.

Les coupures sur les distances de vol de désintégration, comme la signification de la distance de vol candidate B par rapport au vertex primaire et les paramètres d'impact, rejettent la plupart des bruits de fond impliquant des particules provenant du vertex primaire.

Les bruits de fond restants sont construits avec des traces de particules se décomposant après une distance importante. Comme on peut l'observer dans les distributions de temps de décroissance du τ montrées sur la Figure 7 pour différentes fenêtres de masse du B, deux composants différentes peuvent être distinguées :

- une composante avec un temps de décroissance reconstruit exponentiellement décroissant caractéristique des désintégrations τ ou D, constitué principalement de désintégrations B partiellement reconstruites. Cette composante est présente principalement à faible masse et s'étend à l'intérieur de la région du signal.

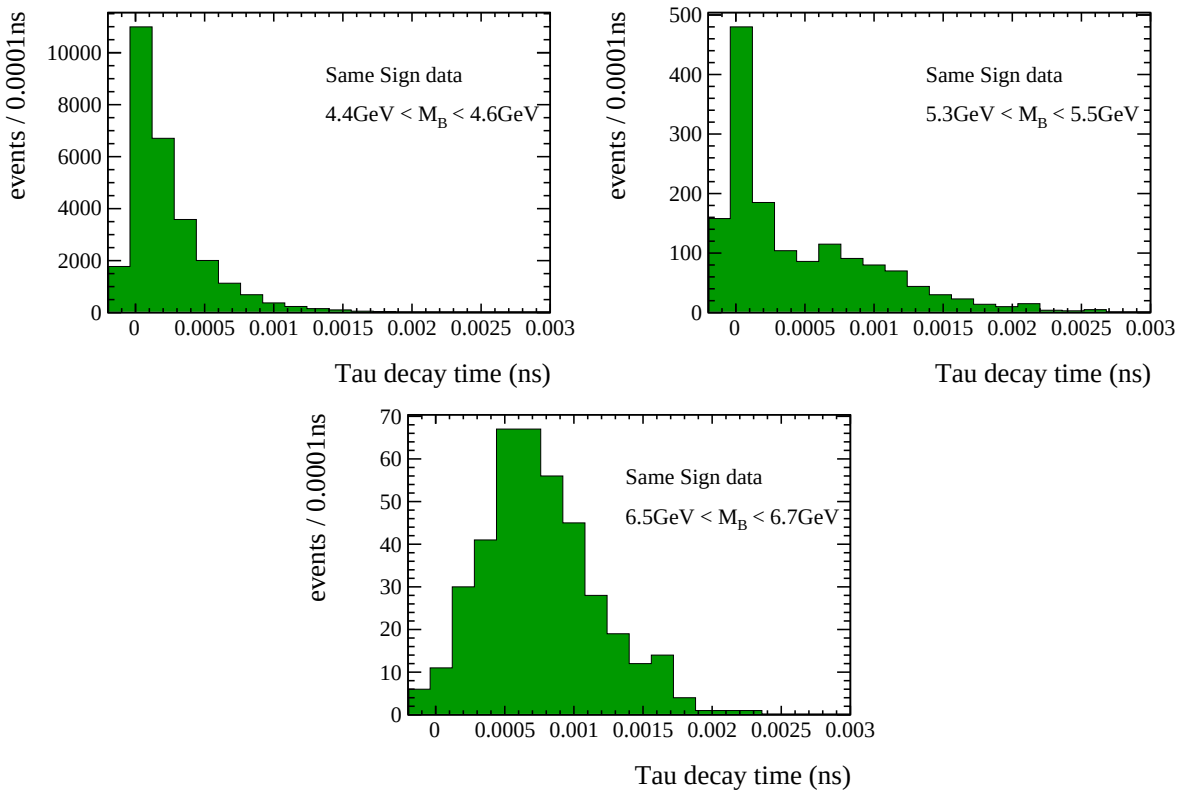

Figure 7: Distribution du temps de vie du τ pour différentes fenêtres de masse du B dans l'échantillon du bruit de fond $B^0_{(s)} \to \tau^\pm\mu^\pm$: [4.4,4.6] MeV (en haut à gauche), [5.3,5.5] MeV (en haut à droite), [6.5,6.7] MeV (en bas).

- une composante du bruit de fond combinatoire avec une distribution de temps de décroissance très large centrée sur $5 \cdot 10^{-4}$ ns. Cette composante est la seule présente en masse élevée alors qu'elle est complètement dominée par le fond partiellement reconstruit dans la région de masse faible.

Les désintégrations B partiellement reconstruites peuvent être divisés en deux catégories en fonction de leur topologie (Figure 8) :

- *Topologie inverse à celle du signal* : Les 3 pions proviennent directement du vertex de désintégration B et du muon d'une désintégration de mésons $D^{(*)}$.

- *Topologie semblable à celle du signal* : les 3 pions proviennent d'un vertex déplacé après le vertex de désintégration B. Ces désintégrations n'ont pas toujours la même topologie que le signal, car le muon peut aussi provenir d'une désintégration τ ou $D^{(*)}$.

Afin de supprimer ces composantes du bruit de fond spécifiques, deux autres étapes de sélection sont appliquées. La première étape consiste en un BDT ciblant le bruit de fond combinatoire. Dans la deuxième étape, une coupure est placée sur le temps de vie du τ pour rejeter les candidates avec une topologie inverse au signal. Une fois la sélection hors ligne complète appliquée, un seul élément du bruit de fond partiellement reconstruit

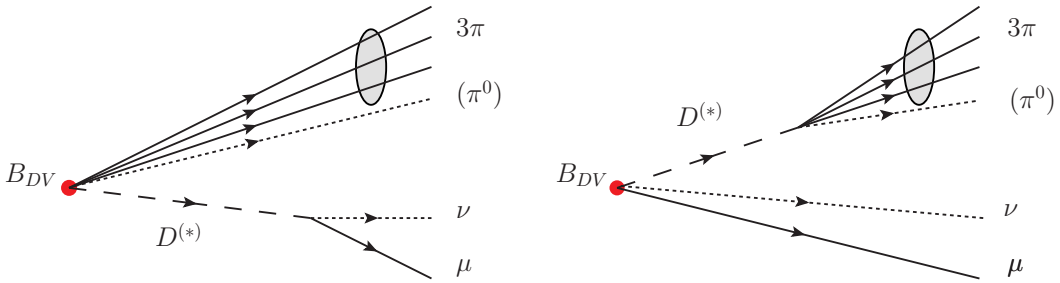

Figure 8: Topologies principales des bruits du fond partiellement reconstruits . Gauche: topologie inverse au signal. Droite: topologie semblable au signal.

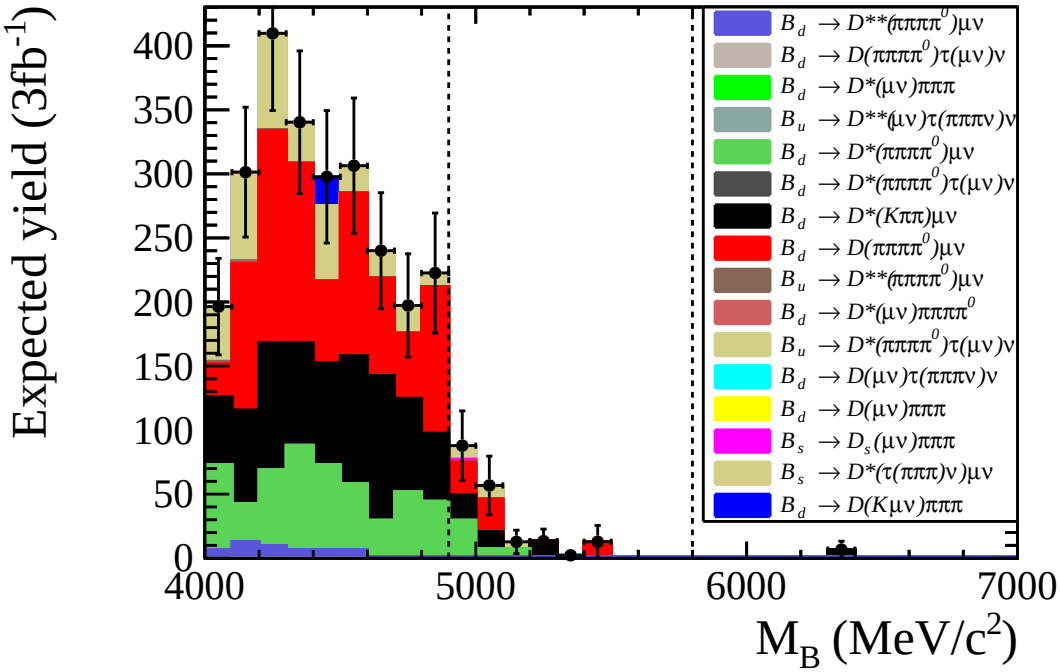

Figure 9: Distributions de masse invariante du B obtenus a partir d'echantillons de bruits de fond exclusifs une fois la sélection hors ligne a eté appliquée. Le graphique montre le *yield* attendu selon une luminosité intégrée de 3 fb^{-1}.

survit: celui où les 3π proviennent d'un vertex déplacé, imitant la signature du signal. Comme montre la Figure 9 aucun bruit de fond se démarque dans la région du signal. La distribution des echantillons du signal auprés du selection est montre dans la Figure 10 et la Figure 11 montre la distribution finale des données sélectionnées.

Les efficacités des differentes étapes de selection du signal sont estimées soit en utilisant la simulation, soit, lorsque la simulation n'est pas assez fiable, des techniques basées sur les donées. En particulier, les techniques utilisées pour calculer l'efficacité du trigger des canaux de signal et de normalisation et la correction de l'efficacité de la sélection hors ligne ont été particulièrement développées et conçues pour cette analyse.

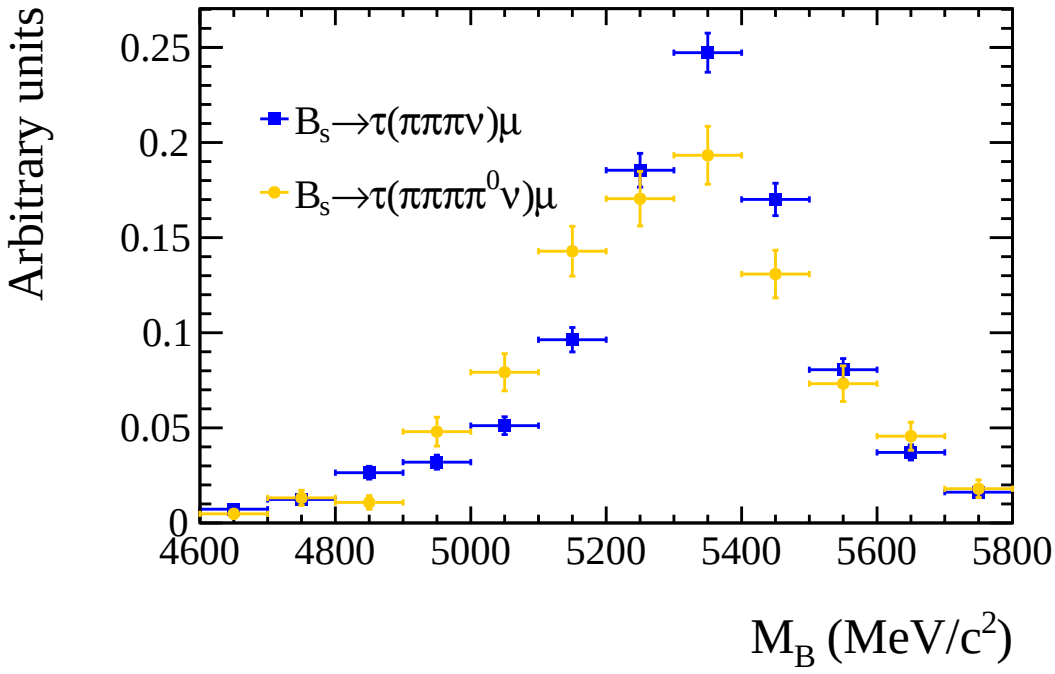

Figure 10: Distributions de masse invariante des echantillons MC $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ et $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^0\nu)\mu^{\mp}$ auprés de la sélection hors ligne.

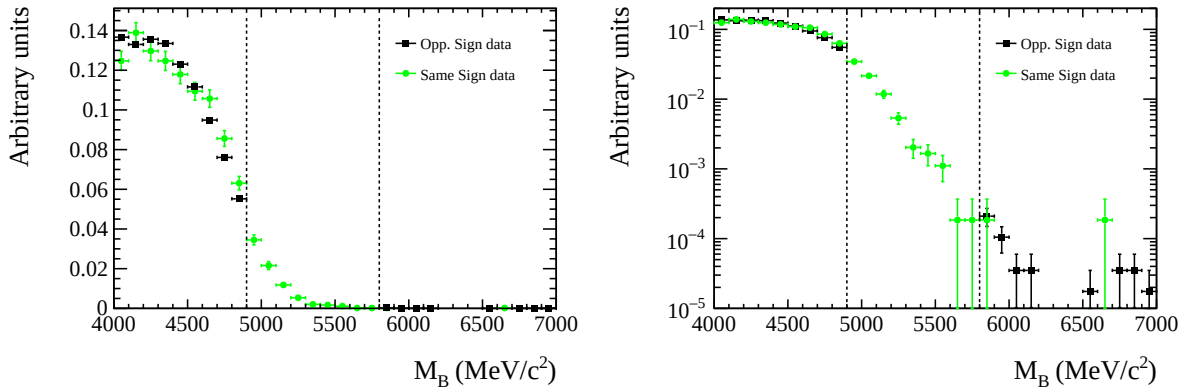

Figure 11: Data B invariant mass distribution once the complete offline selection have been applied. Les deux graphiques montrent les mêmes données, la gauche avec un axe linéaire et la droite avec un axe logarithmique.

À ce stade de l'analyse, un BDT final est formé en intégrant le pouvoir de discrimination restant. La Figure 12 montre la variable de sortie fournie par ce BDT pour les différents échantillons de signaux et de données.

La stratégie d'analyse est complétée par un ajustement simultané à la distribution de masse invariante candidate B sur les différents segments du BDT final. Les données sont

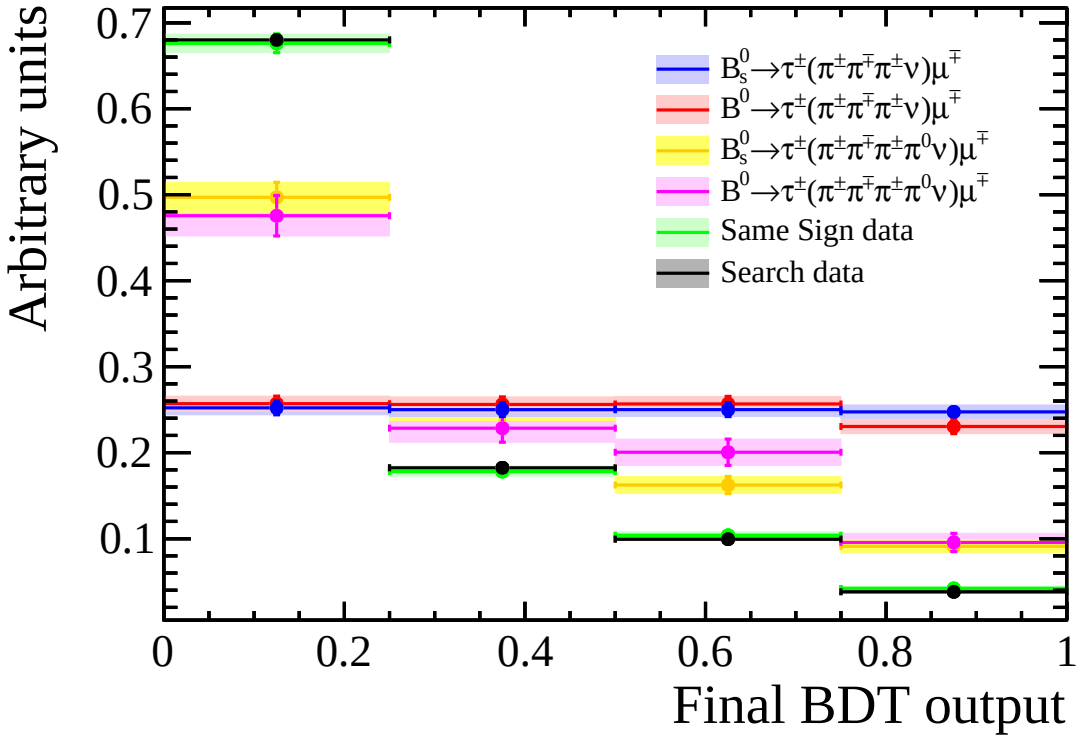

Figure 12: Distribution de l'output du BDT final pour la simulation du signal $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$, $B^0_s \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ et des données appelées comme Same Sign ($B^0_{(s)} \to \tau^\pm\mu^\pm$).

modeliseés de la façon suivante :

$$PDF^{tot} = \sum_{i}^{\text{BDT bins}} \left(\mathbf{N^{sig}} \epsilon_i^{sig} Hyp_i^{sig} + n_i^{bkg} CrystalBall_i^{bkg}(\boldsymbol{\mu}, \sigma_i, \boldsymbol{\alpha}, \boldsymbol{\eta}) \right) \tag{3}$$

où les paramètres en rouge sont laissés libres pendant le processus d'ajustement;

- $\mathbf{N^{sig}}$: Nombre d'événements de signal. Paramètre commun entre les ajustements.
- n_i^{bkg} : Nombre d'événements de bruit du fond à chaque segment du BDT final.
- ϵ_i^{sig} : l'efficacité totale du signal par segment de BDT. Libre avec des contraintes gaussiennes. La largeur de la contrainte gaussienne correspond à l'incertitude sur l'efficacité du signal dans chaque segment de BDT.
- Hyp_i^{sig} : Hypatia PDF décrivant la forme du signal avec les paramètres fixés à partir du fit de la simulation. La largeur de l'hypathia peut varier dans les contraintes gaussiennes de σ correspondant à 12% de la valeur centrale d'efficacité.
- $CrystalBall_i^{bkg}(\boldsymbol{\mu}, \sigma_i, \boldsymbol{\alpha}, \boldsymbol{\eta})$: Crystal Ball PDF décrivant la forme du bruit de fond.

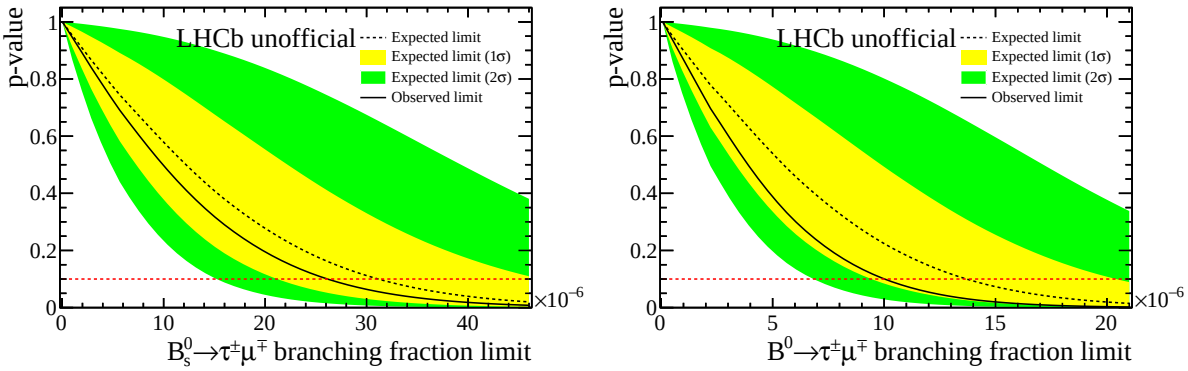

Figure 13: $B_s^0 \to \tau^\pm\mu^\mp$ (gauche) and $B^0 \to \tau^\pm\mu^\mp$ (droite) intervalles CLs utilisés pour évaluer les limites supérieures attendues des rapports d'embranchement.

Le nombre optimal de segments dans lesquels le BDT est divisé est choisi en optimisant la sensibilité d'ajustement tout en gardant l'ajustement non biaisé.

Le nombre d'événement de signal observés sont:

- $\mathrm{N}^{\mathrm{sig}}_{B_s^0\to\tau^\pm\mu^\mp} = -18 \pm 38$,
- $\mathrm{N}^{\mathrm{sig}}_{B^0\to\tau^\pm\mu^\mp} = -63 \pm 57$.

correspondant respectivement à des fluctuations à la baisse de 0.3 et 1.1σ. Par conséquent, aucun excès de signal significatif n'est observé.

La méthode CLs [15] est utilisée pour extraire les limites supérieures des rapports d'embranchement. La méthode CLs compare la probabilité que les données soient bien décrites par des hypothèses de signal plus de bruit de fond ou de bruit de fond seulement. Les graphiques d'exclusion des limite supérieures sont montrés dans la Figure 13. Les limites supérieures des rapports d'embranchement attendues (observées) sont :

- $\mathcal{B}\left(B_s^0 \to \tau^\pm\mu^\mp\right) < 3.0(2.5) \cdot 10^{-5}$ at 90%CL,
- $\mathcal{B}\left(B^0 \to \tau^\pm\mu^\mp\right) < 1.3(1.0) \cdot 10^{-5}$ at 90%CL.

Ces résultats représentent les meilleures limites supérieures à ce jour, étant la première mesure mundiale pour le mode du B_s^0.

Chapter 1

Introduction

1.1 The Standard Model of particle physics

The Standard Model (SM) describes the particles composing the universe and their interactions. It is built within the Quantum Field Theory (QFT) framework, comprising the concept of *field* and describing its dynamics via quantum mechanics and special relativity. Furthermore, it is minimal and assembled from first principles based on symmetries known to be respected, up to now, by nature.

The SM has an enormous predictive power. It has been tested with great precision showing agreement between the predicted observables and the experimental measurements. However, it is known that the SM can not explain everything we see, giving rise to several fundamental questions: Where is the limit of the SM applicability? Are the particles described in the SM truly elementary? Can it be joined with gravity?

In this section, the SM building blocks are explained from a pragmatic point of view. For a more detailed description of the SM, the reader is referred to [16].

Foundations

Particles are understood as oscillations of a dynamic field. A *field* is a quantity defined at every point of space and time $(t, \vec{x})$. As the laws of nature are relativistic, the fields must be invariant under Lorentz transformations, imposed by special relativity. If the field comprises time and space at the same level it will be Lorentz invariant i.e. the field will behave the same way under boosts and/or rotations. Depending on how their representations transform under the Lorentz group, particles can be categorized as:

- **Fermions**: Fields with half integer spin $\Psi_i(t, \vec{x})$

or

- **Bosons**: Fields with integer spin $\phi_i(t, \vec{x})$,

where the *spin* is a quantum number associated to the intrinsic angular momentum of the fundamental particles.

The dynamics of a system is governed by the so-called *Lagrangian* ($\mathcal{L}$). $\mathcal{L}$ is obtained by imposing the principle of minimal action ($\mathcal{S}$) to a path followed by a given field $\varphi(t, \vec{x})$, leading to the following equations:

$$0 = \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \frac{\partial \varphi(x_j,t)}{\partial t}}\right) - \frac{\partial \mathcal{L}}{\partial \varphi(x_j)}. \tag{1.1}$$

In essence, $\mathcal{L}$ describes the difference between the body motion energy (kinetic) and the energy due to the interaction with the system (potential). Therefore, $\mathcal{L}$ is be expressed as the difference of the *free* and the *interaction* energies:

$$\mathcal{L} \approx E_{\text{free}} - E_{\text{int}} \tag{1.2}$$

Within nature, 3 fundamental interactions are known. One of them is gravity, which is not going to be described given as it does not contribute to the physical processes depicted in this document. The other two fundamental interactions are:

- **Electroweak interaction**: It is the unified description of the electromagnetic interaction, represented by *Quantum Electroynamics* (QED), and the *weak interaction*. Within this theory physical particles have an electrical charge (Q), positive, negative or neutral and a weak isospin (T_3), where the sub-index 3 denotes its 3$^{\text{rd}}$ component. Mathematically, the unification is achieved under a $U(1)_Y \times SU(2)_L$ gauge group invariance. $Y = 2(Q - T_3)$ represents the quantum number of hypercharge and L indicates exclusive coupling to states with left chirality.

- **Strong interaction**: It is described by *Quantum Chromodynamics* (QCD). Within this theory particles have three color charges (c). It is mathematically accommodated under $SU(3)_c$ gauge group invariance.

The SM accomodates the electroweak and strong interactions in the same mathematical framework $U(1)_Y \times SU(2)_L \times SU(3)_c$.

The fundamental particle content of the SM, also observed experimentally, is shown in Figure 1.1 and described in the two following sections.

From now on the coordinate indices are implicit. A given 4-vector with space-time dependence is implicitly expressed as $a \equiv a^\mu = a(t, \vec{x})$. The product of two 4-vectors is assumed to be summed over all components: $ab \equiv a^\mu b_\mu$.

Particle content: Fermions

Fermions are the building blocks of the known matter. They are divided into two categories: *leptons* and *quarks*.

Leptons have no color charge but they may have electrical charge. Three kinds of charged leptons exist (e, μ and τ) with $Q = -1$ and each are associated to a neutral neutrino (ν_e, ν_μ and ν_τ).

Quarks have both electrical and color charges and they are classified in six different flavors divided into two families: the *up-type* family (u, c, t) with electrical charge $+2/3$, and the *down-type* (d, s and b) with electrical charge $-1/3$. Fermions are grouped by pairs

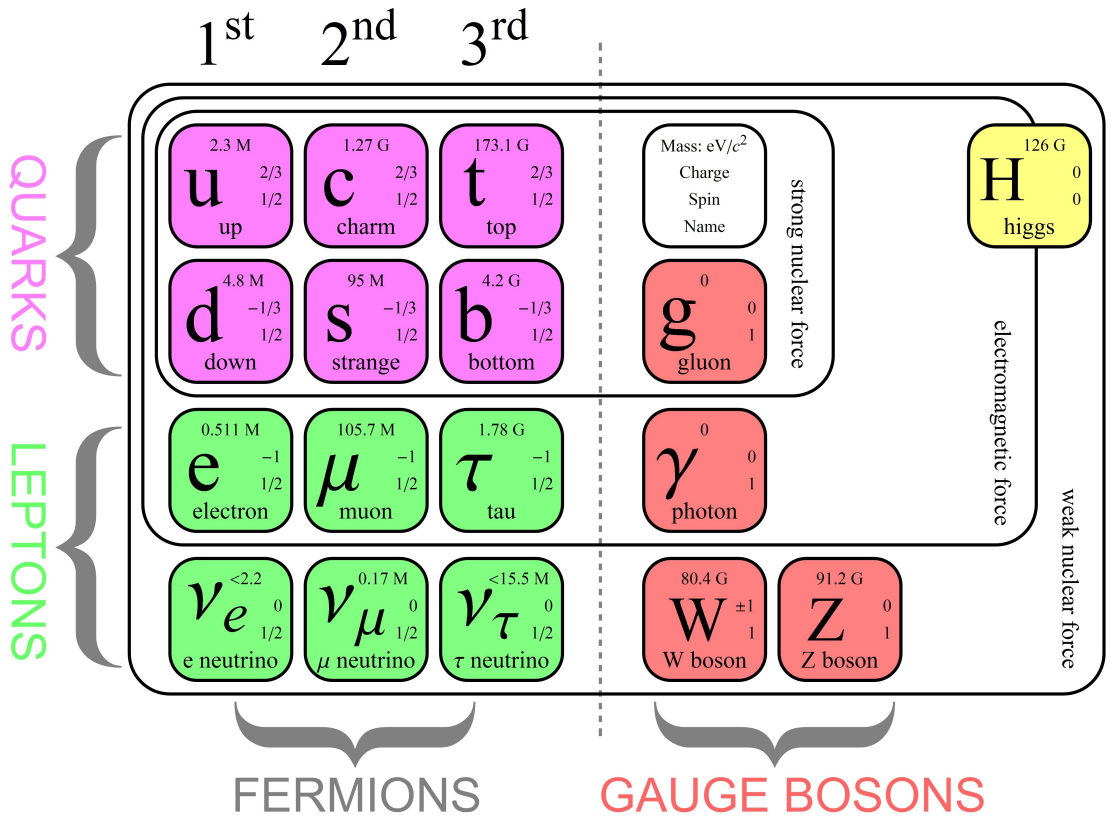

Figure 1.1: Standard Model particle content. The numbers describing each particle, from top to bottom, correspond to: mass (in MeV/c^2 (M) or GeV/c^2 (G)), electric charge and spin.

into three generations, as represented in Figure 1.1. QCD is a binding *strong* force at low energies, consequently quarks are typically seen in colorless aggregate states which are generically called *hadrons*. Two special cases are: *mesons*, formed by quark and anti-quark aggregation, and *baryons* formed by the aggregation of 3 quarks; the main examples of the later are the protons and neutrons which form the atomic nuclei.

Fermions fulfill the Dirac equation :

$$\left(i\partial\!\!\!/ - m\right)\psi(x) = 0, \tag{1.3}$$

where $\partial\!\!\!/ = \gamma^\mu \partial_\mu$ and γ^μ are the Dirac gamma matrices and m is the field mass. The Dirac equation has two solutions, one of them implying negative energy. This negative energy solution is associated to the existence of antimatter. Therefore, for each fermion f an associated antiparticle $\overline{f}$ exists. Despite that a fermion and an anti-fermion have the same mass, the antifermion has color charge, electrical charge and/or chirality opposite to the corresponding fermion.

Fermions and antifermions are mathematically defined by the following fields:

$$\begin{aligned}
\psi(x) &= \frac{1}{(2\pi)^{2/3}} \int \frac{dp}{\sqrt{2p^0}} \sum_{r=-\frac{1}{2},\frac{1}{2}} \left(e^{-ipx} a_r(p) u_r(p) + e^{ipx} b_r^\dagger(p) v_r(p)\right), \\
\bar{\psi}(x) &= \frac{1}{(2\pi)^{2/3}} \int \frac{dp}{\sqrt{2p^0}} \sum_{r=-\frac{1}{2},\frac{1}{2}} \left(e^{-ipx} b_r(p) \overline{u}_r(p) + e^{ipx} a_r^\dagger(p) \overline{v}_r(p)\right),
\end{aligned} \tag{1.4}$$

where $a_r^\dagger(b_r^\dagger)$ and $a_r(b_r)$ are the so-called creation and annihilation operators introduced during the quantization of the field and p is the field 4-momentum. The operators fulfill the anti-commutation rules. r is the spin. Given that u_r, v_r, $\overline{u}_r$ and $\overline{v}_r$ are 4-component spinors, it implies that $\psi(x)$ is a 4-vector column.

Dirac equations have two independent solutions u_r and u_l which correspond to different polarization states related with the intrinsic angular momentum of the particle. Therefore, fermions are generally decomposed between their *left* (L) and *right* (R) *chirality*:

$$\psi = \psi_L + \psi_R. \tag{1.5}$$

A main characteristic of the SM is that only left left-handed fermions and right-handed antifermions couple with the weak interaction, hence all SM fermions must bear left-handed chirality.

From the Dirac equation, the Lagrangian for a free fermion reads as:

$$\mathcal{L}_{\text{free}}^{\text{fermion}} = i\overline{\psi}\partial\!\!\!/\psi - m^2\overline{\psi}\psi. \tag{1.6}$$

Particle content: Bosons

Depending on how bosons transform under the Lorentz group they are classified in *vector* or *scalar*. The SM consists in eight vector bosons so-called *gluons* (g) which act as messengers of the strong interaction; four vector bosons acting as messengers of the electroweak interaction, named *photon* (γ), Z^0 and $W^\pm$; and one scalar boson, the so-called *Higgs* (H) boson, which couples to the SM particles allowing them to acquire mass.

Bosons fulfill the Klein-Gordon equation, which likewise predicts negative energy solutions. The neutral electric charged bosons (H, γ, Z^0 and gluons) coincide with their own antiparticle as their properties remain invariant under charge conjugation. Each gluon has one color and one anti-color. Concerning the W^+ boson, the opposite charge boson W^- is its antiparticle.

Gluons and photons are massless and have two degrees of freedom, while Z^0 and $W^\pm$ are massive and have 3 degrees of freedom. Each degree of freedom corresponds to a polarization state.

A massless vector boson is defined as the following field:

$$A_\mu(x) = \frac{1}{(2\pi)^{2/3}} \int \frac{dp}{\sqrt{2p^0}} e_\mu^n(p) \left(e^{-ipx} a_\lambda(p) + e^{ipx} a_\lambda^\dagger(p) \right), \tag{1.7}$$

where $a^{(\dagger)}$ are the annihilation(creation) operators and e_μ^n (n = 1, 2 and 3) represents the polarization vector. Formally this field has four polarizations, but only two of them correspond to physical degrees of freedom. The corresponding Lagrangian for a massless vector reads as:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \qquad \text{with } F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{1.8}$$

Concerning the massive vector bosons and anti-bosons, they are defined as:

$$U_\mu(x) = \frac{1}{(2\pi)^{2/3}} \int \frac{dp}{\sqrt{2p^0}} \sum_{n=1,2,3} e^n_\mu(p) \left(e^{-ipx} a_n(p) + e^{ipx} b^\dagger_n(p)\right),$$
$$U^*_\mu(x) = \frac{1}{(2\pi)^{2/3}} \int \frac{dp}{\sqrt{2p^0}} \sum_{n=1,2,3} e^n_\mu(p) \left(e^{-ipx} b_n(p) + e^{ipx} a^\dagger_n(p)\right), \tag{1.9}$$

where again e^n_μ are the polarization vectors. The corresponding Lagrangian for a massive vector boson is then:

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu U^*_\nu - \partial_\nu U^*_\mu)(\partial^\mu U^\nu - \partial^\nu U^\mu) + m^2 U^*_\mu U^\mu. \tag{1.10}$$

For a scalar boson field, the Lagrangian reads as:

$$\mathcal{L} = \partial_\mu \phi^* \partial_\mu \phi - m^2 \phi^* \phi. \tag{1.11}$$

Particle content: Evolution of states and interactions

The propagation of a free non-interacting field is described via the *propagator*. One way to find the propagator for a free field is to compute the corresponding expected value of the time ordered (T) fields on the vacuum ($|0\rangle$). For two coordinates x and y and a given field φ, the propagator is defined as:

$$G(x,y) = \langle 0|T\left(\varphi(x)\varphi(y)\right)|0\rangle. \tag{1.12}$$

Applying it to the fermionic and bosonic fields:

$$\langle 0|T\left(\psi(x)\overline{\psi}(y)\right)|0\rangle = \frac{i}{(2\pi)^4} \int dp \frac{e^{-ip(x-y)}(|p| + m)}{p^2 - m^2 + i\epsilon},$$
$$\langle 0|T\left(U_\mu(x)U^*_\nu(y)\right)|0\rangle = \frac{-i}{(2\pi)^4} \int dp \frac{e^{-ip(x-y)}(g_{\mu\nu} - \frac{p_\mu p_\nu}{m^2})}{p^2 - m^2 + i\epsilon},$$
$$\langle 0|T\left(A_\mu(x)A_\nu(y)\right)|0\rangle = \frac{-i}{(2\pi)^4} \int dp \frac{e^{-ip(x-y)} g_{\mu\nu}}{p^2 + i\epsilon}, \tag{1.13}$$

where the factor $i\epsilon$ is introduced to avoid the integral poles in order to achieve finite results.

In QFT the evolution of states is described by a transition amplitude ($\mathcal{M}$), which will relate the initial ($|in\rangle$) and final ($|out\rangle$) states involving the system Lagrangian which, in general, may contain an interaction term:

$$\mathcal{M} \equiv \langle out|S|in\rangle. \tag{1.14}$$

S is the so-called *action*. The system follows the time ordered (involving causality) *minimal action principle*:

$$S = Te^{i\int dx(\mathcal{L})}. \tag{1.15}$$

The modulus squared of $\mathcal{M}$ integrated over all the Lorentz invariant space phase is directly proportional to the differential transition probability ($d\omega$) to evolve from $|in\rangle$ to $|out\rangle$.

$$d\omega = (2\pi)^4\delta^4\left(\sum_i^{\text{particles}} p_i'\right)|\mathcal{M}|^2\prod_{j=1}^{r}\frac{dp_j'}{(2\pi)^3 2E_j'}, \tag{1.16}$$

where δ^4 is a Dirac delta on the incoming and outgoing particle's momentum.

Up to now in the text, only free Lagrangians have been looked at, but non-trivial transitions happen due to interaction of fields. In the SM the interactions are local e.g. the interactions take place in a given point of the space-time called *vertex*. The usual strategy to introduce the interactions is to consider them as perturbations to the free Lagrangian ($\mathcal{L}_{\text{free}} >> \mathcal{L}_{\text{int}}$), therefore making the SM a *low energy theory* in the *interaction picture*. The interaction picture is a useful viewpoint in quantum mechanics to describe situations where we have small perturbations of a well-understood system. For example, for a generic scalar field ϕ:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \sum_{n\geq 3}\frac{\lambda_n}{n!}\phi^n \approx \mathcal{L}_{\text{free}} - \mathcal{L}_{\text{int}} \tag{1.17}$$

where the coefficients λ_n are called *coupling constants* and parametrize the coupling strength between fields. The condition to ensure that the chosen additional terms are 'small' perturbations and they lead to finite physical quantities is that the coupling constants have dimensions of one power of mass (in natural units). If the coupling constants are dimensionless, they are marginal and do not contribute significantly. On the other hand, if the coupling constants are of higher order, they lead to divergences.

The term *coupling constants* is missleading. In fact, the coupling strengths depend on the energy scale. Therefore the SM deals with *running coupling constants*. For QED, the strength of the coupling increases with the energy, whereas for QCD, the strength of the coupling diminishes with the energy producing the so-called *asymptotic freedom* of quarks and gluons.

Feynman rules

Any interaction process amplitude can be summarized using diagrams. The link between the amplitude formula and the diagram is given by the *Feynman rules*. A given process is described by the initial and the outgoing particles represented by propagators. Each propagator has a given field structure associated to it. Where two states interact, a vertex is drawn as the intersection between propagators. Each vertex has associated a coupling constant, depending on the interactions involved between the propagators. Furthermore, internal propagators which connect vertexes also exist and they represent the intermediate, also called *virtual*, particles.

At first order in perturbation theory, only vertices joined by one propagator are drawn. A process with alike diagrams is called a *tree-level* process. An example of a tree-level process can be seen in Figure 1.2 (left).

When more orders in perturbation theory are to be computed, i.e. corrections to the tree-level process, *loops* are introduced to the propagators of the tree-level diagram. The loops correspond to second order physics effects, like for example the radiation of a

photon by an electron propagator. Each loop has associated an integral over a certain phase-space. An example of a one-loop process is illustrated in Figure 1.2 (right).

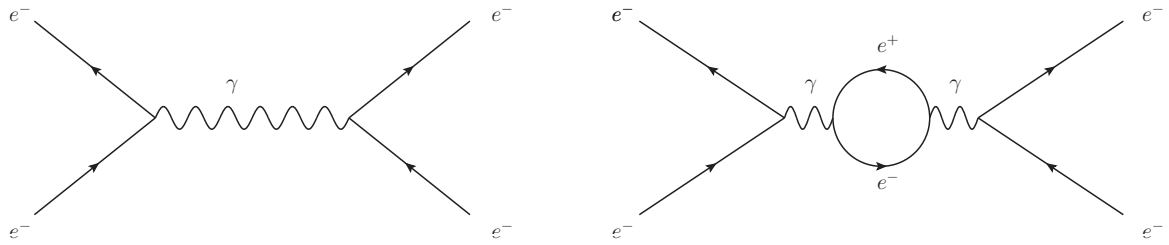

Figure 1.2: Electron scattering via the interchange of a virtual photon. Left: Tree-level diagram. Right: One-loop correction where the photon generates an electron-positron pair.

The Higgs mechanism

A SM Lagrangian can be built using the previously explained ingredients, the kinetic terms and the finite interactions between the different fields. A major problem arises when checking the gauge invariance of the Lagrangian. The mass terms of the type $\mathcal{L}_M = m^2 V_\mu V^\mu$ are not invariant under local group transformations. Therefore, all particles should remain massless and this fact would definitely contradict experience. To solve this important issue involving the particle's mass generation, the so-called Higgs mechanism was postulated.

The basic idea of the Higgs mechanism is that the universe is "filled" with a spinless scalar field with self interactions, the so-called *Higgs field*, that acquires a vacuum expectation value triggering a *Spontaneous Symmetry Breaking* (SSB) without a preferred frame or direction. Within the SSB, mass terms in the Lagrangian are allowed by having a contribution from the Higgs field.

The Higgs field, in complete generality, is introduced as a complex scalar field doublet with two degrees of freedom:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{1.18}$$

where ϕ^0 and ϕ^+ are defined as *Higgs bosons*, and act as messengers of the Higgs field.

The Higgs field is described by the following Lagrangian due to the self-interaction potential $V(\phi)$:

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - V(\phi), \tag{1.19}$$

where $D_\mu = \partial_\mu + ieA_\mu$ is the covariant derivative of the electroweak force $U(1)_Y \times SU(2)_L$. The potential has the following expression:

$$V(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2, \qquad \text{with } \mu^2 < 0, \tag{1.20}$$

where μ and λ are free parameters related to the self-interactions and the gauge couplings respectively.

The Higgs potential has a non-trivial minimum with a vacuum expectation value v. The vacuum expectation value is directly related with the different particle masses. The visualization of the potential shape can be illustrated by using a single scalar field defined as $\phi = (1/\sqrt{2})(\phi_1 + i\phi_2)$ (Figure 1.3) where η and ξ are introduced as perturbations around this minimum.

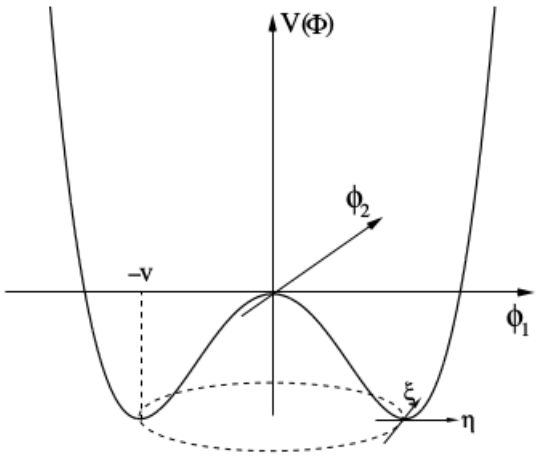

Figure 1.3: Higgs potential shape visualization. Left: In 3 dimensions. Right: In the complex plane.

The chosen (minimal) SM vacuum is:

$$\phi = \frac{1}{\sqrt{2}}(v + h) \tag{1.21}$$

where h denotes the SM Higgs boson.

The strength of couplings between the different particles and the Higgs field determines their masses. Figure 1.4 shows a representation of the relative magnitude of the mass of the different fermions.

Fermions masses and mixing

For each generation of quarks, one left-handed $SU(2)_L$ doublet (left and right) and two right-handed singlets exist, being eigenstates of the weak interaction. After the SSB the quark mass terms contain the so-called *Yukawa couplings*. Such couplings are proportional to the strength of the fermion interaction with the Higgs field and they are described with matrices which, in general, are not diagonal. However, they can be diagonalized by proper unitary matrices. The physical masses of the particles, the experimental observables, are the eigenvalues of the diagonalized matrix. Consequently the observed quarks, are indeed a combination of states.

The flavor oscillation is parametrized by the Cabibbo-Kobayashi-Maskawa (CKM) unitary mixing matrix, V_{CKM}, which relates the up-type quark family (containing the u, c and t flavors) to the down-type family (d, s, b). A representation of the relative

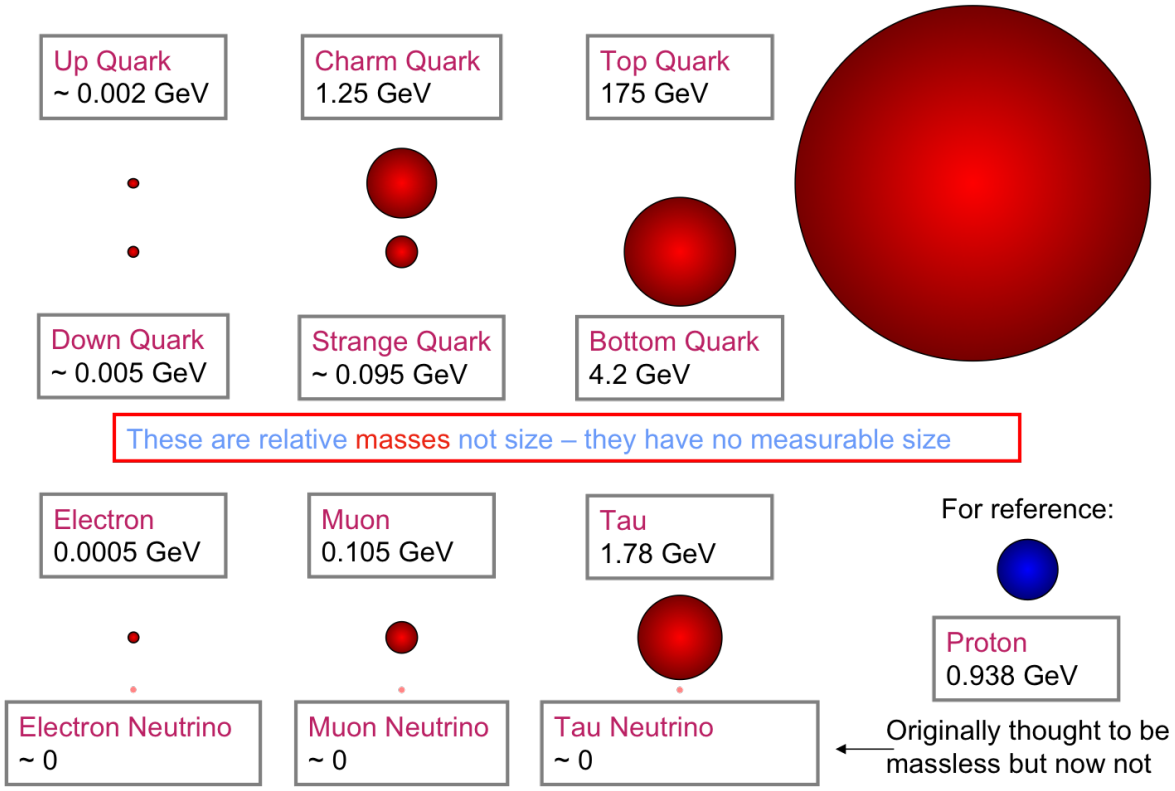

Figure 1.4: Schematic relative magnitud of the mass of the different fermions.

magnitude of the CKM matrix elements is shown in Figure 1.5 (left). Within the SM, this is the only source of flavor changing quark interactions.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.22}$$

A similar scenario to the quarks case is presented for the three lepton families, represented by their charged leptons and the corresponding neutrino. However a major difference stands. In the SM, neutrinos are massless. Consequently, only the coupling to left-handed neutrinos is allowed. With this (accidental) constraint, for the charged leptons, the Yukawa matrices can always be diagonalized without mixing of the interaction picture states. No experimental fact, for the time being, has contradicted the non-mixing in the charged lepton sector.

However, when the neutrinos were discovered to be massive, right-handed neutrinos were introduced, although only in the neutrino sector. In this case, there is not enough freedom to diagonalize the neutrino Yukawa matrix. At this point the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix was introduced, analogous to the CKM matrix, describing the experimentally observed neutrino mass oscillation. A representation of the relative magnitude of the PMNS matrix elements is shown in Figure 1.5 (right).

The main conclusion is that lepton flavor mixing in the SM is only allowed when neutrino masses are introduced. Flavor change in the case of charged leptons, named *Lepton Flavor Violation* (LFV), can only take place at loop level processes via a neutrino

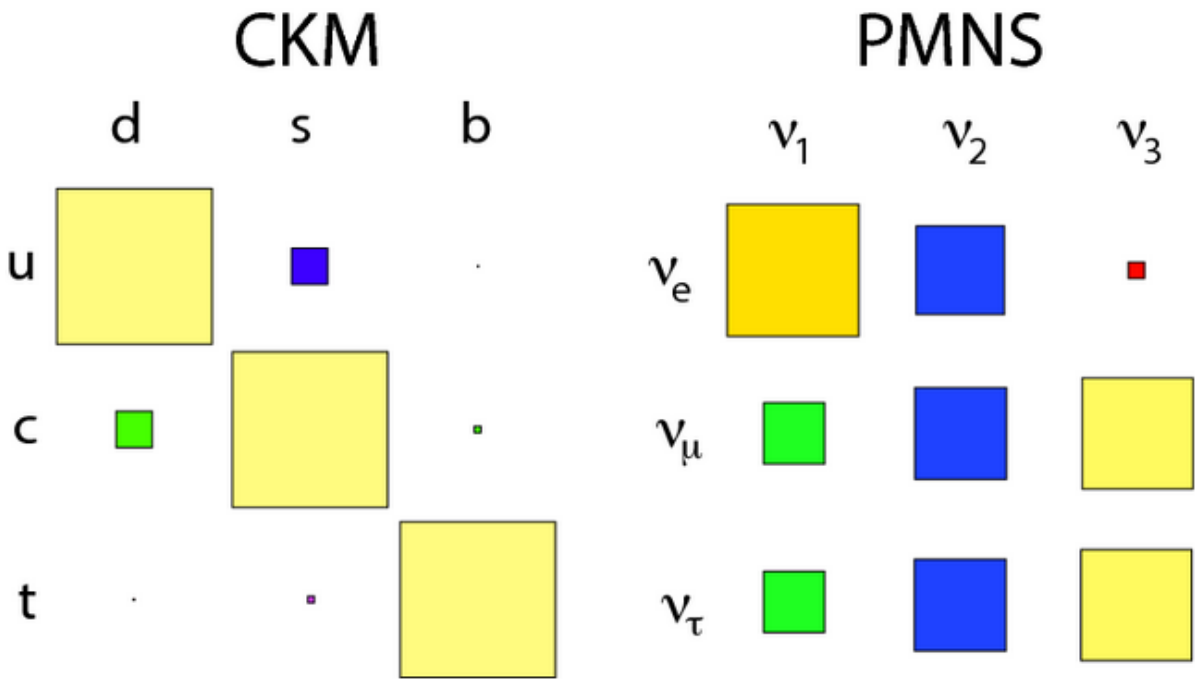

Figure 1.5: Relative magnitude of the CKM and PMNS matrix elements. Left: CKM Matrix. Right: PMNS matrix.

flavor oscillation, suppressing the rate of LFV by a factor of $(m_\nu/m_W)^4 \sim \mathcal{O}(10^{-48})$ (i.e. LVF in the charged sector is negligible).

Consequences of the Standard Model

The structure of the SM lead to accidental consequences. Some of them are:

- **Lepton Flavor Universality:** The coupling of the charged leptons to the electroweak gauge bosons is flavor independent. Meaning that all couplings are of the same magnitude for e, μ and τ. However, hints of deviation from LFU have been observed in the recent years and they are described in Section 1.2.3 and 1.2.4.

- **Lepton number conservation:** As explained before, no mixing in the charged lepton sector (no Lepton Flavor Violation) is expected which is completely different from what is seen in the quark sector. Many measurements, like the search described in the present document, test the validity of this feature of the SM.

- **Flavor Changing Neutral Currents (FCNC) in the quark sector:** Flavor changing interactions are allowed in the quark sector, in contrast with the leptonic one in which they are prohibited. In the quark sector, FCNC are forbidden at tree level but they may occur at the loop level, although they are suppressed by the GIM mechanism [17].

Meson	Quark content	Mass (MeV/c^2)	Mean lifetime (p s)
B^+	$u\bar{b}$	5279.29 ± 0.15	1.638 ± 0.004
B^0	$d\bar{b}$	5279.61 ± 0.16	1.520 ± 0.004
B_s^0	$s\bar{b}$	5366.79 ± 0.23	1.510 ± 0.005
B_c^+	$c\bar{b}$	6275.1 ± 1.0	0.507 ± 0.009

Table 1.1: Overview of the different B-mesons measured properties [13].

Limitations of the Standard Model

Despite the fact that the SM has been tested to a large precision, open questions remain, for example:

- **Hierarchy:** Why three generations and why do they have such different masses?
- **High number of free parameters in the flavor (Yukawa) sector:** There are 16 free parameters in the flavor sector. Which is unusually large if compared with the 3 gauge couplings and the 2 parameters of the Higgs potential.
- **Unified theory:** The SM describes only two of the three known nature interactions not including gravity.
- **Asymmetry between matter and anti-matter:** Our universe is almost exclusively formed by matter despite the *big bang* theory predict the same quantity of matter and anti-matter: what happened to the anti-matter? The SM does not provide a source of such a big degree of asymmetry between matter and anti-matter.
- **Cosmological observations:** From the rotation speed of the galaxies and the gravitational lensing, among others, the existence of *dark matter* is inferred. Nevertheless, the SM does not provide a natural candidate for dark matter. In the same way, from the accelerated expansion of the universe, the existence of *dark energy* is assumed. Equally, the SM does not provide a solution.

Therefore, although it is an evidence that the SM performs extremely well inside its range of applicability, it is also a fact that it cannot explain everything we observe. Another question is, up to which energy level is the SM valid? Trying to address all these questions, since many years fervent theoretical and experimental work is being carried out in order to discover the nature of physics Beyond the SM (BSM).

1.2 Probing the Standard Model with rare *b*-hadron decays

The main interest in studying b-quark interactions is that the b-quark is the heaviest quark which hadronizes before decaying, allowing research about a wide variety of topics. Furthermore, b-hadrons have a relatively long lifetime which greatly helps the experimental search. An overview of the different B-mesons properties is given in Table 1.1.

Specifically, in this document the attention is focused on $b \to q l^+ l^-$ transitions, where $q = s$ or d-quarks, and $l = e$, μ or τ.

The transitions within the down-type quark family (d, s, b) and within the up-type family group (u, c, t) are produced via FCNC (change of flavor but not of charge), processes which are suppressed at tree level in the SM but possible at higher orders in loop perturbative expansion via the interchange of electroweak charged bosons although they suffer from additional suppression by the GIM mechanism [17]. Consequently, the branching fractions of these kinds of processes are typically lower than 10^{-5} and they are therefore called *rare* decays. In many extensions of the SM, contributions from BSM physics are expected to appear at the same level as the SM leading order for these rare processes.

The difficulty in the theoretical computation of such transitions comes from the multiple physical scales involved in FCNC transitions, ranging from strong interaction dynamics in its non-perturbative regime (~ 0.1 GeV) to the mass of the W bosons (~ 80 GeV). The strategy useful to tackle a problem with multiple scales involved is the Effective Field Theory (EFT).

Basics on Effective Field Theory

Natural phenomena can be split according to their scale and different theoretical tools can be used to explain them depending on the concerned scale. Whenever phenomena are spread out over different energy or length scales, an effective description can be valuable, either to simplify calculations, or to actually allow model independent statements that would be otherwise impossible.

An EFT is a physics model including all relevant effects affecting a given scale, but not those that exclusively play a role at much higher or lower energies than the energy scale of interest. Using an EFT, a model independent study is implemented without the need to specify the underlying theory. The SM can be seen as an EFT in the GeV scale of a higher energy (i.e. heavy physics) BSM underlying theory, which is unknown to us.

The EFT key ingredient is the *Operator Product Expansion* (OPE) defined as a sum of *operators* Q_i, each one with specific mass dimension D_i. The coupling of each operator can be differentiated as a dimensionless constant, known as the *Wilson coefficient* C_i, and some powers of a mass scale, for which usually the scale of heavy physics (Λ) is used. The effective operator Q_i describes the long distance physics, namely the particles in the initial and final states, whereas the Wilson coefficients C_i are obtained by integrating out the remaining degrees of freedom of the fundamental underlying theory.

If nothing is known about the underlying theory at scale Λ, the best guess is that it consists of dimensionless couplings $\sim \mathcal{O}(1)$ and masses $\sim \mathcal{O}(\Lambda)$. The combination of these factors should be proportional to the effects mediated by the unknown high energy physics.

The dynamics of a given system are described using an effective *Hamiltonian*. In general for an EFT:

$$\mathcal{H}_{EFT} = \mathcal{H}_{\text{Free and mass terms}} + \sum_i \frac{C_i}{\Lambda^{D_i-4}} Q_i. \tag{1.23}$$

For physical predictions, the Hamiltonian formalism is more suitable than the Lagrangian one, which was used to introduce the SM. Despite the fact that the Lagrangian

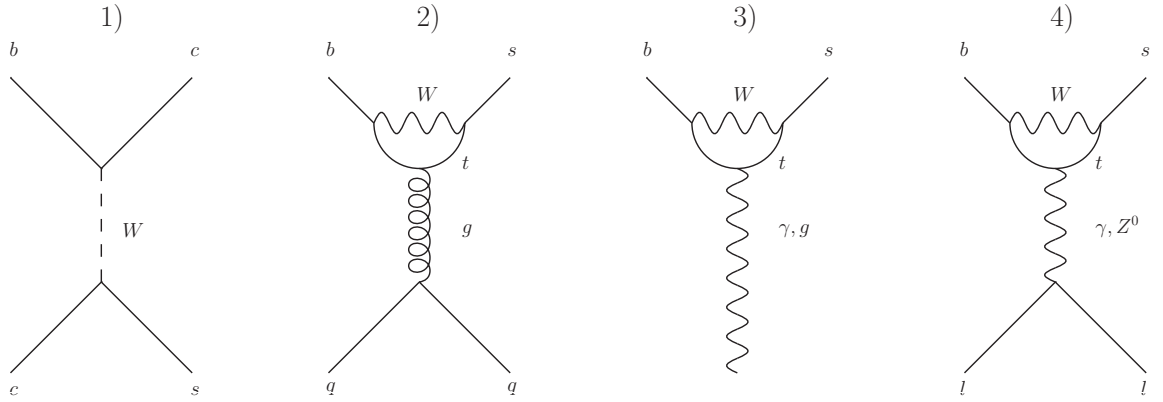

Figure 1.6: SM Feynman diagrams examples corresponding to the different operators: (1) current-current operators, (2) penguin QCD operators, (3) dipole operator and (4) semileptonic penguin operators.

presents more insights to the symmetries of the theory, the Hamiltonian directly encodes the time evolution of the system without need to write explicitly the system action.

1.2.1 Heavy quark expansion

Using the local OPE [18], b-hadron interactions are embedded inside the Wilson coefficients corresponding to the general lowest dimension, i.e. six, effective operators containing the light SM matter fields. The weak interactions are seen as point-like from the QCD scale (Λ_{QCD}), allowing to use perturbation theory.

For SM interactions the effective Hamiltonian for FCNC transitions contains operators contributing to $b \to q\gamma$, $b \to q l^+ l^-$ and $b \to q\nu\overline{\nu}$ being $q = s$ or d quarks. Therefore, the Wilson coefficients result from the computation of all the corresponding SM Feynman diagrams (Figure 1.6). The general effective Hamiltonian is:

$$\mathcal{H}_{eff}^{b\to q} = \frac{4G_F}{\sqrt{2}} \left(V_{ub}V_{uq}^* \sum_{i=1}^{2} C_i Q_i^u + V_{cb}V_{cq}^* \sum_{i=1}^{2} C_i Q_i^c - V_{tb}V_{tq}^* \sum_{i=3}^{10} C_i Q_i - V_{tb}V_{tq}^* C_\nu Q_\nu + h.c. \right), \tag{1.24}$$

containing the following operators [19]:

- The current-current operators mixing up and down-type quarks at tree level by Flavor Changing Charged Currents (FCCC):

$$Q_1^u = (\overline{q}_L \gamma_\mu T^a u_L)(\overline{u}_L \gamma^\mu T^a b_L), \qquad Q_1^c = (\overline{q}_L \gamma_\mu T^a c_L)(\overline{c}_L \gamma^\mu T^a b_L), \tag{1.25}$$

$$Q_2^u = (\overline{q}_L \gamma_\mu u_L)(\overline{u}_L \gamma^\mu b_L), \qquad Q_2^c = (\overline{q}_L \gamma_\mu c_L)(\overline{c}_L \gamma^\mu b_L). \tag{1.26}$$

  Being T^a the generators of the $SU(3)$ gauge group. The Wilson coefficients of Q_1 cancel order by order in the perturbative expansion.

- The so-called penguin QCD operators:

$$Q_3 = (\overline{q}_L \gamma_\mu b_L) \sum_{p}^{u,d,s,c,b} (\overline{p} \gamma^\mu p), \tag{1.27}$$

$$Q_4 = (\overline{q}_L \gamma_\mu T^a b_L) \sum_p^{u,d,s,c,b} (\overline{p} \gamma^\mu T^a p)\,, \tag{1.28}$$

$$Q_5 = (\overline{q}_L \gamma_\mu \gamma_\nu \gamma_\rho b_L) \sum_p^{u,d,s,c,b} (\overline{p} \gamma^\mu \gamma^\nu \gamma^\rho p)\,, \tag{1.29}$$

$$Q_6 = (\overline{q}_L \gamma_\mu \gamma_\nu \gamma_\rho T^a b_L) \sum_p^{u,d,s,c,b} (\overline{p} \gamma^\mu \gamma^\nu \gamma^\rho T^a p)\,. \tag{1.30}$$

- The dipole electromagnetic operator and its chromodynamic counterpart:

$$Q_7 = \frac{e}{16\pi^2} m_b \,(\overline{q}_L \sigma^{\mu\nu} b_R)\, F_{\mu\nu}, \tag{1.31}$$

$$Q_8 = \frac{g_s}{16\pi^2} m_b \,(\overline{q}_L \sigma^{\mu\nu} T^a b_R)\, G^a_{\mu\nu}. \tag{1.32}$$

- And the semileptonic penguin operators:

$$Q_9 = \frac{e}{16\pi^2} \,(\overline{q}_L \gamma_\mu b_L) \sum_l^{e,\mu,\tau} \left(\bar{l} \gamma^\mu l\right), \tag{1.33}$$

$$Q_{10} = \frac{e}{16\pi^2} \,(\overline{q}_L \gamma_\mu b_L) \sum_l^{e,\mu,\tau} \left(\bar{l} \gamma^\mu \gamma_5 l\right), \tag{1.34}$$

$$Q_\nu = \frac{e}{8\pi^2} \,(\overline{q}_L \gamma_\mu b_L) \sum_l^{e,\mu,\tau} (\overline{\nu}_{lL} \gamma^\mu \nu_{lL})\,, \tag{1.35}$$

Being the sub-indices R and L the right and left-handed chiralities:

$$\gamma^\mu q_L = \gamma^\mu \left(\frac{1-\gamma^5}{2}\right) q \text{ and } \gamma^\mu q_R = \gamma^\mu \left(\frac{1+\gamma^5}{2}\right) q. \tag{1.36}$$

By convention the right-handed component of an operator is denoted by a "prime" super-index.

In the SM, scalar (Q_S), pseudo-scalar (Q_P) and tensor operators (Q_T) operators

$$Q_S = \frac{e}{16\pi^2} m_b \,(\overline{q}_L b_R) \left(\bar{l} l\right), \tag{1.37}$$

$$Q_P = \frac{e}{16\pi^2} m_b \,(\overline{q}_L b_R) \left(\bar{l} \gamma_5 l\right), \tag{1.38}$$

$$Q_T = \frac{e}{16\pi^2} m_b \,(\overline{q}_R \sigma^{\mu\nu} b_L) \left(\bar{l} \sigma_{\mu\nu} l\right), \tag{1.39}$$

are highly suppressed due to the small masses of the leptons, and can be neglected even for τ decays.

The dominant operators for $b \to s l^+ l^-$ transitions are semileptonic operators Q_9 and Q_{10}, and the photon contribution (photon pole) encoded in the electromagnetic operator Q_7. The Q_7 dynamics are dominated by the left-handed part [20].

C_9, C_{10} and C_7 are the corresponding Wilson coefficients and they include the electromagnetic coupling constant α_{em}. Their computation is divided in 2 steps given the scale dependence. First, a computation at large scale $\sim M_W$ gives the following results in the SM with an uncertainty of order (m_b^2/M_W^2):

- $C_7(M_W) \approx -0.19$,
- $C_9(M_W) \approx +2$,
- $C_{10}(M_W) \approx -4$,

where the relative sign between C_7 and C_9 depends on the convention on the sign of the covariant derivative.

The second step is to add the SM contribution at low scale (μ) around m_b. At this scale, QCD penguin effects (δC_i) gather importance and C_9 and C_7 acquire contributions of the penguin operators:

- $C_7^{eff}(\mu) = C_7(M_W, \mu) + \delta C_7(\mu) = C_7(M_W, \mu) + \sum_{i=1}^{6} y_i C_i(\mu)$,
- $C_9^{eff}(\mu) = C_9(M_W, \mu) + \delta C_9(\mu) = C_9(M_W, \mu) + \Upsilon(C_i(\mu), q^2)$ with $i = [1, 6]$;

where y_i are scale independent coefficients and Υ a scale dependent linear combination of the Wilson coefficients. C_{10} remains unaffected as it is not relevant below M_W. Having all the ingredients, the amplitude of a $b \to q l^+ l^-$ process in the SM is finally written as:

$$\mathcal{M}\left(b \to q l^+ l^-\right) = \frac{G_F \alpha_{em}}{\sqrt{2}\pi} V_{tq}^* V_{tb} \left\{ C_9^{eff}(\mu, q^2) Q_9 + C_{10} Q_{10} - \frac{2i}{q^2} C_7^{eff}(\mu) Q_7 \right\}. \quad (1.40)$$

$\mathcal{M}$ includes some long distance effects related to the 'energy' (q^2) of the initial and final states, especially in $C_9(q^2)$.

How should BSM physics manifest itself?

BSM physics is expected to have a scale heavier than m_b and to show up as a modification of the SM Wilson coefficients and/or the generation of operators not present in the SM (especially right handed).

Concerning the operators, for $b \to q l^+ l^-$ transitions (assuming Lepton Universality in radiative decays), LFU could be violated. Then the BSM physics contribution would be visible in the semileptonic operators Q_9, Q_{10} and Q_ν, and possibly generate the corresponding right handed operators Q'_9, Q'_{10} and Q'_ν. There could also be contributions of $Q_S^{(\prime)}$, $Q_P^{(\prime)}$ and $Q_T^{(\prime)}$. For LFV theories even semileptonic operators with different lepton flavor could exist as well: $Q_k^{(\prime)}|^{l_i l_j}$ where $k = 9$, 10, S, P, T.

Some Wilson coefficient relations in different BSM scenarios are expected to be, from symmetry arguments, the following ones:

- LFU implies $C_k^{(\prime)}|_q^e = C_k^{(\prime)}|_q^\mu = C_k^{(\prime)}|_q^\tau$
- LFV implies $C_k^{(\prime)}|^{l_i l_j} \neq 0$, for some k.
- Any *weakly* coupled BSM physics implies $C_S|_q^l = -C_P|_q^l$, $C'_S|_q^l = C'_P|_q^l$ and $C_T^{(\prime)}|_q^l = 0$.

In the following section, the experimental and theoretical results of $B_s \to l^+ l^-$ and LFU searches are given before getting in depth into LFV.

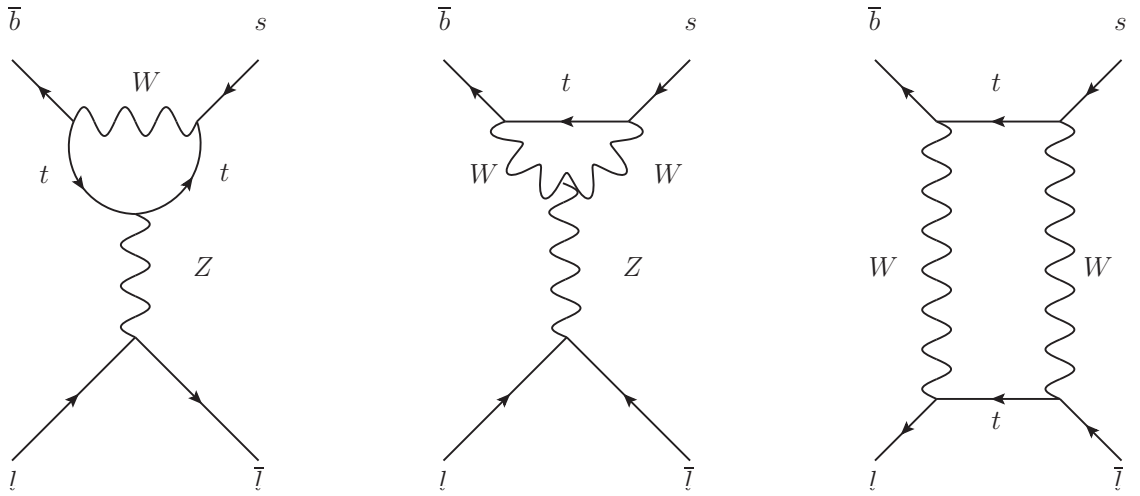

Figure 1.7: SM Feynman diagrams contributing at leading order to the $B_q \to l^+l^-$ branching ratio.

1.2.2 Rare dileptonic B decays

Theory

Dileptonic B decays $B_q \to l^+l^-$, are forbidden at tree level in the SM and CKM suppressed. Furthermore, they have a significant helicity suppression ($\sim (m_l^2/m_{B_q}^2)$). Consequently, they are *very rare*. The Branching Ratio ($\mathcal{B}$) of the process, being f_{B_q} the fragmentation constant of the B meson determined from lattice QCD, is expressed as:

$$\mathcal{B}\left(B_q \to l^+l^-\right)_{SM} = \tau_{B_q} \frac{G_F^2 \alpha_{em}^2}{16\pi^2} f_{B_q}^2 m_l^2 m_{B_q} \sqrt{1 - \frac{4m_l^2}{m_{B_q}^2}} |V_{tb}V_{tq}^*|^2 |C_{10}|^2, \tag{1.41}$$

where τ_{B_q} is the meson lifetime and $V_{ij}^{(*)}$ the CKM matrix elements. It is worth to emphasize that these decays are very clean, e.g. they are free from QCD penguin pollution, given that they exclusively depend on C_{10} with the corresponding Feynman diagrams shown in Figure 1.7.

Given the different lifetimes between the mass of the heavy and light B_s^0 eigenstates, B_s^0 decays have 2 different branching ratios: one before the $B_s^0 - \overline{B}_s$ mixing (Equation 1.41) and the time integrated one $\overline{\mathcal{B}}$, which is the one actually measured:

$$\overline{\mathcal{B}}\left(B_q \to l^+l^-\right) = \frac{1 + \mathcal{A}_{\Delta\Gamma}^{ll} y_q}{1 - y_q^2} \mathcal{B}\left(B_q \to l^+l^-\right), \tag{1.42}$$

where $\mathcal{A}^{ll}$ and y_q are defined as:

$$y_q = \frac{\Gamma_H^q - \Gamma_L^q}{\Gamma_H^q + \Gamma_L^q}, \tag{1.43}$$

$$\mathcal{A}^{ll} = \frac{\Gamma_H^{q,ll} - \Gamma_L^{q,ll}}{\Gamma_H^{q,ll} + \Gamma_L^{q,ll}} \tag{1.44}$$

being $\Gamma^q_{L,H}$ and $\Gamma^{q,ll}_{L,H}$ the B_q total and $B_q \to l^+l^-$ partial width for the light and heavy eigenstates respectively.

The SM predictions for the branching ratios [21] in the B^0 channel are:

$$\mathcal{B}\left(B^0 \to e^+e^-\right)_{SM} = (2.48 \pm 0.21) \cdot 10^{-15}, \tag{1.45}$$

$$\mathcal{B}\left(B^0 \to \mu^+\mu^-\right)_{SM} = (1.06 \pm 0.21) \cdot 10^{-10}, \tag{1.46}$$

$$\mathcal{B}\left(B^0 \to \tau^+\tau^-\right)_{SM} = (2.22 \pm 0.19) \cdot 10^{-8}. \tag{1.47}$$

Concerning the B^0_s channel:

$$\overline{\mathcal{B}}\left(B^0_s \to e^+e^-\right)_{SM} = (8.54 \pm 0.55) \cdot 10^{-14}, \tag{1.48}$$

$$\overline{\mathcal{B}}\left(B^0_s \to \mu^+\mu^-\right)_{SM} = (3.65 \pm 0.23) \cdot 10^{-9}, \tag{1.49}$$

$$\overline{\mathcal{B}}\left(B^0_s \to \tau^+\tau^-\right)_{SM} = (7.73 \pm 0.49) \cdot 10^{-7}. \tag{1.50}$$

Notice that the SM expectations are higher for B^0_s decays than for B^0, being the $b \to s$ transitions less CKM suppressed than the $b \to d$ ones.

Experiment

From the experimental side, no hints for BSM physics have been spotted. All measurements agree with the SM predictions so far.

$B^0_{(s)} \to \mu^+\mu^-$ is the cleanest experimental channel to look at. In 2014, culminating 3 decades of research, the CMS and LHCb collaborations performed a joined analysis [22] achieving the first observation of the $B^0_s \to \mu^+\mu^-$ decay with a 6.2σ statistical significance and the first evidence of the $B^0 \to \mu^+\mu^-$ decay with a significance of 3.2σ, being compatible with the SM at 1.2σ and 2.2σ respectively.

$$\overline{\mathcal{B}}\left(B^0_s \to \mu^+\mu^-\right) = \left(2.8^{+0.7}_{-0.6}\right) \cdot 10^{-9}, \tag{1.51}$$

$$\mathcal{B}\left(B^0 \to \mu^+\mu^-\right) = \left(1.9^{+1.6}_{-1.4}\right) \cdot 10^{-9}, \tag{1.52}$$

The updated values for the last LHCb analysis [23] (still in agreement with the SM predictions) are the following ones:

$$\overline{\mathcal{B}}\left(B^0_s \to \mu^+\mu^-\right) = \left(3.0^{+0.3}_{-0.2}\right) \cdot 10^{-9} \text{ at } 7.8\sigma, \tag{1.53}$$

$$\mathcal{B}\left(B^0 \to \mu^+\mu^-\right) < 3.4 \cdot 10^{-10} \text{ at } 95\%\text{CL}, \tag{1.54}$$

The channel with electrons has a smaller SM branching ratios due to the small mass of this lepton and thus higher helicity suppression. The $\mathcal{B}$ value is beyond the current and near future experimental sensitivities. The best existing limits up to date were set back in 2009 by the CDF collaboration [24]:

$$\overline{\mathcal{B}}\left(B^0_s \to e^+e^-\right) < 2.8 \cdot 10^{-7} \text{ at } 90\%\text{CL}, \tag{1.55}$$

$$\mathcal{B}\left(B^0 \to e^+e^-\right) < 8.3 \cdot 10^{-8} \text{ at } 90\%\text{CL}. \tag{1.56}$$

The current best experimental limit on the tauonic channel has been set by the LHCb collaboration. The $\mathcal{B}$ prediction is the highest one among the 3 leptons but experimentally

it is a highly complex measurement, specially due to the neutrinos from the τ decays which can not be detected and that the search uses the hadronic τ, which is handicapped by the LHCb trigger (optimized for muons). Furthermore there some BSM models, which have not yet been excluded, predict a significant increase of this $\mathcal{B}$ expecting up to a $\mathcal{B}$ ~30 times larger than the SM prediction [25]. The measured limits are:

$$\overline{\mathcal{B}}\left(B_s^0 \to \tau^+\tau^-\right) < 6.3 \cdot 10^{-3} \text{ at } 95\%\text{CL}, \tag{1.57}$$

$$\mathcal{B}\left(B^0 \to \tau^+\tau^-\right) < 2.1 \cdot 10^{-3} \text{ at } 95\%\text{CL}. \tag{1.58}$$

Although these golden modes have not provided any BSM physics insight yet, the experimental accuracy is expected to increase as more data is collected from the LHC in the near future.

1.2.3 Hints of LFU violation in FCNC semileptonic decays

Theory

Particularly powerful probes for new physics are $b \to q l^+ l^-$ transitions of the type $B_q \to H l^+ l^-$, where H is a hadron containing a s or d-quark. The SM contribution is composed by penguin and box diagrams as shown in Figure 1.8. The amplitude for the process $B_q \to K^* l^+ l^-$ momenta and neglecting the m_s mass, is:

$$\begin{aligned}\mathcal{M} = \frac{G_F \alpha_{em}}{2\sqrt{2}\pi} V_{ts}^* V_{tb} \Bigg[&\left\langle \overline{K}^* | \bar{s}\gamma^\mu(1-\gamma_5)b|\overline{B}\right\rangle \left\{ C_9^{eff} \bar{l}\gamma_\mu l + C_{10}\bar{l}\gamma_\mu\gamma_5 l \right\} \\ &-2\frac{C_7^{eff}}{q^2} m_b \left\langle \overline{K}^* | \bar{s} i\sigma_{\mu\nu} q^\nu (1+\gamma_5) b|\overline{B}\right\rangle \bar{l}\gamma_\mu\gamma_5 l \Bigg].\end{aligned} \tag{1.59}$$

In particular, some of the most interesting measurements to perform are ratios (R_H) of the decay rate (Γ) integrated over the squared dilepton invariant mass (q^2). The ratio allows the cancellation of hadronic uncertainties in the theoretical computations, and thus a higher accuracy in the prediction. The SM predictions is close to unity.

$$R_H = \frac{\int \frac{d\Gamma\left(B\to H l_1^+ l_1^-\right)}{dq^2} dq^2}{\int \frac{d\Gamma\left(B\to H l_2^+ l_2^-\right)}{dq^2} dq^2}, \text{ where: } l = e, \mu. \tag{1.60}$$

An interesting case is R_{K^*}, defined as

$$R_{K^*} = \frac{\int \frac{d\Gamma\left(B^0\to K^*\mu^+\mu^-\right)}{dq^2} dq^2}{\int \frac{d\Gamma(B^0\to K^* e^+ e^-)}{dq^2} dq^2}. \tag{1.61}$$

The R_{K^*} SM predictions depend on which strategy is used to compute the QCD effects. SM predictions have been provided by different theory groups and the most recent predictions are reported in Table 1.2.

Another interesting observable is R_K, defined as

$$R_K = \frac{\int \frac{d\Gamma\left(B^+\to K^+\mu^+\mu^-\right)}{dq^2} dq^2}{\int \frac{d\Gamma(B^+\to K^+ e^+ e^-)}{dq^2} dq^2}, \tag{1.62}$$

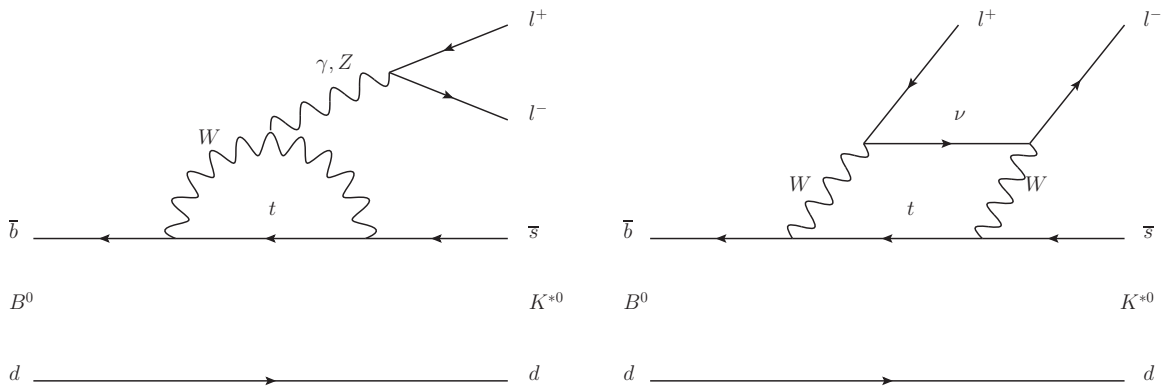

Figure 1.8: SM Feynman diagrams contribution at leading order to $B^0 \to K^* l^+ l^-$ decay.

q^2 (GeV^2/c^4)	$R\,(K^*)$ SM	Fitter package
$[0.045, 1.1]$	0.906 ± 0.028	BIP
	0.922 ± 0.022	CDHMV
	$0.919^{+0.004}_{-0.003}$	EIS
	0.925 ± 0.004	flav.io
	$0.920^{+0.007}_{-0.006}$	JC
$[1.1, 6.0]$	1.000 ± 0.010	BIP
	1.000 ± 0.006	CDHMV
	$0.9968^{+0.0005}_{-0.0004}$	EIS
	0.9964 ± 0.005	flav.io
	$0.996^{+0.007}_{-0.002}$	JC

Table 1.2: SM predictions for R_{K^*} [26].

has been measured in the bin $q^2 = [1, 6]$ GeV$^2/c^4$ and the SM prediction is 1.00 at the 1% level.

Experiment

Unlike the dileptonic decays, FCNC semileptonic measurements have provided surprises. They show tensions, although not yet significant, with respect to the SM expectations. While the deviations from the SM of the individual ratios are about 2.2-2.5σ, the results are seen to point to the same direction (Figure 1.9).

For R_{K^*} LHCb [2] has measured:

$$R_{K^*} = \begin{cases} 0.66^{+0.11}_{-0.07}(stat) \pm 0.03(syst), & \text{in } 0.045 < q^2 < 1.1 \text{ GeV}^2/\text{c}^4 \\ 0.69^{+0.11}_{-0.07}(stat) \pm 0.05(syst), & \text{in } 1.1 < q^2 < 6.0 \text{ GeV}^2/\text{c}^4 \end{cases} \tag{1.63}$$

and $R_K = 0.745^{0.09}_{-0.07}(stat) \pm 0.04(syst)$ in $1 < q^2 < 6$ GeV^2/c^4 [3].

These results point to LFU violation, as they would imply different behavior among the lepton families. A wide spectrum of theoretical models is available to try to explain

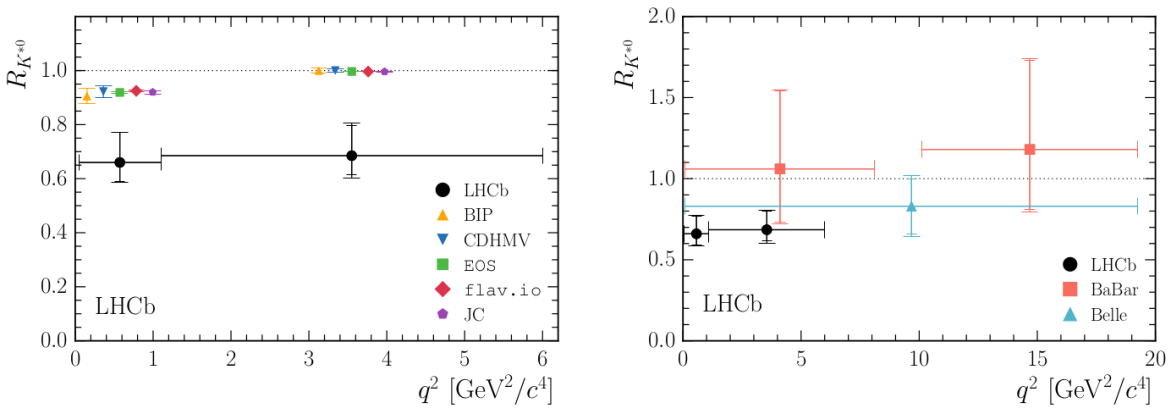

Figure 1.9: $B^0 \to K^* l^+ l^-$ summary of the current theoretical and experimental results.

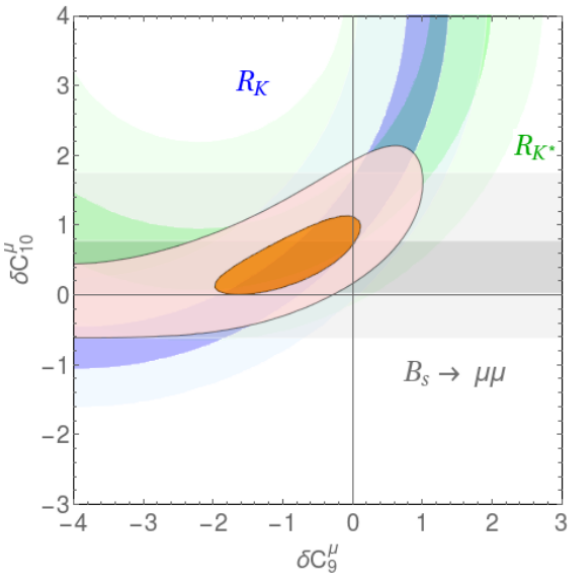

Figure 1.10: Global fits to $R_{K^{(*)}}$ and $\overline{\mathcal{B}}\left(B_s^0 \to \mu^+\mu^-\right)$ for BSM physics contributions to C_9 and C_{10} [27]. The SM expectation lies in the origin of the coordinates non excluded region is shown in orange. The band for $R_{K^{(*)}}$ includes only the [1.1,6] GeV$^2/c^4$ bin [27].

these hints of BSM physics in the lepton sector. Many of these models provide LFV explanation as well. Some of them are detailed in the next section.

Combining the results of $R_{K^{(*)}}$ with $\overline{\mathcal{B}}\left(B_s^0 \to \mu^+\mu^-\right)$, *global fits* are performed in order to constrain the BSM physics contributions on the different Wilson coefficients. An example is in Figure 1.10, where the impact of BSM physics effects on C_9 and C_{10} makes them to deviate almost 4σ from the SM values [27].

1.2.4 Other hints of LFU violation in B decays

Another hint of LFU violation is found in Flavor Changing Charged Currents (FCCC) semileptonic decays. Since they are mediated by tree-level processes, FCCC transitions are *not rare*. They relate the down-type with up-type quarks families. In particular, the interesting transitions are of the type $b \to \bar{c} l^+ \nu_l$ and the experimental observables are ratio of branching ratios including the third lepton generation. The main ones are R_D and R_{D^*}, defined as:

$$R_{D^{(*)}} = \frac{B^0 \to D^{(*)+} \tau^- \nu_\tau}{B^0 \to D^{(*)+} l^- \nu_l}, \tag{1.64}$$

where $l = e, \mu$ for BaBar and Belle, while $l = \mu$ for LHCb.

The SM expectations for $R_{D^{(*)}}$ [28] are:

$$R_D^{SM} = 0.299 \pm 0.003 \tag{1.65}$$

$$R_{D^*}^{SM} = 0.257 \pm 0.003 \tag{1.66}$$

which are not unity due to the mass difference between l and τ.

From the experimental side, the world averages performed by the Heavy Flavor Averaging [29] group between BaBar, Belle and LHCb results, are:

$$R_D^{Exp} = 0.403 \pm 0.040(stat) \pm 0.024(syst) \tag{1.67}$$

$$R_{D^*}^{Exp} = 0.310 \pm 0.015(stat) \pm 0.008(syst), \tag{1.68}$$

which translates to a combined 4.1σ deviation from the SM expectations. The summary of the results is shown in Figure 1.11.

Nowadays not only $R_{D^{(*)}}$ is studied, the computation and measurement of many varieties of FCCC ratios is a trending topic: $R_{J/\psi}$, R_{Λ_b}... which is understandable given the striking tension between the experiment and the SM predictions.

Despite the high significance of the experimental discrepancy with the SM, it has to be kept in mind that computations and measurements are difficult. The SM prediction heavily relies on QCD computations, as in this case the Wilson coefficients QCD uncertainties do not completely cancel with the ratio. In experiment, the analysis is quite complex due to the missing energy from the multiple neutrinos in the final state and the abundant double charm backgrounds (at least in hadronic machines).

Just recently, common explanation for the two so-called *anomalies* of $R_{K^{(*)}}$ and $R_{D^{(*)}}$ has been proposed. However, in terms of model building, these common explanations are in continuous discussion [30].

1.3 Lepton Flavor Violation in $B_{(s)} \to l_i^+ l_j^-$ decays

The hints of LFU violation have increased the interest for LFV searches, as LFV implies by definition differences in the lepton couplings.

LFV is basically forbidden in the SM due to the coincidence between the interaction and mass eigenstates of the charged leptons Yukawa matrices. The only highly suppressed source of LFV comes from the introduction of the neutrino mass oscillations. Therefore, LFV B decays like $B^0_{(s)} \to \tau^\pm \mu^\mp$, can occur in the SM when including neutrino masses

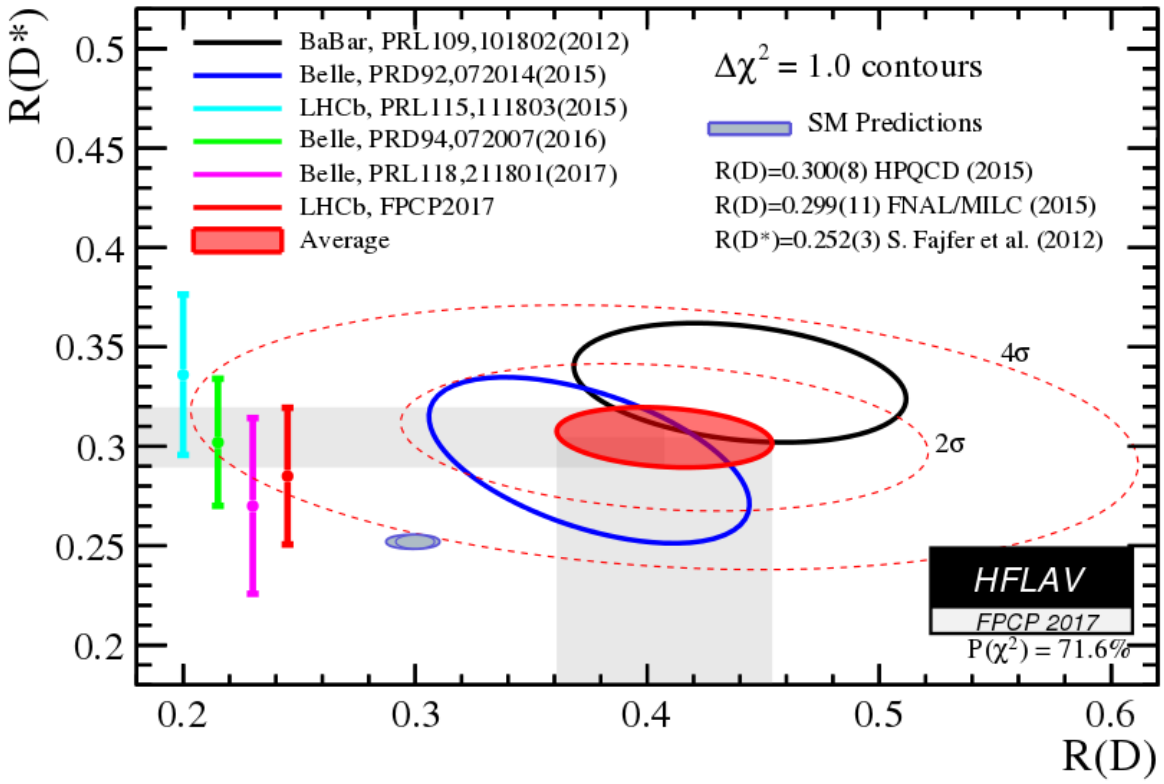

Figure 1.11: Combination of the latest $R\left(D^{(*)}\right)$ results [29].

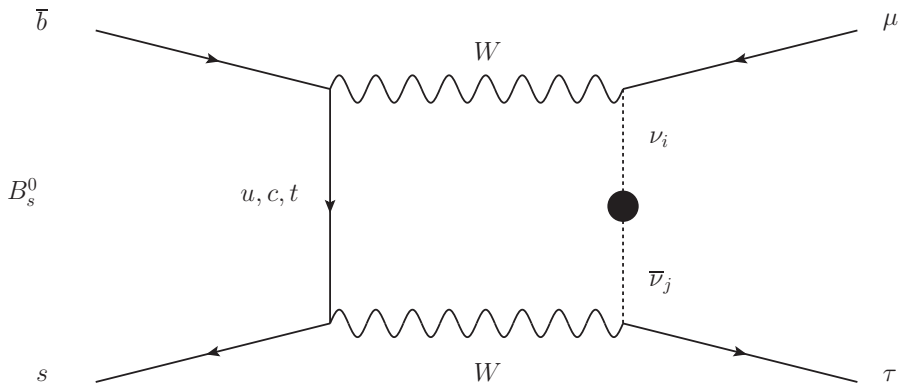

Figure 1.12: SM contribution to the LFV decay $B_s^0 \to \tau^{\pm}\mu^{\mp}$.

via box diagrams when a virtual neutrino exchanged between two charged leptons of different families oscillates, as shown in Figure 1.12. The amplitude of these processes is suppressed by a $(m_\nu/m_W)^4 \sim \mathcal{O}(10^{-48})$ factor, and thus lies beyond the current and future experimental sensitivities. However, in a wide variety of BSM scenarios the LFV rate in the charged sector increases dramatically up to measurable levels.

The purpose of this section is to give a general and conceptual overview of the way LFV is introduced in some BSM models for $b \to s l^+ l^-$ transitions, focusing on $B^0_{(s)} \to \tau^{\pm}\mu^{\mp}$

decays, its branching fraction being predicted between 10^{-4} and 10^{-9} for the different models.

$b \to s l_i^+ l_j^-$ transitions involving BSM physics

The Hamiltonian for the heavy quark expansion is written in Equation 1.24. The effective Hamiltonian for $b \to s l_i^+ l_j^-$ transitions, where the leptons are from different families, can be expressed as:

$$\mathcal{H}_{eff}\left(b \to s l_i^+ l_j^-\right) = \mathcal{H}_{eff}^{SM} + \mathcal{H}_{eff}^{VA} + \mathcal{H}_{eff}^{SP}, \tag{1.69}$$

with $\mathcal{H}_{eff}^{SM}$ as the SM contribution,

$$\begin{aligned}
\mathcal{H}_{eff}^{VA} = -\frac{G_F \alpha_{em}}{\sqrt{2}\pi} \Big[& C_V \left(\bar{s}_L \gamma^\mu b_L\right) \bar{l}_i \gamma_\mu l_j + C_A \left(\bar{s}_L \gamma^\mu b_L\right) \bar{l}_i \gamma_5 \gamma_\mu l_j \\
& + C'_V \left(\bar{s}_L \gamma^\mu b_R\right) \bar{l}_i \gamma_\mu l_j + C'_A \left(\bar{s}_L \gamma^\mu b_R L\right) \bar{l}_i \gamma_5 \gamma_\mu l_j \Big] , \\
\mathcal{H}_{eff}^{SP} = -\frac{G_F \alpha_{em}}{\sqrt{2}\pi} \Big[& C_S \left(\bar{s}_L b_R\right) \bar{l}_i l_j + C_P \left(\bar{s}_L b_R\right) \bar{l}_i \gamma_5 l_j \\
& + C'_S \left(\bar{s}_L b_L\right) \bar{l}_i l_j + C'_P \left(\bar{s}_L b_L\right) \bar{l}_i \gamma_5 l_j \Big] ;
\end{aligned} \tag{1.70}$$

being $C_V^{(\prime)}$, $C_A^{(\prime)}$, $C_S^{(\prime)}$, $C_P^{(\prime)}$ respectively the vector, axial, scalar and pseudo-scalar BSM effective couplings in a general scenario. In the SM, their contribution is negligible.

In terms of the Wilson coefficients, the branching ratio of the LFV decay $B_s^0 \to l_i^+ l_j^-$ is expressed as [5]:

$$\begin{aligned}
& \mathcal{B}\left(B_s^0 \to l_i^+ l_j^-\right) = \\
& \frac{\tau_{B_s^0}}{64\pi^2} f_{B_s^0}^2 G_F^2 m_{B_s^0} \alpha_{em}^2 |V_{tb} V_{ts}^*|^2 \sqrt{\left[1 - \left(\frac{m_{l_i} + m_{l_j}}{m_{B_s^0}}\right)^2\right] \left[1 - \left(\frac{m_{l_i} - m_{l_j}}{m_{B_s^0}}\right)^2\right]} \\
& \left\{ \left[1 - \left(\frac{m_{l_i} + m_{l_j}}{m_{B_s^0}}\right)^2\right] \left| \left(C_9 + C_V - C'_V\right)\left(m_{l_i} - m_{l_j}\right) + \frac{m_{B_s^0}^2}{m_b + m_s}\left(C_S - C'_S\right) \right|^2 \right. \\
& \left. + \left[1 - \left(\frac{m_{l_i} - m_{l_j}}{m_{B_s^0}}\right)^2\right] \left| \left(C_{10} + C_A - C'_A\right)\left(m_{l_i} - m_{l_j}\right) + \frac{m_{B_s^0}^2}{m_b + m_s}\left(C_P - C'_P\right) \right|^2 \right\} ,
\end{aligned} \tag{1.71}$$

where m_b and m_s are the masses of the b and s-quarks. The Wilson coefficients C_9 and C_{10} can implicitly contain a right-handed component: C'_9 and C'_{10}. Notice that, in the dileptonic SM decay, only C_{10} contributes.

A broad spectrum of BSM models predicting LFV via different mechanisms, and in particular giving rise to the decays $B_{(s)}^0 \to \tau^\pm \mu^\mp$, is described in the following lines.

1.3.1 Leptoquark framework

A *Leptoquark* (LQ) is a colored state that can mediate interactions between quarks and leptons carrying both baryon and lepton numbers. Having in some cases fractional electric charge, it appears in many models leading to various new physics effects. In general, a LQ

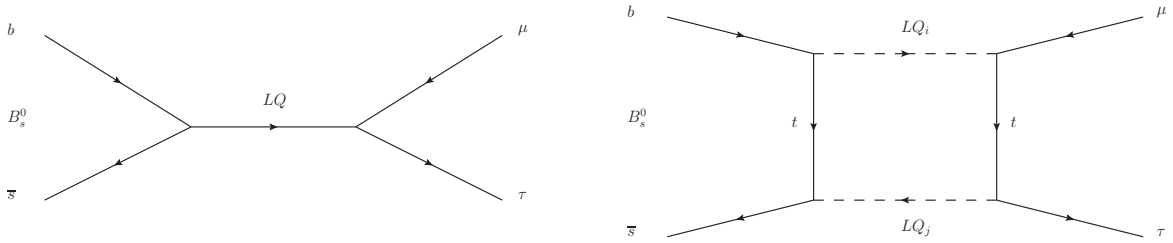

Figure 1.13: Vector LQ (left) and scalar LQ (right) examples of contribution to the LFV decay $B_s^0 \to \tau^{\pm}\mu^{\mp}$.

can be a scalar or a vector field which in turn can come as an $SU(2)$ left singlet, doublet or triplet. LQ's provide a direct coupling between quarks and leptons. Depending on the LQ type, its contribution to LVF can come via tree (vector) or box (scalar) diagrams as those shown in Figure 1.13.

The different types of LQs are specified by their quantum numbers, being the notation: $(SU(3), SU(2)_L)_Y$, where it is indicated if the field is a singlet, doublet or triplet of $SU(3)$ and $SU(2)_L$ respectively, and Y is the so-called hypercharge containing the information of the electric charge and the third weak isospin component. Bearing in mind the experimental results and the different LQ models predictions for the measured $b \to sl^+l^-$ observables, the remaining main accepted models contain:

- Scalar LQ $(3,2)_{1/6}$,
- Vector LQ $(3,1)_{2/3}$,
- Vector LQ $(3,3)_{2/3}$;

having, for the diverse models, different mass couplings and mixing angles.

LQ models can explain quite well the experimental constraints and their use has increased considerably in the theory community during the last years. However, there are detractors to this BSM framework, the main argument being that the LQ field can not be introduced from first principles, hence one need to postulate it by hand.

Scalar LQ $(3,2)_{1/6}$ models [6]

After integrating out the heavy fields, this model gives rise to the chirality flipped operators in the effective Hamiltonian:

$$\left(C_9^{l_1 l_2}\right)' = -\left(C_{10}^{l_1 l_2}\right)' \propto \frac{(g_L)_{sl_1}(g_L)_{bl_2}}{m_{LQ}^2}, \tag{1.72}$$

where g_L are generic coupling matrices and, in this case, the LQ mass is $m_{LQ} = m_{LQ^{2/3}} = m_{LQ^{-1/3}}$. Using the current experimental measurements as constraints, and some assumption for the couplings, the branching ratio of $B_s^0 \to \tau^{\pm}\mu^{\mp}$ can be as large as 10^{-5}, 10^{-6} or 10^{-7} for $m_{LQ} = 1$, 5 or 10 TeV respectively.

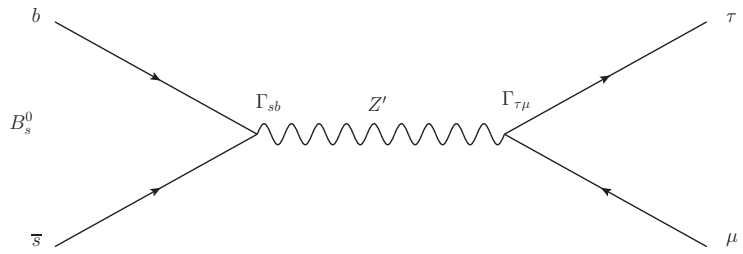

Figure 1.14: Z' boson contribution to the LFV decay $B_s^0 \to \tau^{\pm}\mu^{\mp}$. Γ_{ji} correspond to the coupling amplitudes to the different quarks and leptons.

Vector LQ models [6] [7]

The main characteristic here, on top of the new interactions, is that the LQ currents induce fermion mixing. Consequently, the CKM and PMNS matrices components vary with respect to the SM ones due to the mixing with the charged leptons. In $(3,1)_{2/3}$ models the branching ratio of $B_s^0 \to \tau^{\pm}\mu^{\mp}$ can be as large as 10^{-9} and in $(3,3)_{2/3}$ models as large as 10^{-6}.

1.3.2 Z' models

Many models proposed to explain the $b \to sl^+l^-$ data contain a heavy neutral vector boson (Z') [8] [9] [10] which generates a tree-level contribution to $C_9^{(\prime)}$ and $C_{10}^{(\prime)}$. The Z' field is presumed to be heavier than the SM boson Z^0 and to have tree level couplings of different magnitudes for e, μ and τ leptons and the different quarks. Figure 1.14 shows an example of interaction due to the Z' boson. As the Z' is assumed to be much heavier than the scale of electroweak symmetry breaking, the couplings to neutrinos and to left-handed charged leptons are equal: $\Gamma^L_{l_i l_j} = \Gamma^L_{\nu_i \nu_j}$.

Apart from the mass of the Z', the differences between the various models rely on the value of the right handed couplings Γ_{ij} to leptons. Another important factor is the relative magnitude of the Z' left and right handed coupling to the quarks, as it will affect the computation of the Wilson coefficients which include the QCD penguins. With the current experimental constraints, the different referred models predict a branching ratio of $B_s^0 \to \tau^{\pm}\mu^{\mp}$ between 10^{-8} and 10^{-9}.

However, a document released this current year [31] states that $B_s^0 \to \tau^{\pm}\mu^{\mp}$ is incredibly sensitive to some UV complete Z' models which are used to describe the $R_{K^{(*)}}$ and $R_{D^{(*)}}$ anomalies simultaneously. The branching ratio of $B_s^0 \to \tau^{\pm}\mu^{\mp}$ is $\sim 10^{-4}$ in these kind of models.

1.4 Lepton Flavor Violation experimental searches

Charged LFV could manifest itself not only in $b \to sl^+l^-$ transitions, but also in certain leptons, bosons or other hadrons. Therefore, many searches on LFV are being performed by different collaborations in a wide variety of decay channels. Some of the current experimental upper limits on LFV searches and their future expectations are shown in Table 1.3.

The LFV results displaying the greatest sensitivities are the ones corresponding to muon decays. These kinds of decays are measured in experiments built for these purposes and where the detectors are fully optimized for a given decay. The experiments dedicated to muon decay searches are MEG [32] and SINDRUM (II) in the Paul Scherrer Institute (PSI) in Villigen (Switzerland).

Other experiments study the direct conversion of muons into electrons in fixed target experiments (Au, Al, Ti). Their current sensitivity is 10^{-13} and they plan to greatly improve it up to 10^{-18} in the near future. At Fermilab, reusing part of the Tevatron acceleration complex, is located the experiment Mu2e (II) [33]. The COMET [34] and PRISM [35] experiments are located at KEK in Tsukuba (Japan).

Another area is being exploited with τ LFV decays. These searches were and are done in flavor physics detectors, like LHCb and Belle (II) [36] [37] as τ's need high energies to be massively produced. The studied τ leptons typically are a product of a b-hadron decay.

NA62 [38] is a fixed-target experiment at the CERN in Geneva (Switzerland) dedicated to measurements of rare kaon decays and it performs LFV searches as well.

At CERN in Geneva, general purpose LHC experiments like ATLAS and CMS also perform LFV searches based on decays of the Z^0 and Higgs bosons.

The LHCb experiment has performed LFV searches in various modes. It has results on dileptonic decays and, nowadays, also analyses of modes with one hadron and two different leptons final states are ongoing, although being still in the early stages of the analyses. Many modes will have alternative measures from the Belle II experiment once it records more data.

Concerning the decay search described in this document, an upper limit on the $B^0 \to \tau^{\pm}\mu^{\mp}$ channel has been measured by the BaBar collaboration: $\mathcal{B}\left(B^0 \to \tau^{\pm}\mu^{\mp}\right) < 2.2 \cdot 10^{-5}$ at 90% CL [11], summarized in Section 3.1.1. The first search of the $B_s^0 \to \tau^{\pm}\mu^{\mp}$ channel is performed within the analysis described later in the text.

1.5 Conclusions

The absence of LFV in the charged lepton sector is seen as paradoxical in the SM of particle physics when compared with the quarks mixing. The discovery of the neutrino's mass and their oscillations have given even more room for suspicions about the Lepton Flavor conservation for charged leptons.

The $R_{K^{(*)}}$ and $R_{D^{(*)}}$ anomalies, although not having reached the 5σ significance threshold yet, are the most striking tensions with the SM seen so far. Most of the models proposed explaining such anomalies predict a certain degree of LFV. The anomalies will need other similar measurements as further confirmation. However, if a charged LFV decay is ever observed, it will be an irrefutable proof of BSM physics.

LFV in charged leptons is being searched with increasing fervor, from leptons to higgs to b-hadrons decays. LHCb, in particular, is performing searches in a wide variety of decays, including the third lepton generation.

The effort on LFV searches will keep increasing in the near future in all areas with already many planned upgrades for different experiments. In the case of flavor physics, the data taking of Belle II and the upgrade of LHCb are expected to bring significant insights for LFV searches in the upcoming years.

Process	Current Limit	Next Measurement
$\mu \to e\gamma$	$5.7 \cdot 10^{-13}$ 90%CL [39]	PSI
$\mu \to eee$	$1.0 \cdot 10^{-8}$ 90%CL [40]	
$\mu \to e$	$10^{-12}, 10^{-13}$ 90%CL [41] [42]	PRISM, COMET, Mu2e(II)
$\tau \to e\gamma$	$3.3 \cdot 10^{-8}$ 90%CL [43]	LHCb (Belle II)
$\tau \to eee$	$2.7 \cdot 10^{-8}$ 90%CL [44]	
$\tau \to e\mu\mu$	$2.7 \cdot 10^{-8}$ 90%CL [44]	
$\tau \to e$ had	10^{-8} 90%CL [45]	
$\tau \to \mu\gamma$	$4.4 \cdot 10^{-8}$ 90%CL [43]	LHCb (Belle II)
$\tau \to \mu ee$	$1.8 \cdot 10^{-8}$ 90%CL [44]	
$\tau \to \mu\mu\mu$	$2.1 \cdot 10^{-8}$ 90%CL [44]	
$\tau \to \mu$ had	10^{-8} 90%CL [45]	
$\pi^0 \to e\mu$	$3.6 \cdot 10^{-10}$ 90%CL [46]	NA62
$K_L \to e\mu$	$4.7 \cdot 10^{-12}$ 90%CL [47]	NA62
$K_L \to \pi^0 e\mu$	$7.6 \cdot 10^{-11}$ 90%CL [46]	
$K_L \to \pi^0\pi^0 e\mu$	$1.7 \cdot 10^{-10}$ 90%CL [46]	
$K^+ \to \pi^+ e^- \mu^+$	$6.8 \cdot 10^{-11}$ 90%CL [48]	
$Z \to e\mu$	$7.5 \cdot 10^{-7}$ 95%CL [49]	ATLAS
$h \to e\mu$	$6.1 \cdot 10^{-3}$ 95%CL [50]	CMS
$h \to \mu\tau$	1.8-2.5 $\cdot 10^{-3}$ 95%CL [51] [50]	ATLAS, CMS, LHCb
$D^0 \to e\mu$	$1.3 \cdot 10^{-8}$ 90%CL [52]	LHCb (Belle II)
$B^0 \to e\mu$	$1.3 \cdot 10^{-9}$ 95%CL [53]	LHCb (Belle II)
$B_s^0 \to e\mu$	$6.3 \cdot 10^{-9}$ 95%CL [53]	
$B^0 \to \mu\tau$	$2.2 \cdot 10^{-5}$ 90%CL [11]	
$B_s^0 \to \mu\tau$	-	
$B^+ \to \pi^- \mu^+ \mu^+$	$5.8 \cdot 10^{-8}$ 95%CL [54]	
$B^+ \to K^- \mu^+ \mu^+$	$5.4 \cdot 10^{-8}$ 95%CL [54]	
$B^+ \to Ke\mu$	0.091 90%CL [29]	
$B^+ \to K\mu\tau$	48 90%CL [29]	
$B^+ \to K^*\mu\tau$	-	
$\Lambda_b \to \Lambda^0 e\mu$	-	LHCb

Table 1.3: Some of the current experimental upper limits for different LFV searches.

Chapter 2

The LHCb experiment

2.1 Basics on B physics experiments

The necessary energy to produce b quark pairs can be achieved by colliding particle beams. In particular, the center of mass energy ($\sqrt{s}$) of the colliding beams must be greater than two times the typical mass of a B meson.

B physics detectors must hold, at least, two main properties:

- **Excellent tracking and vertex resolution**: The masses of the B mesons spectrum are very close to each other.
- **π-K separation**: The most abundant final states of B decays need to be correctly differentiated.

Another key factor of the detectors is its trigger. The trigger allows the selection of B meson decays from other kinds of processes which are also produced during the collisions, specially in hadronic machines.

Producing b quarks

Historically, two strategies have been used to produce b quarks at large scale in particle colliders.

- **e^+e^- accelerators**: (Super)KEKB located in Tsukuba (Japan) with the Belle (II) detector [36] [37], and PEP-II at SLAC National Accelerator Laboratory (USA) with the BaBar experiment [55]. These experiments are the so-called B *factories*. Belle and BaBar have finished their data taking period in 2010 and 2008 respectively. Belle 2 is expected to start collecting data at the end of 2018.
- **Proton-(anti)proton (pp ($p\overline{p}$)) accelerators**: Tevatron ($p\overline{p}$) at Fermilab (USA) with the D0 [56] and CDF [57] experiments, and the Large Hadron Collider (LHC) (pp) with LHCb, ATLAS and CMS. The LHC detectors are the only ones in active service nowadays.

To describe the differences between the two colliding strategies, one must have in mind the formula describing the number of b quarks produced (N_b), which can be written

roughly as:

$$N_b = \sigma_{b\bar{b}} \int \mathcal{L} dt, \tag{2.1}$$

where $\sigma_{b\bar{b}}$ is the cross section production of the b quark pairs, and $\int \mathcal{L} dt$ (or L_{int}) the time integrated luminosity which indicates the amount of data collected by the experiment in a given period of time. The *instantaneous luminosity* ($\mathcal{L}$) indicates the *power* of the accelerator, e.g. the data per unit of time that the machine can deliver.

e^+e^- colliders

The e^+e^- colliders KEKB and PEP-II used to run most of the time at the $\Upsilon(4S)$ resonance [13], in order to maximize the production of a $B\overline{B}$ meson pair. The mass of the resonance is just above the threshold of 2 times the B^0 mass, thus they are able to study B^0 and B^+ decays. KEKB took few data running at the $\Upsilon(5S)$ resonance [13], allowing the study of some B_s^0 decays.

B mesons are produced isotropically, thus the detectors generally have a 4π radiants coverage, which simplifies the work with missing particles in the final state. In order to study time dependent processes, the e^- and e^+ need to be collided asymmetrically. The energy difference between the e^- and e^+ beams allows to boost the produced B meson.

The $\sigma_{b\bar{b}}$ at the e^+e^- colliders is relatively low, $\sim 10^{-9}$ nb. The typical instantaneous luminosity $\mathcal{L}$, for instance for KEKB and PEP-II, was 1 to $2 \cdot 10^{34}$ cm^{-2}s^{-1}. The main advantage of colliding fundamental particles is a very clean environment.

In summary: with an e^+e^- machine, due to the relatively small cross section, not many B mesons are generated. However, they are produced in a low background environment and the trigger efficiency is close to 100%.

Proton-proton colliders

The main advantage of using a pp collider is that the $\sigma_{b\bar{b}}$ in this case is huge, being energy dependent. At $\sqrt{s}$ of 7 and 13 TeV the production cross section is $(72.0 \pm 0.3 \pm 6.8)$ μb and $(144 \pm 1 \pm 21)$ μb [58], and it grows roughly proportionally to the center of mass energy.

In proton-proton collisions the $b\bar{b}$ pairs are produced in the processes shown in Figure 2.1. In fact, the proton themselves do not collide, but the gluons and quarks (*partons*) inside. In particular, the b quark production [59] is dominated by gluon interaction processes, although the quark based processes contribution are significant as well:

- **Flavor excitation** contributes with $\sim$54% of the total production. It is produced when gluons of interacting protons interact.
- **Gluon splitting** contributes with $\sim$27% of the total production.
- **Pair production** ($g\bar{g} \to b\bar{b}$ and $q\bar{q} \to b\bar{b}$) contributes with $\sim$16% of the total production.

The resulting b quark angular distribution is not at all isotropic. The produced $b\bar{b}$ are concentrated around the axes of the colliding beams in the forward region, as shown in

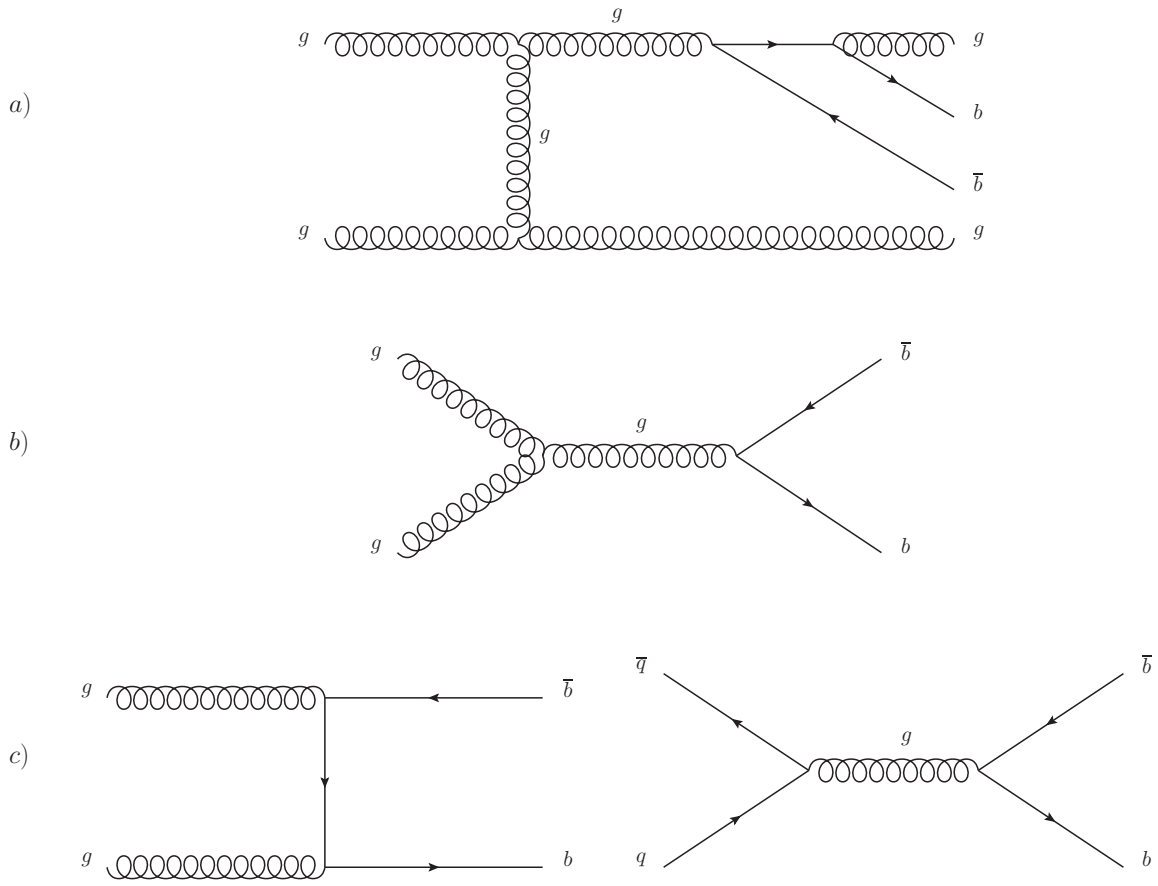

Figure 2.1: $b\bar{b}$ main production mechanisms in proton-proton collisions at 14 TeV. a) Flavor excitation. b) Gluon splitting. c) Pair production.

Figure 2.2. Therefore, the tracks of the B meson daughters are very colinear, and this complicates the B decay reconstruction.

Given the large energy range of the $b\bar{b}$ pairs produced, the b quarks do not hadronize only into $B^0_{(s)}$ and B^+ mesons. In fact, all b hadron types are produced, leading to a rich physics program. However, for a given physics search, large amounts of background is present. The large background, combined with the final state tracks colinearity, causes the degradation of the experiment's trigger efficiency.

2.2 The Large Hadron Collider

The Large Hadron Collider (LHC) is located at the European Center of Nuclear Research (CERN) in Geneva (Switzerland). The LHC is a double ring proton-proton circular accelerator of ∼27 km of circumference and between 50 and 150 m underground. It accelerates protons circulating in opposite directions to an energy of 3.5 to 6.5 TeV corresponding to $\sqrt{s}$ of 7 TeV in 2011, 8 TeV in 2012 (Run I) and 13 TeV from 2015 to end of 2018 (Run II). The analysis described in this document exploits Run I pp collisions data.

The LHC has 4 interactions points, in which the two proton beams cross each other

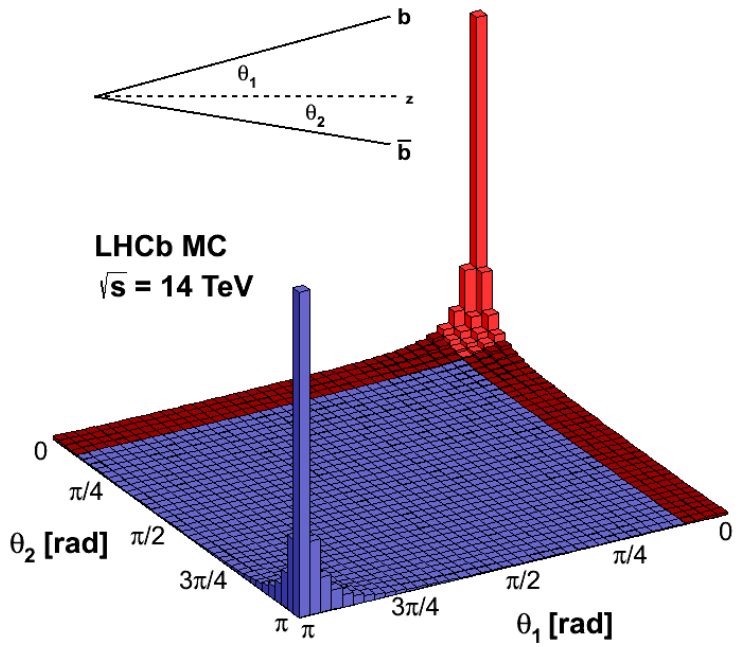

Figure 2.2: $b\bar{b}$ angular distribution and correlation in proton-proton colisions. The red area corresponds to the LHCb detector coverage.

and the collisions take place. This is where the detectors are installed. The four detectors are:

- **ATLAS (A Toroidal Lhc ApparatuS) and CMS (Compact Muon Solenoid) experiments:** They are general purpose detectors aiming to exploit the high luminosity of the LHC in order to discover new particles. ATLAS and CMS are the largest collaborations at CERN (∼3000 physicists each) with an extensive and diverse physics program ranging from SM measurements to the most exotic BSM searches. ATLAS and CMS use different detector concepts. They discovered the Higgs Boson in 2012 finding the final missing piece of the SM.

- **LHCb (Large Hadron Collider beauty) experiment:** LHCb is a B physics detector described in Section 2.3. It has been build by a smaller collaboration (around 1000 physicists at the time of building), and has a rich physics program focused on b and c-hadron studies.

- **ALICE (A Large Ion Collider Experiment) experiment:** It is an experiment dedicated to quark-gluon plasma measurements. It is focused on measurement of lead-lead collisions although it collects pp and p-lead data for normalization purposes.

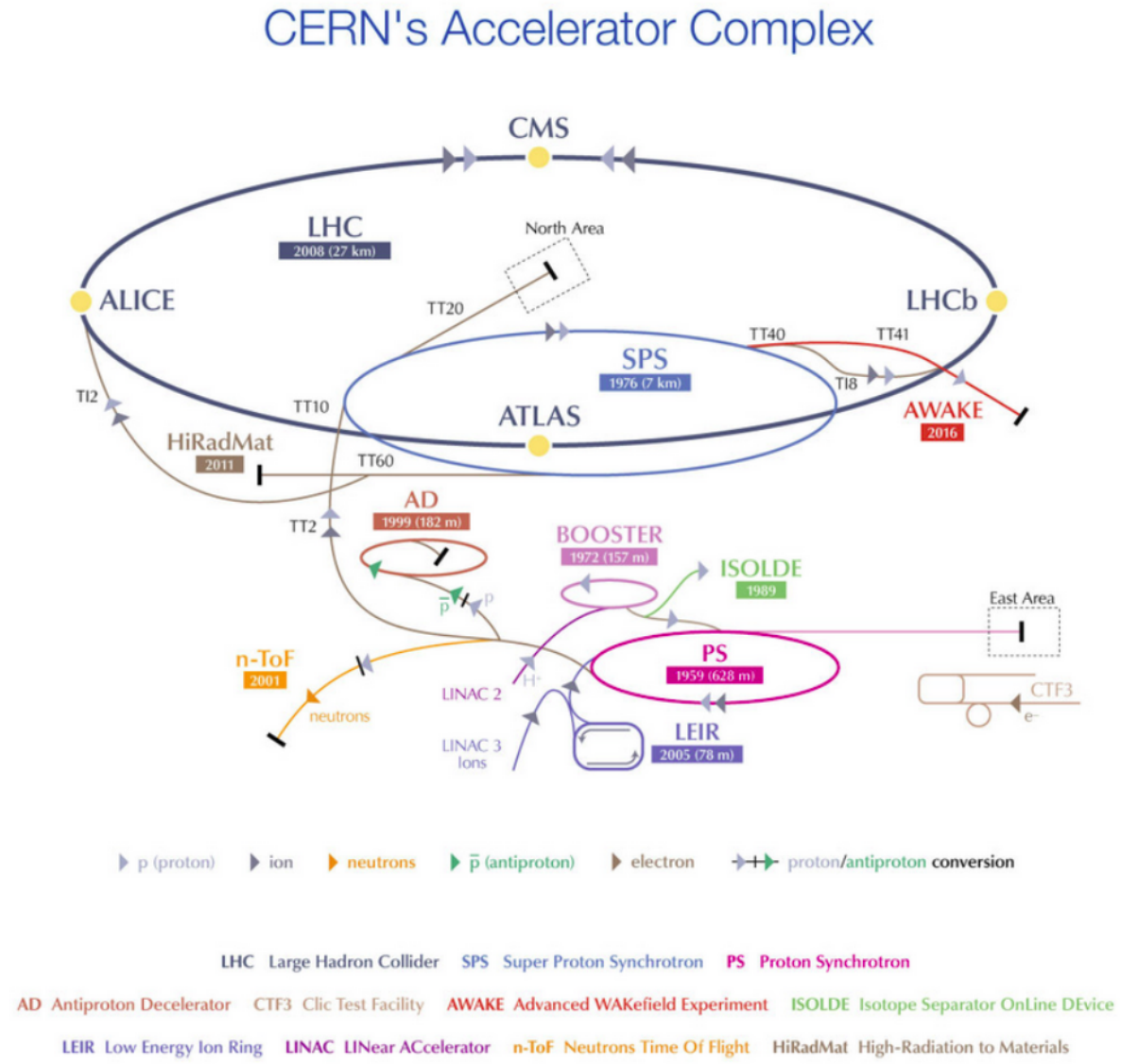

Figure 2.3: CERN accelerator complex.

Accelerating protons in the LHC

The electrons are stripped from hydrogen atoms nuclei using electric fields and the protons are injected in a vacuum pipe to begin the acceleration process, which is divided into several steps. First the LINAC accelerates the protons up to 50 MeV, then they reach 1.4 GeV in the Proton Synchroton Booster. In the Proton Synchroton, the proton beams are accelerated up to an energy of 25 GeV before reaching the Super Proton Synchroton (SPS) which boosts them up to 450 GeV. Finally, protons are injected in the two LHC accelerator rings, clock and anticlockwise, to reach their final energy of (currently) 6.5 TeV. The CERN accelerator complex is shown in Figure 2.3.

During the acceleration process, the protons are organized in bunches. When they reach the LHC, a typical bunch has a population of 1.1 to $1.4 \cdot 10^{11}$ protons and the LHC can cope with 2808 filled bunches in each *fill*. A fill is a period which comprises between the beam injection and the beam dump. The bunches are separated ~7.5 m, which at

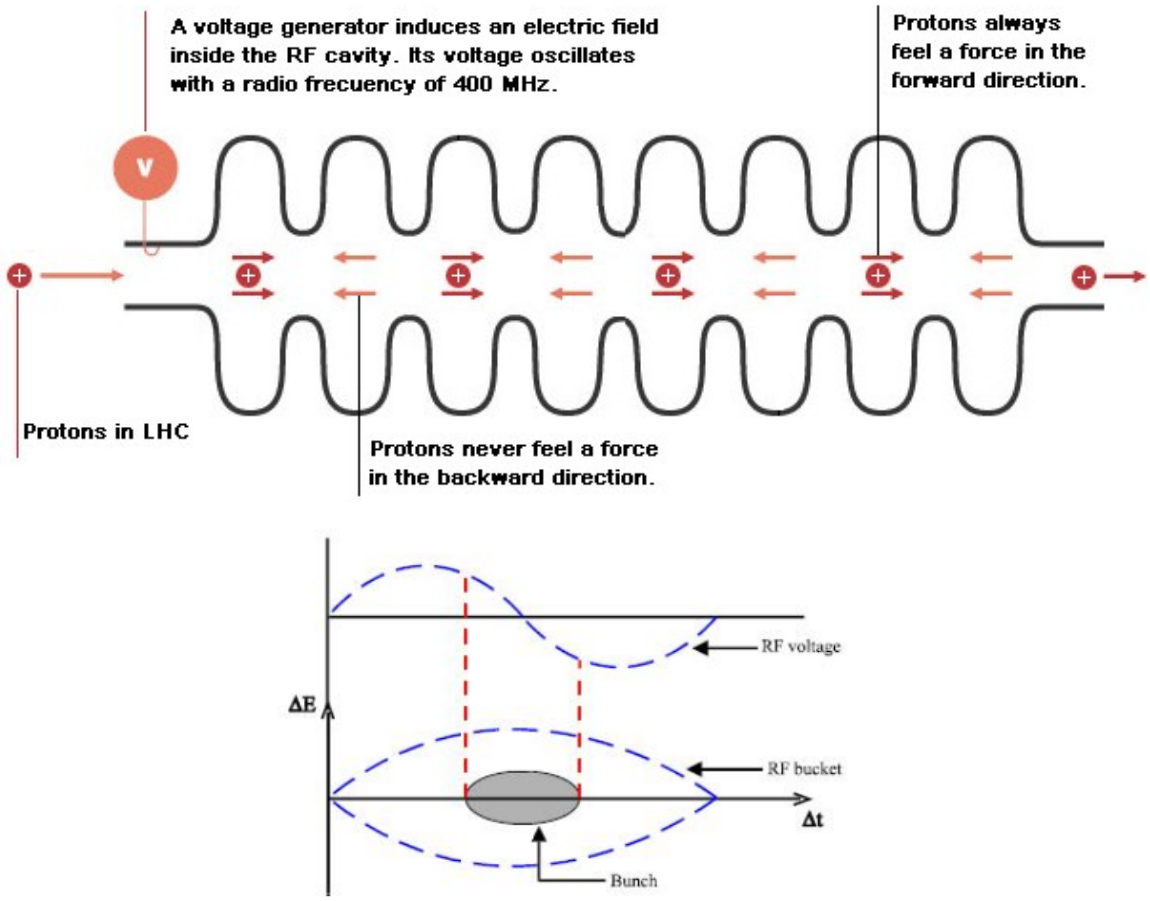

Figure 2.4: RadioFrequency chambers schematic behavior.

6.5 TeV means a collision every 25 ns, thus a crossing frequency of ∼40 MHz.

To accelerate bunches in a synchronized way, a very specific electric field is needed. Such field is provided by the RadioFrequency (RF) chambers. The LHC use eight RF cavities per beam, each delivering 2 MV (an accelerating field of 5 MV/m) at 400 MHz. The RF Cavities generate a longitudinal oscillating voltage, which makes the bunch "surf" along the accelerating electromagnetic waves. A period of the RF oscillations is called *bucket*, as if it was used to carry the bunch. Figure 2.4 illustrates the acceleration procedure.

The proton beams are bent so they follow the LHC circular trajectory. For that purpose 1232 superconducting magnetic dipoles are used. Each dipole is 15 m long and provides a magnetic field of 8.3 T.

To stabilize and focus the beams at the interaction point, superconducting multipoles are used. The most abundant ones are quadrupoles. However sextupoles and octupoles are used to help in beam focusing and counteracting other interactions that each beam suffers as electromagnetic interactions among bunches and electron clouds from the pipe wall.

Parameters	2011	2012	Run II (2015-2018)
$\sqrt{s}$ (TeV)	7	8	14
N	$1.2 \cdot 10^{11}$	$1.2 \cdot 10^{11}$	$1.2 \cdot 10^{11}$
n	1800	1800	2808
f_{rev} (kHz)	11	11	11
σ^*_{xy} μm	60	60	15
$\mathcal{L}$ (cm^{-2}s^{-1})	$3.65 \cdot 10^{33}$	$3.65 \cdot 10^{33}$	10^{34}

Table 2.1: LHC parameters during RunI (2011 and 2012) and RunII data taking periods.

Luminosity and colliding beams

The instantaneous luminosity can be written as:

$$\mathcal{L} = \frac{N^2 n f_{rev}}{4\pi\sigma_{xy}} F, \tag{2.2}$$

where N is the number of protons per bunch, n the number of bunches, f_{rev} the bunch revolution frequency, σ_{xy} the sizes of the beams in the transverse plane to the beam axis, and F is a geometrical factor containing the crossing angle of the colliding beams.

Basically, the beams must be squeezed before the interaction point, minimizing as much as possible the denominator of Equation 2.2, which can be re-written in terms of two quantities, the transverse emittance (ϵ), and the amplitude function (β):

$$\sigma_{xy} = \epsilon\beta. \tag{2.3}$$

The geometry of the beam is described by ϵ and β is determined by the accelerator magnet configuration (the quadrupole magnet arrangement) and power. When β is minimal, it is called β^* (thus σ^*_{xy}), and it corresponds to the maximum instantaneous luminosity an accelerator can provide. In terms of instantaneous luminosity, the LHC is the most powerful accelerator ever built. Table 2.1 contains the parameter values for the LHC Run I and Run II data taking periods.

Compared to ATLAS and CMS, the LHCb case is very particular as the geometrical factor F plays an essential role. High pile-ups (number of interactions per crossing bunch) are to be avoided as the $b\bar{b}$ production in hadronic machines implies a very high co-linearity among the final state tracks. LHCb is designed to nominally have $\sim$1.8 parton interactions per bunch crossing (in practice they are $\sim$2.5). The beams are squeezed normally to avoid parasitic collisions but, to lower the instantaneous luminosity, a luminosity levelling procedure was introduced at the LHCb interaction point by adjusting the transverse overlap of the beams (Figure 2.5).

As a comparison, the nominal instantaneous luminosities with bunch separation of 25 ns beams achieved by ATLAS and CMS are $\sim 4 \cdot 10^{34}$ cm^{-2}s^{-1}, whereas for LHCb it is $\sim 2 - 5 \cdot 10^{32}$ cm^{-2}s^{-1}.

2.3 The LHCb detector

LHCb is a dedicated b and c-quarks physics precision experiment at the LHC that searches for BSM physics through the study of very rare decays of charm and beauty hadrons and

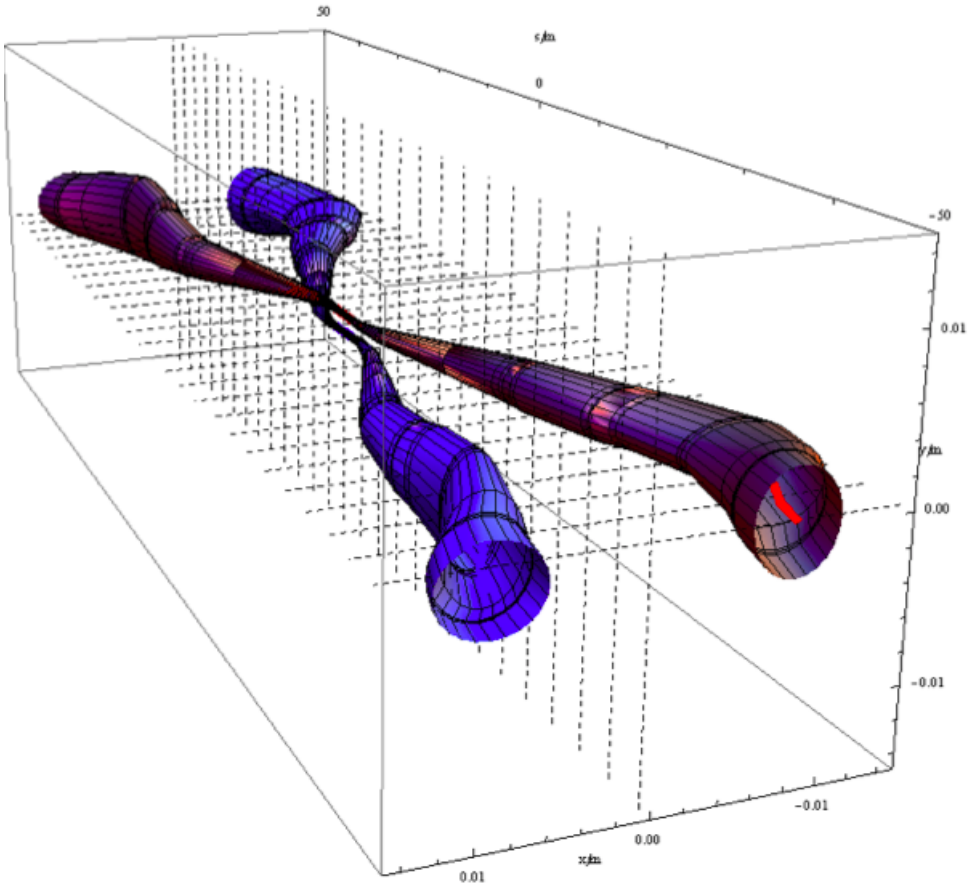

Figure 2.5: Schematic representation of the beams collisions in the interaction point 8 (LHCb) in 2011 (on 2012 the crossing angle is in the vertical plane).

Conditions	Run I	Run II
$\sqrt{s}$ (TeV)	7-8	13
Bunch spacing (ns)	50	25
Pile-up	$\sim$2	1.3-2.4

Table 2.2: LHCb conditions during Run I (2011 and 2012) and Run II (2015-2018) data taking periods.

performs precision measurements of CP-violating observables.

During the different data taking periods of the LHC, LHCb has recorded the time integrated luminosity shown in Figure 2.6 in the conditions reported in Table 2.2. This document is focused mainly in Run I LHCb configuration [60].

The detector design is that of a single arm spectrometer in the forward region, with an angular coverage from approximately 15 mrad to 300(250) mrad in the detector's magnet bending (non-bending) plane. The detector geometry is driven by the angular production of $b\bar{b}$ in the proton-proton collisions as illustrated in Figure 2.2. The pseudo-rapidity (η) acceptance of LHCb, e.g. its angular coverage, is unique when compared to the other LHC detectors, as illustrated in Figure 2.7. The LHCb detector layout in the vertical

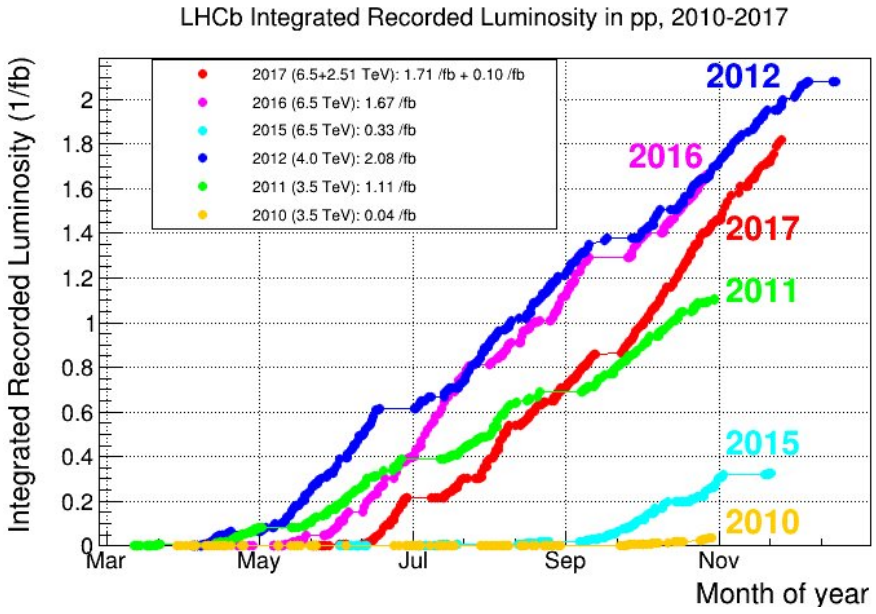

Figure 2.6: LHCb integrated luminosity during the different data taking periods.

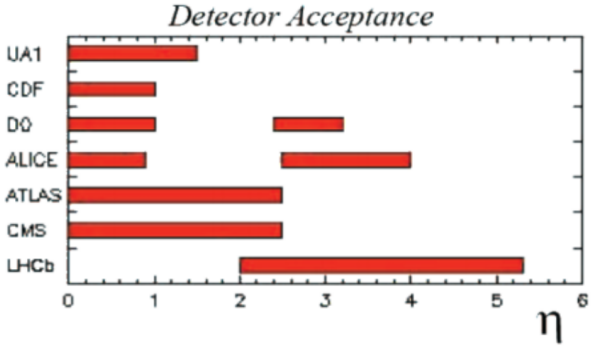

Figure 2.7: Different detectors pseudo-rapidity coverage.

plane is shown in Figure 2.8.

The LHCb detector is divided in different sub-systems depending on their functionality:

- **Tracking system:** In charge of measuring the trajectory and momenta of the charged tracks.
- **Particle IDentification (PID) system:** It disentangles the different final state particles.
- **Trigger and Data AcQuisition (DAQ):** It selects the interesting events and stores the data for the further physical analyses.

Figure 2.9 shows the behavior of the different particle species through the detector's components.

In the following sections, a basic description of the LHCb sub-systems is provided. For a more complete description of the detector's design, the reader is referred to [12].

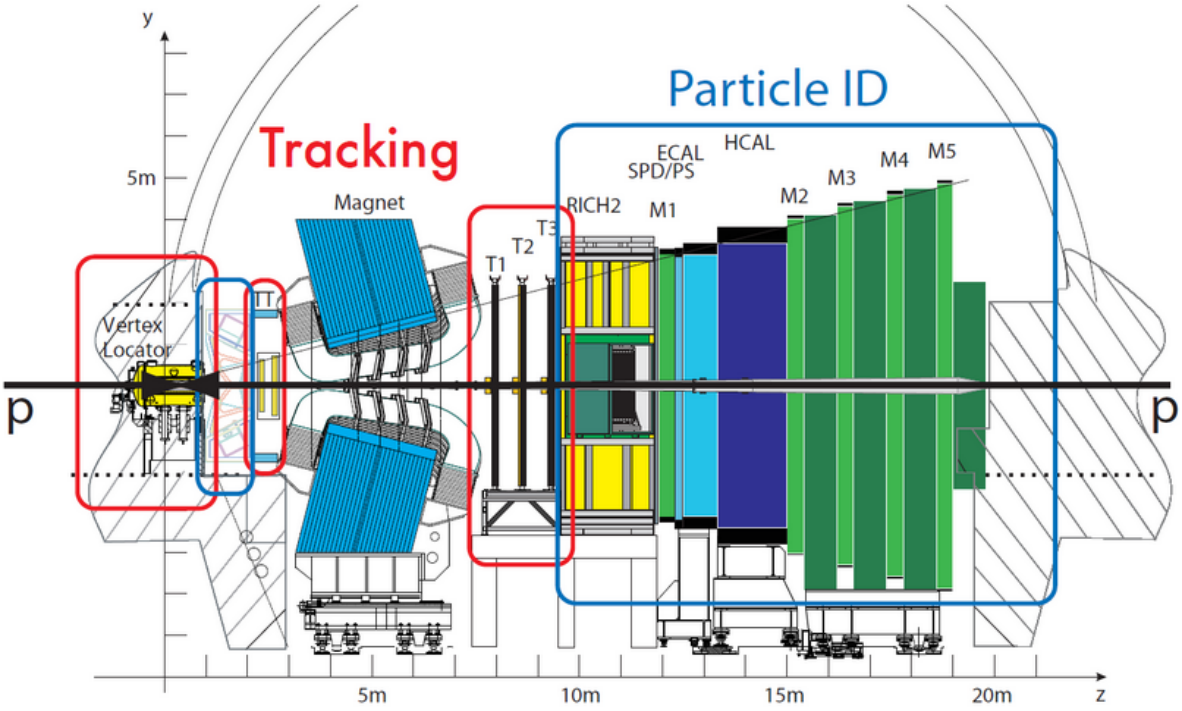

Figure 2.8: LHCb detector layout in the vertical plane.

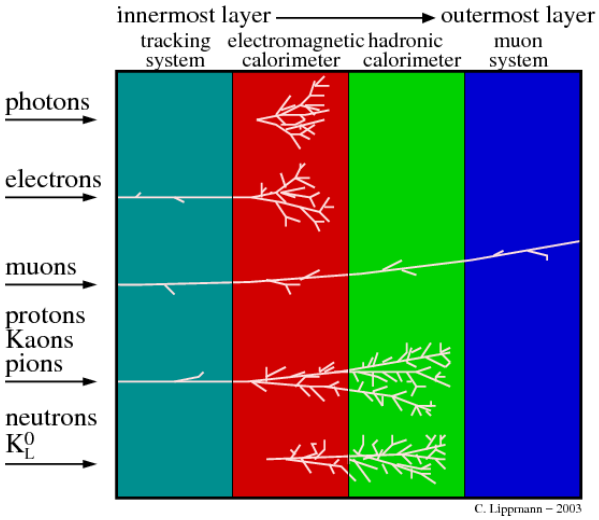

Figure 2.9: Typical particle interactions with a high energy detector.

2.3.1 Tracking system

The tracking system measures the charged particle's trajectories and thus their geometrical and kinematic basic observables. It is composed of the magnet, the VErtex LOcator (VELO) and the Tracking stations (TT, T1, T2, T3), illustrated in Figure 2.10. For further reference, the measured tracks can be divided in different categories depending on

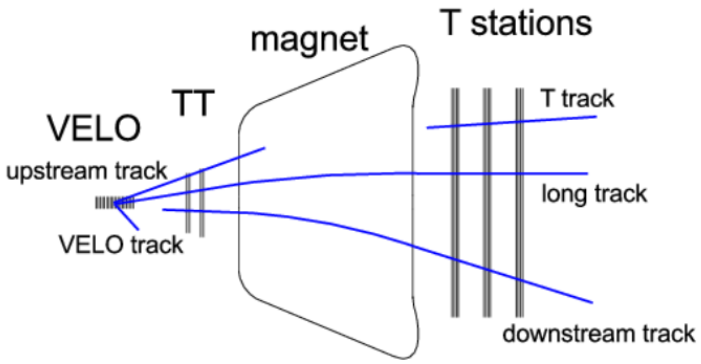

Figure 2.10: LHCb track types classification.

the tracking sub-detectors participating in the detection.

- **Long tracks:** They traverse all the tracking system and are the most useful for physics analysis, as their momentum measurement is very precise.
- **VELO tracks:** They are only reconstructed in the VELO and are used, together with other kind of tracks, for the measurement of the point position of the *pp* interaction, the so-called Primary Vertex (PV).
- **Upstream tracks:** Low momentum tracks seen by the VELO and the TT station which have been bent away by the magnet.
- **T tracks:** Tracks that are only seen in the tracking stations after the magnet. They are due to very long lived particles or material interactions.
- **Downstream tracks:** Tracks that go through the entire tracking system except the VELO. They are used in the studies of long living particles as K_s^0.

The Magnet

The magnet is responsible for curving the trajectory of charged particles. This is necessary to measure the particles momentum. It consists of a warm dipole magnet providing an integrated field of about ∼4 T which deflects charged particles in the horizontal plane. The magnet field has an impact on the trajectory of the LHC beams as well. Three dipole magnets are used to compensate for this effect, thus ensuring a closed orbit for the beams.

A particularity of the LHCb magnet is that its polarity can be changed during a given data taking period. Approximately half of the time the data is taken with the magnet in the *up* configuration and the other half in a *down* one. This feature is fundamental for CP violation searches.

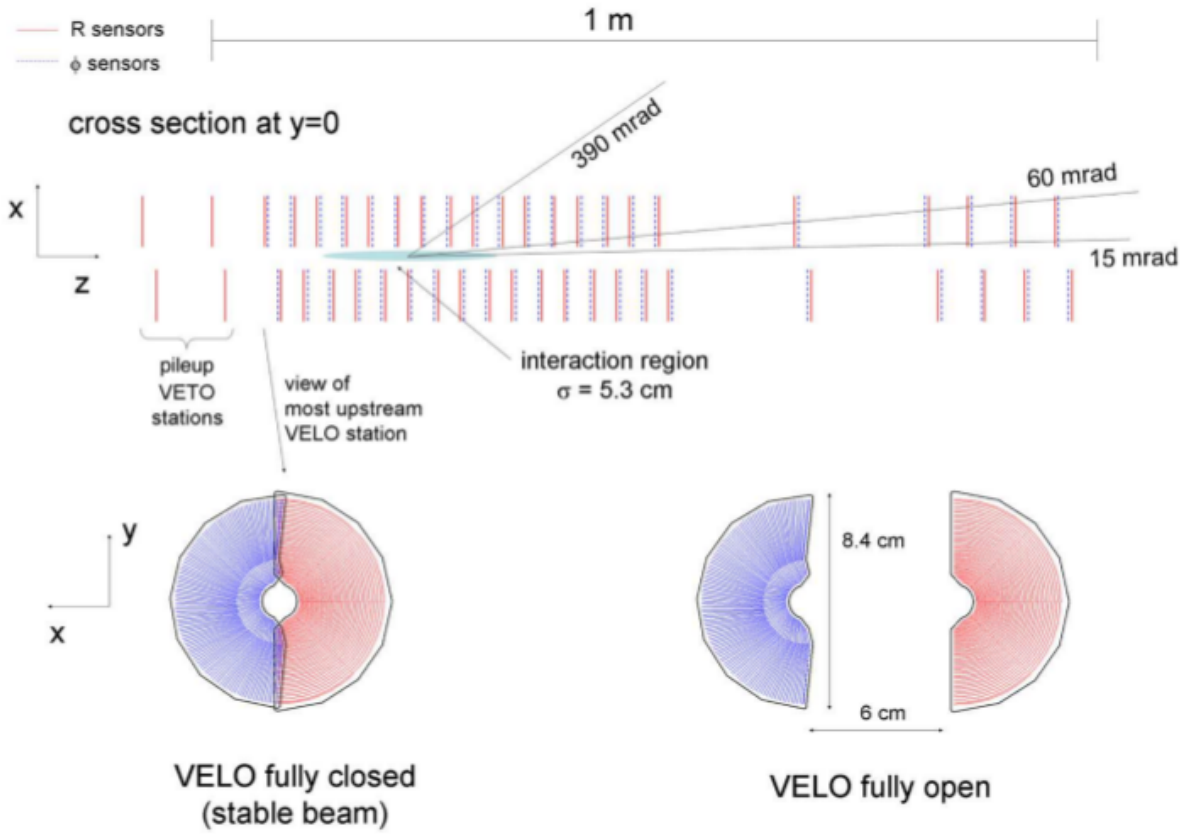

Figure 2.11: LHCb Vertex locator schematic representation.

The Vertex Locator

The VELO, located inside a vacuum tank, is the closest detector to the interaction point and is responsible for finding the PV and particle's decay vertices (also called secondary vertexes) coordinates. It plays a fundamental role to identify b and c hadron decays by measuring their decay time and Impact Parameter (IP) with respect to the PV.

The VELO resolution permits to measure the lifetime of the hadrons produced at the PV with an uncertainty of ~ 50 fs, as the impact parameter resolution is around 20 μm for a track of 3 GeV/c.

The subdetector is schematically shown in Figure 2.11. It consists of 42 silicon modules, also called stations, arranged along the beam, each providing a measurement of the distance (R sensors) and angle (ϕ sensors) coordinates. Tracks must cross at least three VELO stations to be reconstructed. The stations are arranged to ensure that this requirement is usually satisfied for tracks in the LHCb acceptance.

The VELO is split in two halves. Given its proximity to the interaction point, for integrity reasons, it has 2 configurations: during the physics data taking, when the proton beams are stable, the two halves of the VELO overlap slightly being *closed* configuration; otherwise the VELO splits itself separating its 2 halves by 6 cm and remains in *open* configuration.

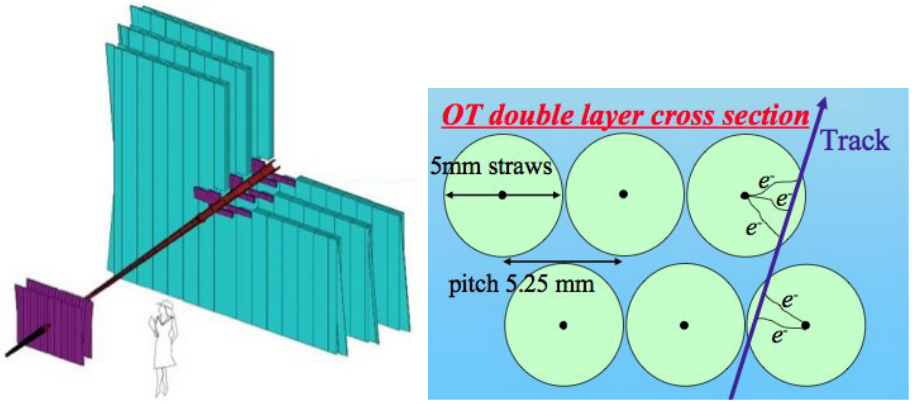

Figure 2.12: LHCb tracking system representation. Left: TT and T-stations, in purple the silicon detectors and in light blue the scintillating drift fibers. Right: scintillating drift fibers.

Tracking stations

The TT and inner parts of the T1, T2 and T3 stations are made of silicon detectors, while the outer parts are made of scintillating drift fibers (Figure 2.12). The silicon detectors are designed to achieve a spatial resolution of 50 μm as here the tracks are more separated than in the VELO.

The TT station consists of four planar detection layers with a width and height of approximately 160 cm and 130 cm, respectively. This amounts to a total active area of approximately 8.4 m^2. The four detection layers are arranged in two half stations separated by approximately 27 cm along the beam axis.

In T1, T2, and T3 (T-stations), the silicon tracker consists of cross-shaped area around the LHC beam pipe. The cross extends over approximately 120 cm in width and 40 cm in height. The remaining T-stations area consists in two layers of scintillating drift fibers with inner diameters of 4.9 mm, filled with a mixture of 70% Ar and 30% CO_2. The straw drift-tubes can achieve momentum resolutions of $\sim$4%.

In the raw data, a track is seen as a set of hits in the tracking stations. To translate the particle interactions with the detector material to actual tracks, a fitter algorithm, based on a Kalman filter [61], is used from downstream to upstream tracking stations. The track trajectories are parametrized using the geometry of the detector, the momentum coordinates and resolution, and the magnetic field. The algorithm basically searches for the detectors hits which are most likely coming from the same track. In order to link the tracks from the tracking stations to the VELO ones, a similar fitter is used.

2.3.2 Particle identification systems

The PID systems are formed by the RICH detectors, the calorimeters and the muon stations at the end of the detector. Each one targets the identification of particular types of particles and thus uses very different detection methods. The PID sub-detectors information is finally combined optimizing this way the particle identification.

The RICH Detectors

RICH is the acronym for Ring Imaging Cherenkov detectors. LHCb has two RICH detectors: the RICH 1 located between the VELO and the TT, and the RICH 2 located downstream right after the T-stations. Their main mission is to distinguish pions from kaons although they can identify other final state particles as protons, muons and electrons. This function is vital for B-physics studies.

As their name indicates, the RICH detectors use Cherenkov radiation to identify the particle type. Cherenkov radiation is emitted when a particle travels through a medium with a speed greater than the speed of light in such medium. The emitted light is characterized by a ring from a light-cone of angle (θ_c), whose cosinus is proportional to the speed of the particle (v):

$$cos\theta_c = \frac{c}{nv}, \tag{2.4}$$

where c is the speed of light in the void and n the refraction index of the medium.

The ring resolution is proportional to $\sigma_\theta/\sqrt{N}$ where σ_θ is the uncertainty on θ_c and N is the number of photons in the ring.

The RICH 1 sub-detector covers the complete LHCb acceptance and uses a mixture of aerogel and C_4F_{10} to identify particles with a momentum range between 1 and 60 GeV/c. The RICH 2 uses CF_4 to identify particles with momenta between 15 and 1000 GeV/c covering an acceptance between 12 and 120 mrad.

The photons from the Cherenkov radiation are detected by a system of mirrors and photodetectors. The schematic view of a RICH detector is shown in Figure 2.13. Finally, a likelihood fit is used. It quantifies how well the measured Cherenkov angle is described by expected ring of the different particle hypotheses (Figure 2.14). Each identification has a certain efficiency as well as a mis-identification (misID) rate, an example is shown in Figure 2.15.

Calorimeters

The calorimeters are important components in high energy physics detectors. Their mission is to capture the energy of the particles interacting with them, allowing the observation of neutral particles (which are invisible for the trackers). Charged particles have a track associated to the energy deposit in the calorimeter whereas for neutral ones no associated track to the energy deposit can be found. Typically, the calorimeters stop all the usual long lived particles except neutrinos (undetectable) and muons (seen by tracking system). For this reason the calorimeters are placed after the tracking system. The hadrons are stopped in the *hadronic calorimeter*, and photons and electrons in the *electromagnetic calorimeter*.

The calorimeters are composed of interleaved layers of dense material and scintillating fibers. The dense material layers heavily interact with the particles triggering showers and the scintillating fibers collect the shower radiation. A shower is a cascade of secondary particles produced as the result of a particle interacting with dense matter. The hadronic showers are longer and wider due to the binding contribution of the strong interactions.

The LHCb electronic calorimeter (ECAL) consists of 66 layers of 2 mm thick lead, 120 μm thick reflecting material and 4 mm thick scintillator tiles. The energy resolution is roughly $1\% + 10\%/\sqrt{E}$ for energies from 15 to 100 GeV/c.

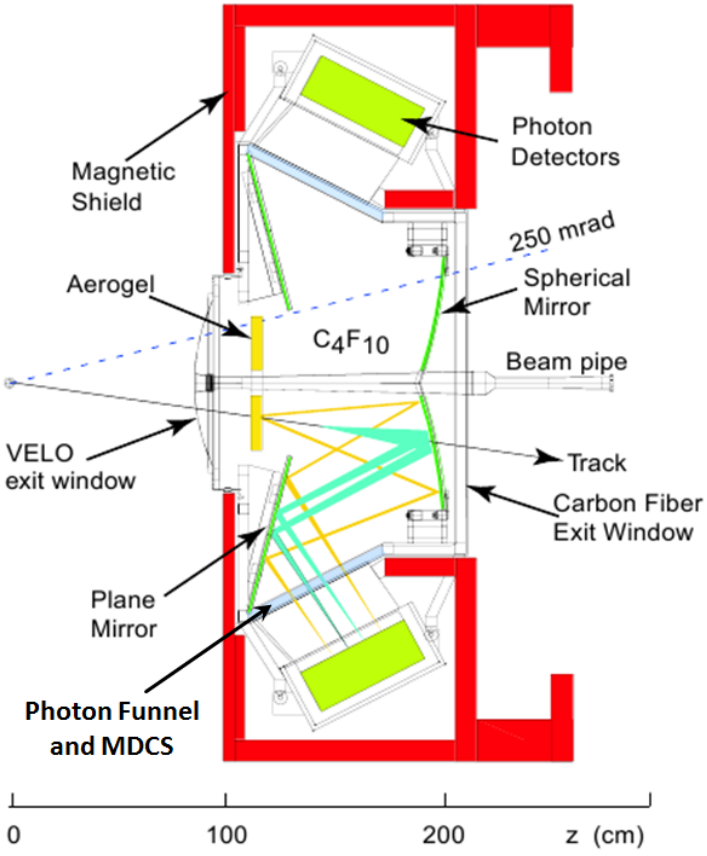

Figure 2.13: LHCb RICH 1 scheme.

The LHCb hadronic camorimeter (HCAL) has a similar structure but the absorber is iron instead of lead. It is divided into square cells of side length 131.3 mm in the inner section and 262.6 mm in the outer section. The resolution varies as a function of the incoming momentum from 23 to 12% for momenta of 15 to 100 GeV/c.

Muon stations

Typically muons traverse the full detector. They do not interact strongly and are massive enough to not emit Bremsstralhung radiation when passing through the calorimeters.

LHCb is excellent on muon triggering, therefore some of the most important LHCb physics measurements contain muons in the final state. LHCb has five muon stations, one upstream (M1) in front of the calorimeters, and four downstream (M2, M3, M4, M5).

M1 mission is to improve the muon momentum measurement resolution in the first stage of the trigger. It consists of 1368 multi-wire proportional chambers and 12 sets of three gas electron multiplier foils in the region closest to the beam pipe where the particle flux is highest.

Stations M2 to M5 use the same M1 detection system configuration but interleaved with 80 cm thick iron absorbers. Their information is used to identify and trace penetrating muons both in the online and offline analyses. The detectors are divided into cells. Each

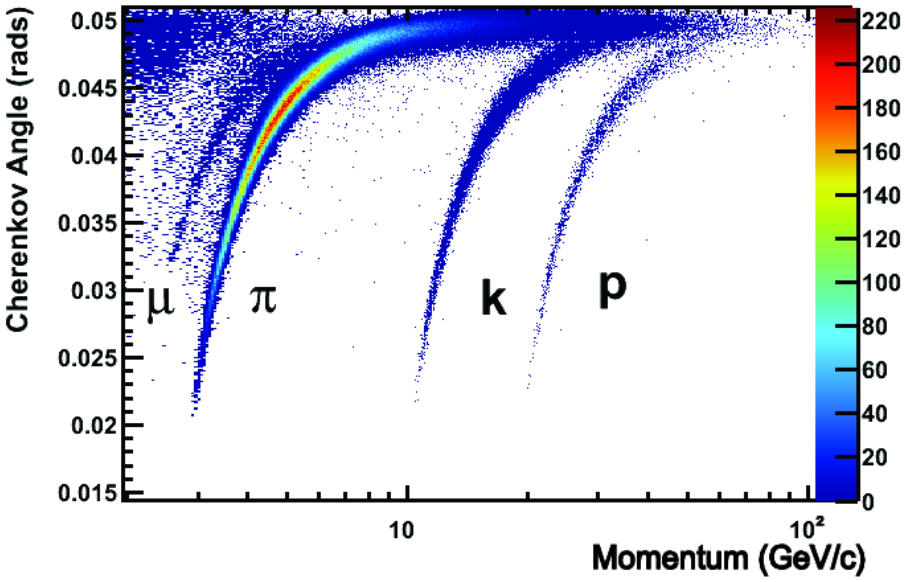

Figure 2.14: Measured angle with respect to the measured particle's momentum used to compute the particle identification likelihood.

cell provides a binary decision to the trigger system, which requires aligned hits above the discriminator threshold in all five stations to fire.

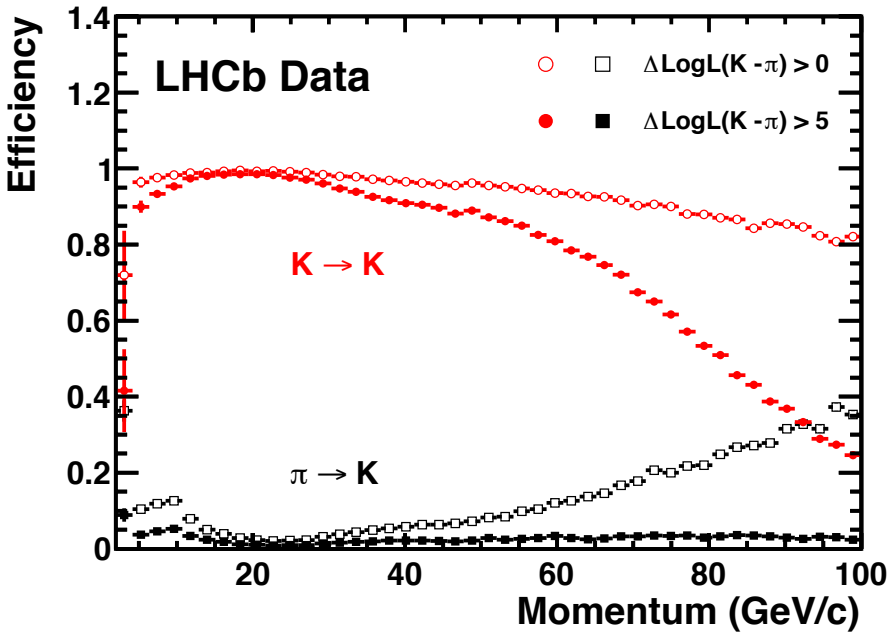

Figure 2.15: Kaon identification efficiency and pion misidentification rate as measured using data as a function of track momentum.

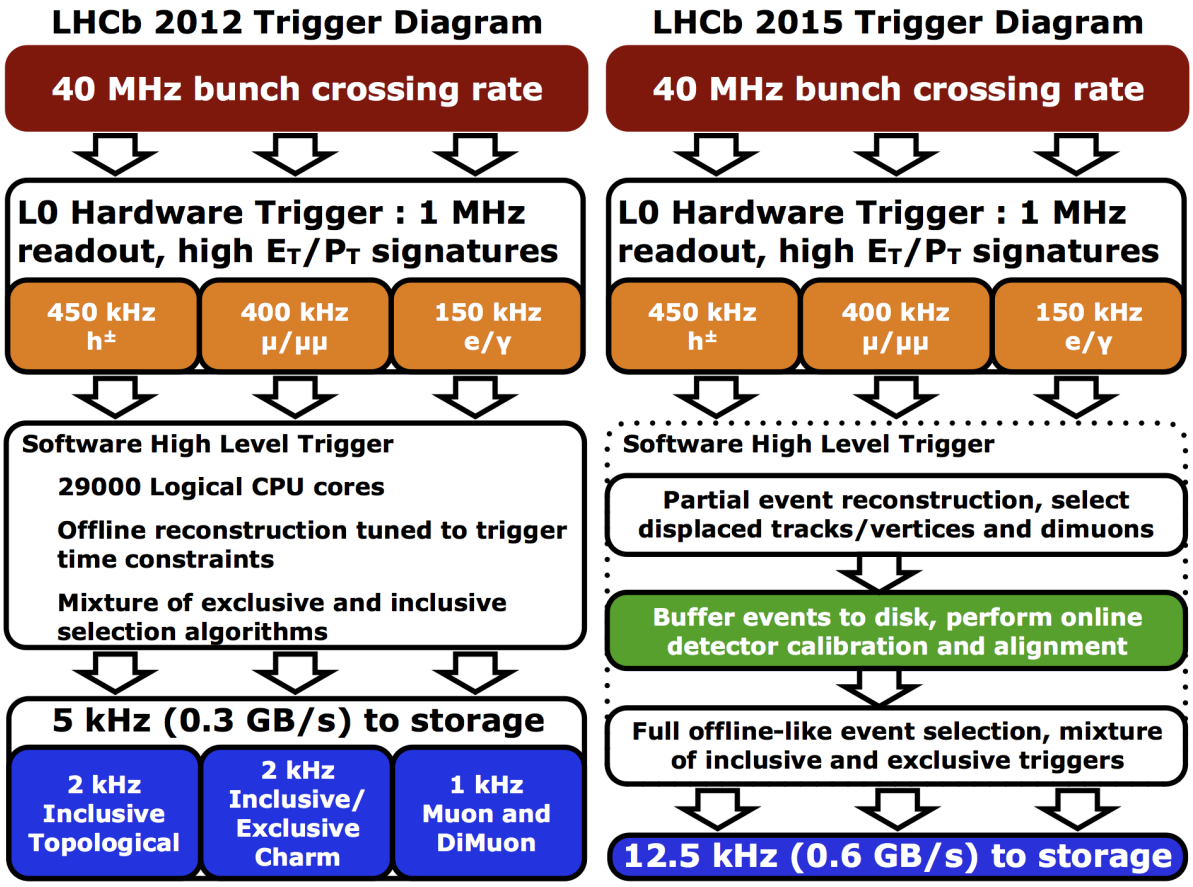

Figure 2.16: Trigger schemes for Run I and Run II data taking periods.

2.3.3 Trigger systems and the stripping process

The role of the trigger system is to collect and reduce the data to accept only interesting events useful for further physics analyses. LHCb has a two-level trigger. The first one is implemented at the hardware level (L0 trigger). The second one is called high level trigger and is implemented at the software level as it uses partial reconstruction (HLT1 and HLT2 triggers).

The LHCb trigger has to deal with a frequency of ∼40 MHz (the LHC bunch crossing frequency). The L0 reduces the data rate to ∼1 MHz and the high level trigger to 5 kHz in Run I and 12.5 kHz in Run II. Figure 2.16 contains the trigger schemes for Run I and Run II data taking periods.

Hardware trigger: L0

Via read-out boards which digitize the subdetector's signals, the L0 trigger receives input from 24 high-speed optical links from two subdetectors at 40 MHz: the calorimeters, which trigger on electrons, photons or hadrons, and the muon system. To reject events with multiple interactions in one bunch crossing, a pile-up veto is also used using VELO information. The L0 gives permission or not to read out the rest of the detector 4 μs after

an interaction.

The L0 trigger is designed to select preferentially events with large transverse energy, or events which have a muon carrying large transverse momentum. The calorimeter information to the trigger is the transverse energy of 2x2 groups of cells. The muon momentum is measured by trying to align the hits in the five muon stations to form a track. Finally, the event is accepted if it fulfills the thresholds in energy or muon momentum imposed by the L0 trigger.

High level trigger: HLT1 and HLT2

The high level trigger searches for events that have a displaced secondary vertex from decaying long-lived particles, which is the case for B-mesons. For this reason, a certain level of online reconstruction is needed. The HLT1 and HLT2 are implemented as software applications running on the online event filter farm of around 30 k CPU where each event can be processed in 30 ms. HLT1 runs in real-time and writes events to the local hard-drives of the farm machines, while HLT2 uses the rest of the available CPU (100% when there is no beam) to process the events written by HLT1. Events accepted by HLT2 are sent to the final storage.

The HLT1 starts with the reconstruction of primary vertices and tracks. The tracks from the muon stations are matched with the ones in the VELO. Two types of tracks are then identified: those which are detached from the PV, and those which can be matched to track segments in the muon detectors. These requirements discriminate around 12 tracks per event for further processing. The chosen tracks are then extrapolated to the tracker stations with a specific track finding algorithm. A track quality algorithm using a Kalman filter [61] is applied to the successfully reconstructed tracks. Finally, requirements on different observables are applied to select the interesting events. Such observables are typically track's transverse momentum and impact parameters with respect to the PV.

The HLT2 refines the track fit for events selected by the HLT1. Subsequently, it performs a fine search using two kinds of selections. On the one hand, it looks for c and b-hadrons produced in the PV which decay in a handful of specific modes. On the other hand the so-called *topologial trigger* is used. The topological trigger looks for multi-track decay vertices detached from the PV and consistent with coming from b-hadrons. All events selected by the HLT2 are written permanently on tape.

Stripping process

The stripping process is applied after the complete offline reconstruction. It is the last data management step before having the data ready for the offline physics analyses. During the process of stripping, the recorded data is separated into *streams*, each of them related with the physics topics to study. The final step is to divide the data streams in *stripping lines* which are exclusive for each (or among very similar) physical analyses in which the candidates are reconstructed. The stripping process is intended to save space, processing time and speed up data access to the analysts.

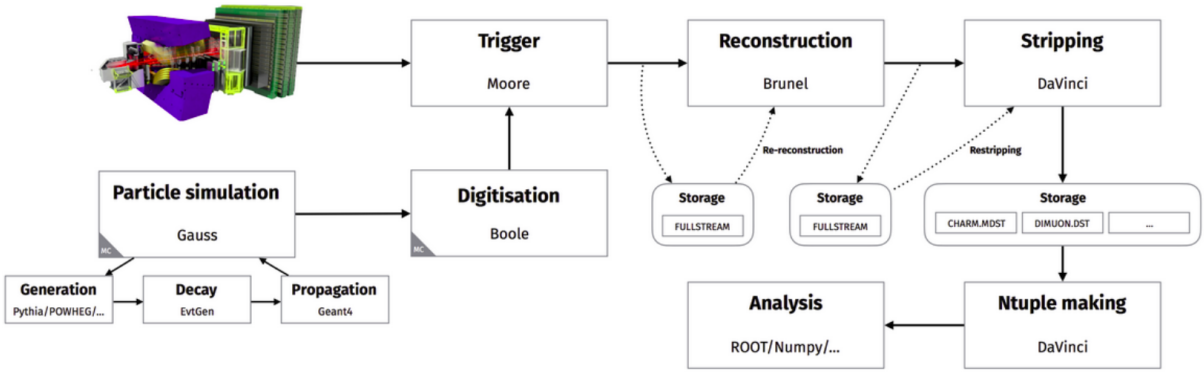

Figure 2.17: LHCb data flow and the software implemented at each step.

2.3.4 Simulation

One of the purposes of the simulation is to reproduce the experimental data taking conditions to understand the data and the detector performance. In other words, one needs to reproduce the *pp* collisions but also the detector's response. Figure 2.17 contains the data flow, either real or simulated, in the LHCb DAQ system.

The structure of high energy collision events and their detection is complex and it cannot be computed from first principles. Simulated events are built using Monte Carlo (MC) generators. MC generators deal with a wide range of physics effects to reproduce properly the *pp* collisions. The main generator packages used in LHCb are `Pythia` [62] [63] to generate *pp* interactions up to hadronization, and `EvtGen` [64] to generate the decay and evolution of all particles. Another fact is that prior to the permanent storage of the simulated event samples, some specific requirements are applied in order to save space and computing time. These requirements are referred as the *generator level cuts*.

To simulate the detector, MC generators are also used. They describe how the different particles traverse the experimental setup. They simulate the geometry of the detector, the tracking through materials, the hit creation, the showers in the calorimeters and the Cherenkov light in the RICH sub-detectors among others. The software package used to simulate the LHCb detector is `Geant4` [65].

Chapter 3

Experimental analysis: Search for the LFV decays $B_s^0 \to \tau^{\pm}\mu^{\mp}$ and $B^0 \to \tau^{\pm}\mu^{\mp}$

3.1 Introduction

The purpose of this analysis is to measure the Branching Ratio ($\mathcal{B}$) of the LFV processes $B^0 \to \tau^{\pm}\mu^{\mp}$ and $B_s^0 \to \tau^{\pm}\mu^{\mp}$, e.g. the fraction of times a $B^0_{(s)}$ meson decays into $\tau\mu$. The Branching Ratio of a process is defined as follows:

$$\mathcal{B}\left(B_Y \to X\right) = \frac{N_{sig}^{obs}}{2\sigma_{b\bar{b}} f_{B_Y} \epsilon_{sig} \int \mathcal{L}dt}. \tag{3.1}$$

Where the external parameters in black are:

- $\sigma_{b\bar{b}}$ is the $b\bar{b}$ pair cross section production in proton-proton collisions,
- f_{B_Y} is the ratio of the fragmentation fractions which quantifies the rate of b quarks hadronizing into a B_Y meson,
- $\int \mathcal{L}dt$ is the time integrated luminosity and quantifies the total recorded data.

The parameters (in red) which need to be measured are:

- N_{sig}^{obs} the observed signal yield, e.g. the number of observed signal events.
- ϵ_{sig} the signal efficiency, which quantifies the signal events lost due to the measurement process.

The branching ratio measurements typically use another auxiliary decay mode named as *normalization channel*. A normalization channel is a physics process which has a precisely measured branching ratio which is used to cancel uncertainties and thus increase the precision of the measurement. The normalization channel final states are typically very close to the signal channel ones and its data have been taken in the same conditions than the data of the signal sample. In particular it allows the cancellation of $\sigma_{b\bar{b}}$ and $\int \mathcal{L}dt$ which bear large uncertainties. It can help to cancel some particular component of the signal efficiency as well.

The goal is to get the highest ϵ_{sig} and the most precise N_{sig}^{obs} while rejecting the background events, namely the events corresponding to other decay processes and polluting the data sample.

To disentangle signal and background, a *selection process* is used and it is based on variables differentiating signal from background events. Signal events are usually simulated from MC methods and, once the selection has been implemented, ϵ_{sig} is extracted. The discriminating variables within the selection must be correctly simulated.

From the selected data, N_{sig}^{obs} is generally obtained by modeling the signal and background components in the most discriminant variable(s). It is known as a *fit process*.

A key part of the measurement is the handling of *uncertainties*, as the significance of the measurement depends on them. The uncertainties on N_{sig}^{obs} and ϵ_{sig} are divided into two categories:

- The *statistical uncertainty* (stat) only depends on the amount of data used in the measurement and is described by Poisson statistics. The higher the amount of data is used, the lower the statistical uncertainty.

- The *systematic uncertainty* (syst) quantifies all the unknowns of the measurement, like for example the effects of the resolution of the sub-detectors and the precision of the techniques used through the measurement.

The searches for BSM physics, at least up to now, have not found signal events in the selected data. In these cases, an *upper limit* on the branching ratio is set. The upper limit value depends on the sensitivity of the measurement.

3.1.1 Description of the BaBar analysis

Prior to detail the LHCb Run I $B^0_{(s)} \to \tau^{\pm}\mu^{\mp}$ analysis, few lines are dedicated to explain the previous analysis on the same channel. The BaBar collaboration published on 2008 the *search for the decays* $B^0 \to l^{\pm}\tau^{\pm}$ *and* $B^+ \to l^+\nu$ $(l = e, \mu)$ *using hadronic tag reconstruction* [11]. Our attention is set in the $B^0 \to \mu^{\pm}\tau^{\pm}$ part. The analysis is performed on a data sample corresponding to 342 fb^{-1} of integrated luminosity recorded by the BaBar detector from the e^+e^- collisions delivered by the PEP-II accelerator.

The difficulty of studying channels involving τ leptons are the undetectable (multiple) neutrinos resulting from the τ decay, and its consequent loss of information. The 6 τ decay modes considered in the analysis are: $\tau^- \to e^-\overline{\nu}_e\nu_\tau$, $\tau^- \to \mu^-\overline{\nu}_\mu\nu_\tau$, $\tau^- \to \pi^-\nu_\tau$, $\tau^- \to \pi^-\pi^0\nu_\tau$, $\tau^- \to \pi^-\pi^0\pi^0\nu_\tau$ and $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$.

The electrons and positrons are fundamental particles, therefore the e^+e^- colliders have precise control of the collisions energy which helps to constraint the missing neutrinos momenta. BaBar analysts introduce in this search a technique which further helps to deduce the information of the missing neutrinos: the *hadronic tag reconstruction*. The $B\overline{B}$ mesons are produced in pairs, being one of the B mesons the subject of the decay search. What the hadronic tag reconstuction does, is to reconstruct the accompanying B in a handful of hadronic modes, $B^{0(+)} \to D^{(*)}(\to X_{\rm had})(\pi^0/\gamma)X'_{\rm had}$, allowing the missing momentum of the neutrino(s) to be fully determined. The resulting increase on the energy resolution provides the extra kinematic handles that permit cleanly distinguish the signal events from the background ones. The dominating background is the so-called *continuum* background of e^+e^- collisions, which consist on $e^+e^- \to ff$ where f represents u, d, s, c

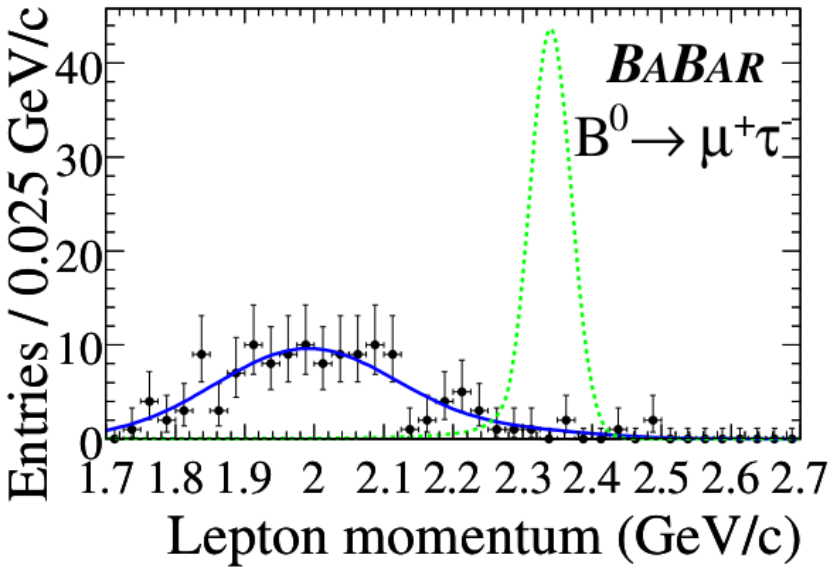

Figure 3.1: $B^0 \to \tau^\pm \mu^\mp$ data fit for the BaBar analysis [11].

or any charged lepton. Performing an offline selection based on kinematic requirements, their obtained signal efficiency, so it can be later compared with the LHCb analysis case, is $(2.7 \pm 0.2) \cdot 10^{-4}$.

The data surviving the selection is fitted with an unbinned maximum likelihood fit on the τ momentum variable distribution. The signal is modelled with a Crystal Ball distribution [66] and the background is modelled with a double Gaussian distribution. The obtained $B^0 \to \mu^\pm \tau^\pm$ signal yield is fully compatible with 0, being 0.01 ± 0.01 observed events. This corresponds to a branching fraction upper limit of $2.2 \cdot 10^{-5}$ at 90% C.L.

3.1.2 LHCb Run I analysis: workflow and analysis strategy

The analysis is performed on data recorded by the LHCb detector during the years 2011 and 2012. The purpose is to perform the measurement of the $\mathcal{B}$ of $B^0_{(s)} \to \tau^\pm \mu^\mp$ processes using the decay mode $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ as normalization channel.

$$\mathcal{B}\left(B^0_{(s)} \to \tau^\pm \mu^\mp\right) \propto \frac{N^{obs}_{B^0_{(s)} \to \tau^\pm \mu^\mp}}{\epsilon^{sig}_{B^0_{(s)} \to \tau^\pm \mu^\mp}} \mathcal{B}_{norm} \frac{\epsilon^{sig}_{norm}}{N^{obs}_{norm}}. \tag{3.2}$$

In absence of signal, as expected in the SM, an upper limit on $\mathcal{B}$ will be set.

The chosen τ decay mode is $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$, where $\mathcal{B}\left(\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu\right) = (9.02 \pm 0.05)\%$. The τ lepton decays via two intermediate resonances ($a^+(1260)$ and $\rho^0(770)$ [13]) which are helpful in the signal selection:

$$\tau^+ \to a^+(1260)\nu \to \rho^0(770)\pi^+\nu \to \pi^+\pi^-\pi^+\nu.$$

The τ decay mode $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu$ with $\mathcal{B}\left(\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu\right) = (4.49 \pm 0.05)\%$ contributes to the amount of signal, although less significantly.

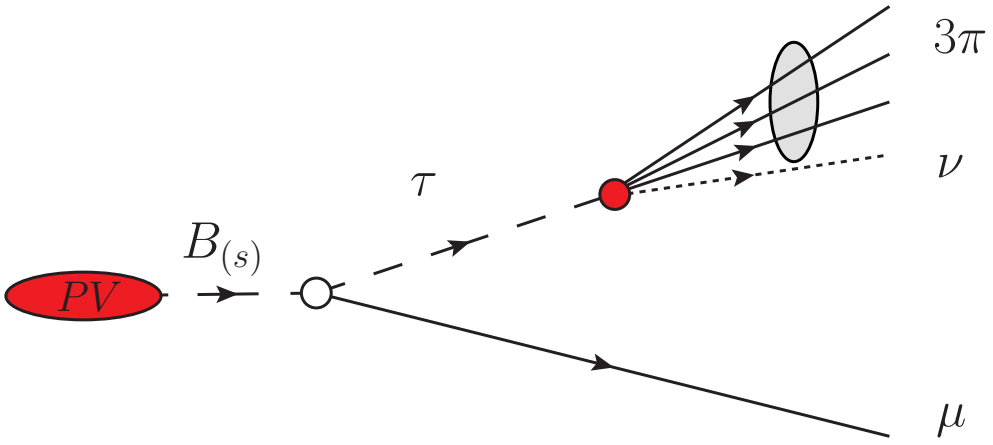

Figure 3.2: $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ process topology.

Figure 3.2 shows the topology of the decay. A $b\bar{b}$ pair is produced by the parton collisions at the Primary Vertex (PV) followed by the hadronization of one of the quarks into a $B^0_{(s)}$ meson. The $B^0_{(s)}$ meson flies until it decays into a μ, which typically traverses all the detector, and a τ, which decays almost inside the Vertex Locator into 3π and a neutrino, which escapes detection. The signal signature presents the following properties:

- 3 pion tracks coming from a common vertex displaced from the PV,
- A muon trajectory not pointing to the PV.

The first step of the analysis is the signal reconstruction (described in Section 3.4). A specific reconstruction technique is used in order to infer the energy of the ν, taking advantage of the known τ vertex position given by the 3π reconstructed vertex. This way, the complete kinematics of the process can be solved up to a two-fold ambiguity. Once the complete reconstruction procedure has been performed, the B invariant mass can be computed. Using the reconstructed B invariant mass with the help of signal simulation samples, a signal region is defined and the data sample is blinded in that region. The blinding technique ensures that the prior prejudices about the final data, i.e. expecting 0 signal, does not affect the techniques used in the analysis and thus preventing a possible biases. The data is then unblinded, only when the analists and the different review committees consider that all analysis steps and involved processes are well understood.

In order to disentangle signal from background, an offline selection consisting of different steps is applied. Data driven and multivariate analysis techniques, such as Boosted Decision Trees (BDT) [14], are used during the selection process. A BDT is a multivariate technique used to optimize the signal and background separation. It generates an output by combining a given set of variables using signal and background *training* samples information. A decision tree (non-boosted) is binary, and it categorizes each event by considering the variables distributions in the training samples. The decision tree has a given number of *nodes* in which it applies a cut in a variable following a given signal background separation criteria. An illustration is shown in Figure 3.3. The *boosting* uses a forest of decision trees as the training is separated in cycles. The basic idea is that the

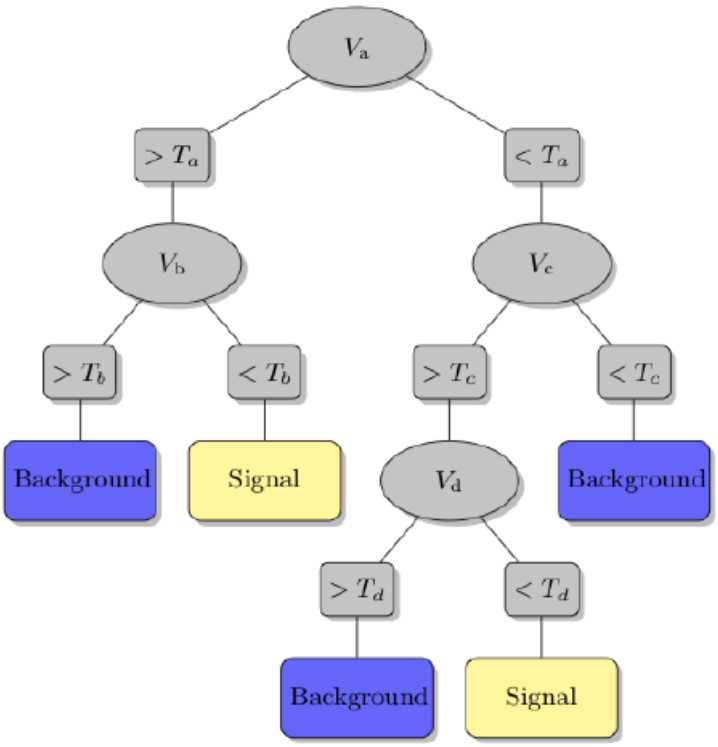

Figure 3.3: Decision tree example scheme. It uses a set of cuts in the variables V_a, V_b, V_c and V_d to separate signal and background.

algorithm teaches itself how to give a certain weight to samples that were mis-classified in a previous learning cycle. The final set of requirements placed on the variables is the combination of the requirements used in the forest of decision trees. The concept is illustrated in Figure 3.4.

First, a preselection is applied at two different levels:

- A cut based preselection (Section 3.5.1) consisting of simple cuts to veto potentially dangerous backgrounds (i.e. $B^0_{(s)} \to D_s\,(\to \pi\pi\pi)\,X$) and on variables providing a high discrimination power between signal and background.

- A BDT based preselection (Section 3.5.2) based on isolation variables, built to provide a high background rejection and reduce the data into a manageable level. The isolation variables look at the presence of undesired tracks in the vicinity of the candidate track and/or vertexes.

Following the preselection, in order to prune the data from identified specific background components, two more selection steps are applied. The first step consists of a BDT targeting the combinatorial background (Section 3.6.1), i.e. background from combining tracks not coming from the same decay chain. In the second step a requirement is placed on the τ decay time to reject a given type of partially reconstructed backgrounds (Section 3.6.2). The partially reconstructed backgrounds are due to missing final states of the decay chain during the reconstruction process.

Once the complete offline selection has been applied, only one partially reconstructed background component survives: partially reconstructed background where the 3π come

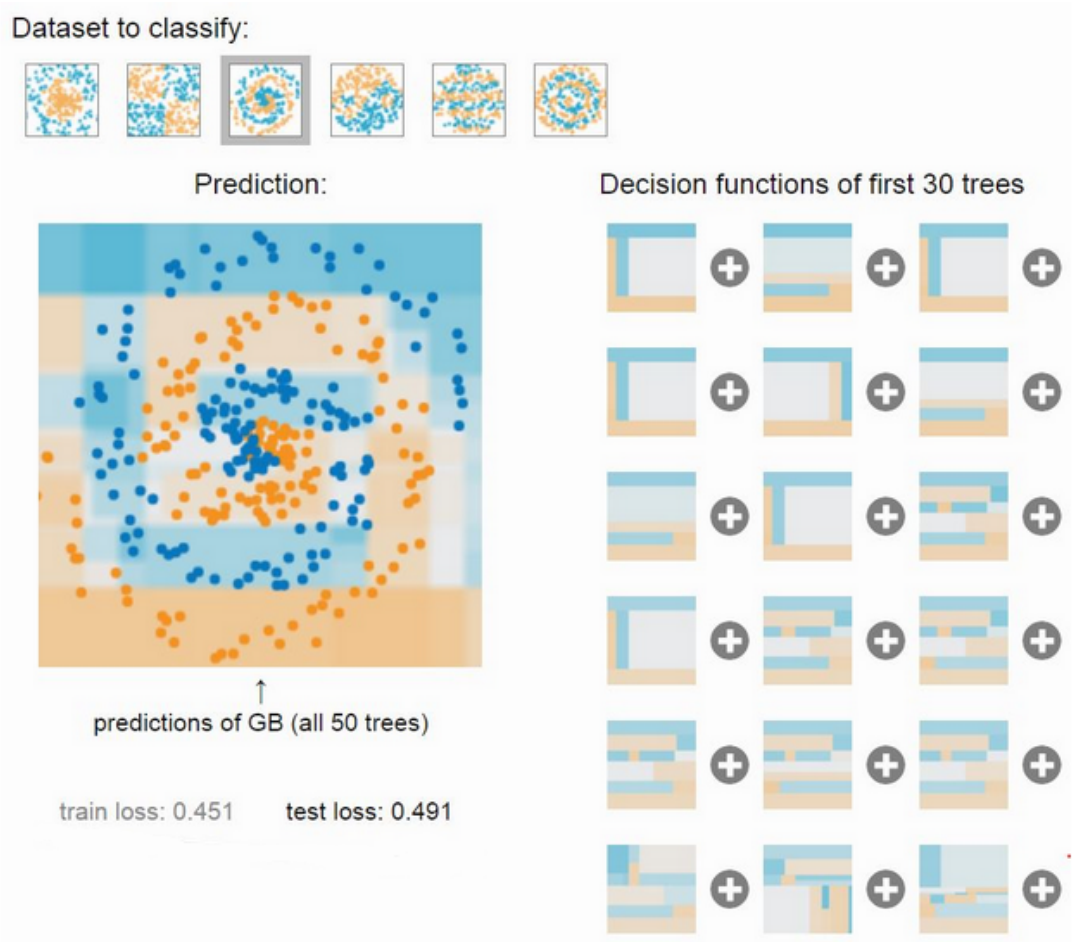

Figure 3.4: Boosted Decision Tree dataset classification example, signal is shown in blue and background in orange. The final classification relies in the optimization "boosting" performed through consecutive decision trees and it is shown on the plot on the left.

from a displaced vertex mimicking the signal signature. No peaking background in the signal region remains, as shown in Section 3.6.3. The signal efficiency is taken from simulation, although corrections are applied using data-driven tools and techniques (Section 3.7).

At this stage of the analysis, and in order to contribute to the fit strategy, a final BDT is trained embedding the remaining discrimination power.

The analysis strategy is completed by a simultaneous fit to the B candidate invariant mass distribution over the different bins of the final BDT (Section 3.8). According to the SM expectations, no signal events should be observed. In this case, the CLs method [15] will be used to extract the upper limits on the branching fractions.

Year	Polarity	$\sqrt{s}$ (TeV)	$\mathcal{L}$ (pb^{-1})
2011	*MagDown*	7	568.06 ± 0.89
2011	*MagUp*	7	422.33 ± 0.74
2012	*MagDown*	8	1026.17 ± 1.68
2012	*MagUp*	8	1033.01 ± 1.68

Table 3.1: Overview of the different data samples and their data taking conditions used in the analysis. The luminosity is calculated from the default luminosity tree provided by *DaVinci*.

3.2 Event samples

3.2.1 Data

The analysis is performed using data corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$ of proton-proton collisions recorded by the LHCb experiment at center-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ during the years 2011 and 2012, respectively. The integrated luminosity used in this analysis, subdivided in terms of data taking conditions, can be found in Table 3.1.

The main samples are formed by candidates selected by the `B2XTau` stripping lines described in Section 3.2.3. Two main kinds of samples are build. On one hand, the Opposite Sign samples (OS), referring to the search data samples where the two daughters of the reconstructed B candidates have opposite charges, are used to build signal and normalization decay candidates. On the other hand, the Same Sign samples (SS) where the two reconstructed B daughters have equal charges, are intended for background studies and have a downscaling factor of 0.5 applied in the stripping.

In addition, a sample of $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ selected by the `BToJpsiK` stripping line is available and it is used on trigger efficiency computations descrived later on in the document.

3.2.2 Simulation

Different Monte Carlo (MC) samples are used to study the properties of signal and background. In order to save computing time, the samples are generated requiring that all particles in the final state be in the LHCb acceptance. In addition, for the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ samples, the following generator cuts are applied: $p_{\text{T}}(\pi) > 250\,\text{MeV}/c$, $P(\pi) > 2\,\text{GeV}/c$ and $p_{\text{T}}(\mu) > 250\,\text{MeV}/c$.

The signal samples correspond to the simulation of the signal events for the main channel $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ and the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$. $B^0_s \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^0\nu)\mu^{\mp}$ events are also available in order to check the contribution of this τ decay channel to the signal efficiency. The B two-body decays are produced following phase-space distributions. The $\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu$ decays are generated by the `Tauola` [67] MC generator using the model tuned on BaBar measurements. The $D^- \to K^+\pi^-\pi^-$ decays are using the `D_DALITZ` model implemented in EvtGen which includes K^* resonances

($K^*(892)$, $K^*(1430)$ and $K^*(1680)$).

In order to study the background properties, several samples of exclusive B decays with at least 3 charged pions and a muon in the final state were produced. The exclusive samples list (in Table 3.6) does not pretend to be exhaustive, it is build to provide qualitative information about the background behavior and categorization.

For some specific studies, the signal candidates are matched to the MC generated particles in order to select only well reconstructed candidates. The truth matching requirements for the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ ($/B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$) mode are :

1. Each signal reconstructed particle must be associated to a MC particle of the same type.

2. The MC particles associated to the reconstructed pions (/or kaons) coming from the τ ($/D$) must share the same τ ($/D$) mother.

3. The MC particle associated to the reconstructed muon (/pion from the B) must share the same B as the τ ($/D$) associated to the mother of the three light hadrons from the τ ($/D$).

Probe muons are used in the trigger efficiency computation and they are taken from simulated events of $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$.

3.2.3 Stripping selection

The $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ (Opposite Sign), $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\pm}$ (Same Sign) and $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ candidates are selected by the `B2XTau_TauMu_TOSLine`, `B2XTau_TauMu_SameSign_TOSLine` and `B2XTau_Dpi_Line` lines, respectively. An overview of the requirements applied on the stripping selection is reported in Table 3.2. $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ candidates are selected by the `BToJpsiK_mmKLine` line.

3.2.4 Trigger requirements

3.2.4.1 $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ channel

The signal candidates are required to satisfy the trigger conditions referenced in Table 3.3. The *TOS* (Trigger On Signal) condition requires that the properties of the tracks composing the signal be enough to fire the specified trigger line is imposed at each level (L0, HLT1 and HLT2). Whereas for HLT2, the signal is defined as the set of particles in the B candidate final state, it is restricted to the muon candidate at the L0 and HLT1 levels. The HLT2 trigger requirements are applied at the stripping level.

3.2.4.2 $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ channel

The candidates for the normalization mode are required to pass the *TOS* conditions presented in Table 3.3 where the signal is the B candidate.

3.2.5 Summary

The statistics available for the samples used in this analysis are summarized in Table 3.4 for the data and in Table 3.5 for the simulated signal, as well as for the control and normalization channels. In Table 3.6 are reported the statistics for the exclusive backgrounds.

For each exclusive mode, the yield corresponding to an integrated luminosity of $\mathcal{L}_{int} = 3\,\text{fb}^{-1}$ is computed using the following formula:

$$\text{Y} = 2\mathcal{B}\sigma_{b\bar{b}} f_x \mathcal{L}_{int} \frac{\epsilon^{\text{strip}} \epsilon^{\text{gen}}}{n_{\text{gen}}}, \tag{3.3}$$

where $\mathcal{B}$ is the Branching Ratio of the corresponding decay, f_x the fragmentation fraction ($x = \{d, s, u\}$), ϵ^{strip} and ϵ^{gen} the stripping and generator efficiencies and n_{gen} the number of generated events in the corresponding sample.

		$B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$		$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$
cut	on	value	on	value
P	π	$> 2000\,\text{MeV}/c$	π/K	$> 2000\,\text{MeV}/c$
PT		$> 250\,\text{MeV}/c$		$> 250\,\text{MeV}/c$
MIPCHI2DV		> 16		> 16
TRCHI2DOF		< 3		< 3
TRGHOSTPROB		< 0.3		< 0.3
PROBNNpi		> 0.55	π	> 0.55
PIDK		-	K	> -5
P	μ	$> 6000\,\text{MeV}/c$	π	$> 2000\,\text{MeV}/c$
PT		$> 1000\,\text{MeV}/c$		$> 250\,\text{MeV}/c$
MIPCHI2DV		> 16		> 16
TRCHI2DOF		< 3		< 3
TRGHOSTPROB		< 0.3		< 0.3
PIDmu		> 0		-
hasMuon		yes		-
PT	B	$> 5000\,\text{MeV}/c$	B	$> 5000\,\text{MeV}/c$
M		$[2000, 7000]\,\text{MeV}/c^2$		$[2000, 7000]\,\text{MeV}/c^2$
MCORR		$< 10000\,\text{MeV}/c^2$		$< 10000\,\text{MeV}/c^2$
FDCHI2 (τ/D)		< 4000		< 4000
IPCHI2 (μ/π)		< 200		< 200
BPVVD		$< 35\,\text{mm}$		$< 35\,\text{mm}$
$\text{PT}_{(\mu+3\pi/\pi+3\pi)}$		$> 2500\,\text{MeV}/c$		$> 2500\,\text{MeV}/c$
childs VCHI2 ($\tau\mu/D\pi$)		< 12		< 12
childs MIPCHI2DV ($\tau\mu/D\pi$)		> 50		> 50
M	τ	$[400, 2100]\,\text{MeV}/c^2$	D	$[1750, 2080]\,\text{MeV}/c^2$
At least 1 daughter w. PT		$> 800\,\text{MeV}/c$		$> 800\,\text{MeV}/c$
AMAXDOCA		$> 0.2\,\text{mm}$		$> 0.2\,\text{mm}$
PT		$> 1000\,\text{MeV}/c$		$> 1000\,\text{MeV}/c$
BPVDIRA		> 0.99		> 0.99
VCHI2		< 16		< 16
FDCHI2		> 16		> 16
VDRHO		$[0.1, 7]\,\text{mm}$		$[0.1, 7]\,\text{mm}$
VDZ		$5\,\text{mm}$		$5\,\text{mm}$
ABSID daughter		π		π or K
HLT2 (Topo*BodyBBDT/ TopoMu/SingleMuon)	B	yes		no
hasMuon		yes		-

Table 3.2: Stripping requirements for $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ and the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ as implemented in StrippingB2XTau v21 and v21r1.

$B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ channel	
L0	μ `L0Muon_TOS`
HLT1	μ `TrackMuon_TOS` or μ `SingleMuonHighPT_TOS`
HLT2	B `TopoMu[2/3/4]BodyBBDT_TOS`
$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ channel	
L0	B `L0Hadron_TOS`
HLT1	B `TrackAllL0_TOS`
HLT2	B `Topo[2/3/4]BodyBBDT_TOS`

Table 3.3: Trigger lines for the signal and normalisation modes.

Sample	Year	Strip. cand.	Trig. cand
$B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ channel			
`B2XTau_TauMu_TOSLine` (OS)	2011/2012	24184600	-
`B2XTau_TauMu_SameSign_TOSLine` (SS)	2011/2012	4820815	3429586
$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ channel			
`B2XTau_Dpi_Line`	2011/2012	1836019	447658
$B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ channel			
`BToJpsiK_mmKLine`	2011/2012	24930723	-

Table 3.4: Used data samples.

Sample	Year	Evt.type	Sim. version	Gen.	Strip.	trig.
$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$	2011	13110004	Sim09b/Reco14c	615387	14183	11381
$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$	2012	13110004	Sim09b/Reco14c	1098681	24707	19899
$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$	2011	11110004	Sim09b/Reco14c	626085	14246	11297
$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$	2012	11110004	Sim09b/Reco14c	1013159	22128	17741
$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$	2011	13110408	Sim09c/Reco14c	714521	14688	11805
$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$	2012	13110408	Sim09c/Reco14c	909887	17810	14386
$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$	2012	11110408	Sim09c/Reco14c	438007	8698	6994
$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$	2012	11110408	Sim09c/Reco14c	1065137	20603	16780
$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$	2011	11264001	Sim09b/Reco14c	506999	8782	1699
$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$	2012	11264001	Sim08e/Reco14c	1026667	16233	2874
$B^\pm \to J/\psi\,(\to \mu^\pm\mu^\mp)K^\pm$	2011	12143001	Sim09a/Reco14c	9497433	2901523	-
$B^\pm \to J/\psi\,(\to \mu^\pm\mu^\mp)K^\pm$	2012	12143001	Sim08e/Reco14a	7421329	2037148	-

Table 3.5: Simulated MC signal samples for the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$, $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ and $B^\pm \to J/\psi\,(\to \mu^\pm\mu^\mp)K^\pm$ channels.

	Sample	Evt.type	Gen.	Strip.	Trig.	Exp. Yield
$D(\mu)3\pi$ type	$B^0 \to (D^- \to \mu^- \overline{\nu}_\mu)\pi^+\pi^-\pi^+$	11872010	411998	2778	2482	806
	$B_s^0 \to (D_s^- \to \mu^- \overline{\nu}_\mu)\pi^+\pi^-\pi^+$	13272000	414500	2954	2683	6630
	$B^0 \to (D^*(2010)^- \to (D^- \to \mu^- \overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+$	11872400	414998	2776	2444	262
	$B^0 \to (D^*(2010)^- \to (D^- \to \mu^- \overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+\pi^0$	1774400	414500	1497	1301	320
	$B^0 \to (D^- \to K^0\mu^- \overline{\nu}_\mu)\pi^+\pi^-\pi^+$	11872000	829998	3238	2788	73056
$D(3\pi)\mu$ type	$B^0 \to (D^*(2010)^- \to (D^- \to \pi^+\pi^-\pi^-\pi^0)\pi^0)\mu^+\nu_\mu$	11772410	832995	6395	5598	47335
	$B^0 \to (D^- \to \pi^-\pi^+\pi^-\pi^0)\mu^+\nu_\mu$	11772400	832998	6689	5979	85046
	$B^0 \to (D^{-(*)} \to K^-\pi^+\pi^-)X\mu^+\nu_\mu$	11874042	11874042	21757	18486	144310
	$B^0 \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)\mu^+\nu_\mu$	11774410	838748	4291	3670	13907
	$B^+ \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)\mu^+\nu_\mu$	12673400	837995	4441	3826	4406
$D(3\pi)\tau(\mu)$ type	$B^0 \to (D^- \to \pi^-\pi^+\pi^-\pi^0)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	11572400	827497	2188	1854	1975
	$B^0 \to (D^*(2010)^- \to (D^- \to \pi^-\pi^+\pi^-\pi^0)\pi^0)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	11572410	414000	983	812	933
	$B^+ \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	12673410	824496	1691	1347	405
$D(\mu)\tau(3\pi)$ type	$B^0 \to (D^- \to \mu^-\overline{\nu}_\mu)(\tau^+ \to \pi^+\pi^+\pi^-\overline{\nu}_\tau)\nu_\tau$	11574000	413250	1273	1058	50
	$B^+ \to (D^{**} \to (D^- \to \mu^-\overline{\nu}_\mu)X)(\tau^+ \to \pi^+\pi^-\pi^+\overline{\nu}_\tau)\nu_\tau$	12675000	834995	3173	2515	162
$D(\tau(3\pi))\mu$ type	$B_s^0 \to (D_s^- \to (\tau^- \to \pi^-\pi^-\pi^+\nu_\tau)\overline{\nu}_\tau)\mu^+\nu_\mu$	13574000	417749	3197	2815	17064

Table 3.6: Simulated exclusive MC samples. All samples are produced with 2012 data taking conditions. The expected yield corresponds to an integrated luminosity of $\int \mathcal{L} = 3\,\text{fb}^{-1}$ and it is computed using Eq. 3.3.

3.3 Normalization channel

The mode $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ has a topology very similar to that of the signal, i.e. same number of tracks in the final state and three light hadrons coming from a displaced vertex. It is used as a normalization channel for the estimation of the Branching Ratio of the $B^0_{(s)} \to \tau^\pm\mu^\mp$ modes. The $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ event selection is described in Section 3.3.1 and the extraction of the event yield is presented in Section 3.3.2. The $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ mode is also used as a control channel to check the agreement of the variables used in this analysis between data and simulation, reported in the Section 3.7.4.

3.3.1 Event selection

The stripping and trigger requirements used for the normalization channel are shown in Tables 3.2 and 3.3 respectively. To remove most of the remaining background, the used stripping requirement on the kaon PID has been tightened and other requirements are applied on kinematic variables. They are listed in Table 3.7. The offline selection efficiency of the different requirements on MC and data is reported in Table 3.8. The signal and background separation is illustrated in Figure 3.5.

Variable	Associated to	Value
PID requirements		
PID K	K from D	> 15
Kinematic requirements		
MINIP	B	$< 0.05\,\text{mm}$
DiraAngle		< 0.0085
ENDVERTEX χ^2		< 5
FD OWNPV		> 2
PT		$> 5050\,\text{MeV}/c$
TAU χ^2		< 10
DOCA		< 0.05
M	D	$[1850, 1900]\,\text{MeV}/c^2$
ENDVERTEX χ^2		< 6
FD χ^2 OWNPV		> 200
ORIVX		< 5
PT	K from D	$> 750\,\text{MeV}/c$
MINIP	π's from D	$> 0.2\,\text{mm}$

Table 3.7: Offline selection requirements for the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$.

Requirements on	Efficiency (%)		
	MC 2011	MC 2012	Data
PID	52.51 ± 0.49	57.64 ± 0.36	5.75 ± 0.17
B	68.64 ± 0.46	68.37 ± 0.34	6.24 ± 0.18
D	61.73 ± 0.48	61.38 ± 0.36	4.74 ± 0.16
π and K	56.65 ± 0.49	58.85 ± 0.36	14.91 ± 0.26

Table 3.8: $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ offline selection efficiency of the different requirements for MC signal and data samples. The efficiencies are computed sequentially, the quoted value is the efficiency of the corresponding requirement once the previous ones have been applied. The first efficiency is computed on the triggered and stripped events.

Parameter	Value
Mean	5281.0 ± 0.4
Width	15.1 ± 0.4
$\alpha 1$	1.2 ± 0.2
n1	40 ± 36
$\alpha 2$	-1.0 ± 0.3
n2	39 ± 34

Table 3.9: Parameters of the of the two CB functions composing the signal PDF in the fit for the signal MC simulation of the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$.

3.3.2 Event yield

The B invariant mass distribution after the selection process is fitted using an unbinned maximum likelihood fit. In an unbinned fit the data is modelled using Probability Density Functions (PDF). The unbinned fit finds the optimal PDF's parameters best describing the data. To do so, in this analysis, the `Minuit` algorithm [68] is used. The sample is evaluated on a event per event basis to construct a *likelihood*:

$$L = \prod_{i}^{\text{events}} PDF(\text{event}_i). \tag{3.4}$$

The likelihood is evaluated in different steps by changing the PDF parameters in order to maximize $-\log(L)$.

The fit is performed in the B invariant mass range $[5200, 5400]\,\text{MeV}/c^2$.

The signal is modelled by a the sum of two Crystal Ball functions (CB) [66] that share common mean and width and have independent tail parameters. One CB is accounting for the tail at low mass and the other one at high mass. The fit on MC is shown in Figure 3.6 and the fitted parameters are reported in Table 3.9.

The $D^{\pm}\pi^{\mp}$ invariant mass distribution in Opposite-Sign data is modeled with the same function as the simulation (sum of two CB), and an exponential distribution accounting

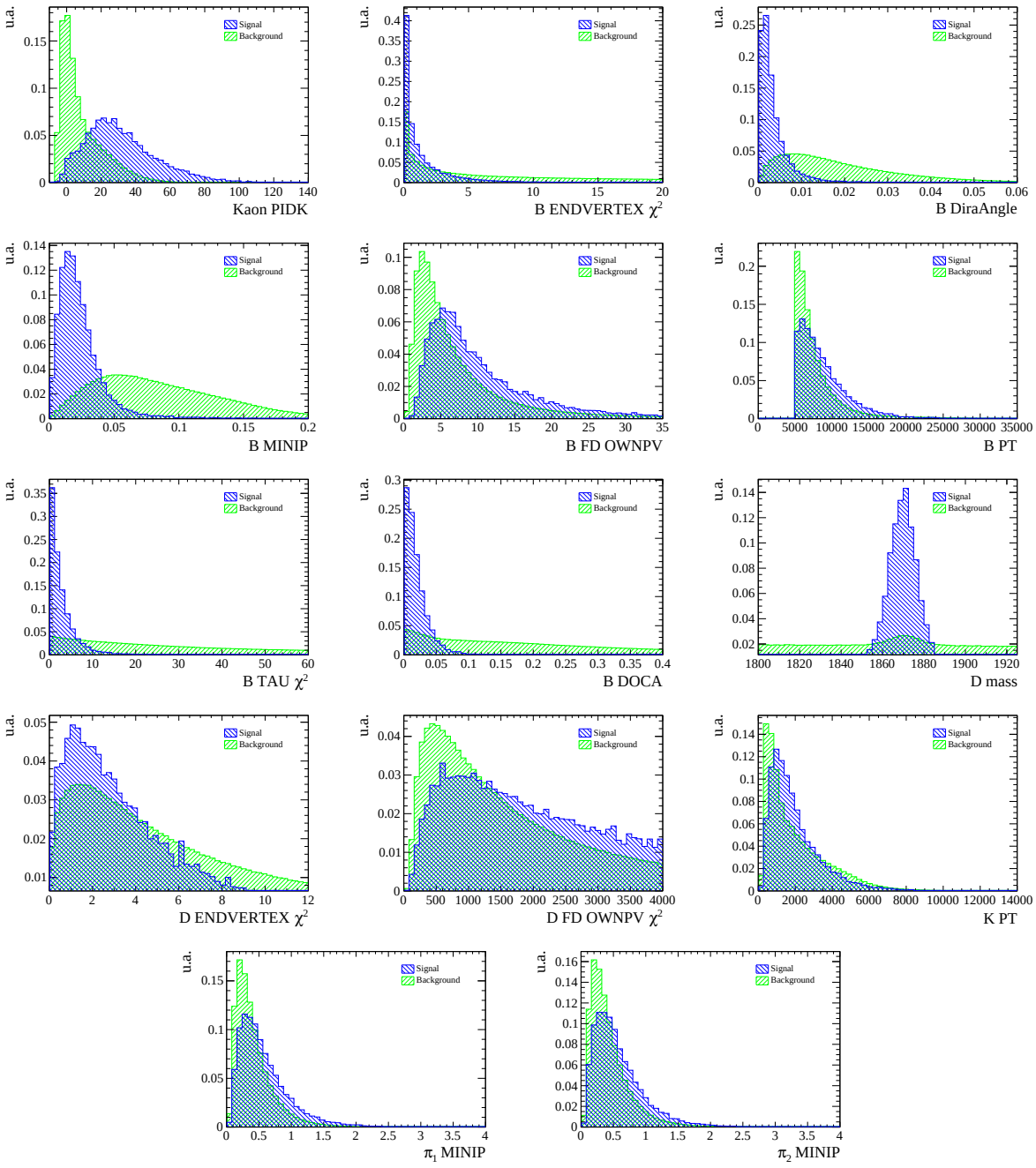

Figure 3.5: Variables distributions used in the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ offline selection signal and background separation. Signal is drawn in blue and background (Same Sign data) in green.

for the combinatorial background. The $n1$ and $n2$ tail parameters of the signal PDF are fixed to the values obtained with the fit to the simulation.

The fit result is shown in Figure 3.7 and the parameters of the total PDF are reported in Table 3.10. The yield of the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ mode is: 22588 ± 176 events.

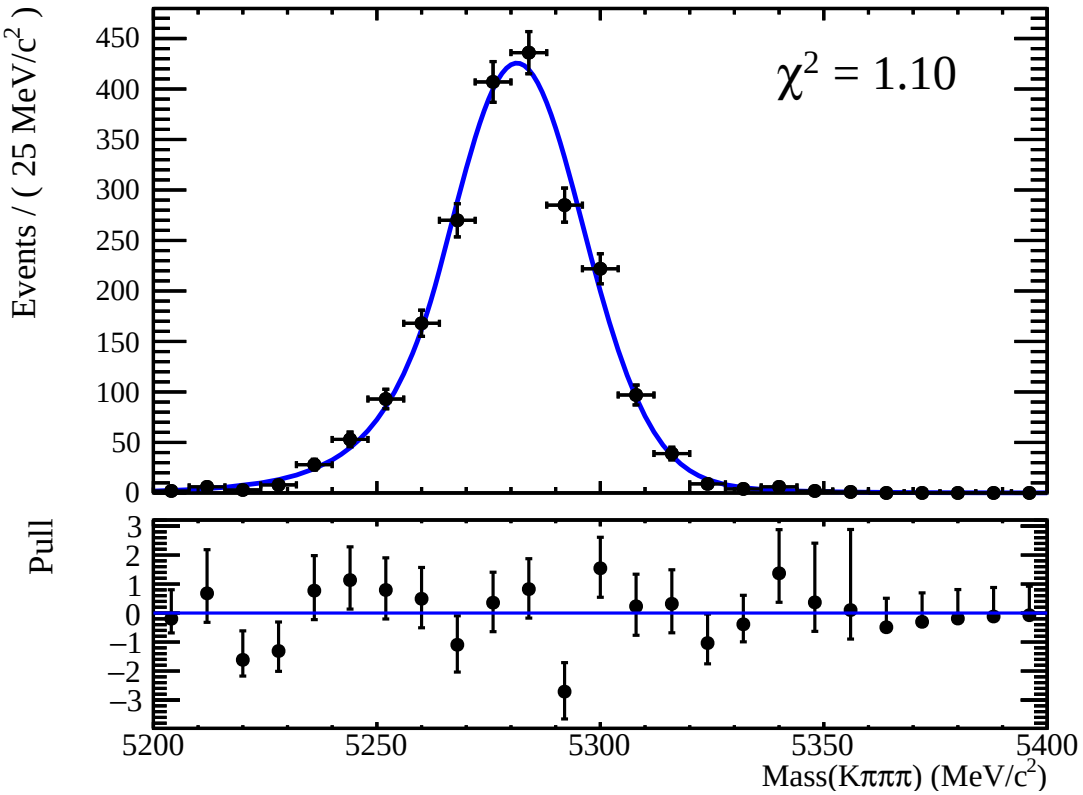

Figure 3.6: $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ 2011 and 2012 Monte-Carlo signal simulation fit after the offline selection requirements.

Parameter	Value
Crystal ball PDF	
Yield	22588 ± 176
Mean	5284.3 ± 1.2
Width	17.4 ± 0.2
$\alpha 1$	1.1 ± 0.3
n1	40
$\alpha 2$	-1.2 ± 0.4
n2	39
Exponential PDF	
Yield	559 ± 95
Exponent	$(-1.2 \pm 0.2) \cdot 10^{-2}$

Table 3.10: $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ parameters of the total PDF (signal + background) coming from the fit to the B invariant mass distribution of Opposite Sign data.

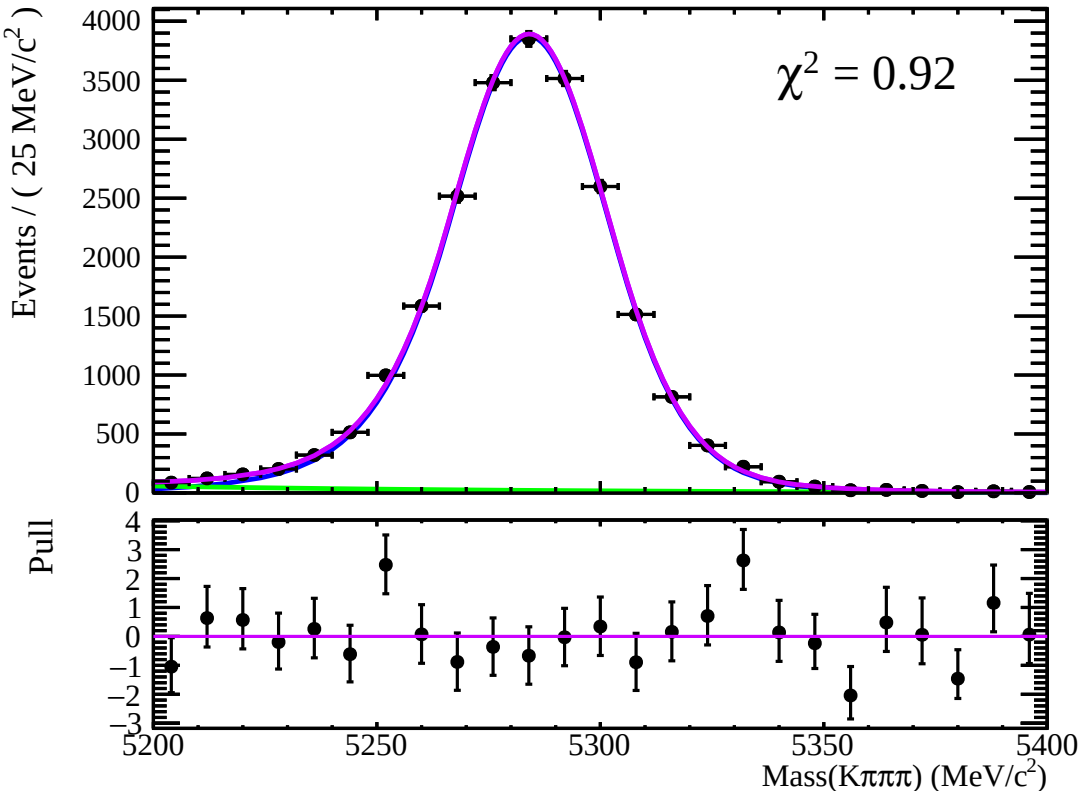

Figure 3.7: $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ Opposite Sign data fit on the B invariant mass after the offline selection requirements. The total PDF is represented in purple, the signal in blue, and the background in green.

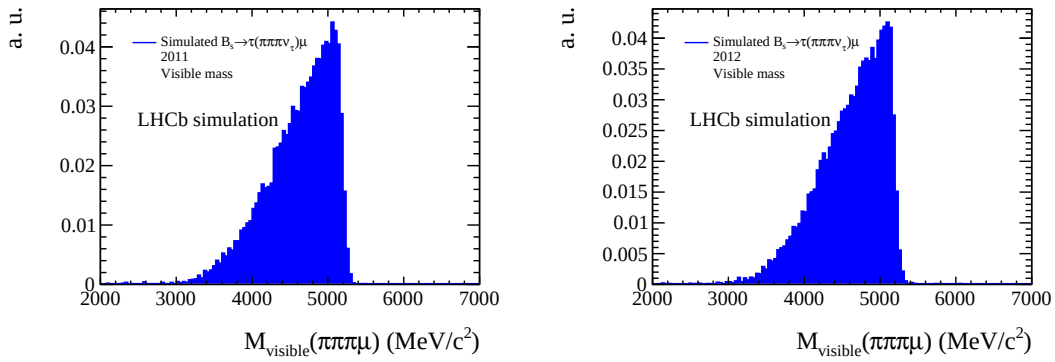

Figure 3.8: $B_s^0 \to \tau^\pm\mu^\mp$ MC signal: distributions of the visible mass (M_{vis}) for 2011 and 2012 simulations.

3.4 B invariant mass computation

As the neutrino from the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ decay escapes detection, the visible B invariant mass (M_{vis}) computed using the reconstructed 4-momenta of the three pions and the muon does not peak at the measured B mass value (Figure 3.8). However, using the reconstructed primary vertex, the τ decay vertex and the known τ mass, there are

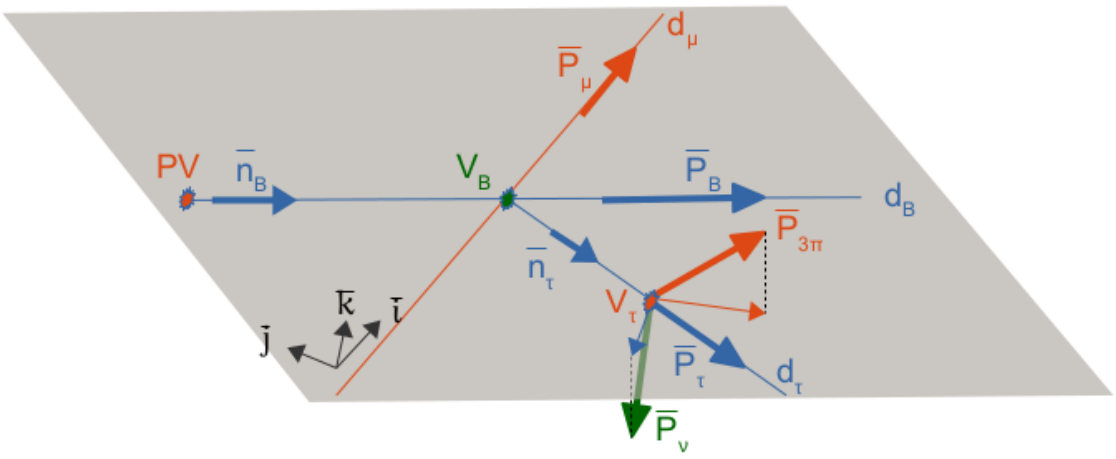

Figure 3.9: $B_s^0 \to \tau^\pm \mu^\mp$ Decay plane given by the μ momentum and the B flight direction ($\hat{n}_B$).

enough constraints to determine the neutrino momentum and thus the B invariant mass up to a 2-fold ambiguity.

3.4.1 Analytic reconstruction of the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ decay

The B candidate decay vertex position (V_B) and the neutrino 4-momentum (P_ν) are not measured experimentally, leaving 6 degrees of freedom in the description of the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ decay.

These 6 unknown quantities (V_B and P_ν components) can be determined using the following constraints (Figure 3.9):

1. The ν momentum component in the direction orthogonal to the decay plane defined by the B two-body decay balances the one of the momentum of the 3 pions system,
2. The B decay vertex is on the muon trajectory.
3. The momentum of the 3 pions and neutrino system is aligned with the τ direction given by the B and τ decay vertices positions.
4. The momentum of the 3 pions, neutrino and muon system is aligned with the B direction given by the primary vertex and the B decay vertex positions.
5. The invariant mass of the 3 pions and neutrino system adds up to the true τ mass.

Because the last requirement leads to a second order equation, there are in general two solutions for the neutrino momentum and these solutions are not always real (they may contain an imaginary part in case of negative discriminant).

The B mass can then be determined with a 2-fold ambiguity using the two neutrino momenta satisfying the constraints. These computed B masses are labeled $M_B^\oplus$ and $M_B^\ominus$ depending on whether the discriminant is added or subtracted when deriving the neutrino momentum. The detailed computation is given in Appendix A. The $M_B^\oplus$ and $M_B^\ominus$ distributions are shown in Figure 3.10.

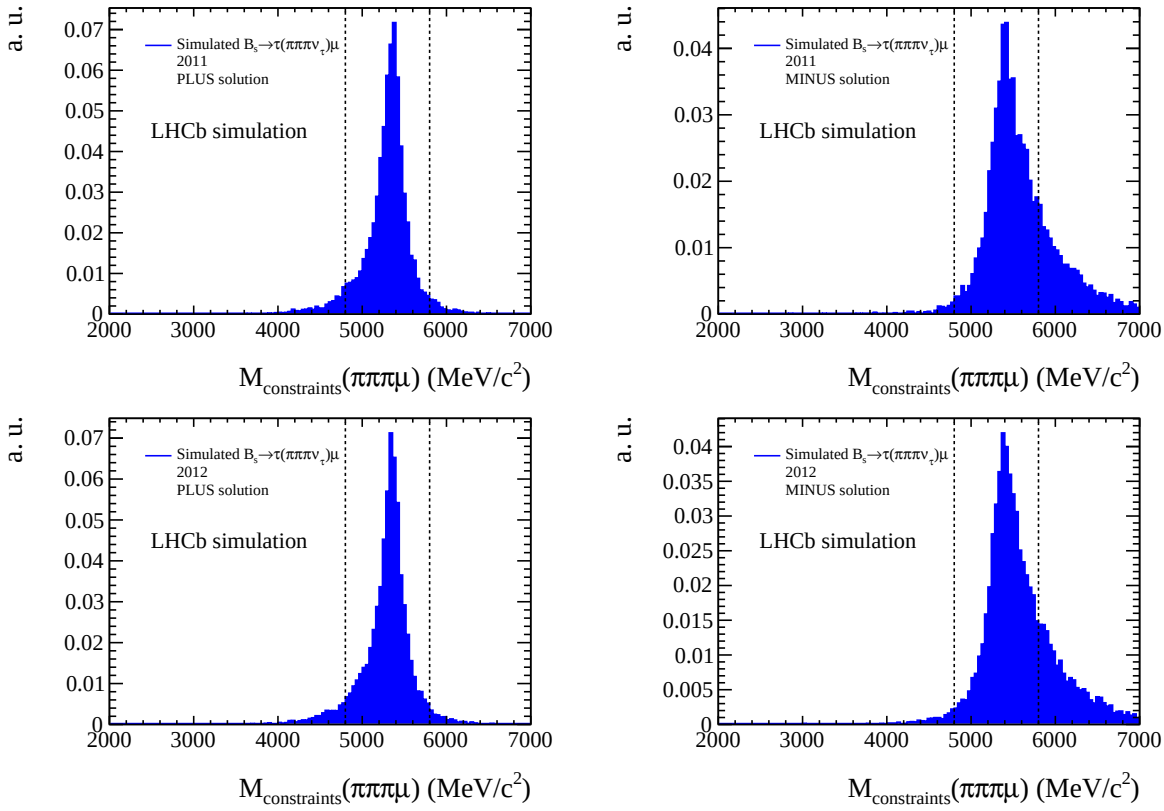

Figure 3.10: $B_s^0 \to \tau^\pm\mu^\mp$ 2011 (top) and 2012 (bottom) MC signal samples $M_B^\oplus$ (left) and $M_B^\ominus$ (right) mass reconstructions. The dashed lines represent the blinded region in the search data.

3.4.2 Choice of the B invariant mass variable

The fraction of candidates where the neutrino momentum can be determined thanks to a positive discriminant, is shown in Table 3.11 for the data samples as well as for the simulated signal and background samples. While more than 30% of signal events are lost, the fraction is much larger in background, making a positively definite discriminant a good requirement to reject background.

The performance of the 3 reconstruction methods (M_{vis}, $M_B^\oplus$ and $M_B^\ominus$) can be seen on the Receiver Operating Characteristic (ROC) curve on Figure 3.11. This ROC curve presents the simulated $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ signal efficiency against background rejection where the background is modelled with the Same Sign $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ data. It is built after reordering the bins of the mass distributions according to the following figure of merit (FoM):

$$FoM = \frac{N_{bin}^{sig}}{N_{bin}^{bkg}}, \text{if } N_{bin}^{bkg} \text{ is } \neq 0. \tag{3.5}$$

Each bin in the given variable (e.g. M_{vis}) is given a value for the FoM. Since the mass distribution of the signal candidate is peaking and that the tail of the mass distribution extends over the signal peak, the ROC curve can not simply be constructed scanning an upper or a lower mass cut. Therefore, the mass bins are reordered for both distributions according to the figure of merit given in Equation 3.5 and the ROC curve is build scanning this reordered histogram from the most to the least discriminating bins.

	Sample	Efficiency (%)
MC Signal	$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011	68.68 ± 0.43
	$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012	68.37 ± 0.33
	$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011	69.01 ± 0.44
	$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011	68.00 ± 0.35
	$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012	90.11 ± 0.27
MC Data	Same Sign data	47.80 ± 0.0
MC $D(\mu)3\pi$ type	$B^0 \to (D^- \to \mu^-\overline{\nu}_\mu)\pi^+\pi^-\pi^+$	24.06 ± 0.92
	$B_s^0 \to (D_s^- \to \mu^-\overline{\nu}_\mu)\pi^+\pi^-\pi^+$	21.77 ± 0.86
	$B^0 \to (D^*(2010)^- \to (D^- \to \mu^-\overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+$	22.37 ± 0.91
	$B^0 \to (D^*(2010)^- \to (D^- \to \mu^-\overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+\pi^0$	16.37 ± 1.10
	$B^0 \to (D^- \to K^0\mu^-\overline{\nu}_\mu)\pi^+\pi^-\pi^+$	12.04 ± 0.68
MC $D(3\pi)\mu$ type	$B^0 \to (D^*(2010)^- \to (D^- \to \pi^+\pi^-\pi^-\pi^0)\pi^0)\mu^+\nu_\mu$	43.79 ± 0.71
	$B^0 \to (D^- \to \pi^-\pi^+\pi^-\pi^0)\mu^+\nu_\mu$	37.82 ± 0.68
	$B^0 \to (D^{-(*)} \to K^-\pi^+\pi^-)X\mu^+\nu_\mu$	21.96 ± 0.33
	$B^0 \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)\mu^+\nu_\mu$	41.91 ± 0.88
	$B^+ \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)\mu^+\nu_\mu$	40.86 ± 0.86
MC $D(3\pi)\tau(\mu)$ type	$B^0 \to (D^- \to \pi^-\pi^+\pi^-\pi^0)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	36.07 ± 1.23
	$B^0 \to (D^*(2010)^- \to (D^- \to \pi^-\pi^+\pi^-\pi^0)\pi^0)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	44.56 ± 1.91
	$B^+ \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	39.73 ± 1.46
MC $D(\mu)\tau(3\pi)$ type	$B^0 \to (D^- \to \mu^-\overline{\nu}_\mu)(\tau^+ \to \pi^+\pi^+\pi^-\overline{\nu}_\tau)\nu_\tau$	45.31 ± 1.65
	$B^+ \to (D^{**} \to (D^- \to \mu^-\overline{\nu}_\mu)X)(\tau^+ \to \pi^+\pi^-\pi^+\overline{\nu}_\tau)\nu_\tau$	40.62 ± 1.04
MC $D(\tau(3\pi))\mu$ type	$B_s^0 \to (D_s^- \to (\tau^- \to \pi^-\pi^-\pi^+\nu_\tau)\overline{\nu}_\tau)\mu^+\nu_\mu$	50.61 ± 1.03

Table 3.11: Efficiency obtained with the analytic mass reconstruction (physical solutions of $M_B^\oplus$) and the DTF applied on the data and simulated samples.

3.4.3 Decay Tree Fitter

The Decay Tree Fitter (DTF) is an algorithm implemented in the `DecayTreeFitter` package of the LHCb software. It parameterizes a complete decay chain in terms of vertex positions, decay lengths and momentum parameters and it fits these parameters simultaneously taking into account the relevant constraints, such as the measured parameters of the final state tracks and photons, 4-momentum conservation at each vertex etc. To perform the fit efficiently a Kalman filter is used, the procedure is described in [69].
The Decay Tree Fitter is used to refit the full $B_{(s)}^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ decay including the neutrino momentum. The neutrino momentum used in the DTF algorithm is initialized

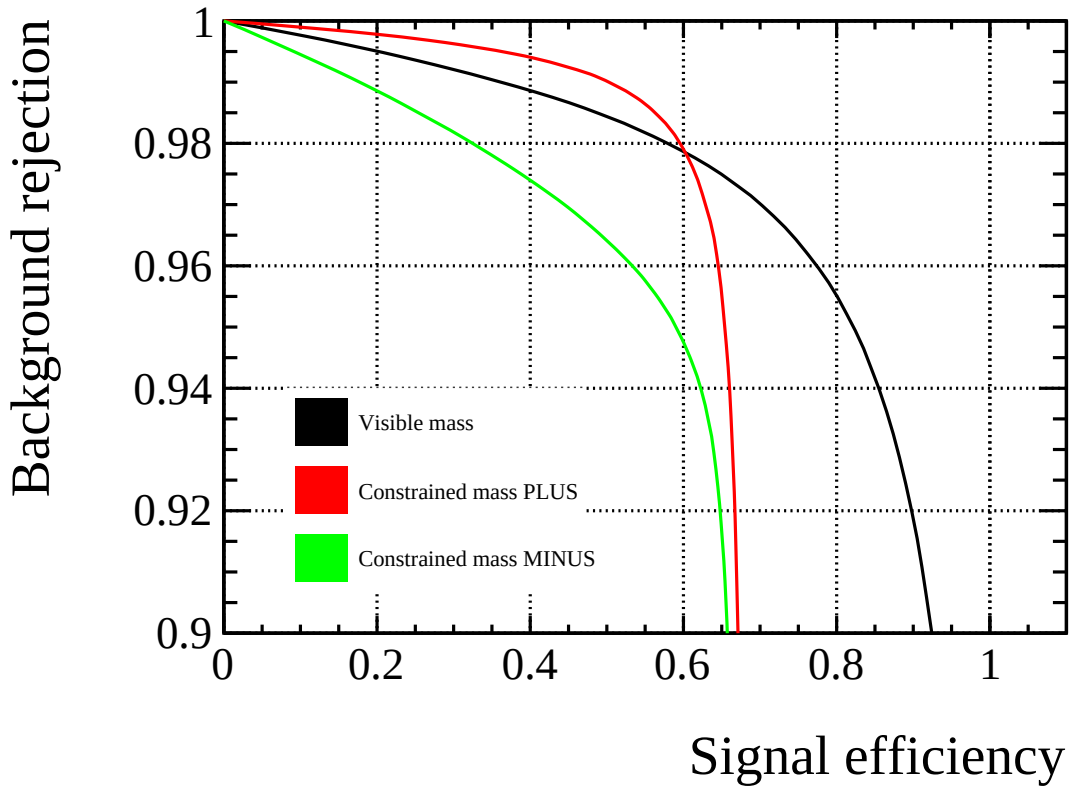

Figure 3.11: Background rejection in function of the signal efficiency for the different B reconstructed invariant mass variables for the processes $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$.

Sample	Type	Year	previous # cand.	# sel. cand.	Efficiency(%)
$B^0_s \to \tau^{\pm}\mu^{\mp}$	MC	2011	11381	7817	68.68 ± 0.43
$B^0_s \to \tau^{\pm}\mu^{\mp}$	MC	2012	19899	13604	68.37 ± 0.33
$B^0 \to \tau^{\pm}\mu^{\mp}$	MC	2011	11297	7796	69.01 ± 0.44
$B^0 \to \tau^{\pm}\mu^{\mp}$	MC	2012	17741	12064	68.00 ± 0.35
Same Sign	DATA	2011/2012	3429586	1639359	47.80 ± 0.03

Table 3.12: Data and $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ Monte-Carlo signal samples statistics before and after the reconstruction procedure. The previous number of candidates (previous # cand) values are taken from tables 3.4 and 3.5.

to the values corresponding to the $M_B^{\oplus}$ solution of the analytic method. The estimated B invariant mass from the DTF algorithm does not change significantly from the $M_B^{\oplus}$ value. However, it allows to access additional quantities such as the uncertainty on the reconstructed B mass or the χ^2 of the fit that can be used in the signal selection. The cases where the Decay Tree Fitter does not converge add $\sim 1\%$ inefficiency to the mass reconstruction. From now on the 4-body invariant mass calculated using DTF is referred to as M_B or the B invariant mass. Information about the change of statistics due to the reconstruction process in Same Sign data and signal samples can be seen on Table 3.12.

3.4.4 Signal blinding

The M_B region between $4900\,\mathrm{MeV}/c^2$ and $5800\,\mathrm{MeV}/c^2$ is blinded until the selection requirements and the fit strategy have been defined and validated.
The blinded region will be referred to as 'signal region' in the following.

3.5 Preselection

The preselection is made of 2 steps. The first is based on simple cuts and is described in Section 3.5.1. The second uses a multivariate classifier and is described in Section 3.5.2.

3.5.1 Cut-based preselection

The cut-based selection rejects very obvious background allowing the next selection steps to focus on the rejection of the background more difficult to distinguish from signal with higher efficiency.
The cuts applied in this preselection step are described below and are applied sequentially, i.e. the efficiencies mentioned for a given cut refer to the efficiency of the cut after the previous ones have been applied.

- $M_B > 4000\,\mathrm{MeV}/c^2$: Discards the low mass background region. This cut reduces the Same Sign and Opposite Sign samples by $(46.06 \pm 0.02)\%$ and $(40.58 \pm 0.02)\%$ respectively for $(99.41 \pm 0.07)\%$ signal efficiency.

- `totCandidate`==1: Events with more than one candidate are more likely to be background and are removed. The candidate multiplicity requirement efficiency evaluated on the simulated $B_s^0 \to \tau^\pm\mu^\mp$ signal sample is $(98.23 \pm 0.11)\%$ whereas the background rejections are $(36.90 \pm 0.05)\%$ and $(23.46 \pm 0.02)\%$ in the Same Sign and Opposite Sign samples respectively.

- $M_{\pi^+\pi^-} > 550\,\mathrm{MeV}/c^2$: The $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$ decay mostly proceeds via the $\tau \to a_1\nu$ channel where the a_1 particle decays to 3 pions through various resonances (mainly ρ^0). In some background, like the one due to D_s^+ decays, 2 of the 3 pions often come from the decay of a light resonance like the η. In order to remove this component a cut is placed on the $\pi^+\pi^-$ invariant mass. The Dalitz plot of the 3 pion system is shown in Figure 3.12. The discarded region, referred to as the *Dalitz plane control region* in the rest of the document, contains only $(1.21 \pm 0.09)\%$ of the $B_s^0 \to \tau^\pm\mu^\mp$ signal and $(28.61 \pm 0.06)\%$ of the Same Sign data. The events in this region will be used as a control sample.

- $M_\tau < 1800\,\mathrm{MeV}/c^2$: Veto for the process $D \to \pi\pi\pi$. This cut removes $(0.01 \pm 0.01)\%$ of the $B_s^0 \to \tau^\pm\mu^\mp$ signal and $(0.19 \pm 0.01)\%$ of the Same Sign data.

- τ `DeltaMassOneTrack` $> 1000\,\mathrm{MeV}$: This variable is defined as the difference between the invariant mass of the system formed by the 3 pions and an extra track and the invariant mass of the 3 pion system alone. The extra track is chosen among all other long tracks as the one giving the smallest change to the τ candidate vertex when added to the 3 pions. When the selected extra track is the one from the actual muon belonging to the candidate, the τ `DeltaMassOneTrack` is exactly the

Sample	Type	Year	previous # cand.	# sel. cand.	Efficiency (%)
$B_s^0 \to \tau^\pm\mu^\mp$	MC	2011	7817	6750	86.35 ± 0.39
$B_s^0 \to \tau^\pm\mu^\mp$	MC	2012	13604	11695	85.97 ± 0.30
$B^0 \to \tau^\pm\mu^\mp$	MC	2011	7796	6665	85.49 ± 0.40
$B^0 \to \tau^\pm\mu^\mp$	MC	2012	12064	10156	84.18 ± 0.33
OS (non-blind reg.)	DATA	2011/2012	8729757	1124506	12.88 ± 0.01
Same Sign	DATA	2011/2012	1639359	145585	8.88 ± 0.02
SS (non-blind reg.)	DATA	2011/2012	1563541	131649	8.42 ± 0.02

Table 3.13: Data and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ MC signal samples statistics before and after the cut-based preselection requirements. The Opposite Sign data sample is blinded. The previous number of candidates (previous # cand) values are taken from Table 3.12.

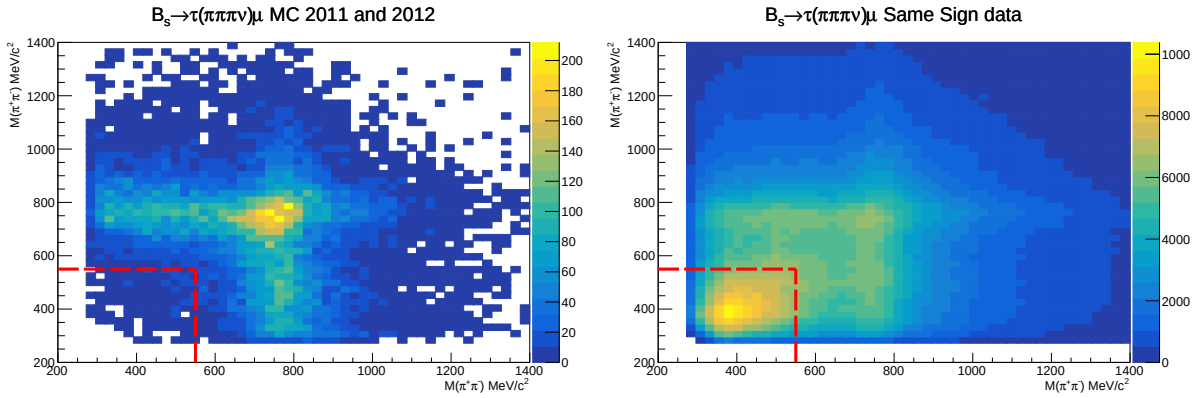

Figure 3.12: Dalitz plane for the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ simulation and Same Sign data samples. The low $M_{\pi^+\pi^-}$ region (inside the red dashed line) is discarded in the preselection.

B candidate visible mass and is fully correlated with the reconstructed mass (see Figure 3.14). On the other hand, if the selected track corresponds to a soft charged particle (e.g. coming from a D^* decay), τ `DeltaMassOneTrack` peaks at low values (see Figure 3.13). The cut placed to remove this component keeps $(88.71 \pm 0.26)\%$ of the signal and removes $(53.67 \pm 0.08)\%$ of the Same Sign data.

The total efficiency of these cuts is ∼86% for the simulated signal samples while rejecting ∼91% of the Same Sign events. The efficiencies for the relevant samples used in the analysis are detailed in Table 3.13. The M_B invariant mass distributions after these cuts are shown in Figure 3.15.

3.5.2 BDT-based preselection

The next step of the preselection uses isolation variables which aim at rejecting background candidates coming from decays with additional charged or neutral particles in the final state. These variables are combined in a BDT, which will be referred to as isolation-based BDT in the following. The different variables are described below and their distributions for the simulated signal and Same Sign samples are shown in Figure 3.16.

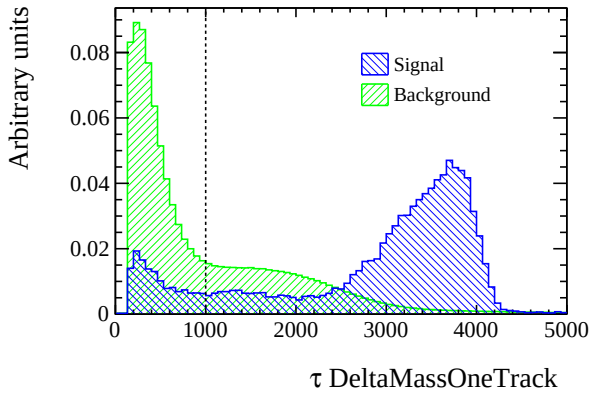

Figure 3.13: τ DeltaMassOneTrack distribution on the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulation sample and in the Same Sign data sample. The black dashed line indicates the requirement value.

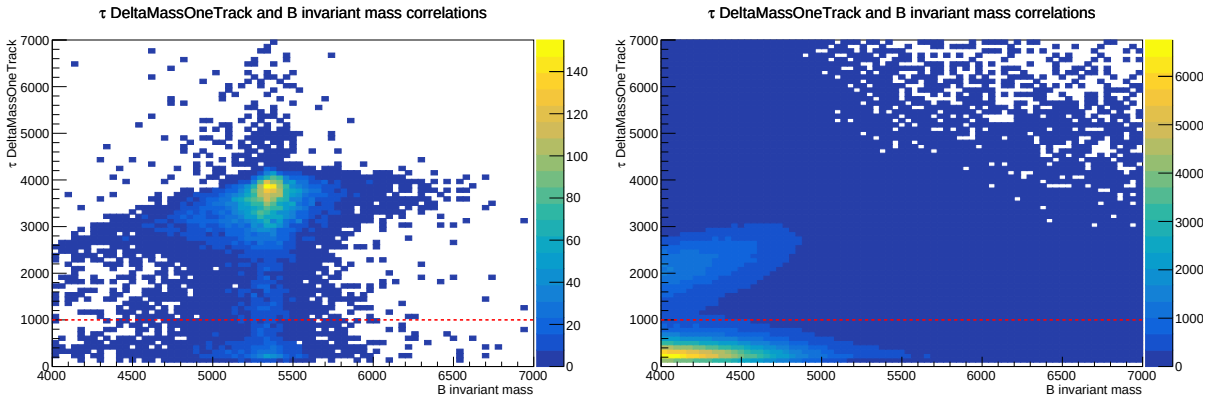

Figure 3.14: Correlation between τ DeltaMassOneTrack and B mass variables on the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulation sample (left) and in the Same Sign data sample (right). The red dashed line indicates the requirement value.

- *Vertex isolation variables*: Value of the smallest change in the B or τ decay vertex χ^2 when an extra track around the vertex is added.

    1. B `SmallestDeltaChi2OneTrack`
    2. τ `SmallestDeltaChi2OneTrack`

- *Track cone isolation variables*: Variables which involve track properties in a cone around a given track direction defined by η or ϕ angles in the LHCb reference system.

    3. B, τ and μ `0_50_cc_mult`: Number of charged tracks inside a cone of $\sqrt{\phi^2 + \eta^2} < 0.5$ around the particle direction.
    4. B, τ and μ `0_50_nc_sPT`: Scalar sum of the PT of the neutral objects inside a cone of $\sqrt{\phi^2 + \eta^2} < 0.5$ around the particle direction.

- *Track isolation variables.*

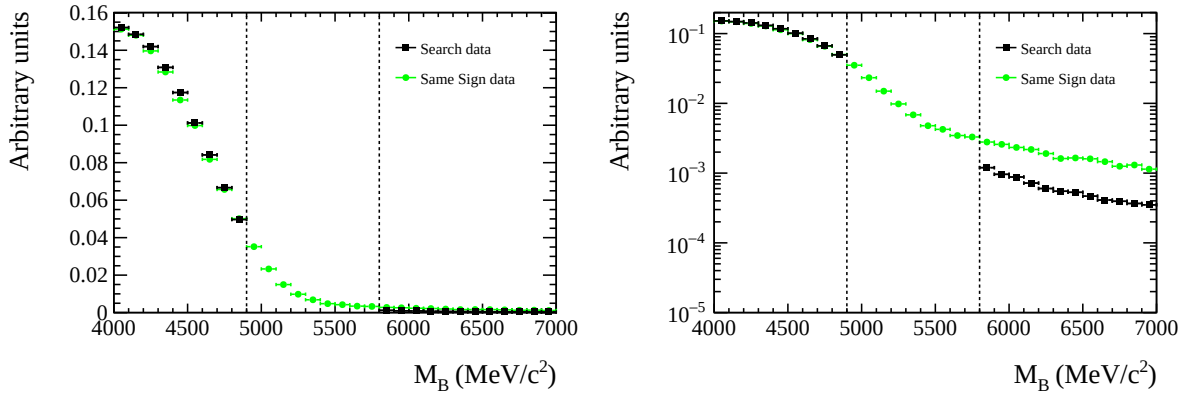

Figure 3.15: Data B invariant mass distribution once the cut-based preselection requirements have been applied. The Opposite Sign data is shown in black and the Same Sign data in green. The two plots show the same data, the left one in linear and the right one in logarithmic scale.

5. μ `isolation_giampi_nopi`: Number of tracks forming a compatible vertex with the muon candidate. The 3 pions coming from the τ candidate are excluded from this counting. Therefore, this variable in signal should peak at 0.

- *Track isolation involving Boosted Decision Trees* [70]: For each long track in the event which is not part of the signal candidate a BDT response is calculated. The tracks with high degree of isolation, 'signal-like' tracks, will lie at low BDT values.

6. τ, μ and $\sum\pi$ `BDTiso1_1`: Number of long tracks with a BDT value smaller than -0.09.
7. τ, μ and $\sum\pi$ `BDTiso3`: The sum of the BDT response of the long tracks with BDT value lower than -0.05 plus the BDT minimim value of the long tracks with BDT value between [-0.05,0]. Defining $x = (-\infty, -0.05)$ and $y = [-0.05, 0]$, the variable is computed the following way:

$$BDTiso3 = \min\{BDT(y)\} + \sum_{BDT(x)<-0.05} BDT(x) \tag{3.6}$$

The training samples used to train the isolation-based BDT are the MC samples of $B_s^0 \to \tau^\pm\mu^\mp$ simulated with 2011 and 2012 conditions for signal and the Same Sign data for background. The BDT is trained with the Adaptative Boosting algorithm using the `TMVA` package [71]. The relevant parameters of the BDT training are: `NTrees`=300, `MinNodeSize`=2.5%, `MaxDepth`=2 and `nCuts`=20. The MC signal events in the signal samples are truth-matched. In order to avoid biases due to applying the BDT on a sample used for training, the samples are partitioned in two subsets and two BDTs are trained independently on each subset. In each subset, a maximum of 10k Same Sign events are used in order to match the size of the training signal sample which uses the full available MC statistic. Each BDT output is then applied on the subset it has not been trained and flattened. The flattening consists of rescaling the isolation-based BDT output such that the distribution of the B_s^0 signal sample is uniform between 0 and 1. The BDT output distributions for signal and background samples are shown in Figure 3.17.

Sample	Type	Year	previous # cand.	# sel. cand.	Efficiency (%)
$B_s^0 \to \tau^\pm \mu^\mp$	MC	2011	6750	2879	42.65 ± 0.60
$B_s^0 \to \tau^\pm \mu^\mp$	MC	2012	11695	4413	37.73 ± 0.45
$B^0 \to \tau^\pm \mu^\mp$	MC	2011	6665	2638	39.58 ± 0.60
$B^0 \to \tau^\pm \mu^\mp$	MC	2012	10156	3665	36.09 ± 0.48
OS (non-blind reg.)	DATA	2011/2012	1124506	135088	12.01 ± 0.03
Same Sign	DATA	2011/2012	145585	12918	8.87 ± 0.07
SS (non-blind reg.)	DATA	2011/2012	131649	11836	8.99 ± 0.08

Table 3.14: Data and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ MC signal samples statistics before and after the cut on the isolation-based BDT output. The Opposite Sign data sample is blinded. The previous number of candidates (previous # cand) values are taken from Table 3.13.

The isolation-based BDT performance is illustrated in the ROC curve shown in Figure 3.18. A lower cut on the isolation-based BDT output is placed at 0.6, keeping $\sim$40% of signal efficiency and rejecting $\sim$92% of background. This working point has been chosen *ad hoc*. The efficiencies of this cut for the various samples used in the analysis are given in Table 3.14. For completeness Figure 3.19 contains the profile plot of the BDT output against the B invariant mass for signal and Same Sign samples.

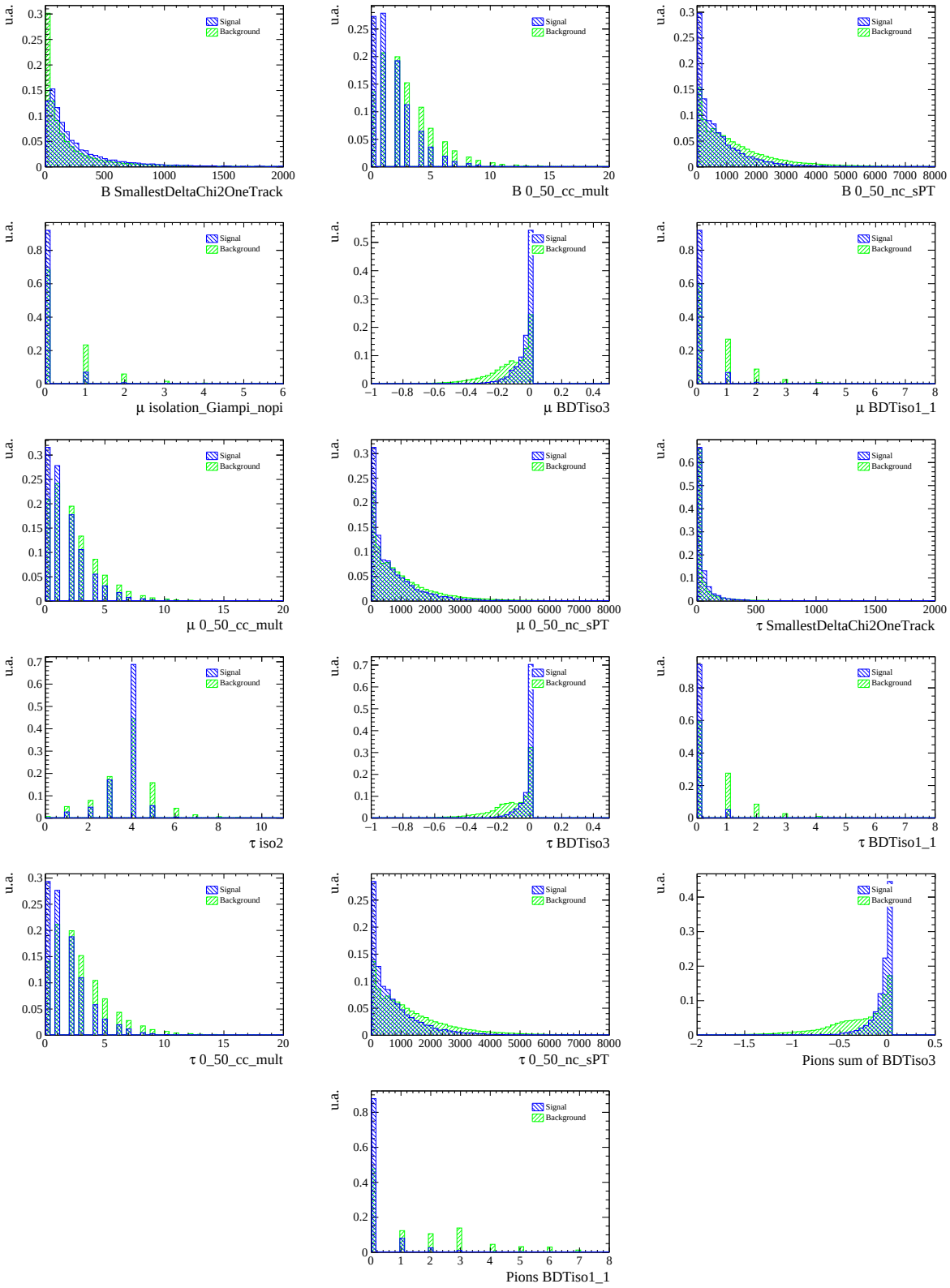

Figure 3.16: Isolation-based BDT input variables distributions. Signal is drawn in blue and background (Same Sign data) in green.

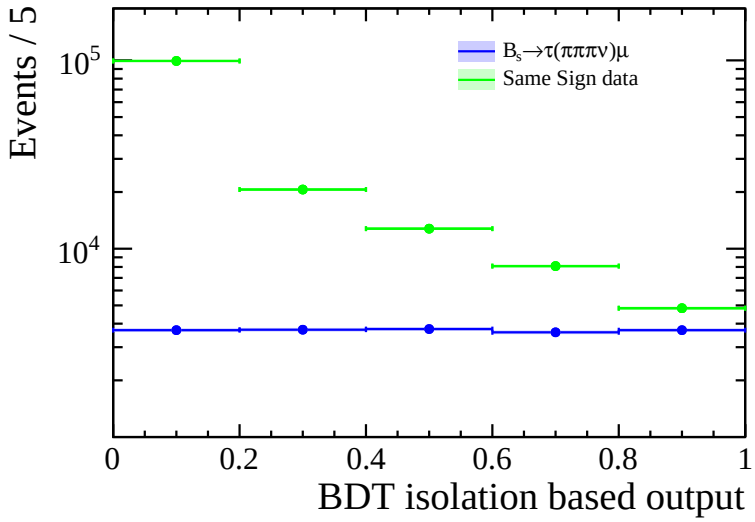

Figure 3.17: Isolation-based BDT ouput distributions for $B_s^0 \to \tau^{\pm}\mu^{\mp}$ signal simulation (2011 and 2012) and Same Sign data.

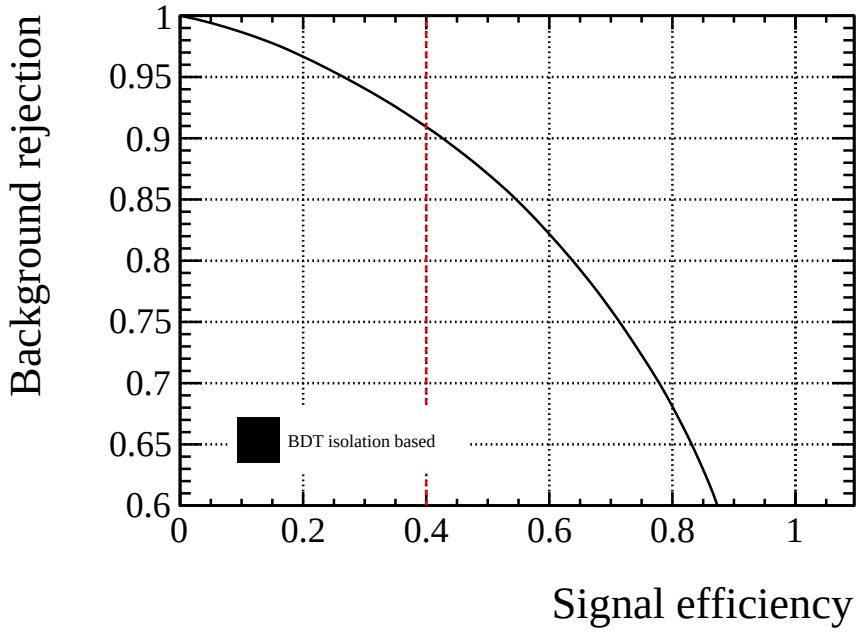

Figure 3.18: Isolation-based BDT efficiency for $B_s^0 \to \tau^{\pm}\mu^{\mp}$ simulated samples -Vs- rejection of Same Sign data. The red line corresponds to the remaining signal efficiency when the requirement is applied on the BDT output.

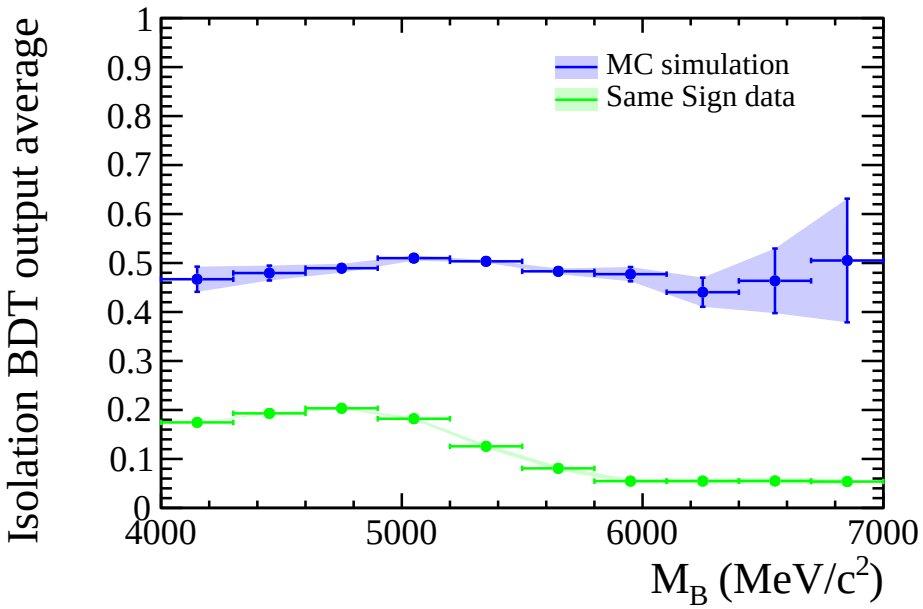

Figure 3.19: Isolation-based BDT output profile with respect to B invariant mass $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulation samples (2011 and 2012) in blue and the Same Sign data sample in green.

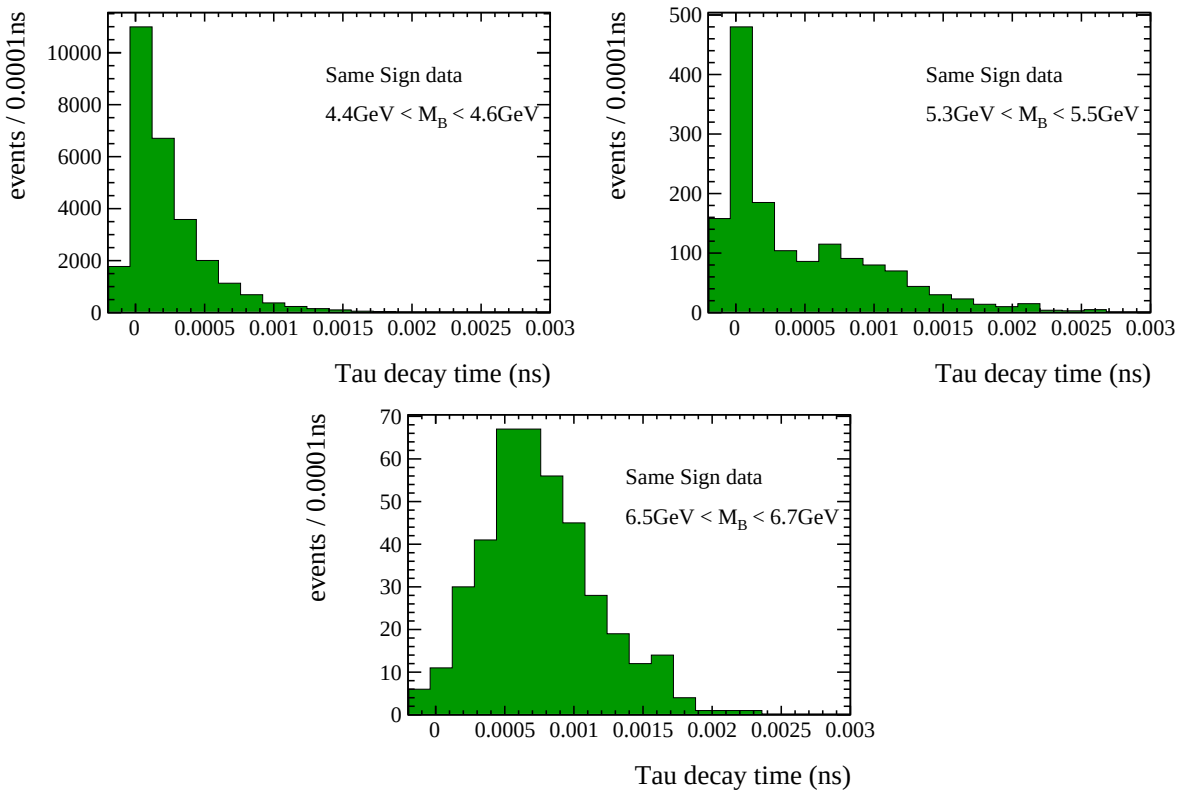

Figure 3.20: $B_{(s)} \to \tau(\pi\pi\pi\nu)\mu$ Same Sign data τ lifetime distribution in different B mass spectrum ranges: [4.4,4.6] MeV (top-left), [5.3,5.5] MeV (top-right), [6.5,6.7] MeV (bottom)

3.6 Specific background rejection

The cuts on the decay flight distances, like the B candidate flight distance significance with respect to the primary vertex, and impact parameters applied in the stripping discard most of the background candidates involving particles originating from the primary vertex. The remaining background candidates are build with tracks from particles decaying after a sizeable distance. The aim of this section is to identify specific background components and reject them when possible.

As it can be observed in the τ decay time distributions shown in Figure 3.20 for different B mass ranges, two different components can be distinguished:

- a component with an exponentially falling reconstructed τ decay time characteristic of τ or D decays, consisting mostly of partially reconstructed B decays. This component is present mainly at low mass and extends inside the signal region.
- a combinatorial background component with a very broad decay time distribution centred around $5 \cdot 10^{-4} ns$. This component is the only one present at high mass while it is completely dominated by the partially reconstructed background in the low mass region.

The simulated samples of exclusive B background decays presented in Section 3.2.2 have been generated specifically to study partially reconstructed and possibly peaking

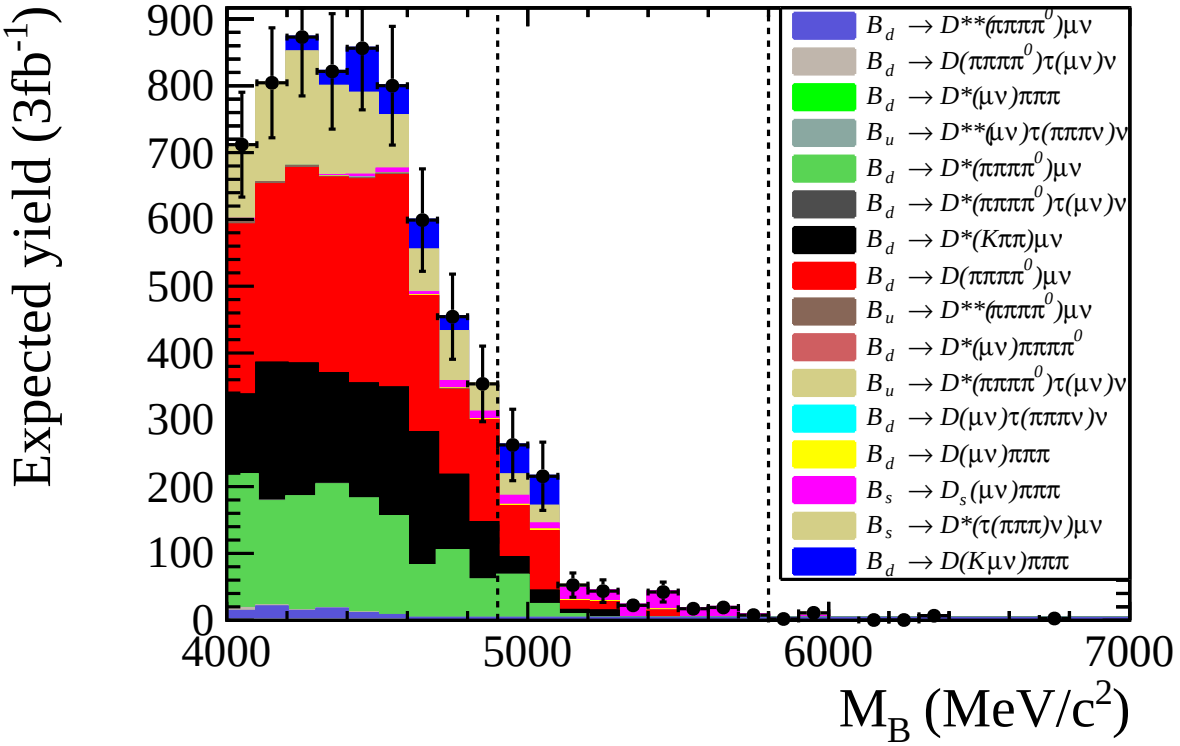

Figure 3.21: M_B distributions of the exclusive samples once the complete preselection is applied. The figure shows the expected yield of the exclusive samples corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$.

backgrounds. Their mass distributions, rescaled to the expected yields in $3\,\text{fb}^{-1}$, are shown in Figure 3.21. These samples can be divided in two categories depending on their topology (see Figure 3.22):

- *Reverse topology*: the 3 pions come directly from the B decay vertex and the muon from a $D^{(*)}$ meson decay.
- *Signal-like topology*: the 3 pions come from a displaced vertex after the B decay vertex. These decays do not always have strictly the same topology as the signal, as the muon can come from a τ or $D^{(*)}$ decay as well.

The rest of the section presents the rejection of the combinatorial background (Section 3.6.1) and of the reverse topology partially reconstructed background (Section 3.6.2). The remaining background with signal-like topology is discussed in Section 3.6.3.

3.6.1 Combinatorial background

The combinatorial background is characterized by candidates made with particles not originated from a common ancestor and it is efficiently removed using the following variables related to the B meson decay vertex:

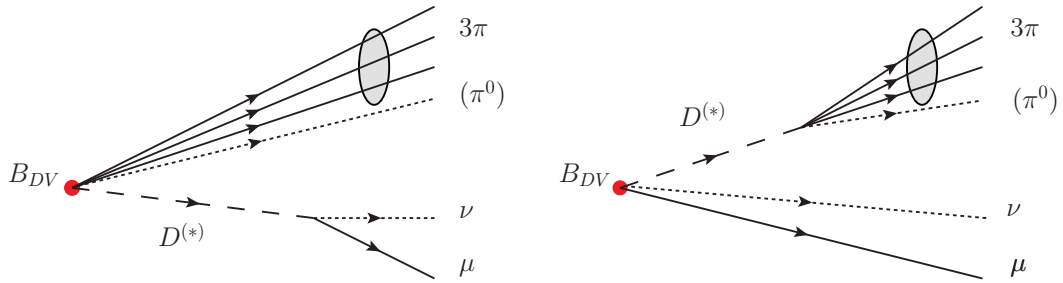

Figure 3.22: Partially reconstructed backgrounds main topologies. Left: Signal reverse topology. Right: Signal-like topology.

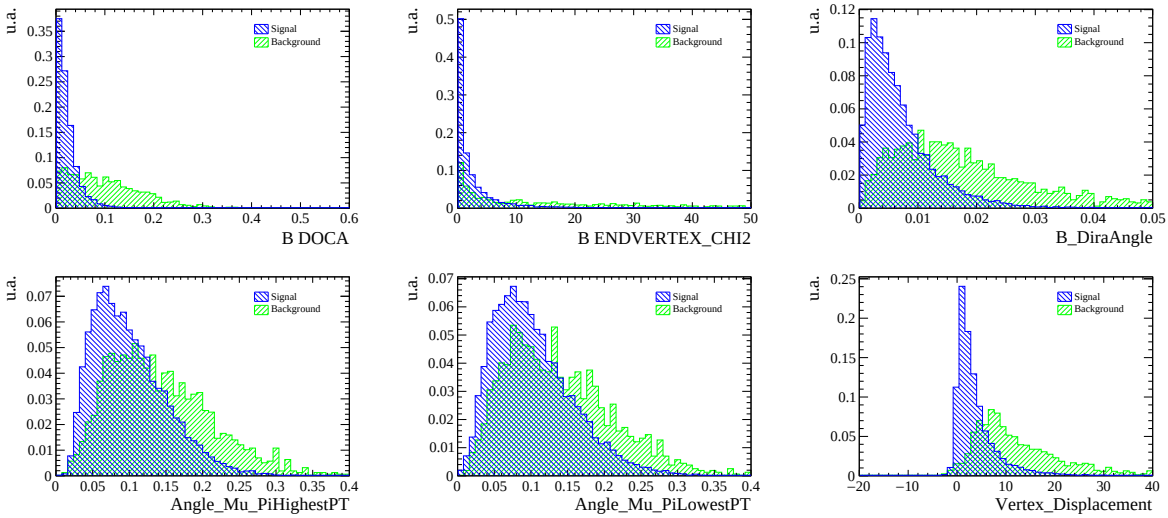

Figure 3.23: Combinatorial background BDT input variables distributions. Signal on the complete B mass spectrum (blue) and Same Sign samples on the high M_B side-band (green).

- B `DOCA`: distance of closest approach between the τ and the μ reconstructed candidates.
- B `ENDVERTEX_CHI2`: χ^2 of the B decay vertex reconstruction fit.
- B `DiraAngle`: Angle between the B reconstructed momentum and the B direction of flight from the best PV to the B decay vertex.
- `Angle_Mu_PiHighestPT`: Angle between the μ and the π (coming from the tau candidate) with highest transverse momentum.
- `Angle_Mu_PiLowestPT`: Angle between the μ and the π (coming from the tau candidate) with lowest transverse momentum.
- `Vertex_Displacement`: Difference in the Z component between the τ and the B decay vertex positions.

These variables are combined in a BDT, referred to as the combinatorial background BDT in the rest of the document. The BDT is trained using the MC samples of $B_s^0 \to \tau^{\pm}\mu^{\mp}$ simulated with 2011 and 2012 conditions as proxy for the signal. The signal events are truth-matched. The BDT is trained with the Adaptative Boosting algorithm using the `TMVA` package [71]. The relevant parameters of the BDT training are: `NTrees`=300, `MinNodeSize`=2.5%, `MaxDepth`=2 and `nCuts`=20. The background sample is taken from the upper side-band ($M_B > 6200\,\text{MeV}$) of the Same Sign data where the combinatorial background completely dominates. Due to the extremely low statistics available in the upper Same Sign data side-band, only the cut based preselection requirements are applied and the cut on the isolation based BDT is omitted when building the training samples. The distributions of the BDT input variables in the training samples are shown in Figure 3.23. In order to avoid biases due to applying the BDT to a sample on which it was trained, the BDT is folded using the same procedure as for the isolation based BDT. Due to the lack of statistics, the folding is here of order 10. In each of the 10 BDT trainings, a maximum of 2k signal MC events are used to match the size of the training background sample which uses the full available statistic in the Same Sign data upper side-band. Each combinatorial background BDT output is applied on the subset it has not been trained on and flattened between 0 and 1 in the B_s^0 MC signal.

The combinatorial background BDT performance is summarized by the ROC curve shown in Figure 3.24. The optimal cut on the BDT is evaluated by maximizing the so-called Punzi Figure of Merit (FoM) defined as:

$$\text{FoM}_{Punzi} = \frac{\epsilon_{\text{sig}}}{\sqrt{N_{\text{bkg}}} + 5/2} \tag{3.7}$$

where ϵ_{sig} and N_{bkg} are the signal efficiency and the background yield for a given cut on the BDT output. As shown in Figure 3.25, the maximum of the FoM is found at 0.30, therefore the cut on the combinatorial background BDT is placed at this BDT output value, thus keeping 70% of signal efficiency. The corresponding rejection is ∼95% for the Same Sign data upper side-band and ∼87% for the Opposite Sign data upper side-band. The efficiencies of this cut for the various samples used in the analysis are given in Table 3.15. The B invariant mass distribution on data once both the isolation based BDT and the combinatorial background BDT requirements have been applied can be seen on Figure 3.26 (to be compared with Figure 3.15). Notice that in Figure 3.26 the shapes between the two data samples are in better agreement than in Figure 3.15.

3.6.2 Reverse topology partially reconstructed background

Among the exclusive background samples that have been simulated for this analysis, the ones with reverse vertex topologies are:

- $B \to D^-(\to \mu^-\overline{\nu}_\mu)\pi^+\pi^-\pi^+$
- $B_s^0 \to D_s^-(\to \mu^-\overline{\nu}_\mu)\pi^+\pi^-\pi^+$
- $B \to D^*(\to D^-(\to \mu^-\overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+$
- $B \to D^*(\to D^-(\to \mu^-\overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+\pi^0$

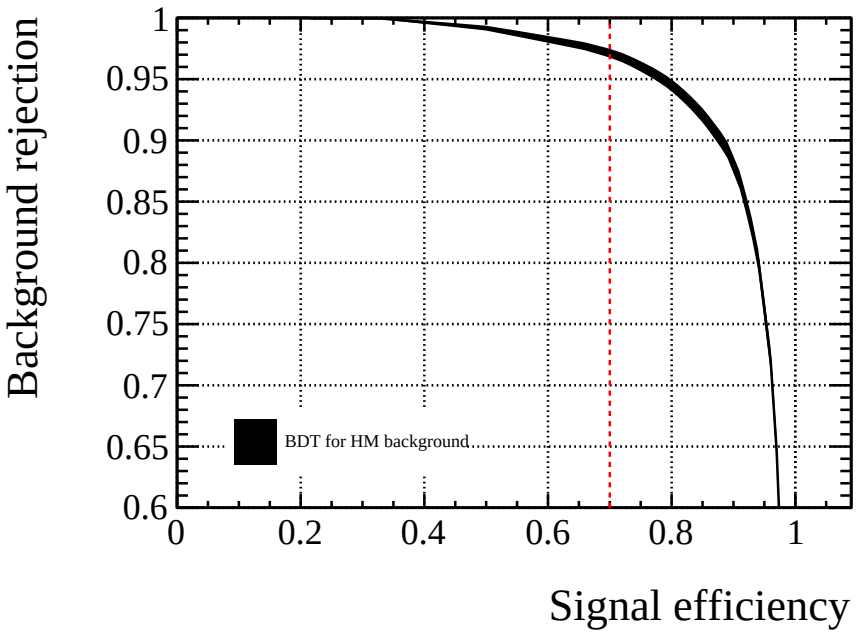

Figure 3.24: ROC curve of the combinatorial background BDT. The plot shows the signal efficiency on the whole B mass spectrum against the background rejection in the high B mass region (B mass $> 6200\,\text{MeV}$).

- $B \to D^-(\to K^0\mu^-\overline{\nu}_\mu)\pi^+\pi^-\pi^+$

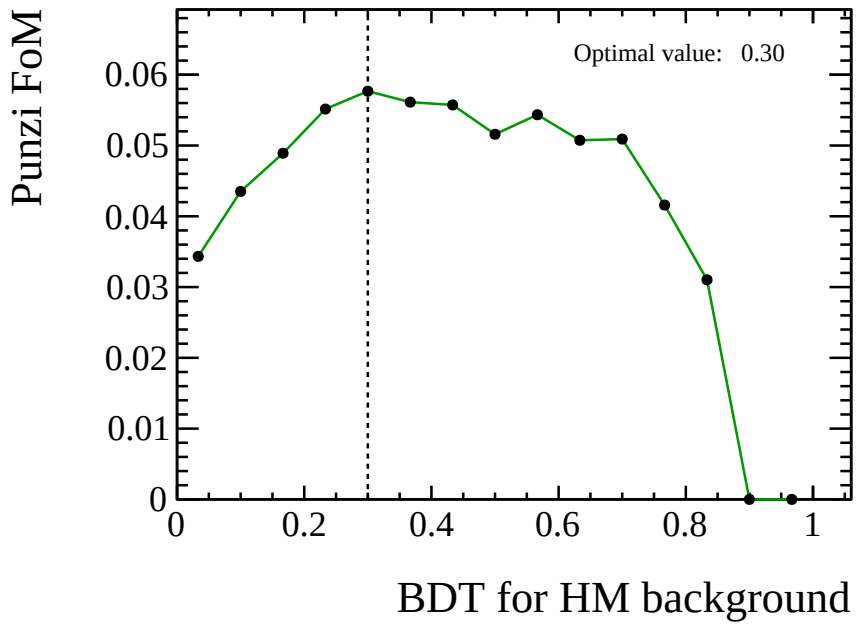

Figure 3.25: Punzi Figure of Merit of the combinatorial background BDT.

Sample	Type	Year	previous # cand.	# sel. cand.	Efficiency (%)
$B_s^0 \to \tau^\pm \mu^\mp$	MC	2011	2879	2006	69.68 ± 0.86
$B_s^0 \to \tau^\pm \mu^\mp$	MC	2012	4413	3068	69.52 ± 0.69
$B^0 \to \tau^\pm \mu^\mp$	MC	2011	2638	1868	70.81 ± 0.89
$B^0 \to \tau^\pm \mu^\mp$	MC	2012	3665	2549	69.55 ± 0.76
OS (non-blind reg.)	DATA	2011/2012	135088	77324	57.24 ± 0.13
Same Sign	DATA	2011/2012	12918	8152	63.11 ± 0.42
SS (non-blind reg.)	DATA	2011/2012	11836	7403	62.55 ± 0.44

Table 3.15: Data and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ MC signal samples statistics before and after the requirements the BDT for the combinatorial background. The Opposite Sign data sample is blinded. The previous number of candidates (previous # cand) values are taken from Table 3.14.

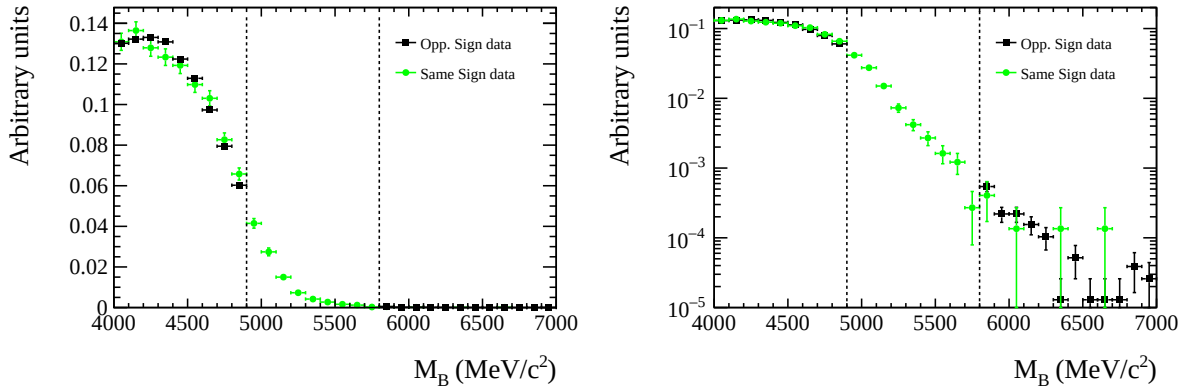

Figure 3.26: Data B invariant mass distribution once the requirements on the complete preselection and the combinatorial background BDT have been applied. The Opposite Sign data is shown in black and the Same Sign data in green. The two plots show the same data, the left one with a linear axis and the right one with a logarithmic axis.

As shown in Figure 3.21 and emphasized on Figure 3.27, some of these modes are peaking in the signal region. Their rejection is therefore crucial in order to remove spurious signal excesses.

The main feature of this background component is that the 3 pions come directly from the B decay vertex, making the variables related with the τ decay time particularly discriminant. Signal events have on average high τ decay time whereas the reverse topology partially reconstructed background events have small or negative τ decay time values (see Figure 3.28). The best variable optimizing the difference between signal and the reverse topology partially reconstructed background is the τ decay time significance.

- τ `decay time significance`: τ decay time estimated by the B and τ vertex positions divided by its uncertainty.

The performance of a cut on the τ decay time significance, estimated on the mix of the four samples of exclusive modes with the reverse vertex topology is shown in the ROC curves presented on Figure 3.29. A lower cut is placed at 1.8, which corresponds to the

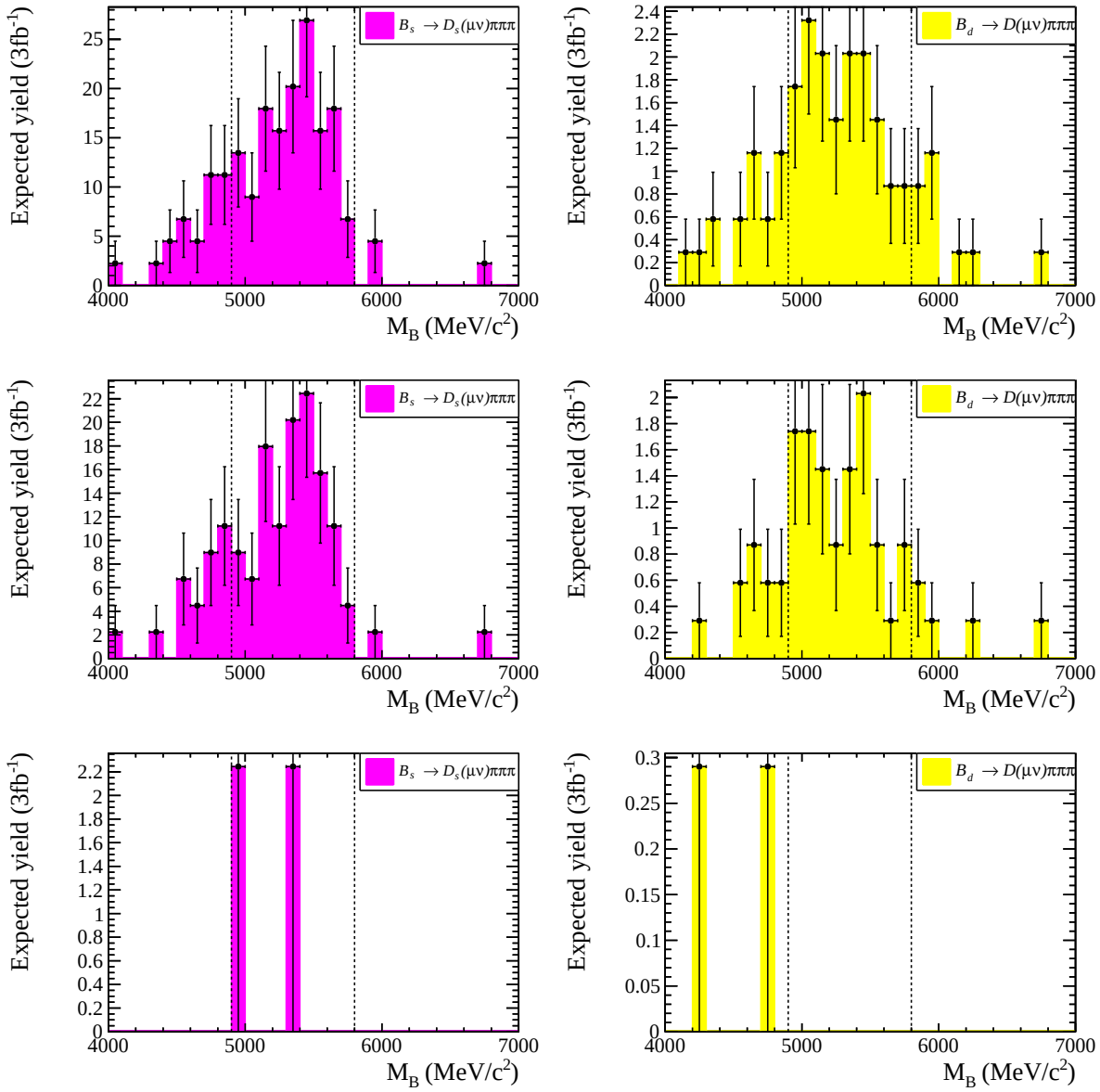

Figure 3.27: Two peaking exclusive samples with reverse topology B invariant mass distributions. Top: Once the complete preselection is applied. Middle: Complete preselection and combinatorial background BDT applied. Bottom: When the requirements on the preselection, the combinatorial background BDT and the τ decay time significance are applied. The figure shows the expected yield of the exclusive samples corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$.

maximum of the Punzi FoM (see Figure 3.29). It keeps around 80% of the remaining signal while rejecting almost the totality of the reverse topology background. The τ lifetime significance requirement performances on data and signal samples are shown in Table 3.16. The B invariant mass distribution on data, once the τ lifetime significance requirement is applied, can be seen on Figure 3.30.

Concerning the exclusive samples, the number of expected events for an equivalent integrated luminosty of $3\,\text{fb}^{-1}$ after the complete offline selection are reported in Table 3.17.

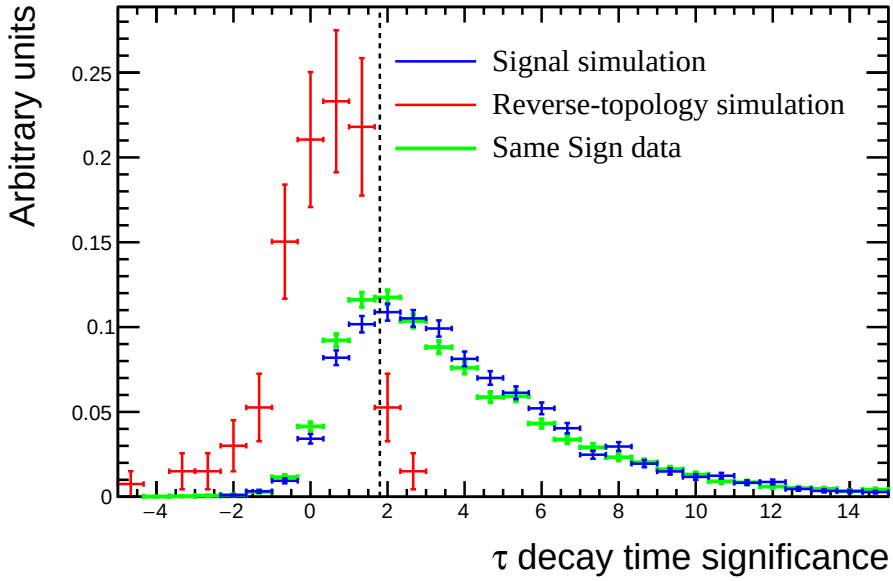

Figure 3.28: τ decay time significance variable distribution. The signal distribution is shown in blue and a mixture of 4 exclusive samples with reverse topology in red. The dashed black line represents the variable selection value.

Sample	Type	Year	previous # cand.	# sel. cand.	Efficiency (%)
$B_s^0 \to \tau^\pm \mu^\mp$	MC	2011	2006	1490	74.28 ± 0.98
$B_s^0 \to \tau^\pm \mu^\mp$	MC	2012	3068	2346	76.47 ± 0.77
$B^0 \to \tau^\pm \mu^\mp$	MC	2011	1868	1400	74.95 ± 1.00
$B^0 \to \tau^\pm \mu^\mp$	MC	2012	2549	1920	75.32 ± 0.85
OS (non-blind reg.)	DATA	2011/2012	77324	56752	73.40 ± 0.16
Same Sign	DATA	2011/2012	8152	5846	71.71 ± 0.50
SS (non-blind reg.)	DATA	2011/2012	7403	5421	73.23 ± 0.51

Table 3.16: Data and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ MC signal samples statistics before and after the requirement on τ decay time significance. The Opposite Sign data sample is blinded. The previous number of candidates (previous # cand) values are taken from Table 3.4 and 3.15.

The table also contains the offline selection efficiency for each exclusive mode. It can be seen that the contribution of the modes corresponding to the reverse topology have been drastically reduced (Figure 3.27).

	Sample	Sample events	Expected events	Offline sel. Eff. (%)
$D(\mu)3\pi$ type	$B^0 \to (D^- \to \mu^- \overline{\nu}_\mu)\pi^+\pi^-\pi^+$	2.0 ± 1.4	0.6 ± 0.4	0.07 ± 0.05
	$B_s^0 \to (D_s^- \to \mu^- \overline{\nu}_\mu)\pi^+\pi^-\pi^+$	2.0 ± 1.4	4.5 ± 3.2	0.07 ± 0.05
	$B^0 \to (D^*(2010)^- \to (D^- \to \mu^- \overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+$	1.0 ± 1.0	0.1 ± 0.1	0.04 ± 0.04
	$B^0 \to (D^*(2010)^- \to (D^- \to \mu^- \overline{\nu}_\mu)\pi^0)\pi^+\pi^-\pi^+\pi^0$	0.0	$< 3 \cdot 10^{-3}$	0.00
	$B^0 \to (D^- \to K^0\mu^- \overline{\nu}_\mu)\pi^+\pi^-\pi^+$	1.0 ± 1.0	22.6 ± 22.6	0.03 ± 0.03
$D(3\pi)\mu$ type	$B^0 \to (D^*(2010)^- \to (D^- \to \pi^+\pi^-\pi^-\pi^0)\pi^0)\mu^+\nu_\mu$	71.0 ± 8.4	525.5 ± 62.4	1.11 ± 0.13
	$B^0 \to (D^- \to \pi^-\pi^+\pi^-\pi^0)\mu^+\nu_\mu$	74.0 ± 8.6	940.9 ± 109.4	1.11 ± 0.13
	$B^0 \to (D^{-(*)} \to K^-\pi^+\pi^-)X\mu^+\nu_\mu$	117.0 ± 10.8	776.0 ± 71.7	0.54 ± 0.05
	$B^0 \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)\mu^+\nu_\mu$	15.0 ± 3.9	48.6 ± 12.6	0.35 ± 0.09
	$B^+ \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)\mu^+\nu_\mu$	5.0 ± 2.2	5.0 ± 2.2	0.11 ± 0.05
$D(3\pi)\tau(\mu)$ type	$B^0 \to (D^- \to \pi^-\pi^+\pi^-\pi^0)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	0.0	$< 5 \cdot 10^{-4}$	0.00
	$B^0 \to (D^*(2010)^- \to (D^- \to \pi^-\pi^+\pi^-\pi^0)\pi^0)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	1.0 ± 1.0	0.9 ± 0.9	0.10 ± 0.10
	$B^+ \to (D^{**} \to (D^- \to \pi^+\pi^-\pi^-\pi^0)X)(\tau^+ \to \mu^+\nu_\mu\overline{\nu}_\tau)\nu_\tau$	0.0	$< 3 \cdot 10^{-2}$	0.00
$D(\mu)\tau(3\pi)$ type	$B^0 \to (D^- \to \mu^-\overline{\nu}_\mu)(\tau^+ \to \pi^+\pi^+\pi^-\overline{\nu}_\tau)\nu_\tau$	10.0 ± 3.2	0.4 ± 0.1	0.79 ± 0.25
	$B^+ \to (D^{**} \to (D^- \to \mu^-\overline{\nu}_\mu)X)(\tau^+ \to \pi^+\pi^-\pi^+\overline{\nu}_\tau)\nu_\tau$	2.0 ± 1.4	0.1 ± 0.1	0.06 ± 0.04
$D(\tau(3\pi))\mu$ type	$B_s^0 \to (D_s^- \to (\tau^- \to \pi^-\pi^-\pi^+\nu_\tau)\overline{\nu}_\tau)\mu^+\nu_\mu$	71.0 ± 8.4	379.0 ± 45.0	2.22 ± 0.26

Table 3.17: Exclusive background samples final number of events in the samples, final number of expected events for an equivalent luminosity of $3\,\text{fb}^{-1}$ and offline selection efficiency.

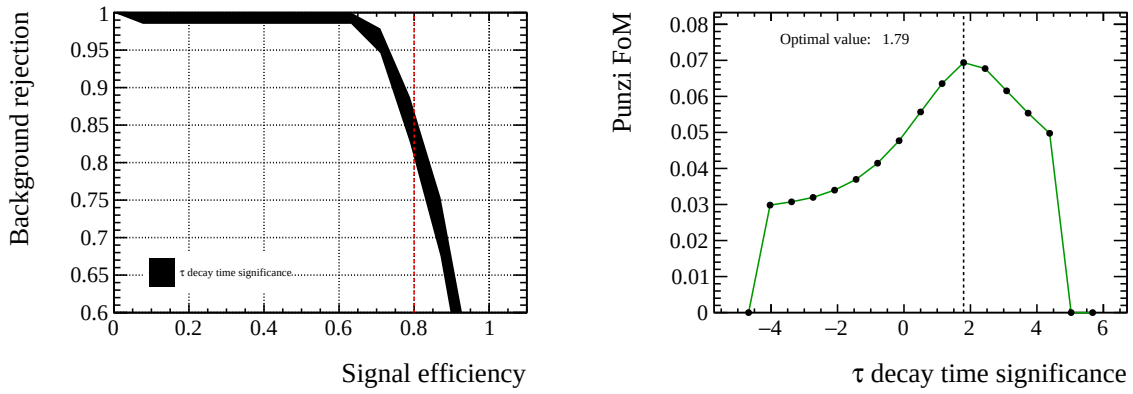

Figure 3.29: Left: ROC curve for the τ decay time significance variable. The plot shows the signal efficiency on the whole B mass spectrum against the background rejection. The background is composed by a mixture of 4 exclusive samples with reverse topology.
Right: Punzi FoM of the τ decay time significance variable. The background properties are extracted from a mixture of 4 exclusive samples with reverse topology.

3.6.3 Remaining background components

Once the complete offline selection is applied, the main remaining background is composed of partially reconstructed decays with signal-like topology.
Furthermore, after the complete offline selection, the backgrounds peaking in the signal region have been reduced to a negligible level. This conclusion is supported by the B invariant mass distributions on various samples.

On the one hand, simulated background samples have been examined to identify possible peaking remaining background components:

1. *Exclusive background distributions*: Figures 3.21 and 3.31 contain the B invariant mass distributions for the expected yield according to an integrated luminosity of $3\,\text{fb}^{-1}$ of the exclusive samples, respectively, after the preselection and after the offline selection requirements. It can be seen that all the known peaking contributions have been rejected. The remaining exclusive decays correspond to signal-like partially reconstructed backgrounds.

2. *Inclusive b background distribution*: Figure 3.32 shows the B invariant mass distributions once the complete offline selection is applied on the inclusive background sample. The remaining events have been investigated looking at the Monte-Carlo truth information. Matching the reconstructed tracks with the underlying MC particles, it turns out that all candidates are originating from a signal-like partially reconstructed B decay of the form: $B^{(0)} \to D^{(0/+/*)}\mu\nu(\pi^{(+/0)}/\gamma)$. The remaining event invariant mass distribution follows the behavior of signal-like partially reconstructed background.

On the other hand, data background samples can be also used to extrapolate the behavior of the opposite-sign data in the blinded region.

3. *Same Sign and Opposite Sign (blinded) data distributions*: Same Sign data can only reproduce background up to 5-body decays. Figure 3.30 contains the M_B of the

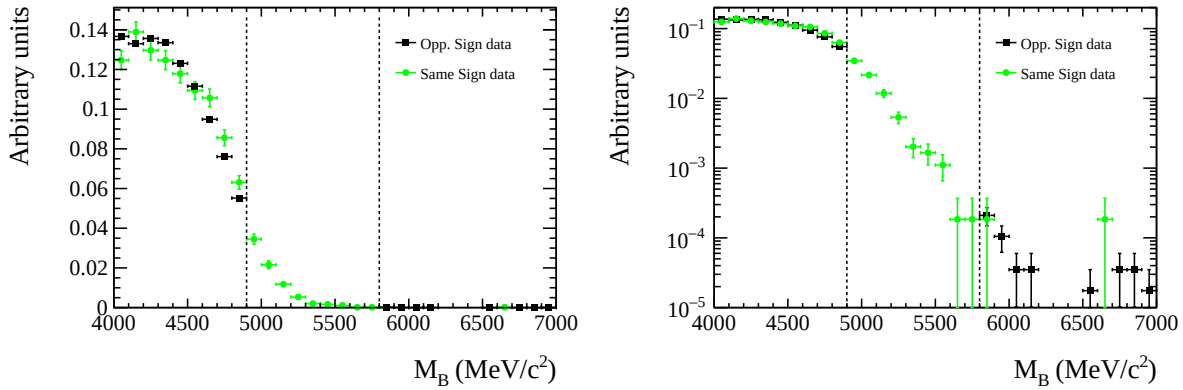

Figure 3.30: Data B invariant mass distribution once the complete offline selection have been applied. The Opposite Sign data is shown in black and the Same Sign data in green. The two plots show the same data, the left one with a linear axis and the right one with a logarithmic axis.

Same and Opposite Sign (blinded) data once the offline selection has been applied. It can be seen that the two selected data samples are compatible in the non-blinded region and that the Same Sign data shows a smooth falling behavior inside the signal region similar to the one shown by the simulated signal-like partially reconstructed backgrounds.

4. *Dalitz plane control region distribution*: The Dalitz plane control region ($M_{\pi^+\pi^-} < 550\,\mathrm{MeV}/c^2$), defined as in Figure 3.12 in Sec. 3.5.1, is rejected during the preselection as it contains mostly background. Therefore, it can be unblinded to check which kind of background is remaining in this rejected subsample. The comparison of the M_B distributions in this region and in the Same Sign and Opposite Sign data samples are shown in Figure 3.33. The Dalitz control region distribution also shows the same smooth falling behavior in the signal region than the other samples. Furthermore, compatibility between the Dalitz control region, Same Sign and Opposite Sign data, respectively, in the non-blinded region is observed.

It can be seen that all distributions representing the remaining background after the complete selection have a wide peak at low B mass and a smooth falling behaviour in the signal region. There is no evidence for possible peaking backgrounds in the signal region.

3.7 Selection efficiency

The signal efficiency is estimated using signal MC simulation except for the contributions of tracking, Particle Identification (PID), trigger efficiency and badly simulated offline selection variables, which are estimated using data driven techniques due to imperfections of the simulation. Thus, the PID and Hlt2 requirements are removed from the stripping selection used on the simulated signal samples to allow for a custom treatment.

The MC signal efficiency and corrections are computed on truth-matched events to avoid biases in the final results. The truth-matching requirements have been described in

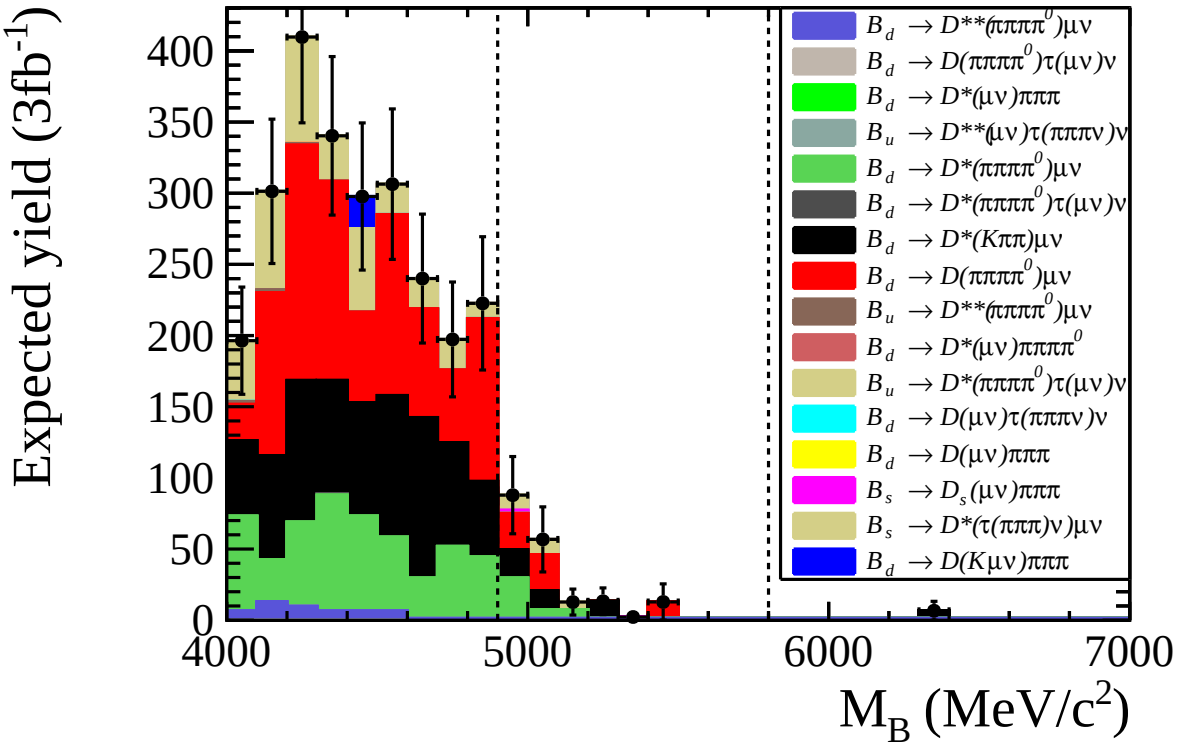

Figure 3.31: Exclusive samples B invariant mass distributions once the complete offline selection is applied. The plot shows the expected yields according to an integrated luminosity of $3\,\mathrm{fb}^{-1}$

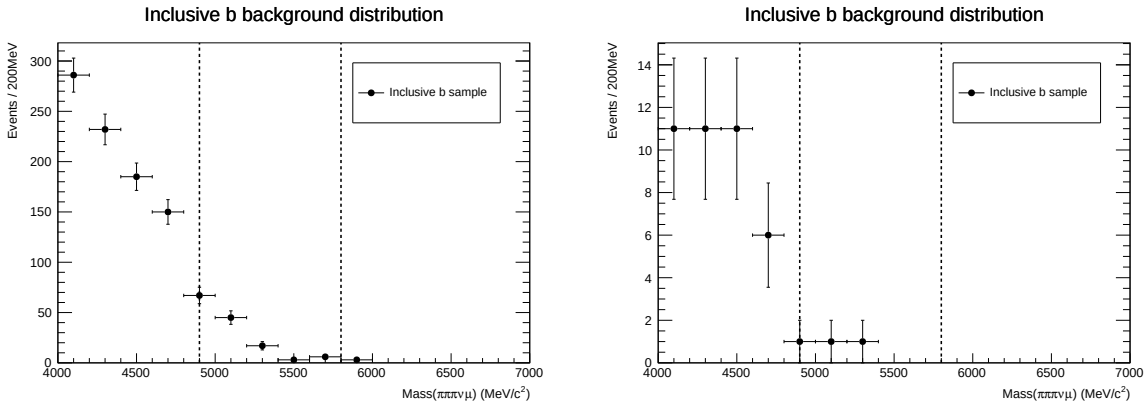

Figure 3.32: B invariant mass distribution of the $b\bar{b}$ inclusive sample. Left: After the cut-based preselection. Right: once the complete offline selection is applied.

Section 3.2.2.

The overall selection efficiency is decomposed into the following contributions:

$$\epsilon^{\text{Total}} = \epsilon^{\text{Acc}} \times \epsilon^{\text{Reco+Strip}} \times \epsilon^{\text{PID}} \times \epsilon^{Trigger} \times \epsilon^{\text{Sel}} \tag{3.8}$$

- $\boldsymbol{\epsilon^{\text{Acc}} = \frac{N^{\text{accepted}}}{N^{\text{generated}}}}$**:** The geometrical acceptance efficiency contains the fraction of

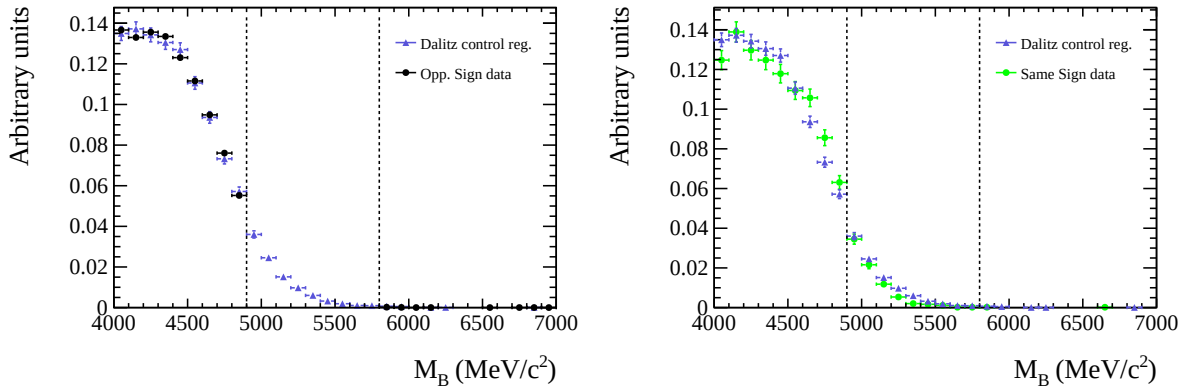

Figure 3.33: Left: B mass distribution of the Dalitz control region (unblinded) and the Opposite-Sign (blinded) data once the offline selection have been applied.
Right: B mass distribution of the Dalitz control region (unblinded) and the Same-Sign data once the offline selection have been applied.

candidates falling inside detector acceptance over the amount of generated candidates. The value of this efficiency is provided by the simulation working group and the different values are shown in Table 3.18.

- $\boldsymbol{\epsilon^{\text{Reco+Strip}} = \frac{N^{\text{Reco+Strip}}}{N^{\text{accepted}}}}$**:** Track reconstruction, vertexing and stripping (without PID requirements) efficiency. It evaluates the fraction of remaining candidates after the reconstruction and stripping requirements over the candidates within the detector acceptance. Corrections on the efficiency must be added due to the mismodeling of the tracking system in the simulations. The corrections are discussed in Section 3.7.1.

- $\boldsymbol{\epsilon^{\text{PID}} = \frac{N^{\text{PID}}}{N^{\text{Reco+Strip}}}}$**:** Efficiency of the requirements on the PID variables. It evaluates the remaining candidates after the applications of the PID requirements over the number of stripped candidates. Corrections and systematic uncertainties are discussed in Section 3.7.2.

- $\boldsymbol{\epsilon^{\text{Trigger}} = \frac{N^{\text{Trigger}}}{N^{\text{PID}}}}$**:** Efficiency of the trigger line requirements ($L0$, $HLT1$ and $HLT2$). It evaluates the number of triggered candidates over the number of candidates passing the PID requirements. Corrections and systematic uncertainties are discussed in Section 3.7.3.

- $\boldsymbol{\epsilon^{\text{Sel}} = \frac{N^{\text{Selected}}}{N^{\text{Trigger}}}}$**:** Efficiency of the subsequent offline selection. It evaluates the number of candidates passing the offline selection over the triggered events. Corrections are applied looking at the Data-MC agreement of the variables used in the offline selection, reported in 3.7.4.

3.7.1 Tracking reconstruction efficiency correction

The simulated tracking efficiency does not reproduce well the data efficiency and is corrected using the track-based correction factors provided by the LHCb tracking working

Sample	Year	ϵ^{Acc}
$B_s^0 \to \tau^\pm \mu^\mp$	2011	0.07767 ± 0.00019
$B_s^0 \to \tau^\pm \mu^\mp$	2012	0.08024 ± 0.00017
$B^0 \to \tau^\pm \mu^\mp$	2011	0.07703 ± 0.00014
$B^0 \to \tau^\pm \mu^\mp$	2012	0.07958 ± 0.00016
$B^0 \to D^- \pi^+$	2011	0.15817 ± 0.00041
$B^0 \to D^- \pi^+$	2012	0.16090 ± 0.00040

Table 3.18: Generation acceptance efficiency for $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC signal samples.

group [72]. These factors ($\mu_i \pm \sigma_i$) are data over simulated efficiencies ratios determined using $J/\psi \to \mu\mu$ events in bins of track momentum (p) and pseudorapidity (η). They are shown in Figure 3.34.

For each of the four candidate tracks, an average correction factor is computed as:

$$f^{track} = \sum_i^{bins} \frac{N^{\text{tracks}}_{\text{bin i}}}{N^{\text{tracks}}_{\text{total}}} \mu_i \pm \sqrt{\sum_i^{bins} \left(\frac{N^{\text{tracks}}_{\text{bin i}}}{N^{\text{events}}_{\text{total}}} \sigma_i \right)^2} \tag{3.9}$$

where the average is computed over the $N^{\text{tracks}}_{\text{total}}$ candidates passing the modified stripping selection (with no HLT2 and PID requirements). The simulated efficiency is corrected by the multiplication of the correction factors among the four tracks which are presented in Table 3.19. The uncertainty associated to the per track correction factors exclusively corresponds to the size of the samples used in their computation, therefore their uncertainty propagation is taken as a systematic uncertainty. As recommended by the tracking working group, an extra systematic uncertainty of 0.4% per track is assigned to each track correction factor.

The resulting correction factors used to correct the signal efficiencies are at the few per mille level for the data taken in 2011 and go up to a few per cent in 2012. As a cross-check, the overall corrections have also been computed for events passing the full selection. They are shown in Table 3.19 as well and are found to be compatible with the ones computed after stripping.

3.7.2 PID efficiency correction

Strong PID requirements are applied in the stripping selection for $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ on each of the pions and the muon:

$$\texttt{ProbNNpi}(\pi)>0.55 \text{ and } \texttt{PIDmu}(\mu)>0.0 \text{ and } \texttt{hasMuon}(\mu)==1.$$

Furthermore, PID requirements are applied during the stripping and offline selection of the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$:

$$\texttt{ProbNNpi}(\pi)>0.55 \text{ and } \texttt{PIDK}(K)>15.0.$$

Due to imperfect modeling of the underlying event in the simulation, which affects the PID performance, the efficiencies of the PID requirements differ for data and simulation.

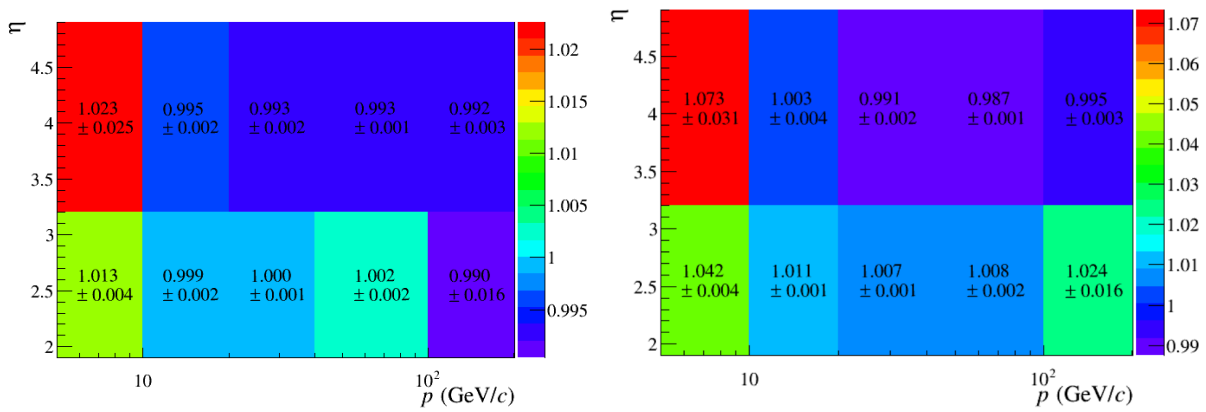

Figure 3.34: Data-Simulation track reconstruction efficiency ratio in bins of pseudo-rapidity and momentum for 2011 (left) and 2012 (right) provided by the tracking working group.

Sample	Year	f^{Track} (after stripping)	f^{Track} (after complete selection)
$B_s^0 \to \tau^\pm\mu^\mp$	2011	$1.004 \pm (0.003 \pm 0.008)$syst	$1.005 \pm (0.003 \pm 0.008)$syst
$B^0 \to \tau^\pm\mu^\mp$	2011	$1.004 \pm (0.003 \pm 0.008)$syst	$1.006 \pm (0.003 \pm 0.008)$syst
$B_s^0 \to \tau^\pm\mu^\mp$	2012	$1.048 \pm (0.004 \pm 0.008)$syst	$1.053 \pm (0.004 \pm 0.008)$syst
$B^0 \to \tau^\pm\mu^\mp$	2012	$1.049 \pm (0.004 \pm 0.008)$syst	$1.054 \pm (0.004 \pm 0.008)$syst
$B^0 \to D^-\pi^+$	2011	$1.001 \pm (0.003 \pm 0.008)$syst	$1.002 \pm (0.003 \pm 0.008)$syst
$B^0 \to D^-\pi^+$	2012	$1.040 \pm (0.004 \pm 0.008)$syst	$1.042 \pm (0.004 \pm 0.008)$syst

Table 3.19: Correction factors for $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC signal samples.

Thus, the PIDCalib tool [73] is used to compute the efficiency corresponding to the set of PID requirements in the selection. It uses a set of custom MC and data *calibration samples* selected with the same PID requirements as the one applied in the signal selection. It estimates the efficiencies in bins of particle's momentum (p) and pseudorapidity (η) to account for difference in the kinematics of the calibration and signal samples. These calibration samples are sets of specific decays used internally by the software to compute the efficiency maps. The calibration samples used for the pions PID are from decays containing D^*, and for the muons PID from J/ψ decays.

Then, the efficiency maps are applied to the *reference samples*, which are the samples of which is desired to know the PID efficiency. The reference samples given as input to the PIDCalib tool are $B \to \tau^\pm\mu^\mp$ and $B^0 \to D^-\pi^+$ candidates selected by the modified stripping where the PID requirements have been removed. The PIDCalib tool computes the efficiency corresponding to a specific PID requirement given the true type of the particle on which the requirement is applied.

The PIDCalib package recipe advises to add as a systematic uncertainty $1/\sqrt{N}$, where N is the number of candidates of the reference sample. In addition, the effect of the binning scheme used to build the efficiency maps is estimated by running the PIDCalib tools with bins reduced or enlarged by a factor 2. The difference between these two

extremes is taken as another systematic uncertainty. Finally, the cut and count efficiency statistical uncertainty is assigned as a statistical uncertainty to the corrected efficiency.

The resulting computed efficiencies are shown in Table 3.20.

Sample	Year	Mag	cut&count (%)	PIDCalib (%)			
			$\epsilon_{PID} \pm \delta_{\text{stat}}$	$\epsilon_{PID}^{\text{Bins0}}$	$\epsilon_{PID}^{\text{Bins0x2}}$	$\epsilon_{PID}^{\text{Bins0/2}}$	$\epsilon_{PID} \pm \delta_{\text{stat}} \pm \delta_{\text{syst}}$
$B_s^0 \to \tau\mu$ Tauola	2011	Dw	77.17 ± 0.41	75.79	76.05	75.16	$75.79 \pm 0.41 \pm 1.05$
	2011	Up	76.69 ± 0.42	74.84	75.04	74.33	$74.84 \pm 0.42 \pm 0.98$
	2012	Dw	77.74 ± 0.30	75.91	76.15	75.35	$75.91 \pm 0.30 \pm 0.83$
	2012	Up	78.20 ± 0.32	75.61	75.91	75.02	$75.61 \pm 0.32 \pm 0.89$
$B^0 \to \tau\mu$ Tauola	2011	Dw	77.23 ± 0.41	75.90	76.17	75.30	$75.90 \pm 0.41 \pm 1.02$
	2011	Up	76.40 ± 0.42	75.11	75.28	74.52	$75.11 \pm 0.42 \pm 1.02$
	2012	Dw	77.93 ± 0.33	76.06	76.28	75.38	$76.06 \pm 0.33 \pm 0.96$
	2012	Up	77.61 ± 0.33	75.71	75.99	75.05	$75.71 \pm 0.33 \pm 0.94$
$B^0 \to D(K\pi\pi)\pi$	2011	Dw	53.97 ± 0.73	53.59	53.34	54.27	$53.59 \pm 0.46 \pm 1.23$
	2011	Up	53.15 ± 0.73	52.56	52.34	53.14	$52.56 \pm 0.45 \pm 1.17$
	2012	Dw	59.51 ± 0.61	54.92	54.67	55.41	$54.92 \pm 0.36 \pm 0.92$
	2012	Up	60.12 ± 0.54	55.01	54.61	55.65	$55.01 \pm 0.40 \pm 1.08$

Table 3.20: Efficiency of the PID requirements of $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ and $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC samples.

3.7.3 Trigger efficiency

The trigger efficiency is computed using data driven techniques. Two trigger conditions are used in order to get the different trigger samples:

- **Trigger On Signal (TOS)**: The *TOS* condition that requires that the signal properties are enough to fire the specified trigger line is imposed at each level (L0, HLT1 and HLT2).
- **Trigger Independent of Signal (TIS)**: The *TIS* condition that requires that the underlying event properties alone (not accounting for the signal candidates) are enough to fire the specified trigger line is imposed at each level (L0, HLT1 and HLT2).

3.7.3.1 $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ normalization channel trigger efficiency

The overall efficiency for the L0, HLT1 and HLT2 trigger requirements is estimated on the stripped $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ candidates prior to applying any offline selection cuts. The *TisTos* technique [74] is used to extract the efficiency on data and the simulation is used to assess the associated systematic uncertainty. It relies on a set of *TIS* events which are events triggered independently of the signal candidate as a trigger unbiased sample to estimate the trigger efficiency.

The trigger requirements applied in the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ selection are:

`L0Hadron_TOS(`B`)` and `TrackAllL0_TOS(`B`)` and `Topo[2/3/4]BodyBBDT_TOS(`B`)`.

Therefore, the *TisTos* method applied here is simplified with respect to the one described in [74]. The *TisTos* efficiencies are evaluated on a *TIS* sample in bins of the B transverse momentum (p_{T}) as:

$$\epsilon_i^{TisTos} = \frac{n_i^{TIS\&TOS}}{n_i^{TIS}} \tag{3.10}$$

where n_i^{TIS} is the number of candidates in the p_{T} bin i satisfying the following *TIS* conditions:

`L0Global_TIS(`B`)` and `Hlt1Phys_TIS(`B`)` and `Hlt2Phys_TIS(`B`)`

and $n_i^{TIS\&TOS}$ is the number of candidates in the p_{T} bin i passing both *TIS* and *TOS* conditions. Both n_i^{TIS} and $n_i^{TIS\&TOS}$ are obtained by fitting a Gaussian parametrizing the signal and an exponential accounting for the background to the reconstructed B mass distributions in each p_{T} bin.

The obtained ϵ_i^{TisTos} efficiencies are presented in Figure 3.35.

The overall efficiency is obtained by applying the ϵ_i^{TisTos} efficiencies on the selected *TOS* sample:

$$\varepsilon^{meas} = \frac{\sum_i^{bins} n_i^{TOS}}{\sum_i^{bins} (n_i^{TOS}/\epsilon_i^{TisTos})} \tag{3.11}$$

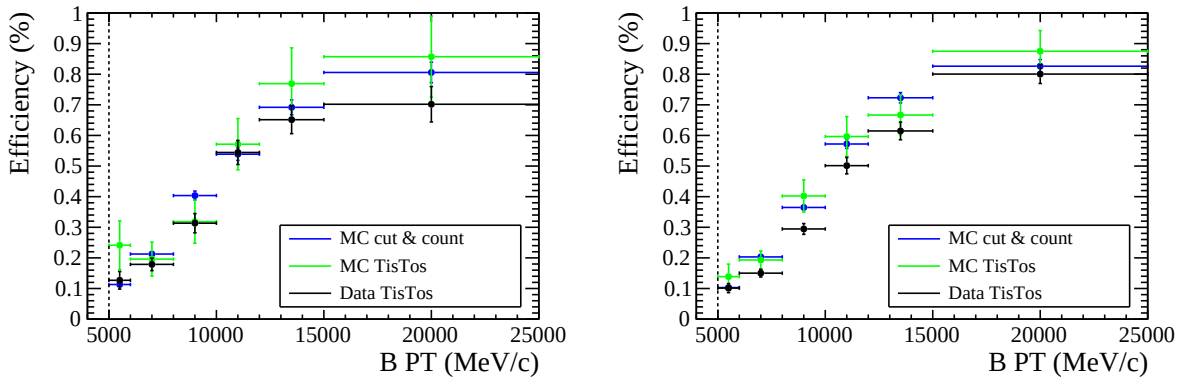

Figure 3.35: $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ 2011 (left) and 2012 (right) trigger efficiency per bin. In black, the measured *TisTos* efficiency in the data; in blue, the true efficiency observed in the simulation; in green, the measured efficiency in the simulation using the *TisTos* efficiencies from the MC.

where the n_i^{TOS} are the number of triggered candidates in each p_{T} bin i obtained by fitting a Gaussian and an exponential to the reconstructed B mass distribution. The measured efficiencies are $(29.6\pm1.9)\%$ and $(26.5\pm1.2)\%$ for 2011 and 2012 samples respectively.

The method is validated using the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ simulated samples. The efficiencies measured in the simulation using the method described above for the data are compared to the true MC efficiencies defined as:

$$\varepsilon^{true} = \frac{N^{TOS}}{N^{Tot}}, \tag{3.12}$$

where N^{TOS} is the number of candidates satisfying the trigger requirements and N^{Tot} is the total number of candidates. The true and measured efficiencies in the 2011 and 2012 samples are given in Table 3.21. The quadratic sum of the differences on the true and the

sample	$\varepsilon^{true}(\%)$	$\varepsilon^{meas}(\%)$	$\varepsilon^{true}(\%) - \varepsilon^{meas}(\%)$
MC 2011	32.3 ± 0.6	33.7 ± 4.6	1.4 ± 4.6
MC 2012	31.6 ± 0.5	33.5 ± 2.7	1.9 ± 2.8

Table 3.21: The $B^0 \to D^-\pi^+$ MC trigger true and measured efficiencies and their differences.

MC map efficiencies central values and their corresponding uncertainty on the differences are taken as the systematic uncertainties associated to the efficiency estimation. The final trigger efficiencies are then:

- $\varepsilon_{2011} = (29.6 \pm 1.9 \pm 4.8)\%$ and
- $\varepsilon_{2012} = (26.5 \pm 1.2 \pm 3.4)\%$.

3.7.3.2 $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ L0 and HLT1 trigger efficiency

The overall efficiency for the L0 and HLT1 trigger requirements is estimated on the stripped $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ candidates prior to applying any offline selection cuts.

The signal selection requires that the μ candidate fires the `L0Muon` channel as well as the `Hlt1TrackMuon` or `Hlt1SingleMuonHighPT` lines. The efficiency of this requirement is estimated using trigger efficiency maps made with probe muons extracted from $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ decays. The maps are constructed in bins of the muon transverse momentum p_{T} (μ) and impact parameter $IP(\mu)$. The simulation is used to assess the associated systematic uncertainty.

Trigger efficiency mapping
The probe muons used to build the maps are taken from $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ decays selected by the `BToJpsiK` stripping line. To ensure that the probe muons are representative of the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ signal muon, the cuts used to select the muon in the `StrippingB2XTau_TauMu_TOSLine` stripping line are applied:

- `P`$> 6000\,\mathrm{MeV}/c$,
- `PT`$> 1000\,\mathrm{MeV}/c$,
- `IPCHI2_OWNPV`> 16,
- `TRACK_CHI2NDOF`< 3,
- `TRACK_GhostProb`< 0.3,
- `PIDmu`> 0.
- `hasMuon`$== 1$.

In addition, constraints on the B and J/ψ invariant masses are also applied to reduce the background:

- $2600\,\mathrm{MeV}/c^2 < M_{J/\psi} < 3200\,\mathrm{MeV}/c^2$ and
- $4500\,\mathrm{MeV}/c^2 < M_B < 6000\,\mathrm{MeV}/c^2$,

and all candidates are required to have fired one Hlt2 physics line (`Hlt2Phys_TOS(`B`)` condition).

To select trigger unbiased muons, the candidates are required to be *TIS* w.r.t. to the L0 and HLT1 triggers. The following requirements are applied:

`L0Hadron_TIS(`B`)` or `L0Electron_TIS(`B`)` or `L0Photon_TIS(`B`)`, and
`TrackAllL0_TIS(`B`)` or `TrackMuon_TIS(`B`)` or `SingleMuonHighPT_TIS(`B`)` or
`TrackPhoton_TIS(`B`)`

The efficiency is computed in 5 bins of p_{T} (μ) and 4 bins of $IP(\mu)$. The cuts applied by the L0 and HLT1 trigger are summarized in Table 3.22 and the binning scheme is chosen such that thinner bins are used around the cut values to better catch the efficiency curve turn-round:

- p_{T} (μ) : $[0., 1000., 1750., 2250., 3500., 4800.]$ MeV
- $IP(\mu)$: $[0., 0.107, 0.12, 0.22]$ mm

	2011	2012
	`L0Muon`	
$L0$PT	$> 1480\,\text{MeV}/c$	$> 1760\,\text{MeV}/c$
	`HLT1TrackMuon`	
`Track`PT	$> 1000\,\text{MeV}/c$	$> 1000\,\text{MeV}/c$
`Track`P	$> 8000\,\text{MeV}/c$	$> 3000\,\text{MeV}/c$
`Track`χ^2	< 2.0	< 2.5
`Track`IP	$> 0.1\,\text{mm}$	$> 0.1\,\text{mm}$
	`HLT1SingleMuonHighPT`	
`Track`PT	$> 4800\,\text{MeV}/c$	$> 4800\,\text{MeV}/c$
`Track`P	$> 8000\,\text{MeV}/c$	$> 3000\,\text{MeV}/c$
`Track`χ^2	< 4.0	< 3.0

Table 3.22: Muon requirements placed in the triggers `L0Muon`, `HLT1TrackMuon` and `HLT1SingleMuonHighPT`.

In each bin, the efficiency is computed as:

$$\epsilon_i = \frac{n_i^{probe}}{N_i^{probe}} \tag{3.13}$$

where N_i^{probe} is the total number of muons in bin i and n_i^{probe} is the number of muon satisfying the L0 and HLT1 trigger conditions defined above. The number of muons are extracted from a fit to the B invariant mass in each bin where signal is parametrized using a double tailed Crystal Ball distribution and the background is described by an exponential. The obtained efficiencies are presented in Figure 3.36.

Efficiency estimation

The L0 and HLT1 trigger efficiency for the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ candidates is computed as:

$$\varepsilon = \frac{\sum_i^{bins} n_i\epsilon_i}{\sum_i^{bins} n_i} \tag{3.14}$$

where n_i is the number of $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ candidates with a muon in bin i.

The uncertainty $\delta\varepsilon$ on ε is given by:

$$\delta\varepsilon^2 = \frac{1}{N^2}\left[\sum_i^{bins}(\epsilon_i - \varepsilon)^2\,\delta n_i^2 + \sum_i^{bins} n_i^2\delta\epsilon_i^2\right] \tag{3.15}$$

where $N = \sum_i^{bins} n_i$ is the total number of events in the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ sample, and δn_i and $\delta\epsilon_i$ are the uncertainties on the number of events and the mapped efficiency in each bin respectively.

When the efficiency maps obtained with $B^\pm \to J/\psi(\to \mu^\pm\mu^\mp)K^\pm$ data are applied to the $B^0_s \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ simulated sample, the resulting efficiencies are $(69.58 \pm 0.36)\%$ and $(73.93 \pm 0.22)\%$ for the 2011 and 2012 samples. For the $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ mode, $(69.58 \pm 0.36)\%$ and $(73.79 \pm 0.23)\%$ for the 2011 and 2012 samples is obtained.

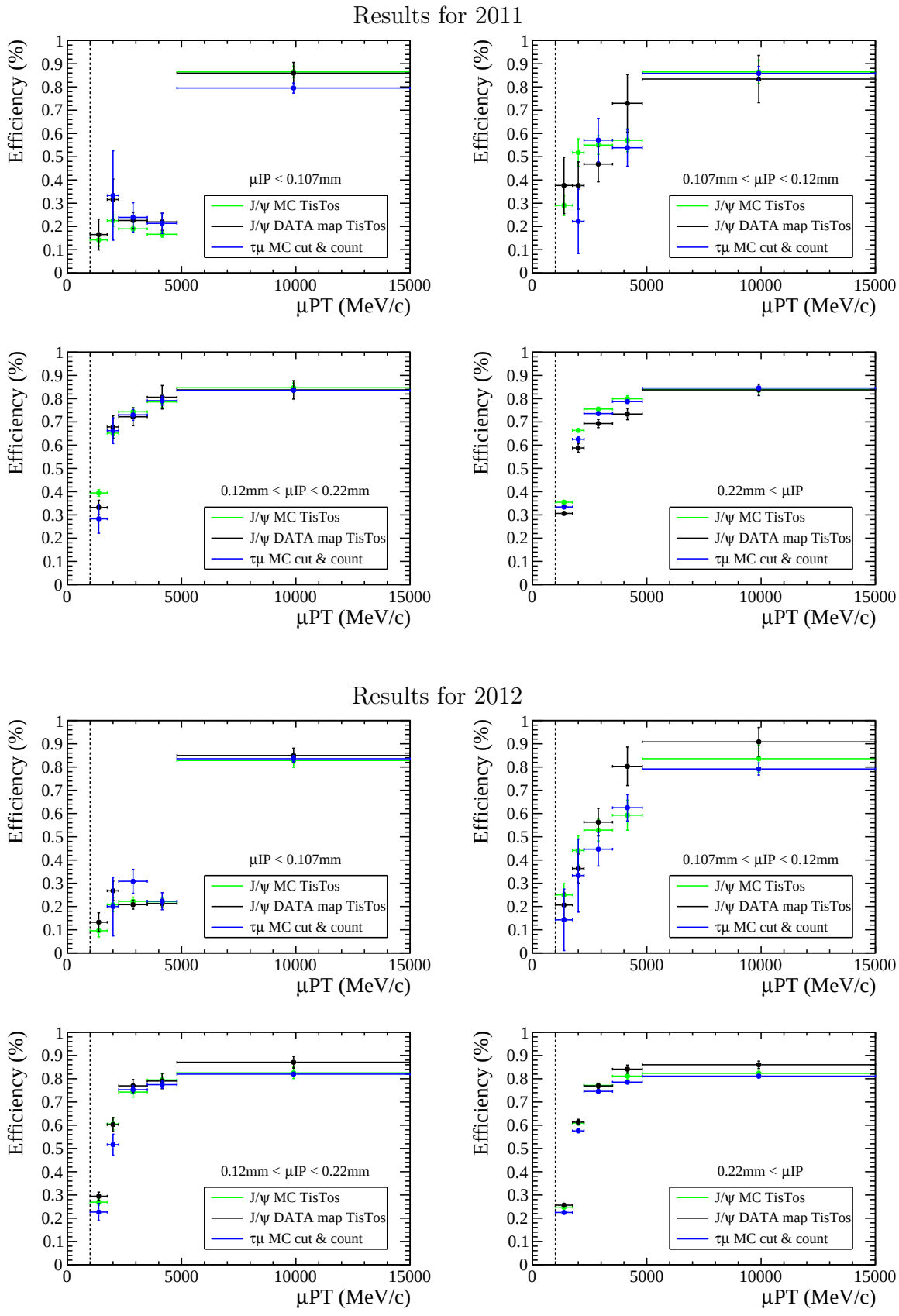

Figure 3.36: Results of the $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ muon L0 and HLT1 *TisTos* efficiency data map for the different bins (black). The $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ MC cut and count efficiency per bin (blue) and the $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ *TisTos* efficiency per bin (green) are also shown.

Systematic uncertainty
The method is validated using a simulated sample of $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ to compute the efficiency maps. The efficiencies ε^{MCcorr} obtained using these maps are compared to the true MC efficiencies defined as:

$$\varepsilon^{true} = \frac{N^{TOS}}{N^{Tot}} \tag{3.16}$$

where N^{TOS} is the number of candidates satisfying the trigger requirements and N^{Tot} is the total number of candidates. The true and MC corrected efficiencies in the 2011 and 2012 samples are given in Table 3.23. Their differences are taken as the systematic

sample	$\varepsilon^{true}(\%)$	$\varepsilon^{MCcorr}(\%)$	$\varepsilon^{true}(\%) - \varepsilon^{MCcorr}(\%)$
$B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011	74.04 ± 0.34	73.48 ± 0.17	$+0.56 \pm 0.38$
$B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012	71.71 ± 0.27	72.00 ± 0.18	-0.29 ± 0.32
$B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011	73.54 ± 0.34	73.40 ± 0.17	$+0.14 \pm 0.38$
$B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012	71.02 ± 0.29	71.90 ± 0.19	-0.88 ± 0.35

Table 3.23: The $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ trigger true and MC corrected efficiencies and their differences. ε^{true} is defined in Equation 3.16 and ε^{MCcorr} is the efficiency estimated using efficiency maps made with the $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ simulated sample.

uncertainties associated to the efficiency estimation. The final trigger efficiencies are then:

- $\varepsilon^{L0+HLT1}_{B_s^0 \to \tau^{\pm}\mu^{\mp},2011} = (69.58 \pm 0.36 \pm 0.68)\%$,
- $\varepsilon^{L0+HLT1}_{B_s^0 \to \tau^{\pm}\mu^{\mp},2012} = (73.99 \pm 0.22 \pm 0.43)\%$,
- $\varepsilon^{L0+HLT1}_{B^0 \to \tau^{\pm}\mu^{\mp},2011} = (69.58 \pm 0.36 \pm 0.41)\%$,
- $\varepsilon^{L0+HLT1}_{B^0 \to \tau^{\pm}\mu^{\mp},2012} = (73.79 \pm 0.23 \pm 0.94)\%$.

3.7.3.3 $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ HLT2 trigger efficiency

The HLT2 trigger efficiencies are taken directly from the simulation. They are estimated on the stripped $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ sample on which the L0 and HLT1 trigger conditions are applied prior to any other offline selection cuts. The agreement between the HLT2 behaviour in data and simulation is checked with the $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ control samples. The obtained efficiencies are reported in Table 3.24 and are compatible. The quadratic sum of the difference between data and simulation and the difference uncertainty is taken as the systematic error on the HLT2 efficiency estimation.

The HLT2 trigger efficiencies are

- $\varepsilon^{HLT2}_{B_s^0 \to \tau^{\pm}\mu^{\mp},2011} = (91.67 \pm 0.25 \pm 0.55)\%$,
- $\varepsilon^{HLT2}_{B_s^0 \to \tau^{\pm}\mu^{\mp},2012} = (95.68 \pm 0.15 \pm 0.46)\%$,
- $\varepsilon^{HLT2}_{B^0 \to \tau^{\pm}\mu^{\mp},2011} = (91.18 \pm 0.26 \pm 0.55)\%$,
- $\varepsilon^{HLT2}_{B^0 \to \tau^{\pm}\mu^{\mp},2012} = (95.94 \pm 0.15 \pm 0.46)\%$.

year	simulation (%)	data (%)	data - simulation (%)
2011	91.23 ± 0.07	91.49 ± 0.49	0.26 ± 0.49
2012	94.21 ± 0.07	94.49 ± 0.36	0.28 ± 0.37

Table 3.24: The HLT2 trigger efficiencies (B `TopoMu[2/3/4]BodyBBDT_TOS` requirements) observed with the $B^{\pm} \to J/\psi(\to \mu^{\pm}\mu^{\mp})K^{\pm}$ control channel in data and simulation (in %).

3.7.4 Offline selection efficiency

The offline selection efficiency is estimated with the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulated samples. However, the non-perfect simulation of the variables used in the offline selection can bias the estimated efficiency. The data-MC agreement of the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ selection is checked using the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC and data samples. The ratio of data and MC $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ distributions of relevant variables are used to re-weight the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ samples and the offline selection efficiencies are measured on the re-weighted samples.

3.7.4.1 Individual re-weighting

The data-MC agreement of all variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ selection but the one related to the τ decay model (studied in Section 3.7.4.3) is checked. Their distributions in the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ samples are shown in Appendix B. For each variable, a set of weights $(\omega_i \pm \delta\omega_i)$ are computed as the ratio of the numbers of data and MC $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ candidates in each bin.

For each variable, these weights are used to compute a corrected offline $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ selection efficiency:

$$\varepsilon = \frac{\sum_i^{bins} \omega_i n_i}{\sum_i^{bins} \omega_i N_i} \tag{3.17}$$

where $N_i \pm \delta N_i$ is the number of stripped and triggered $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ candidates in bin i and $n_i \pm \delta n_i$ is the number of $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ candidates satisfying the full offline selection.

The uncertainty $\delta\varepsilon$ on ε is given by:

$$\delta\varepsilon^2 = \frac{\sum_i^{bins} \omega_i^2 (1-2\varepsilon)\, \delta n_i^2 + \omega_i^2 \varepsilon^2 \delta N_i^2 + (n_i - N_i\varepsilon)^2\, \delta\omega_i^2}{\sum_i^{bins} (\omega_i N_i)^2} \tag{3.18}$$

It can be written as the quadratic sum of 2 terms, $\delta\varepsilon_{stat}$ due to the limited $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ sample size and $\delta\varepsilon_{weights}$ containing the uncertainties on the weights, due to the limited statistics of the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ data and MC samples:

$$\begin{aligned} \delta\varepsilon^2_{stat} &= \frac{\sum_i^{bins} \omega_i^2 (1-2\varepsilon)\, \delta n_i^2 + \omega_i^2 \varepsilon^2 \delta N_i^2}{\sum_i^{bins} (\omega_i N_i)^2} \\ \delta\varepsilon^2_{weights} &= \frac{\sum_i^{bins} (n_i - N_i\varepsilon)^2\, \delta\omega_i^2}{\sum_i^{bins} (\omega_i N_i)^2}. \end{aligned} \tag{3.19}$$

The efficiencies corrected with the weights of each variable are shown in Figure 3.37. They are compared to the raw offline selection efficiencies which are:

- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011: $(19.47 \pm 0.46\ (\text{stat}))\%$,
- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: $(17.51 \pm 0.33\ (\text{stat}))\%$,
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011: $(18.34 \pm 0.45\ (\text{stat}))\%$,
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: $(16.22 \pm 0.34\ (\text{stat}))\%$.

3.7.4.2 Iterative re-weighting

Data-MC agreement for a given variable is quantified using a χ^2 test:

$$\chi^2/\text{ndf} = \frac{1}{N_{bins}} \sum_i^{bins} \frac{\left(a_i^{MC} - a_i^{Data}\right)^2}{\left(\delta a_i^{MC} + \delta a_i^{Data}\right)^2}, \tag{3.20}$$

where $a_i^{MC} \pm \delta a_i^{MC}$ and $a_i^{Data} \pm \delta a_i^{Data}$ are the numbers of $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC and data candidates in each bin i of the given variable. The distributions of these χ^2/ndf for all variables are shown in Figure 3.38 for the 2011 and 2012 samples.

Due to correlations between the variables, the data-MC agreement of a sample re-weighted using a given variable can impact the data-MC agreement for other variables. The $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ corrected efficiency is obtained by re-weighting iteratively the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ samples. At each iteration, the product of the weights of all previous iterations is applied, the χ^2/ndf of all variables are re-evaluated and the variable with the worst χ^2/ndf is used to re-weight the sample and compute a new corrected efficiency.

The efficiencies obtained at each iteration are show on Figures 3.39. The corrected efficiencies are taken as the ones corresponding to the beginning of the plateau, e.g. after only 1 iteration in 2011 and 4 in 2012. The variables used in the iterative process are summarized in Table 3.25. The final χ^2/ndf distributions are shown in Figure 3.38.

Computing the uncertainty of the efficiency computed with this iterative process is impossible since all weights are computed with the same samples and are correlated. Therefore, the associated systematic uncertainty is taken as the spread of the potential next re-weighted efficiencies (corresponding to the gray crosses in Figure 3.39) when 1 and 4 iterations are done, respectively for 2011 and 2012.

Simulation 2011		**Simulation 2012**	
Iteration	Variable	Iteration	Variable
1	τ `BDTiso3`	1	τ `BDTiso1_1`
		2	μ `0_50_nc_sPT`
		3	`totCandidates`
		4	`Pions sum of BDTiso3`

Table 3.25: Variables used in the iterative re-weighting process.

The results, including the systematic uncertainties (Equation 3.19), are:

- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011: $(18.09 \pm 0.44\ (\text{stat}) \pm 0.26\ (\text{syst}))\%$,
- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: $(15.14 \pm 0.32\ (\text{stat}) \pm 0.29\ (\text{syst}))\%$,
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011: $(17.01 \pm 0.43\ (\text{stat}) \pm 0.26\ (\text{syst}))\%$,
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: $(13.82 \pm 0.32\ (\text{stat}) \pm 0.29\ (\text{syst}))\%$.

3.7.4.3 τ decay model effect

In the cut-based preselection a requirement is placed in the invariant mass of opposite sign pions. The τ decay model used in the analysis is `Tauola` tuned with BaBar data. Two other `Tauola` models, in this case tuned with CLEO data, are used to check how the τ decay modelization affects the selection efficiency. Figure 3.40 contains the invariant mass of the opposite sign pions after stripping for the 3 Models [75]: `TauolaBaBar`, `TauolaCleoStd` and `TauolaCleoIntrincate`. The difference is minimal. When the offline selection efficiency is computed with the CLEO tuned models the relative change in efficiency is $\sim 0.71\%$ and $\sim 0.30\%$ for 2011 and 2012 MC samples, respectively. The effects of the τ decay models is negligible compared to the uncertainties due to the corrections described in Section 3.7.4.

3.7.5 MC-truth matching inefficiency correction

Tables 3.26 and 3.27 contain the detailed efficiency computations and corrections of the signal 2011 and 2012 simulation samples of $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ respectively, this computation is done exclusively on truth-matched events.

- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011: 34 non-matched events corresponding to 2.3% of the sample;
- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: 81 non-matched events corresponding to 3.4% of the sample;
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: 40 non-matched events corresponding to 2.8% of the sample;
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012: 64 non-matched events corresponding to 3.3% of the sample.

Even though remaining non-matched events are not properly matched, they are indeed signal events. Being successful in surviving the complete selection, their behavior of these non-matched events is clearly that of signal. This fact is ratified by their M_B distribution as well, shown in Figure 3.41.

The signal unmatched event are then expected to have same selection efficiency than the truth-matched ones. Therefore the estimated signal efficiencies of Tables 3.26 and 3.27 are scaled to account for the signal unmatched events according to:

$$\epsilon = \epsilon_{\text{matched}} \left(1 + \frac{N_{\text{unmatched}}}{N_{\text{matched}}}\right) \tag{3.21}$$

where N_{matched} and $N_{\text{unmatched}}$ are the number of matched and unmatched candidates.

Neglecting the uncertainty on $\frac{N_{\text{unmatched}}}{N_{\text{matched}}}$ which is of the order of a few per mill, the uncertainty on the estimated signal efficiency is taken as:

$$\delta\epsilon = \delta\epsilon_{\text{matched}} \left(1 + \frac{N_{\text{unmatched}}}{N_{\text{matched}}}\right). \tag{3.22}$$

The results for the different $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ samples are:

- $\epsilon^{B^0_s \to \tau^{\pm}(\to\pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}}_{2011} = (1.61 \pm 0.04 \text{ stat} \pm 0.04 \text{ syst}) \cdot 10^{-4}$
- $\epsilon^{B^0_s \to \tau^{\pm}(\to\pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}}_{2012} = (1.55 \pm 0.03 \text{ stat} \pm 0.04 \text{ syst}) \cdot 10^{-4}$
- $\epsilon^{B^0 \to \tau^{\pm}(\to\pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}}_{2011} = (1.49 \pm 0.04 \text{ stat} \pm 0.03 \text{ syst}) \cdot 10^{-4}$
- $\epsilon^{B^0 \to \tau^{\pm}(\to\pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}}_{2012} = (1.37 \pm 0.03 \text{ stat} \pm 0.04 \text{ syst}) \cdot 10^{-4}$

Table 3.28 contains the detailed efficiency computation for the $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ selection. In this case, all events at the end of the selection are truth-matched therefore the selection efficiency remains the same than the estimated efficiency quoted in the corresponding tables.

3.7.6 Selection efficiency summary

The total Run I efficiency is the weighted average of the 2011 and 2012 efficiencies where the weights are proportional to the number of $b\bar{b}$ produced in LHCb each year, i.e. to the recorded luminosity given in Table 3.1 and the $b\bar{b}$ cross section which is assumed to vary linearly from the centre mass energy of 7 TeV to 8 TeV [76].

The total efficiencies for the decays $B^0_s \to \tau^{\pm}\mu^{\mp}$ and $B^0 \to \tau^{\pm}\mu^{\mp}$ are:

- $\epsilon^{B^0_s \to \tau^{\pm}(\to\pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}} = (1.57 \pm 0.03 \text{ (stat)} \pm 0.03 \text{ (syst)}) \cdot 10^{-4}$
- $\epsilon^{B^0 \to \tau^{\pm}(\to\pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}} = (1.40 \pm 0.03 \text{ (stat)} \pm 0.03 \text{ (syst)}) \cdot 10^{-4}$

The total $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ efficiency averaged from the 2011 and 2012 effiencies is:

- $\epsilon^{B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+} = (1.89 \pm 0.08 \text{ (stat)} \pm 0.19 \text{ (syst)}) \cdot 10^{-4}$

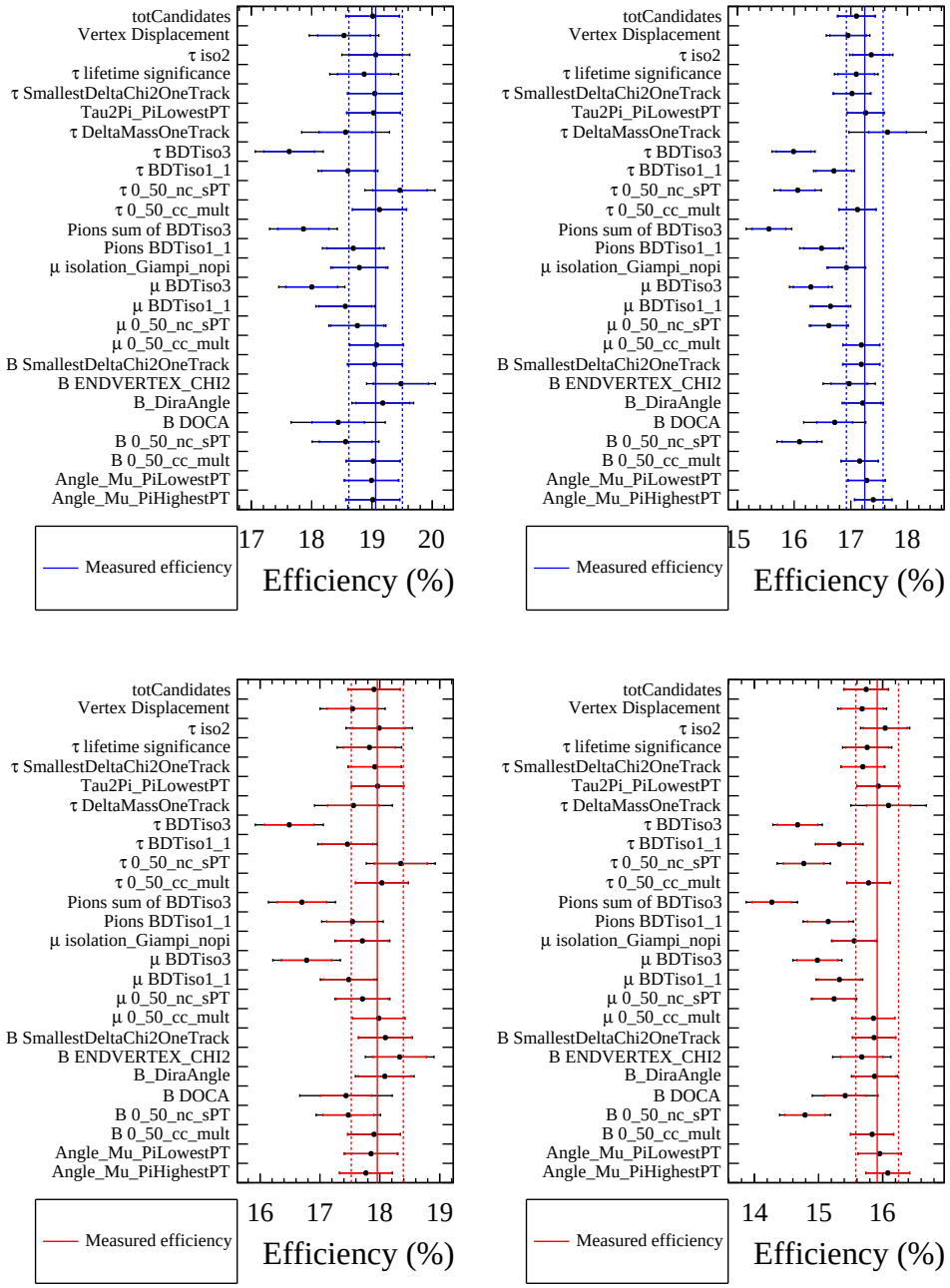

Figure 3.37: Individual re-weighting results. Top-left in blue: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011. Top-right in blue: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012. Bottom-left in red: $B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011. Bottom-right in red: $B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012. The black dots represent the offline selection efficiency when the weights of the corresponding variable are applied. The colored continuous line corresponds to the selection efficiency computed without weights and the colored dashed represents its statistical uncertainty. The colored error bars correspond to the statistical uncertainty of each re-weighted efficiency and the black error bars to full uncertainty.

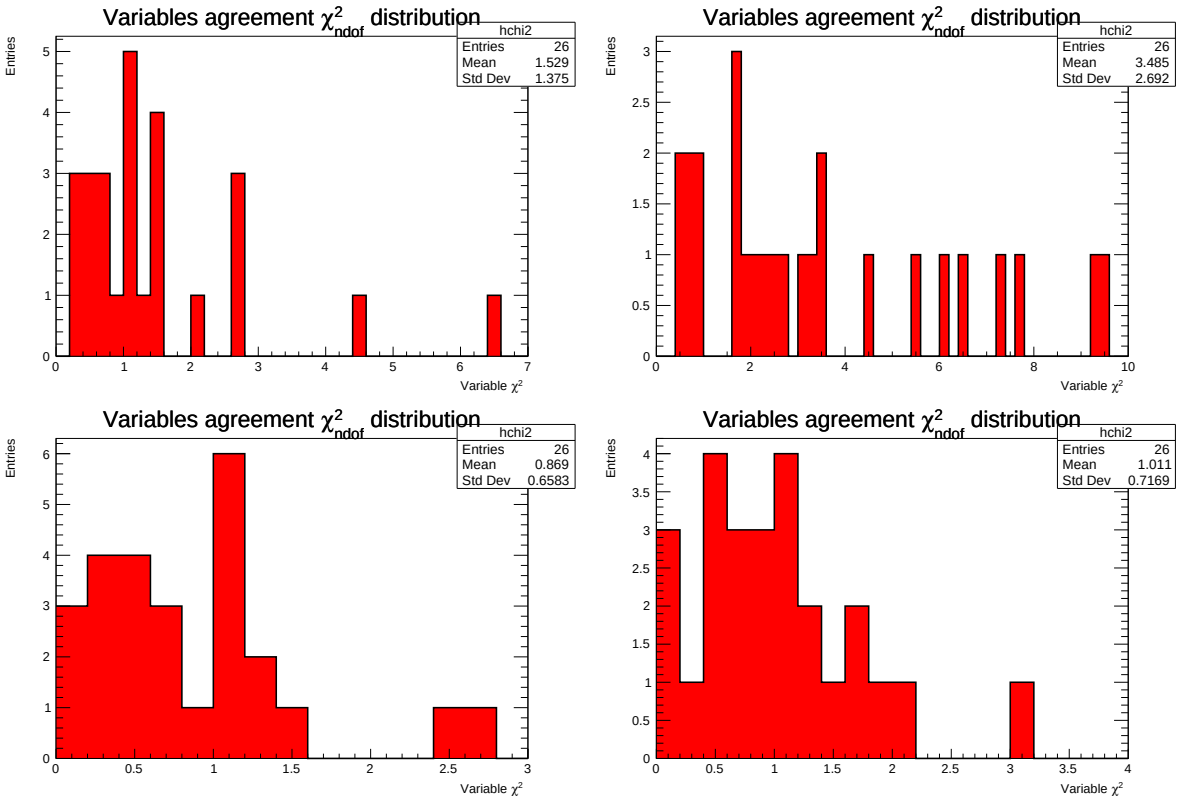

Figure 3.38: χ^2_{ndof} of the data-MC agreement for the different variables used in the offline selection process. Top-left: 2011 before the iterative re-weigting. Top-right: 2012 before the iterative re-weigting. Bottom-left: 2011 after 1 iterative re-weigting iteration. Bottom-right: 2012 after 4 iterative re-weigting iteration.

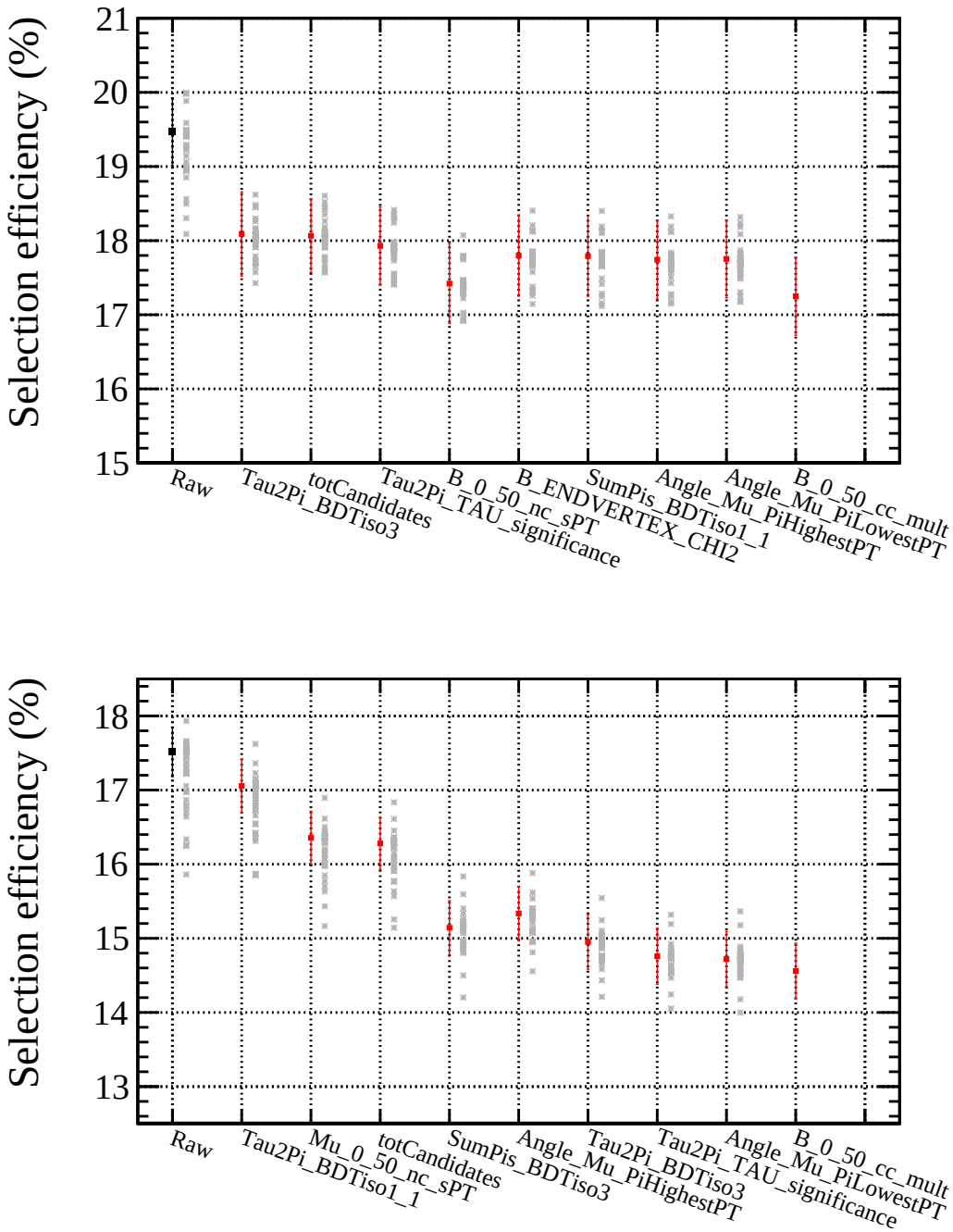

Figure 3.39: Efficiencies obtained in the iterative re-weighting process in 2011 (top) and 2012 (bottom). The point in black is the MC uncorrected efficiency. The red points are the efficiencies obtained at each iteration steps. The names of the variable added at each iteration are shown in the X-axis label. The gray crosses are the efficiencies obtained by re-weighting each variable at each iteration step.

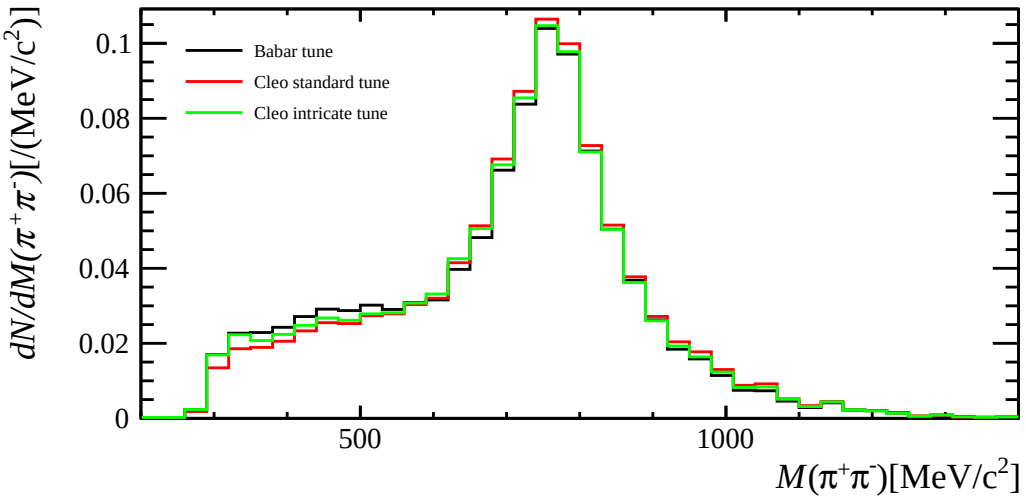

Figure 3.40: Tau decay models comparison of the invariant mass of the opposite sign pions distribution.

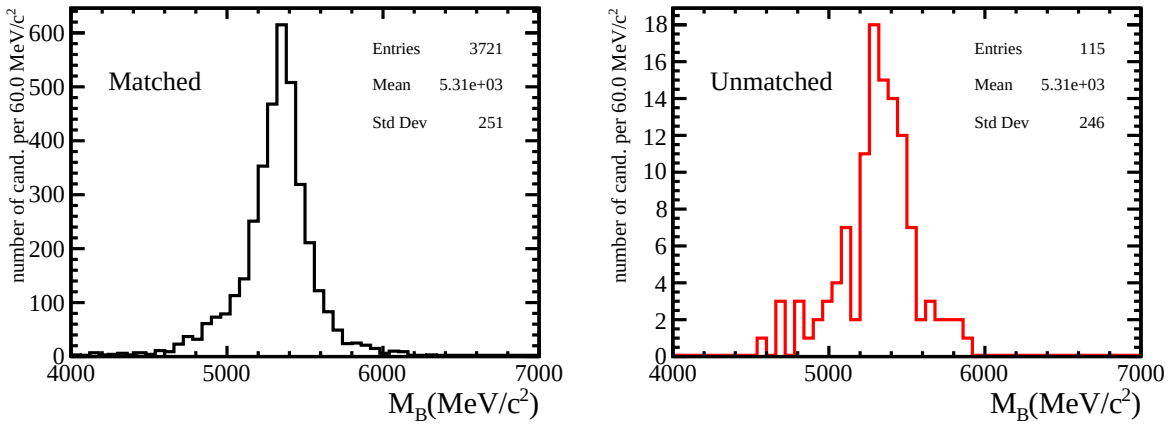

Figure 3.41: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012 MC sample matched (left) and non-matched (right) events M_B distribution once the complete selection have been applied.

$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011 Sim09b ID:13110004

Requirements	Remaining Yield	Cut & count efficiency	Estimated efficiency
Acceptance	615387	0.0777 ± 0.0002	$0.0777 \pm 0.0002 \pm 0.0000$
Reco+Strip	20630	0.0335 ± 0.0002	$0.0337 \pm 0.0002 \pm 0.0003$
PID	15871	0.7693 ± 0.0029	$0.7532 \pm 0.0029 \pm 0.0072$
$L0$ and $HLT1$	11751	0.7404 ± 0.0035	$0.6958 \pm 0.0036 \pm 0.0068$
$HLT2$	10772	0.9167 ± 0.0025	$0.9167 \pm 0.0025 \pm 0.0055$
Decay reconstruction	7479	0.6943 ± 0.0044	$0.6943 \pm 0.0044 \pm 0.0000$
Selection	1456	0.1947 ± 0.0046	$0.1809 \pm 0.0044 \pm 0.0026$
Total	1456	$(1.84 \pm 0.05) \cdot 10^{-4}$	$(1.58 \pm 0.04 \pm 0.04) \cdot 10^{-4}$

$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012 Sim09b ID:13110004

Requirements	Remaining Yield	Cut & count efficiency	Estimated efficiency
Acceptance	1098681	0.0802 ± 0.0002	$0.0802 \pm 0.0002 \pm 0.0000$
Reco+Strip	34900	0.0318 ± 0.0002	$0.0333 \pm 0.0002 \pm 0.0003$
PID	27206	0.7795 ± 0.0022	$0.7577 \pm 0.0022 \pm 0.0061$
$L0$ and $HLT1$	19511	0.7172 ± 0.0027	$0.7399 \pm 0.0022 \pm 0.0043$
$HLT2$	18668	0.9568 ± 0.0015	$0.9568 \pm 0.0015 \pm 0.0046$
Decay reconstruction	12932	0.6927 ± 0.0034	$0.6927 \pm 0.0034 \pm 0.0000$
Selection	2265	0.1751 ± 0.0033	$0.1514 \pm 0.0032 \pm 0.0029$
Total	2265	$(1.65 \pm 0.03) \cdot 10^{-4}$	$(1.50 \pm 0.03 \pm 0.04) \cdot 10^{-4}$

Table 3.26: $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011 and 2012 efficiency computation on truth-matched events.

$B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011 Sim09b ID:11110004

Requirements	Remaining Yield	Cut & count efficiency	Estimated efficiency
Acceptance	626085	0.0770 ± 0.0001	$0.0770 \pm 0.0001 \pm 0.0000$
Reco+Strip	20674	0.0330 ± 0.0002	$0.0332 \pm 0.0002 \pm 0.0003$
PID	15882	0.7682 ± 0.0029	$0.7551 \pm 0.0029 \pm 0.0072$
$L0$ and $HLT1$	11681	0.7355 ± 0.0035	$0.6958 \pm 0.0036 \pm 0.0041$
$HLT2$	10651	0.9118 ± 0.0026	$0.9118 \pm 0.0026 \pm 0.0055$
Decay reconstruction	7416	0.6963 ± 0.0045	$0.6963 \pm 0.0045 \pm 0.0000$
Selection	1360	0.1834 ± 0.0045	$0.1701 \pm 0.0043 \pm 0.0026$
Total	1360	$(1.67 \pm 0.05) \cdot 10^{-4}$	$(1.45 \pm 0.04 \pm 0.03) \cdot 10^{-4}$

$B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012 Sim09b ID:11110004

Requirements	Remaining Yield	Cut & count efficiency	Estimated efficiency
Acceptance	1013159	0.0796 ± 0.0002	$0.0796 \pm 0.0002 \pm 0.0000$
Reco+Strip	31324	0.0309 ± 0.0002	$0.0324 \pm 0.0002 \pm 0.0003$
PID	24361	0.7777 ± 0.0023	$0.7589 \pm 0.0023 \pm 0.0067$
$L0$ and $HLT1$	17303	0.7103 ± 0.0029	$0.7379 \pm 0.0023 \pm 0.0094$
$HLT2$	16600	0.9594 ± 0.0015	$0.9594 \pm 0.0015 \pm 0.0046$
Decay reconstruction	11445	0.6895 ± 0.0036	$0.6895 \pm 0.0036 \pm 0.0000$
Selection	1856	0.1622 ± 0.0034	$0.1382 \pm 0.0032 \pm 0.0029$
Total	1856	$(1.46 \pm 0.03) \cdot 10^{-4}$	$(1.32 \pm 0.03 \pm 0.04) \cdot 10^{-4}$

Table 3.27: $B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011 and 2012 efficiency computation on truth-matched events.

$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ 2011 Sim09b ID:11264001			
Requirements	Remaining Yield	Cut & count efficiency	Estimated efficiency
Acceptance	506999	0.1582 ± 0.0004	$0.1582 \pm 0.0004 \pm 0.0000$
Reco+Strip	9431	0.0186 ± 0.0002	$0.0186 \pm 0.0002 \pm 0.0002$
PID	5051	0.5356 ± 0.0051	$0.5307 \pm 0.0051 \pm 0.0085$
$L0$, $HLT1$ and $HLT2$	1629	0.3225 ± 0.0066	$0.2957 \pm 0.0187 \pm 0.0481$
Selection	758	0.4653 ± 0.0124	$0.4653 \pm 0.0124 \pm 0.0000$
Total	758	$(2.36 \pm 0.09) \cdot 10^{-4}$	$(2.15 \pm 0.15 \pm 0.35) \cdot 10^{-4}$
$B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ 2012 Sim08e ID:11264001			
Requirements	Remaining Yield	Cut & count efficiency	Estimated efficiency
Acceptance	1026667	0.1609 ± 0.0004	$0.1609 \pm 0.0004 \pm 0.0000$
Reco+Strip	16948	0.0165 ± 0.0001	$0.0172 \pm 0.0001 \pm 0.0001$
PID	10143	0.5985 ± 0.0041	$0.5496 \pm 0.0041 \pm 0.0070$
$L0$, $HLT1$ and $HLT2$	3204	0.3159 ± 0.0046	$0.2648 \pm 0.0118 \pm 0.0343$
Selection	1411	0.4404 ± 0.0088	$0.4404 \pm 0.0088 \pm 0.0000$
Total	1411	$(2.21 \pm 0.06) \cdot 10^{-4}$	$(1.77 \pm 0.09 \pm 0.23) \cdot 10^{-4}$

Table 3.28: $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ 2011 and 2012 efficiency computation on truth-matched events.

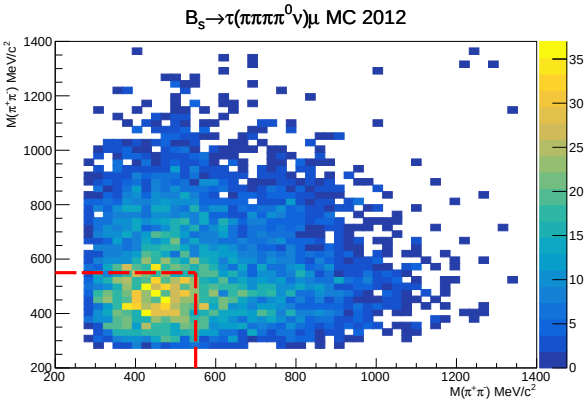

Figure 3.42: Dalitz plane for the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ simulation. The low $M_{\pi^+\pi^-}$ region (inside the red dashed line) is discarded in the preselection. The $\tau \to 3\pi\pi^0$ decay does not proceed through the resonances observed in the $\tau \to 3\pi$ mode.

3.7.7 $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ efficiency contribution

The offline selection process is optimized for the charged τ decay mode. However the $\tau \to 3\pi\pi^0$ mode brings a small but non-negligible contribution to the signal efficiency given that the effects of the hard cut on $M_{\pi^+\pi^-}$ (Figure 3.42) and the neutral isolation variables used in the preselection are compensated by the higher reconstruction efficiency (Table 3.11).

Tables 3.29 and 3.30 contain the signal efficiency computation of the $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu$ samples on truth-matched events.

As the kinematics between $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ are very similar, the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ corrections and relative systematics for truth-matched events are used for the π^0 mode. Denoting the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ cut and count efficiencies as

$$\begin{aligned} \epsilon^{mode(\tau^\pm\to\pi^\pm\pi^\mp\pi^\pm\nu)}_{Cut\&Count} &= a' \pm b'(stat) \pm 0(syst), \\ \epsilon^{mode(\tau^\pm\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)}_{Cut\&Count} &= d' \pm e'(stat) \pm 0(syst), \end{aligned} \tag{3.23}$$

and the corrected efficiency of the $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$ modes (2011 and 2012) as

$$\epsilon^{mode(\tau^\pm\to\pi^\pm\pi^\mp\pi^\pm\nu)}_{Corrected} = a \pm b(stat) \pm c(syst), \tag{3.24}$$

The $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ corrected efficiency are obtained by rescaling the corrected efficiency of the corresponding charged mode and the uncertainties are scaled such that the relative statistical uncertainty (e'/d') of the original sample and the relative systematic uncertainty (c/a) of the correction factor are preserved:

$$\epsilon^{mode(\tau^\pm\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)}_{Corrected} = \frac{a}{a'}d' \pm \frac{a}{a'}e'(stat) \pm \frac{d'}{a'}c(syst), \tag{3.25}$$

The estimated efficiencies are then scaled using the non-matched events as depicted in Section 3.7.5 using the corresponding number of unmatched events of the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ samples after the complete selection. The M_B distribution for the matched and non-matched events is shown in Figure 3.43.

$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2011 Sim09c ID:13110408		
Requirements	Cut and count	
	Remaining Yield	Efficiency
Acceptance	714521	0.0686 ± 0.0003
Reco+Strip	13667	0.0191 ± 0.0002
Trigger	11081	0.8108 ± 0.0034
Decay reconstruction	10277	0.9274 ± 0.0025
Offline Selection	783	0.0762 ± 0.0026
Total	783	$(7.52 \pm 0.27) \cdot 10^{-5}$
$B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012 Sim09c ID:13110408		
Requirements	Cut and count	
	Remaining Yield	Efficiency
Acceptance	909887	0.0710 ± 0.0003
Reco+Strip	16533	0.0182 ± 0.0001
Trigger	13452	0.8136 ± 0.0030
Decay reconstruction	12294	0.9139 ± 0.0024
Offline Selection	808	0.0657 ± 0.0022
Total	808	$(6.30 \pm 0.22) \cdot 10^{-5}$

Table 3.29: $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2011 and 2012 signal efficiency computation on truth matched events.

- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2011: 30 non-matched events corresponding to 3.7% of the sample;
- $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012: 25 non-matched events corresponding to 3.0% of the sample;
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012: 12 non-matched events corresponding to 2.7% of the sample;
- $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012: 28 non-matched events corresponding to 3.2% of the sample.

The resulting efficiencies per year are:

- $\epsilon_{2011}^{B_s^0 \to \tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp} = (6.70 \pm 0.23\ (\text{stat}) \pm 0.14\ (\text{syst})) \cdot 10^{-5}$,
- $\epsilon_{2012}^{B_s^0 \to \tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp} = (5.90 \pm 0.20\ (\text{stat}) \pm 0.14\ (\text{syst})) \cdot 10^{-5}$,
- $\epsilon_{2011}^{B^0 \to \tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp} = (5.76 \pm 0.28\ (\text{stat}) \pm 0.12\ (\text{syst})) \cdot 10^{-5}$,

$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2011 Sim09c ID:11110408		
Requirements	Cut and count	
	Remaining Yield	Efficiency
Acceptance	438007	0.0680 ± 0.0004
Reco+Strip	8108	0.0185 ± 0.0002
Trigger	6546	0.8074 ± 0.0044
Decay reconstruction	6062	0.9261 ± 0.0032
Offline Selection	417	0.0688 ± 0.0033
Total	417	$(6.47 \pm 0.32) \cdot 10^{-5}$
$B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012 Sim09c ID:11110408		
Requirements	Cut and count	
	Remaining Yield	Efficiency
Acceptance	1065137	0.0698 ± 0.0004
Reco+Strip	19116	0.0179 ± 0.0001
Trigger	15678	0.8202 ± 0.0028
Decay reconstruction	14413	0.9193 ± 0.0022
Offline Selection	855	0.0593 ± 0.0020
Total	855	$(5.61 \pm 0.19) \cdot 10^{-5}$

Table 3.30: $B^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2011 and 2012 signal efficiency computation on truth matched events.

- $\epsilon_{2012}^{B^0\to\tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp} = (5.24 \pm 0.18\ (\text{stat}) \pm 0.14\ (\text{syst})) \cdot 10^{-5}$.

The total efficiency is the weighted average of the 2011 and 2012 efficiencies where the weights are proportional to the number of $b\bar{b}$ produced in LHCb each year, i.e. to the recorded luminosity given in Table 3.1 and the $b\bar{b}$ cross section which is assumed to vary linearly from the centre mass energy of 7 TeV to 8 TeV [76].

The total efficiencies for the decays $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ and $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ are:

- $\epsilon^{B_s^0\to\tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp} = (6.15 \pm 0.16\ (\text{stat}) \pm 0.10\ (\text{syst})) \cdot 10^{-5}$,
- $\epsilon^{B^0\to\tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp} = (5.41 \pm 0.15\ (\text{stat}) \pm 0.10\ (\text{syst})) \cdot 10^{-5}$.

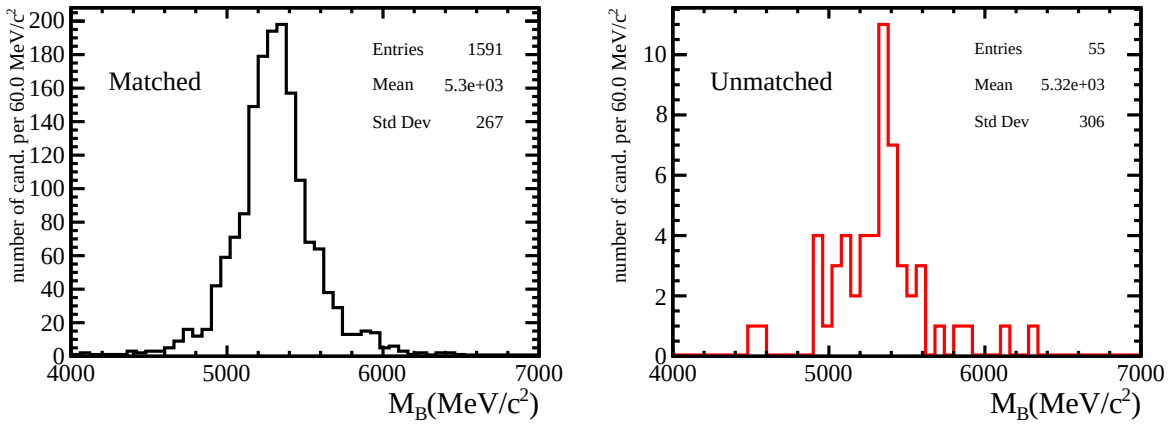

Figure 3.43: $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012 MC sample matched (left) and non-matched (right) events M_B distribution once the complete selection have been applied.

3.7.8 Normalization factor

The normalization factors α are directly proportional to the $B^0_{(s)} \to \tau^\pm\mu^\mp$ branching ratios and are defined as:

$$\alpha\left(B^0_{(s)} \to \tau^\pm\mu^\mp\right) = \mathcal{B}\left(B^0 \to D^-\pi^+\right)\mathcal{B}\left(D^- \to K^+\pi^-\pi^-\right)\frac{f_{B^0}}{f_{B^0_{(s)}}}\frac{\epsilon_{B\to D\pi}}{N^{obs}_{B\to D\pi}} \cdot \frac{1}{\epsilon_{B^0_{(s)}\to\tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp}\mathcal{B}\left(\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu\right) + \epsilon_{B^0_{(s)}\to\tau^\pm(\to\pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp}\mathcal{B}\left(\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu\right)}. \tag{3.26}$$

The $B^0_{(s)} \to \tau^\pm\mu^\mp$ branching ratios are then written simply as:

$$\mathcal{B}\left(B^0_{(s)} \to \tau^\pm\mu^\mp\right) = \alpha\left(B^0_{(s)} \to \tau^\pm\mu^\mp\right) N^{obs}_{B\to\tau\mu}. \tag{3.27}$$

The external informations used to compute the normalization factors are :

- $\mathcal{B}\left(\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu\right) = (9.02 \pm 0.05)\%$ [13],
- $\mathcal{B}\left(\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu\right) = (4.49 \pm 0.05)\%$ [13],
- $\mathcal{B}\left(B^0 \to D^-\pi^+\right) = (0.252 \pm 0.013)\%$ [13],
- $\mathcal{B}\left(D^- \to K^+\pi^-\pi^-\right) = (8.98 \pm 0.28)\%$ [77],
- $f_{B^0_s}/f_{B^0} = 0.259 \pm 0.015$ [78].

Using these inputs as well as the signal yield obtained in Section 3.3.2 and the selection efficiency for the $B^0 \to D^-\pi^+$ mode given in Section 3.7.6, the normalization factors for the B^0_s and B^0 channels are:

- $\alpha\left(B^0_s \to \tau^\pm\mu^\mp\right) = (4.31 \pm 0.19 \text{ stat} \pm 0.57 \text{ syst}) \cdot 10^{-7}$,
- $\alpha\left(B^0 \to \tau^\pm\mu^\mp\right) = (1.25 \pm 0.06 \text{ stat} \pm 0.15 \text{ syst}) \cdot 10^{-7}$.

The main difference between the two normalization factors comes from the b fragmentation fraction between B_s^0 and B^0. The fragmentation fraction has to be used in the B_s^0 case. This fact leads to a difference on the uncertainty of the two normalization factors, 10% in the B^0 case and 15% in the B_s^0 one.

3.8 Simultaneous fit

The signal yield is extracted from a simultaneous fit to the M_B distribution in bins of a final BDT that brings additional discrimination power between signal and background. The final BDT is described in Section 3.8.1. The full fit strategy, presented in Section 3.8.2, is validated using pseudo-experiments as shown in Section 3.8.3. Finally, the expected upper-limits are estimated with the CLs method using Same Sign data as a proxy for the data shape in Section 3.8.4.

3.8.1 Final BDT

3.8.1.1 Input variables

As the signal-like topology background mimics the signal signature in the detector, no single discriminating observable can be found to reject completely this background component. However, some observables, not strongly correlated with the B invariant mass, still hold some discriminating power against this kind of background:

1. `BDT_TMVA_Isolation_Based`: The BDT isolation based output after the cut placed in the offline selection.

- Variables related to the B meson candidate:

2. `MINIPCHI2`: Smallest significance of the impact parameters computed against all primary vertexes.
3. `DOCA`: Distance of closest approach between the 3π and the μ candidates reconstructed tracks.
4. `BPVVDCHI2`: B candidate flight distance significance with respect to the primary vertex.
5. `BDFplus_chi2`: χ^2 of the Decay Tree Fitter reconstruction fit.

- Variables related to the μ candidate:

6. `MINIPCHI2`: Smallest significance of the track impact parameter with respect to all primary vertexes.
7. `IP_OWNPV`: Impact parameter with respect to the primary vertex associated to the B candidate.

- Variables related to the τ candidate:

8. `M`: 3 pions invariant mass (a_0 resonance).
9. `FD_OWNPV`: Distance between the τ decay and the primary vertex.

10. `DOCAMAX`: Maximun distance of closest approach between the 3 pions.
11. `Vertex_Displacement`: Difference in the Z component between the τ and the B decay vertexes positions.

- Variables related to the τ daughter candidates:

12. $M_{\pi^+\pi^-}$: Invariant mass of the Opposite Sign pions (ρ^0 resonance).
13. `PiLowestPT`: Lowest transverse momentum among the 3 pions.
14. `Nu_BDFplus_P`: Reconstructed neutrino momentum.
15. `Angle_3Pi_Nu`: Angle between the 3π system and the neutrino directions.

These variables are not (or very weakly) correlated with the B invariant mass. Their distributions for signal MC and Same Sign background after the complete selection are shown in Figure 3.44.

3.8.1.2 Training

A final BDT is built with the variables listed above. It is trained on simulated $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ MC samples as signal and Same Sign data as background. The BDT is trained with the Adaptative Boosting algorithm using the `TMVA` package [71]. The relevant parameters of the BDT training are: `NTrees`=300, `MinNodeSize`=2.5%, `MaxDepth`=2 and `nCuts`=20. In order to avoid biases by applying the BDT on a sample used for training and also due to the lack of statistics in the background samples, the data and signal MC are partitioned in 10 subsets and 10 BDT's are trained independently. In each BDT training, 4.5k background events (from 9 subsets) are used in order to match the size of the signal sample which uses the full available statistics on 9 subsets. Finally, each BDT is applied on the subset it has not been trained on and the output is re-scaled such that the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ signal sample is flat between 0 and 1.

The BDT performance is illustrated as a ROC curve in Figure 3.46. For completeness, the profile plot of the BDT output with respect the B invariant mass for signal and background is shown in Figure 3.47. In the signal sample profile, some correlation between M_B and the final BDT output in signal is observed. However, no dangerous correlation is observed in the background sample profile which could sculpt a fake peak in the signal region.

3.8.1.3 Signal BDT output distribution

For signal and background samples the BDT output distribution is shown in Figure 3.45. The BDT output is built to be flat between 0 and 1 for the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ MC sample. The background samples peak at 0. Therefore, higher values of the BDT output are more sensitive to signal.

The most discriminant variables used in the BDT training are the τ invariant mass and the invariant mass of the opposite sign pions ($M_{\pi^+\pi^-}$). In Figures 3.12 and 3.42, which show the Dalitz plane for the different samples, it can be seen that the $B_{(s)}^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ do not share the same resonant structures. Partially for this reason the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ BDT output distribution resembles more that of background rather than that of $B_{(s)}^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ signal. Taking into

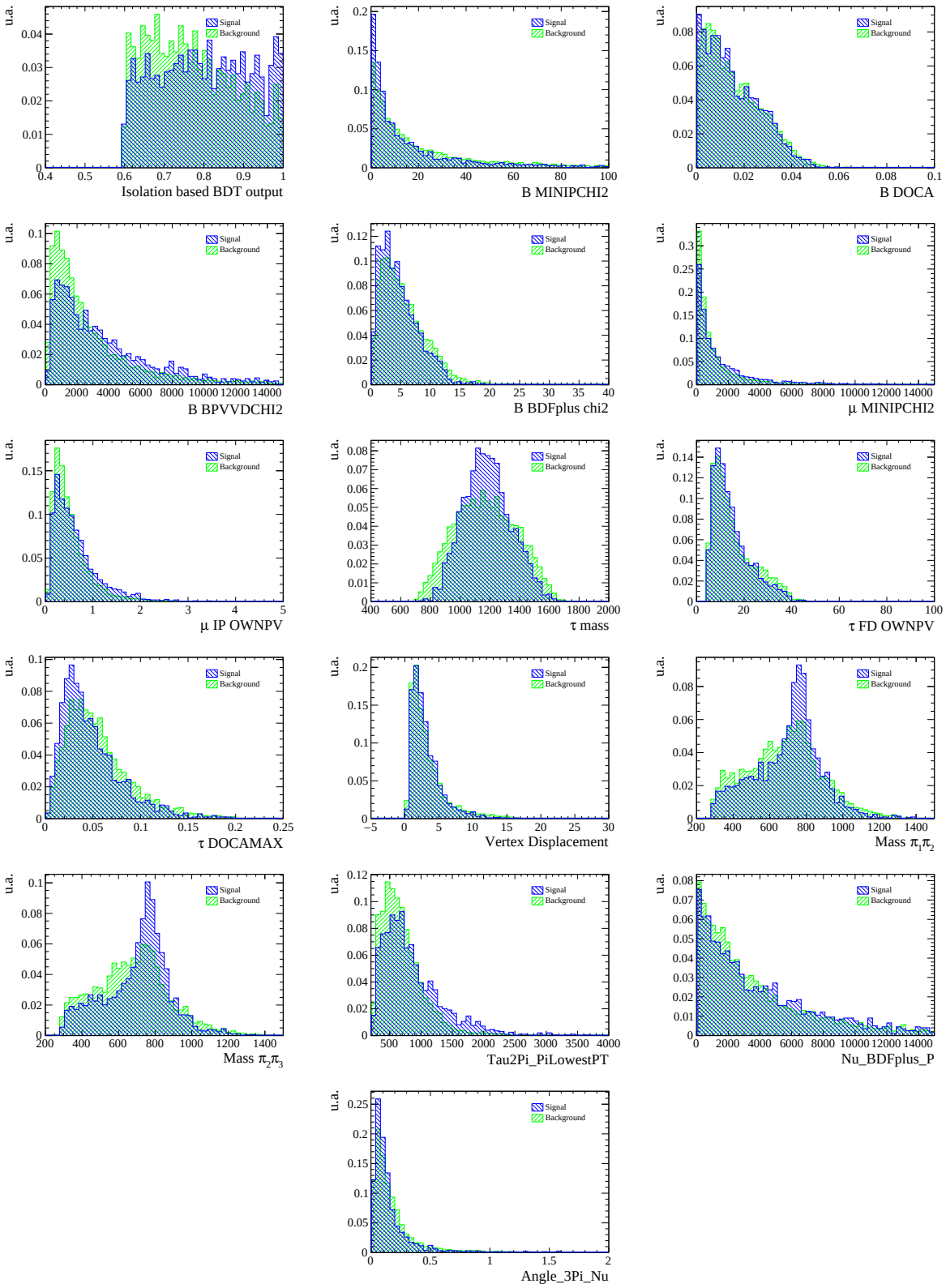

Figure 3.44: Final BDT input variables distributions. B_s^0 MC signal (blue) and Same Sign sample (green).

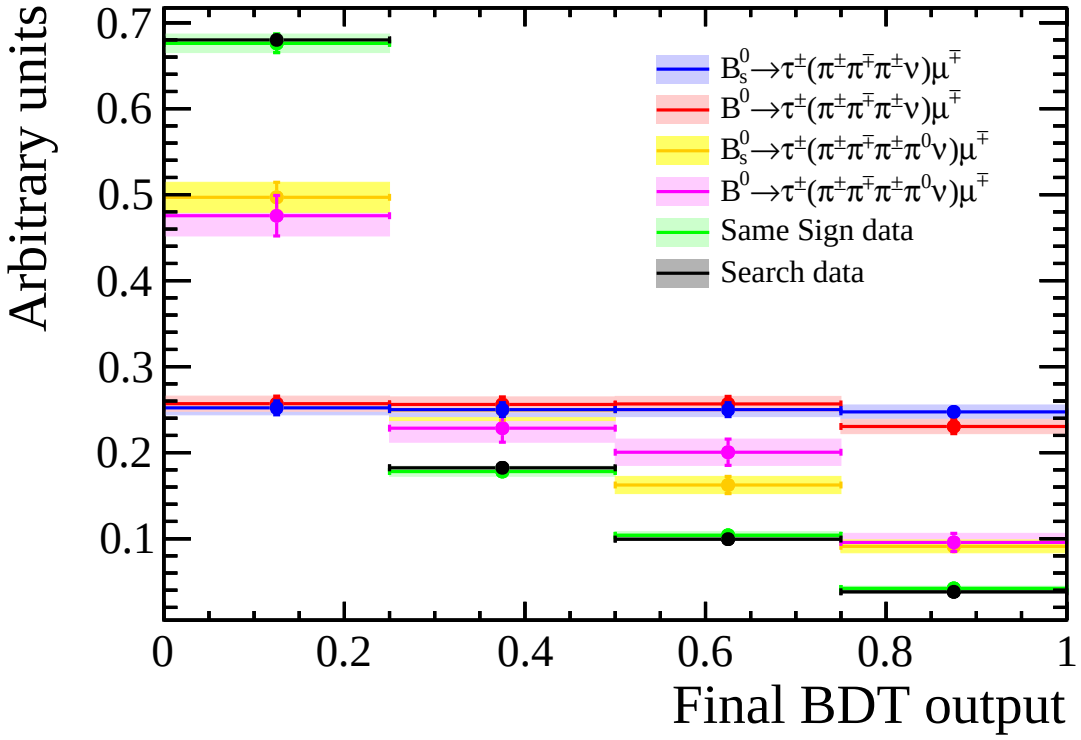

Figure 3.45: Final BDT ouput distribution for $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ signal simulation, $B^0_s \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ 2012 signal simulation and Same Sign data.

account the selection efficiencies and the τ decay Branching Ratios, the contribution of $B^0_s \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ to the total number of $B^0_s \to \tau^\pm\mu^\mp$ in each bin is: ∼21.9%, ∼10.9%, ∼6.7% and ∼3.7% from lower to upper bins. Hence, the contribution of the neutral mode in the most sensitive BDT bin is low.

3.8.1.4 Signal efficiency per BDT bin systematic uncertainty

The discrepancies between data and simulation for the variables used in the offline selection and the final BDT may affect the BDT output distribution. Following the method introduced in section 3.7.4.1, this effect is evaluated by recomputing the efficiency per bin on samples re-weighted according to the MC-data comparison made with $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ samples. This is done independently for each variable used in the offline selection and in the final BDT. The results are shown for B^0_s in Figures 3.48 and for B^0 in Figure 3.49. All recomputed efficiencies are in agreement with the nominal efficiencies within $\pm 1\sigma$ and no systematic uncertainty is assigned to this effect.

The choice of the $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$ decay model used in the simulation affects the distribution of the $M_{\pi^+\pi^-}$ and M_τ variables and thus affects the BDT output. This effect is evaluated by recomputing the efficiencies per bin using different $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$ decay models following the method described in section 3.7.4.3. The relative change in the efficiencies is of the order of ∼0.40%, which is very small compared to the statistical uncertainty and this effect is neglected.

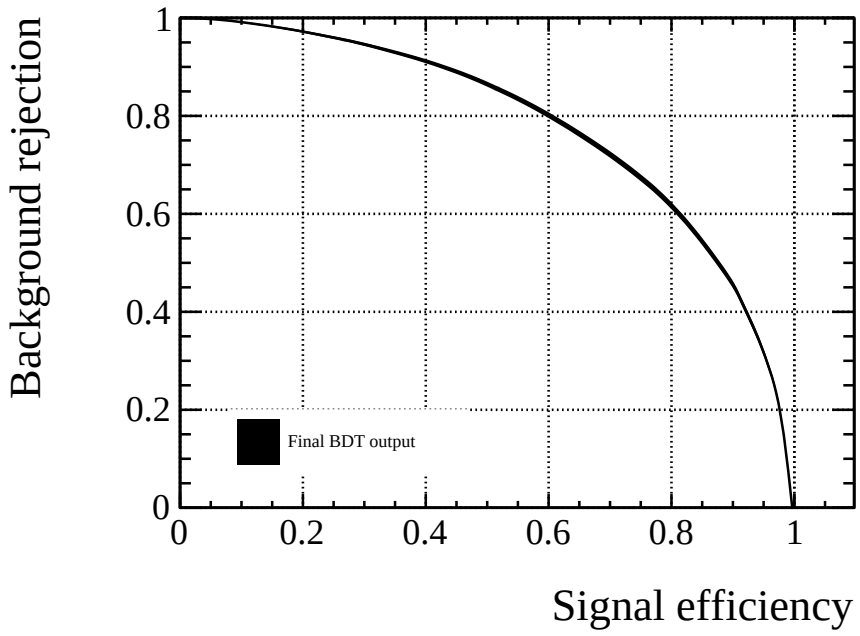

Figure 3.46: ROC curve of the final BDT. The figure shows the signal efficiency with respect to the Same Sign data rejection.

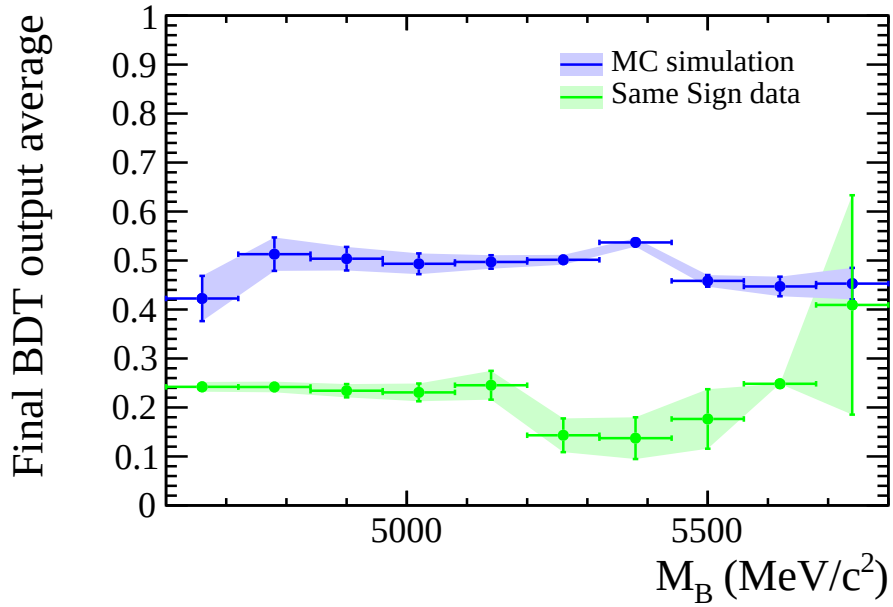

Figure 3.47: Final BDT variable profile with respect to B invariant mass $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulation samples (2011 and 2012) in blue and the Same Sign data sample in green.

	$B_s^0 \to \tau^\pm\mu^\mp$	$B^0 \to \tau^\pm\mu^\mp$
$\epsilon^{\text{Bin1}}(\%)$	27.71 ± 0.85 stat	28.01 ± 0.91 stat
$\epsilon^{\text{Bin2}}(\%)$	25.00 ± 0.78 stat	25.27 ± 0.82 stat
$\epsilon^{\text{Bin3}}(\%)$	24.08 ± 0.74 stat	24.89 ± 0.81 stat
$\epsilon^{\text{Bin4}}(\%)$	23.19 ± 0.72 stat	21.81 ± 0.76 stat

Table 3.31: $B^0_{(s)} \to \tau^\pm\mu^\mp$ signal efficiency in 4 bins of the final BDT output.

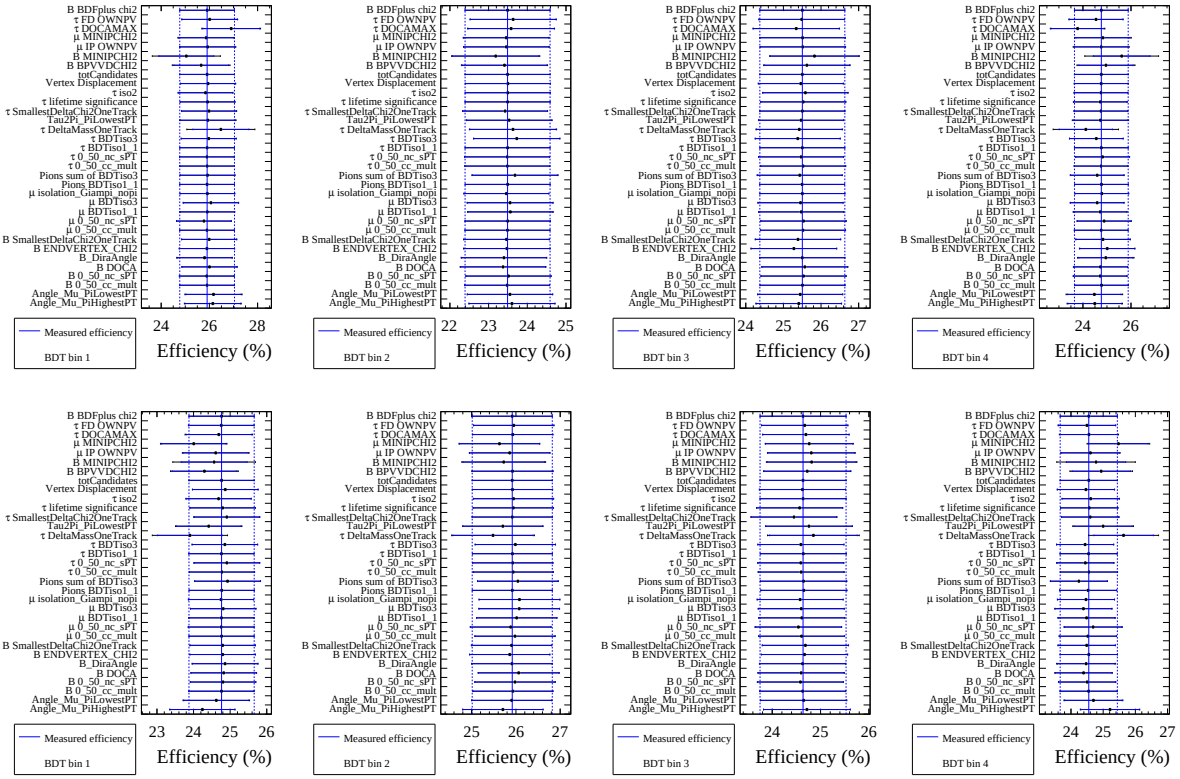

Figure 3.48: Individual re-weighting results in 4 bins of the final BDT output. Top line: $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2011. Bottom line: $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ 2012. The black dots represent the offline selection efficiency when the weights of the corresponding variable are applied. The blue continuous line corresponds to the selection efficiency computed without weights and the blue dashed line represents its statistical uncertainty. The colored error bars correspond to the statistical uncertainty of each re-weighted efficiency and the black error bars to the total uncertainty.

3.8.1.5 Signal efficiency per BDT bin

The $B^0_{(s)} \to \tau^\pm\mu^\mp$ signal efficiency per bin is estimated from the efficiencies per bin of the $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B^0_{(s)} \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ modes taking into account their selection efficiencies and their corresponding τ decay Branching Ratios. They are reported in Table 3.31

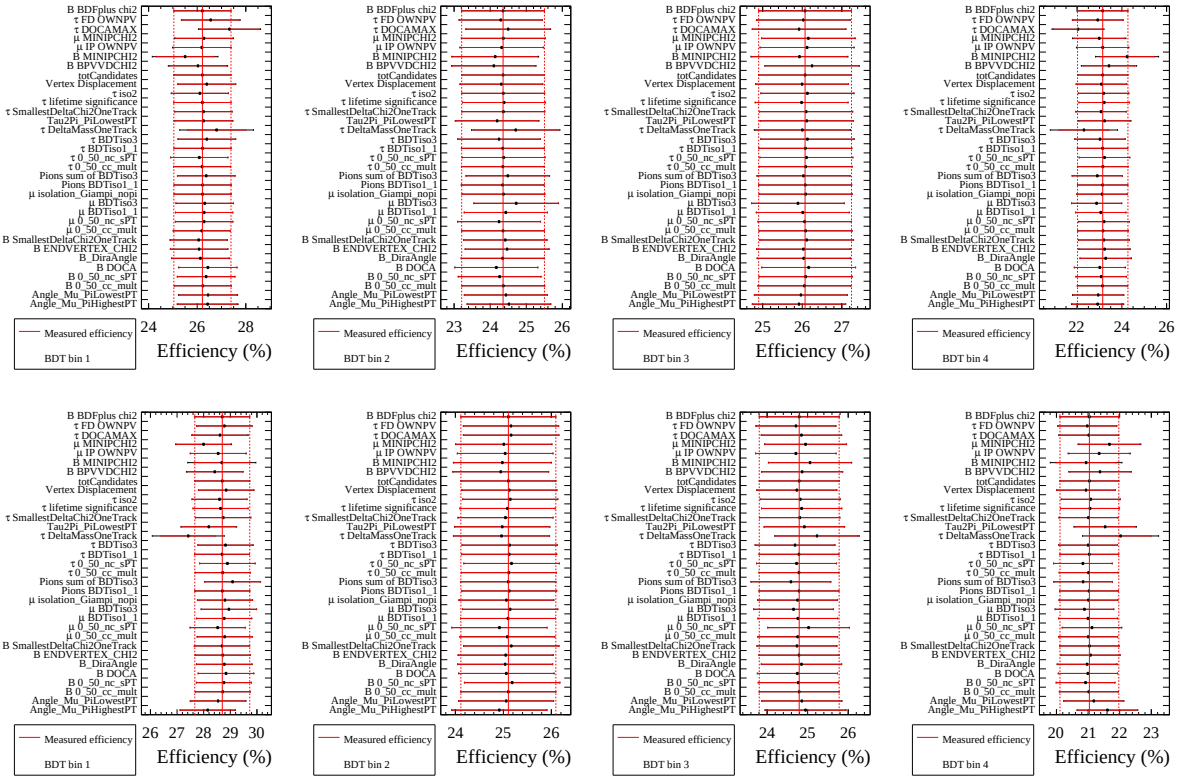

Figure 3.49: Individual re-weighting results in 4 bins of the final BDT output. Top line: $B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2011. Bottom line: $B^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ 2012. The black dots represent the offline selection efficiency when the weights of the corresponding variable are applied. The red continuous line corresponds to the selection efficiency computed without weights and the red dashed line represents its statistical uncertainty. The coloured error bars correspond to the statistical uncertainty of each re-weighted efficiency and the black error bars to the total uncertainty.

3.8.2 Fit strategy

The M_B invariant mass distribution is fitted simultaneously in bins of the final BDT output (Figure 3.45). All fits share only one parameter: the total signal event yield. In each bin, the total signal yield is multiplied by the expected fraction of events in this bin, i.e. the final BDT efficiency in this bin.

The fit performance is quantified using the sensitivity ($\mathcal{S}$) defined as

$$\mathcal{S} = 1.6 \cdot \delta N_{obs} \cdot \alpha \Rightarrow 95\% CL, \tag{3.28}$$

where the 1.6 factor comes from the upper bound of the $]-\infty, 1.6\sigma]$ interval giving 95% of the area under a gaussian. The sensitivity corresponds to the estimation of the limit on the $B^0_{(s)} \to \tau^{\pm}\mu^{\mp}$ branching ratio. It depends on the uncertainty on the fitted signal yield and on the normalization factor.

A simultaneous likelihood fit consists in a set of unbinned maximum likelihood fits sharing some of their fitted parameters. In general, it performs better than a single unbinned fit as it contains more information. In our case, it allows to improve by $\sim 10\%$ the sensitivity on the $\mathcal{B}(B^0_{(s)} \to \tau^{\pm}\mu^{\mp})$ with respect to a single unbinned fit with an optimal

cut on the final BDT output.

The M_B invariant mass distributions for the 2011 and 2012 data samples are merged and the fit is performed in the range between $4600\,\mathrm{MeV}/c^2$ and $5800\,\mathrm{MeV}/c^2$. Finally, motivated by the limited separation between B_s^0 and B^0 signals (Sec. 3.8.2.1), the simultaneous fit is performed twice. Once to extract the $B_s^0 \to \tau^\pm\mu^\mp$ signal assuming the absence of $B^0 \to \tau^\pm\mu^\mp$ events and another time to extract the $B^0 \to \tau^\pm\mu^\mp$ yield assuming no $B_s^0 \to \tau^\pm\mu^\mp$ signal.

3.8.2.1 Signal shape

The signal mass shape is extracted from the mass distributions of the $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$ simulated samples.

Figure 3.50 shows the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ normalized M_B distributions for the MC candidates passing the full selection in 4 bins of the final BDT. The M_B distributions get narrower at high BDT values. This correlation helps slightly the fit sensitivity, as the signal is more peak-like in the most sensitive bin. In each BDT bin, the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ distributions are compatible enough to allow the full M_B distribution to be described by a single line shape.

In order to get the total $B_s^0 \to \tau^\pm\mu^\mp$ signal M_B distribution the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ and the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ MC samples of events passing the full selection are merged. They are weighted such that they contribute in proportion of the expected yields of their corresponding mode. The weight for given sample is computed as:

$$\omega_s = \frac{\mathcal{L}_s}{\mathcal{L}_{tot}} \frac{\mathcal{B}(\tau \to f_s)\varepsilon_s}{N_s^{sel}} \tag{3.29}$$

where $\mathcal{L}_{tot}$ is the total luminosity, $\mathcal{L}_s$ is the luminosity of the corresponding sample (2011 or 2012), $\mathcal{B}(\tau \to f_s)$ is the branching ratio of the tau decay mode used in the sample, ε_s is the total selection efficiency for this sample and N_s^{sel} is the total number of selected events.

The $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$ mass shape are described in each bin of the final BDT by a double-sided Hypatia function [79]. For each channel, 4 Hypatia PDF's (Probability Density Function) are fitted simultaneously to the M_B distributions in the 4 BDT bins. The mean of the PDF's, different for the 2 channels, is shared among the 4 Hypatia. The a, a', n and n' tail parameters are let free while the λ, β and the ζ parameters are fixed. The result of the fit is shown in Figure 3.51 and the parameter values are given in Table 3.32.

The M_B reconstruction method, described in Section 3.4, is mainly sensitive to the decay vertices position resolution. In order to check how the Hypatia shape is affected by the vertex measurement precision, new sets of events are created from the the $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\nu)\mu^\mp$ simulated samples where the B and τ decay vertex true positions are smeared simultaneously. Different samples are produced where the smearing is done with different widths ranging from 90% to 110% of the resolution estimated in the original sample. The M_B reconstruction method is applied to these modified MC samples and the fit procedure is repeated with all parameters fixed to their nominal values but the widths. The relation between the vertex position resolution and the Hypatia width is shown in Figure 3.52 and exhibits a linear correlation between the two quantities.

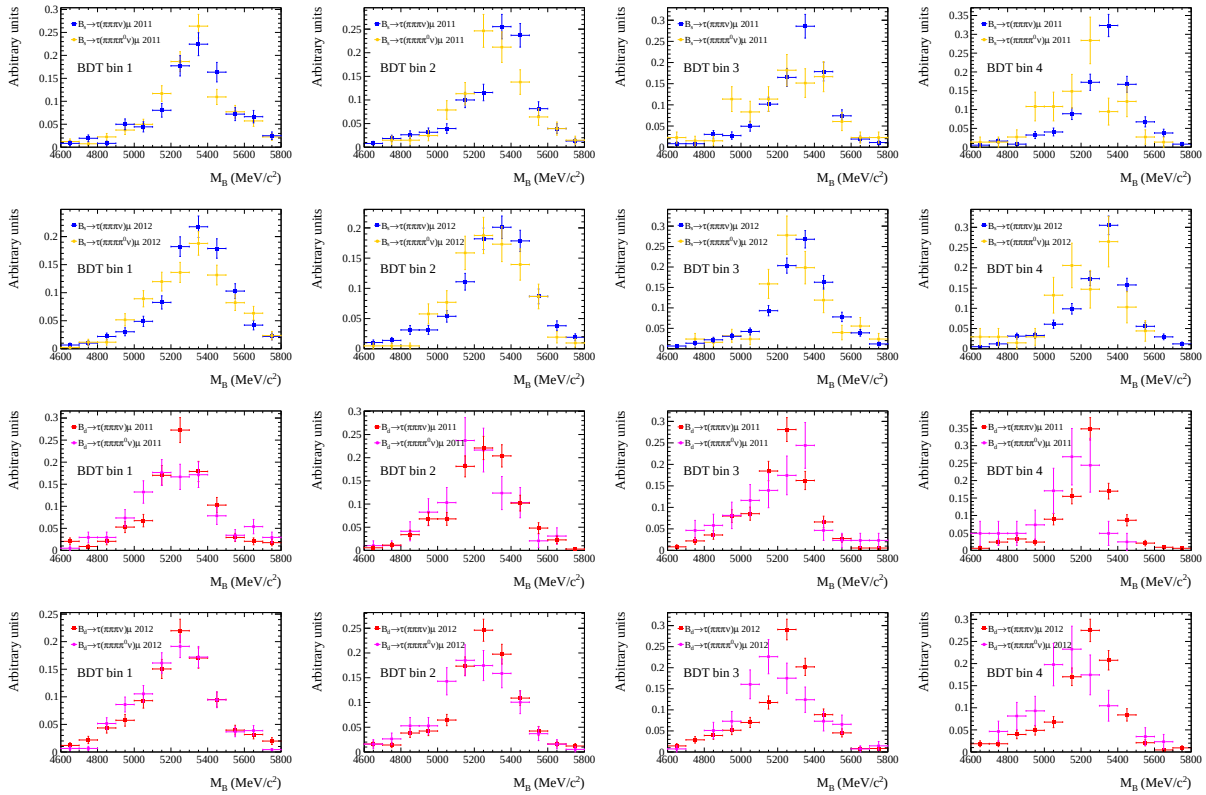

Figure 3.50: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^0\nu)\mu^{\mp}$ MC and $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ MC normalized shapes comparison in 4 bins of the final BDT output. When the corresponding signal efficiency and τ decay Branching Ratio values are taken into account, the $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^0\nu)\mu^{\mp}$ contribution is mainly in the first BDT bins.

The range of the possible variation in the simulated vertex resolution is inferred using the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC and data samples. The angle between the D momentum and the direction given by the line formed by the B and D decay vertices is compared in the data and MC samples. The relative difference between the width of the distributions of this angle in data and MC is about 6%. This is used as an upper bound to the possible variation of the resolution on the B and τ decay vertex positions in the $B_s^0 \to \tau^{\pm}\mu^{\mp}$ simulated samples. This maximum variation translates into a 12% systematic uncertainty on the Hypatia width.

The Hypatia shapes obtained on the simulated samples are used in the data fit with parameters fixed from MC except for the Hypatia width which is left free to vary within Gaussian constraints of 12% of its central value.

3.8.2.2 Background shape

As the M_B mass distributions in Opposite Sign (blinded) and Same Sign samples in the non-blinded region are compatible (Figure 3.30), the Same Sign data is used as a control sample to study the background shape in the different BDT bins.

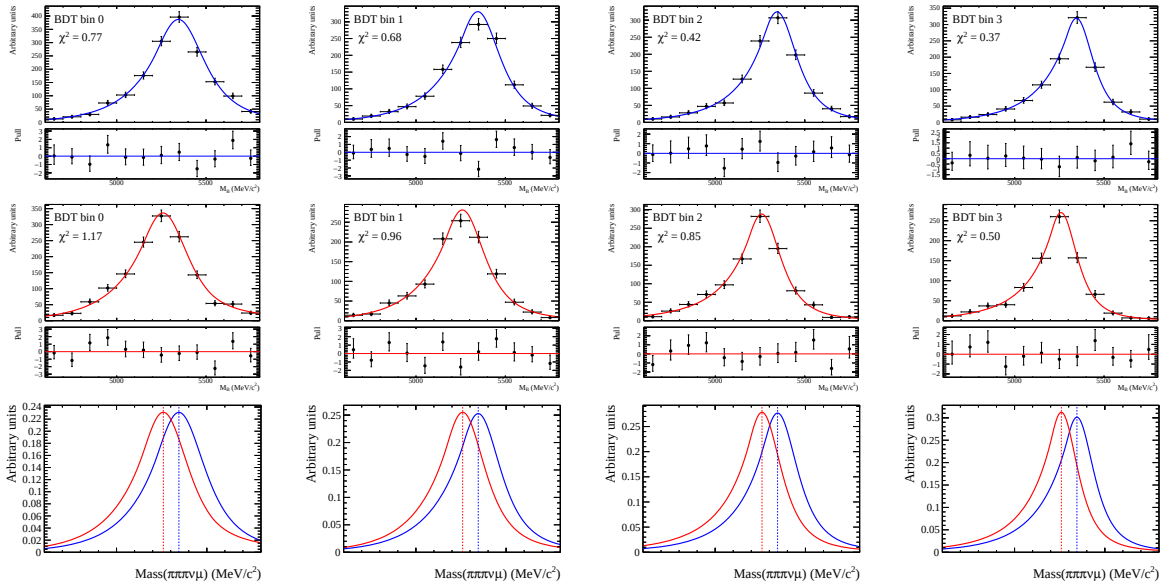

Figure 3.51: Top-line: $B_s^0 \to \tau^{\pm}\mu^{\mp}$ MC simulation B invariant mass simultaneous fit in 4 BDT bins. Middle-line: $B^0 \to \tau^{\pm}\mu^{\mp}$ MC simulation B invariant mass simultaneous fit in 4 BDT bins. Bottom-line: PDF separation between $B^0 \to \tau^{\pm}\mu^{\mp}$ (red) and $B_s^0 \to \tau^{\pm}\mu^{\mp}$ (blue).

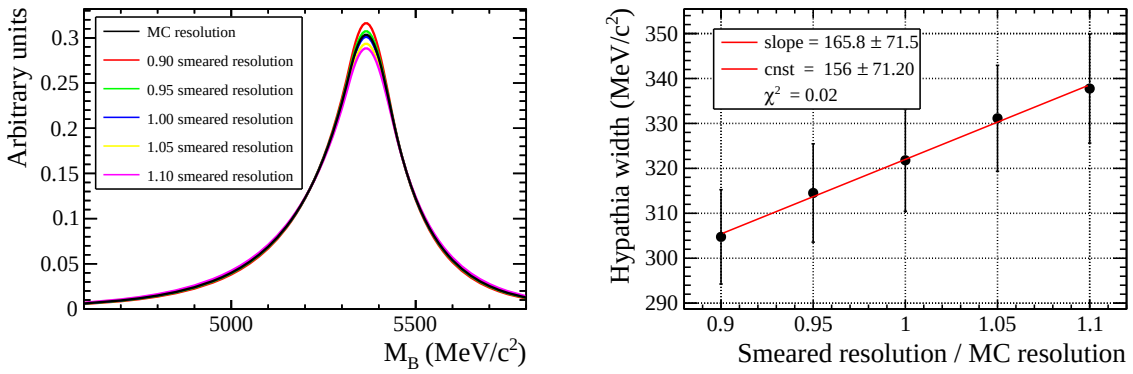

Figure 3.52: Left: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ Hypatia signal fits for different vertex position resolutions smearing. The MC resolution (black) is compared to smearings ranging from 0.90 to 1.10 of the MC vertex position resolution. Right: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ fitted Hypatia width with respect to the vertices position smeared resolution.

Common shape in each bin

As show on Figure 3.53, the M_B mass distributions of Same Sign data in each bin of the final BDT are well described by a common Gaussian PDF.

However, there are no guarantee that this feature remains in the Opposite sign data where, with more statistics, some difference between the bins may become significant. Therefore, the background shape is modelled using independent PDF in each bin. In the case the background shapes of the Opposite Sign data in each BDT bins are compatible, some gain in the fit sensitivity is expected and this potential improvement is discussed in

Parameter	Bin 1	Bin 2	Bin 3	Bin 4
$B_s^0 \to \tau^\pm\mu^\mp$				
Mean (MeV/c^2)	5345 ± 3			
Width(MeV/c^2)	526 ± 16	449 ± 15	398 ± 14	339 ± 13
λ	-1.1	-1.1	-1.1	-1.1
β	0	0	0	0
ζ	0	0	0	0
a	0.22 ± 0.02	0.17 ± 0.01	0.17 ± 0.01	0.14 ± 0.01
n	59.9 ± 0.1	73.5 ± 0.1	9.0 ± 3.2	11.3 ± 4.7
a'	1.8 ± 2.2	1.2 ± 2.7	1.7 ± 3.1	5.1 ± 4.1
n'	0.7 ± 8.8	1.3 ± 8.6	3.0 ± 8.9	3.6 ± 9.6
$B^0 \to \tau^\pm\mu^\mp$				
Mean (MeV/c^2)	5259 ± 3			
Width(MeV/c^2)	510 ± 22	449 ± 22	383 ± 22	316 ± 6
λ	-1.1	-1.1	-1.1	-1.1
β	0	0	0	0
ζ	0	0	0	0
a	0.21 ± 0.02	0.17 ± 0.01	0.18 ± 0.02	0.13 ± 0.01
n	60.0 ± 0.2	73.6 ± 0.2	6.0 ± 4.4	6.6 ± 2.1
a'	0.49 ± 0.1	1.8 ± 0.8	2.0 ± 1.3	3.1 ± 2.0
n'	2.1 ± 1.2	1.3 ± 7.2	3.1 ± 7.6	3.6 ± 8.0

Table 3.32: Hypatia PDF parameters for the signal MC samples: $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$.

Parameter	Bin 1	Bin 2	Bin 3	Bin 4
Same Sign Data				
Mean (MeV/c^2)	4300 ± 110	4583 ± 73	4455 ± 148	4518 ± 174
Width(MeV/c^2)	369 ± 32	275 ± 30	311 ± 50	273 ± 64

Table 3.33: Gaussian PDF parameters for the Same Sign Data independent fits.

section 3.9.2.

Independent shape in each bin

Unbinned maximum likelihood fits with a Gaussian PDF with floating parameters are performed independently on the M_B mass distributions of Same Sign data in each bin of the final BDT. The fitted PDF drawn on top of the Same Sign data on Figure 3.54 shows that it provides a good description of Same Sign data. The fitted parameters are reported in Table 3.33.

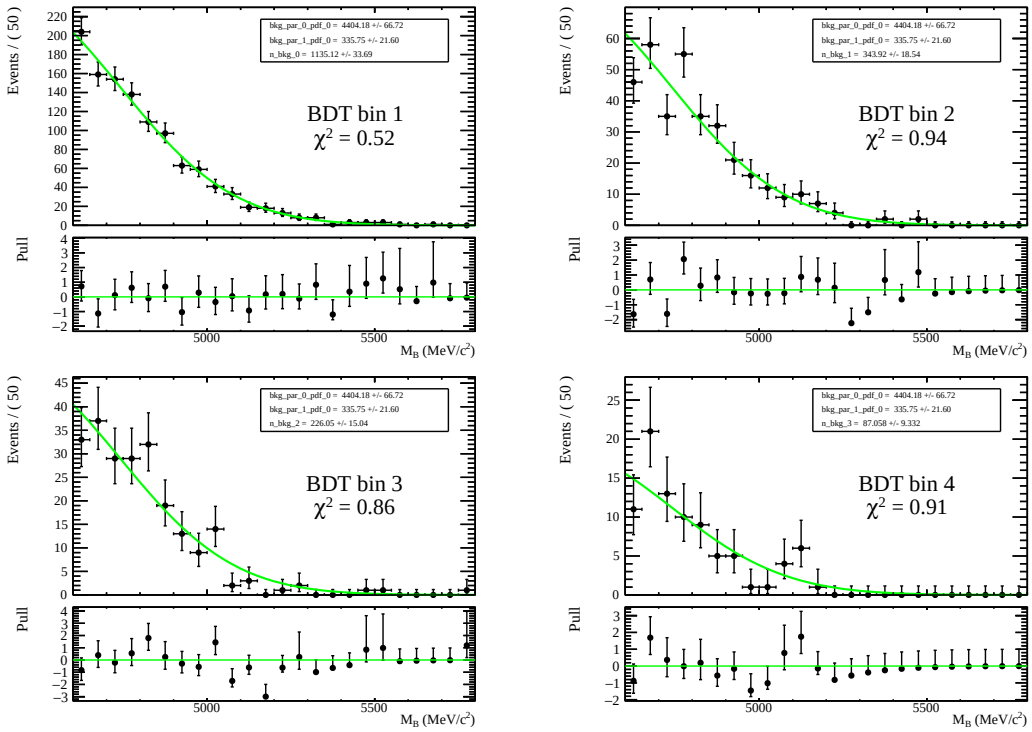

Figure 3.53: Unbinned fits of Gaussian PDF's with common parameters to 4 BDT bins of the Same Sign data.

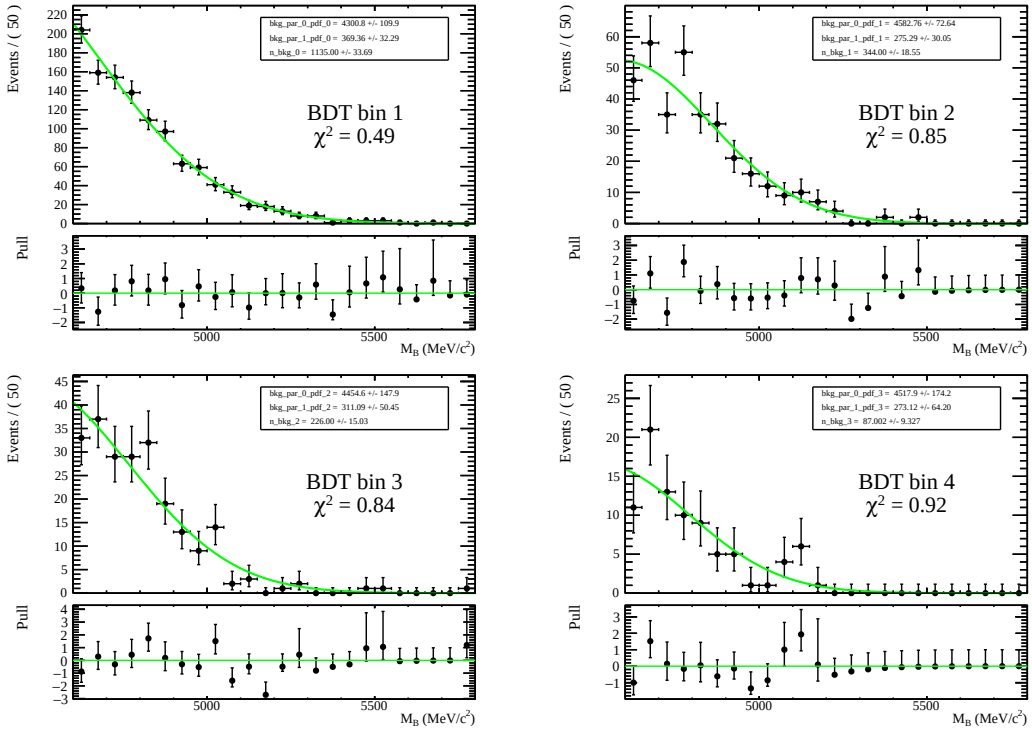

Figure 3.54: Same Sign data independent unbinned fits in 4 BDT bins. In each BDT bin, the data shape is fitted with a Gaussian PDF with floating parameters.

3.8.2.3 Total PDF

The simultaneous fit uses the following PDF:

$$PDF^{tot} = \sum_{i}^{\text{BDT bins}} \left(\mathbf{N^{sig}} \epsilon_i^{sig} Hyp_i^{sig} + n_i^{bkg} Gaus_i^{bkg}(\mu_i, \sigma_i) \right) \tag{3.30}$$

where:

- $\mathbf{N^{sig}}$: signal yield (free). Common parameter between the fits.
- n_i^{bkg}: background yield in each BDT bin (free).
- ϵ_i^{sig}: signal efficiency per BDT bin taken from MC. Floating with Gaussian constraints. The width of the Gaussian constraint corresponds to the total uncertainty of the per bin signal efficiency reported in Table 3.31.
- Hyp_i^{sig}: Hypatia PDF describing the signal shape with the parameters fixed from MC fit. The hypathia width is allowed to vary within Gaussian constraints of σ corresponding to 12% of the efficiency central value.
- $Gaus_i^{bkg}(\mu_i, \sigma_i)$: Gaussian PDF describing the background shape (parameters free).

The parameters in red are left free during the fit process.

3.8.3 Fit validation and optimization

3.8.3.1 Pseudo-experiments generation and fit

The fit validation is performed using pseudo-experiments (also referred to as 'toys'). In each pseudo-experiment, background M_B mass distributions are generated in bins of the final BDT according to the shape and yield expected in data. Since the signal data sample is blinded, the shape of the M_B and BDT background distributions are taken from Same Sign data. The expected background yield N_{Gen}^{Bkg} is naively estimated from the number of events in Same Sign data scaled by the ratio of Opposite Sign and Same Sign data in the non-blinded region and is found to be ~ 14000 events.

For some specific studies, $B^0 \to \tau^{\pm}\mu^{\mp}$ or $B_s^0 \to \tau^{\pm}\mu^{\mp}$ signal is also incorporated in the pseudo-experiments adding the M_B distributions generated according to the Hypatia function describing the simulated B_s^0 or B^0 signal.

The total number of events generated in each toy simulation is given by:

$$N_{Gen}^{Tot} = \sum_{i}^{\text{BDT bins}} \left(P(N_{Gen}^{Sig}) \epsilon_i^{Sig} + P(N_{Gen}^{Bkg} \epsilon_i^{Bkg}) \right) \tag{3.31}$$

where :

- ϵ_i^{Sig} and ϵ_i^{Bkg} are the fractions of events in the i^{th} BDT bin in the simulated signal sample and in the Same Sign data respectively.
- $P(n)$ is a random number generated according to a Poisson distribution with mean n.

The simultaneous fit is applied to the M_B mass distributions of each toy. Any fit bias is quantified by looking at the *pull* of the PDF parameters. For a given parameter a generated with the value a_{Gen} and fitted to $a_{Fit} \pm \delta a_{Fit}$, the pull is defined as:

$$a^{Pull} = \frac{a_{Fit} - a_{Gen}}{\delta a_{Fit}}. \tag{3.32}$$

The values of the signal yield N^{Sig} parameter of interest are expected to be zero following the SM expectations. For this parameter an alternate *pull* definition is implemented, which takes into account the information of the asymmetric parameter uncertainty given by the `Minos` strategy:

$$\begin{aligned} \text{If } N^{Sig}_{Fit} < N^{sig}_{Gen} \text{ then:} \qquad & N^{Sig}_{Pull} = \frac{N^{Sig}_{Gen} - N^{Sig}_{Fit}}{\delta N^{Sig}_{Fit}(\texttt{Minos}+)} \\ \text{otherwise:} \qquad & N^{Sig}_{Pull} = \frac{N^{Sig}_{Fit} - N^{Sig}_{Gen}}{\delta N^{Sig}_{Fit}(\texttt{Minos}-)} \end{aligned} \tag{3.33}$$

The *pull* follows a normal distribution ($N(0,1)$) for unbiased sets of fits.

All studies presented in the rest of this section are performed with sets of 1000 pseudo-experiments.

3.8.3.2 BDT binning optimization

The number of BDT bins in the fit is chosen to maximize the fit sensitivity while keeping the fit stable.

Fit instability causes the appearance of an asymmetric tail in the distributions of the signal yield pulls when no signal is present. Figures 3.55 and 3.56 show the fitted signal yield pull distributions of the toys corresponding to simultaneous fit strategies using 2 to 7 BDT bins. It can be observed that the signal pull distribution, in the case of low number of divisions, is compatible with a $N(0,1)$. On the other hand, in high number of divisions, a tail appears affecting pull shape compatibility with a $N(0,1)$ distribution. This effect is related to the number of events in the last BDT bin. If the number of events in the last BDT bin is not large enough, when the likelihood is minimized, on some of the toys a minimum of the likelihood can be found with sizable negative value and small uncertainty.

The magnitude of the tail in the signal yield pull distribution can be quantified by calculating the amount of pseudo-experiments outside 3σ of the ideal $N(0,1)$. Figure 3.57 shows the signal yield pull quality vs the number of BDT bins. The fit procedure on the B^0 channel is more robust against potential biases than the B^0_s case. This fact is due to the position of the signal peak, which for the B^0 is located in a region with slightly higher background level where the total PDF is less probable to reach a negative value.

The more divisions of the BDT output, the better the $\mathcal{B}$ limit sensitivity as more information is used in the fit. (Figure 3.58).

The optimum number of bins is the maximum number of bins for which the fit is not biased. A high threshold on the amount of pseudo-experiments outside 3σ of the ideal $N(0,1)$ is set to be lower than $\sim$1%. Therefore, the optimum number of BDT output divisions is 4.

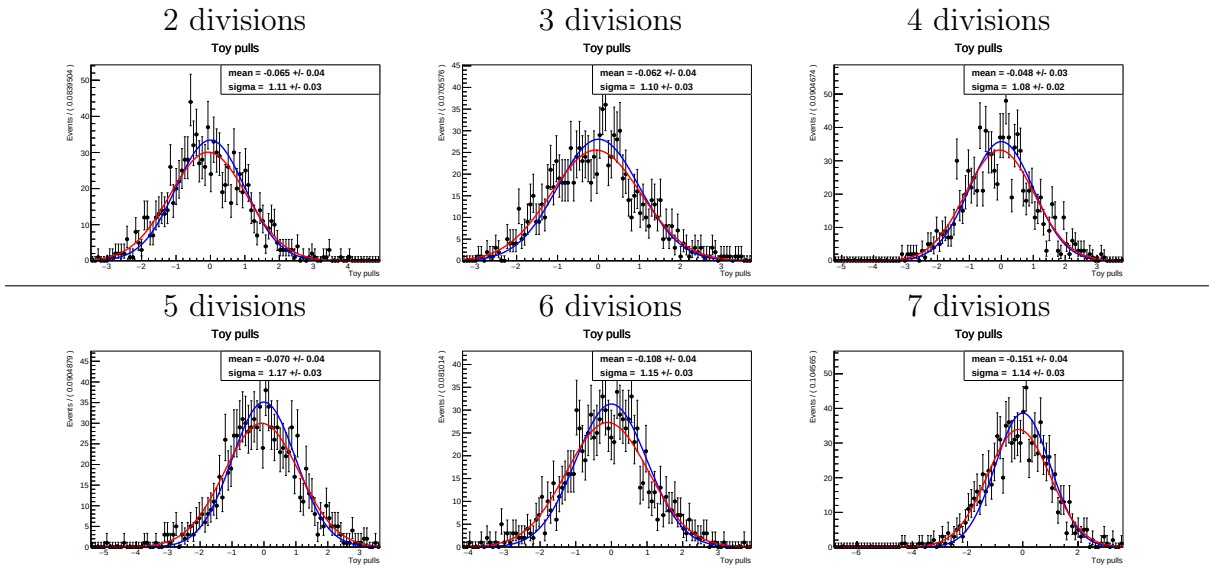

Figure 3.55: $B_s^0 \to \tau^{\pm}\mu^{\mp}$ signal yield pull distributions of 1000 pseudo-experiments for different numbers of BDT bins. The red lines correspond to the Gaussian PDF fits to the toys distributions. The blue lines correspond to a N(0,1).

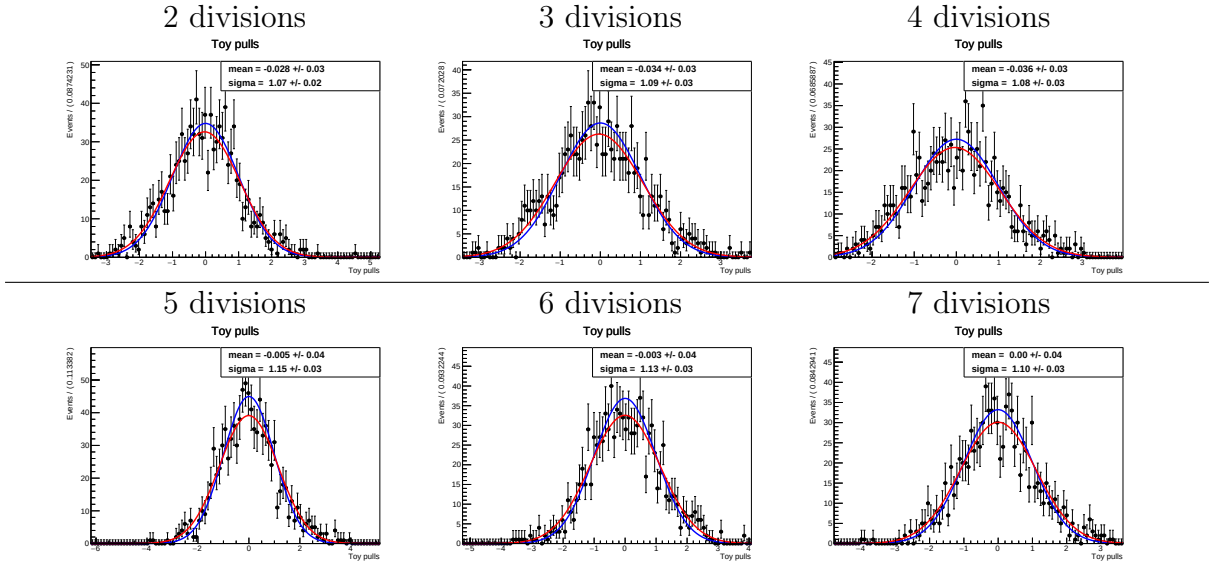

Figure 3.56: $B^0 \to \tau^{\pm}\mu^{\mp}$ signal yield pull distributions of 1000 pseudo-experiments for different numbers of BDT bins. The red lines correspond to the Gaussian PDF fits to the toys distributions. The blue lines correspond to a N(0,1).

3.8.3.3 Fit validation in absence of signal

A set of 1000 pseudo-experiments are generated with no signal ($N_{Gen}^{Sig} = 0$) and fitted with 4 BDT bins as described in Section 3.8.3.1. The fit convergence rate is excellent, only ∼1% of the fits are not convergent. The distributions of the pulls of the parameters describing the background (N^{Bkg}, μ and σ of the Gaussian) are shown in Figure 3.59

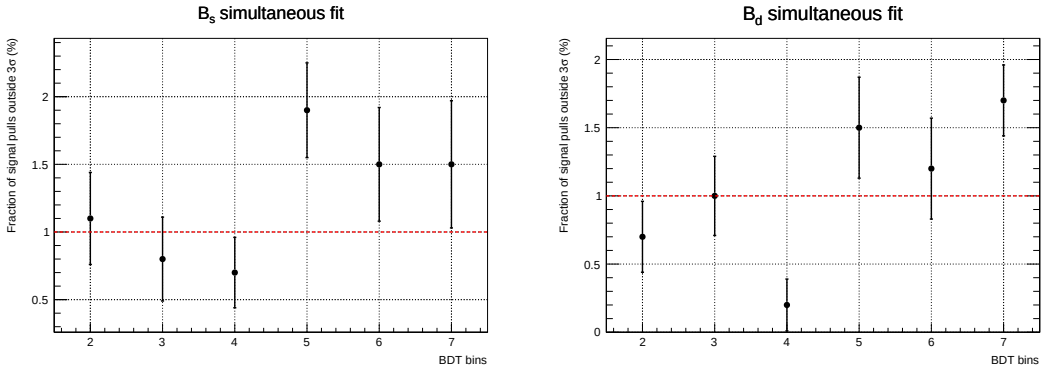

Figure 3.57: Number of pseudo-experiments with signal yield pull more than 3σ away from the $N(0,1)$ distribution as a function of the number of BDT divisions in the simultaneous fit. Left: $B_s^0 \to \tau^{\pm}\mu^{\mp}$. Right: $B^0 \to \tau^{\pm}\mu^{\mp}$.

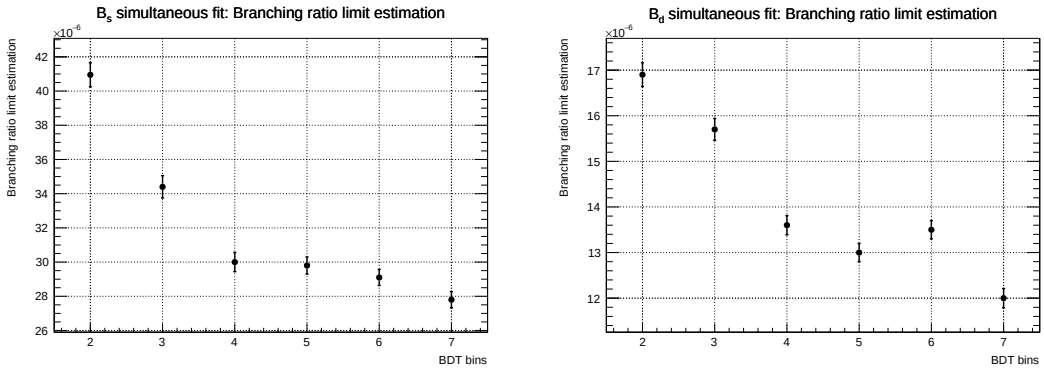

Figure 3.58: Expected $\mathcal{B}$ upper limit estimated as a function of the number of BDT divisions. Left: $B_s^0 \to \tau^{\pm}\mu^{\mp}$. Right: $B^0 \to \tau^{\pm}\mu^{\mp}$.

for the B_s^0 mode and on Figure 3.60 for the B^0. At high BDT bins, where the statistics is smaller, the pulls of both μ and σ are biased. However, since the Gaussian which is fitted is truncated due to the limited range of the fit, these two parameters are highly correlated and the bias which is observed does not affect the integral of the fitted truncated Gaussian (background yield, N^{Bkg}). The mean and width of the pull distributions of all fit parameters are summarized in Table 3.34.

The average uncertainty of the signal yield δN is taken as the width of a Gaussian fitting the distribution of the signal yields, giving:

- $\delta N(B_s^0 \to \tau^{\pm}\mu^{\mp}) = 41$
- $\delta N(B^0 \to \tau^{\pm}\mu^{\mp}) = 64$

It can be observed that the uncertainty on the B^0 yield is greater than that on the B_s^0 yield as the B peak position is located in a M_B mass region containing more background events.

Using the formula in Eq. 3.28, these uncertainties are turned into the following fit sensitivities:

		B_s^0		B^0	
Division	Parameter	Pull mean	Pull width	Pull mean	Pull width
-	N_{sig}	-0.05 ± 0.03	1.08 ± 0.02	-0.04 ± 0.03	1.08 ± 0.03
1	n_{bkg}	0.00 ± 0.03	1.00 ± 0.02	0.00 ± 0.03	1.00 ± 0.02
	Gaus μ	0.04 ± 0.04	1.08 ± 0.03	0.06 ± 0.04	1.09 ± 0.03
	Gaus σ	-0.02 ± 0.03	1.06 ± 0.02	-0.04 ± 0.03	1.07 ± 0.03
2	n_{bkg}	0.00 ± 0.03	1.00 ± 0.02	0.00 ± 0.03	1.01 ± 0.02
	Gaus μ	0.10 ± 0.04	1.10 ± 0.03	0.12 ± 0.04	1.12 ± 0.03
	Gaus σ	-0.06 ± 0.03	1.07 ± 0.03	-0.08 ± 0.04	1.08 ± 0.03
3	n_{bkg}	0.01 ± 0.03	1.00 ± 0.02	0.01 ± 0.03	1.01 ± 0.02
	Gaus μ	0.12 ± 0.04	1.12 ± 0.03	0.04 ± 0.05	1.20 ± 0.04
	Gaus σ	0.06 ± 0.03	1.08 ± 0.02	0.03 ± 0.04	1.11 ± 0.03
4	n_{bkg}	0.00 ± 0.03	1.01 ± 0.02	0.02 ± 0.03	1.01 ± 0.02
	Gaus μ	0.28 ± 0.04	1.36 ± 0.03	0.22 ± 0.05	1.45 ± 0.05
	Gaus σ	-0.13 ± 0.04	1.34 ± 0.03	-0.18 ± 0.04	1.22 ± 0.03

Table 3.34: Pull results for the parameters of the simultaneous fit .

- $\mathcal{S}(B_s^0 \to \tau^{\pm}\mu^{\mp}) = 3.0 \cdot 10^{-5}$
- $\mathcal{S}(B^0 \to \tau^{\pm}\mu^{\mp}) = 1.4 \cdot 10^{-5}$

Despite the larger error on the fitted yields, the sensitivity for B^0 is better than for B_s^0 thanks to the much smaller normalization factor (see Section 3.7.8).

Looking closely at the signal yield distributions, shown in Figure 3.61, tiny biases of -3.3 ± 1.0 in B_s^0 and -2.6 ± 2 in B^0 fits are observed. They correspond respectively to $\sim$8% and $\sim$4% of the statistical uncertainty on the signal yields. This bias is intrinsic of the fit strategy and it depends on the background shape, being small with the Same Sign data shape, it may be larger in the unblinded Opposite Sign data. The bias will be tackled, as discussed on Section 3.9.1, when the Opposite sign data is unblinded and the search data shape is known accurately.

3.8.3.4 Fit validation with signal

Signal events can be injected in the pseudo-experiments to validate the fit strategy in case signal is present in data. Generating 1000 toys with a given number of signal events, a measured $\mathcal{B}$ can be obtained by using the fitted signal yield. Figure 3.62 contains the predicted $\mathcal{B}$ for a given number of injected signal events as a function of the $\mathcal{B}$ computed with the fitted values. The correspondence between the predicted and the measured values is one to one, hence the fit strategy is unbiased even for a small but measurable amount of signal.

3.8.4 CLs method and expected upper limit

The CLs method [15] is used to extract the $B_s^0 \to \tau^{\pm}\mu^{\mp}$ and $B^0 \to \tau^{\pm}\mu^{\mp}$ upper limits on the branching ratios ($\mathcal{B}$). The method compares how likely is the data to be well described by a signal plus background ($s+b$) or a background only (b) hypotheses. Different $\mathcal{B}$ hypotheses are tested in the region of interest. For each scanned $\mathcal{B}$, likelihood distributions for $s+b$ and b hypotheses are used to compute a p-value defined as:

$$p = \frac{\text{Prob (reject } H_{s+b}|H_{s+b} = \text{TRUE)}}{(1 - \text{Prob (reject } H_b|H_b = \text{TRUE))}}. \tag{3.34}$$

The upper limit is the value of $\mathcal{B}$ having the p-value corresponding to the desired confidence level. The expected limit presented below is computed using the `AsymptoticCalculator` of the `RooStats` package that is based on the asymptotic formulae from [80].

The PDF used in the fit is rewritten in terms of the $\mathcal{B}$ and the normalization factor computed in Section 3.7.8:

$$PDF^{tot} = \sum_{i}^{\text{BDT bins}} \left(\mathcal{B}\alpha^{-1}\epsilon_i^{sig} Hyp_i^{sig} + n_i^{bkg} Gaus_i^{bkg}(\mu_i, \sigma_i) \right) \tag{3.35}$$

The uncertainty on the normalization factor is taken into account in the fit as Gaussian constraints on the α parameter.

As the Opposite Sign data is blinded, the model used in the CLs is taken from a Gaussian fit to the Same Sign data scaled to 14000 events. The expected upper limit (Figure 3.63) is obtained using the following scan parameters:

- $B_s^0 \to \tau^{\pm}\mu^{\mp}$: 50 scan points in range $[5.0 \cdot 10^{-7}, 4.0 \cdot 10^{-5}]$
- $B^0 \to \tau^{\pm}\mu^{\mp}$: 50 scan points in range $[7.0 \cdot 10^{-6}, 2.5 \cdot 10^{-5}]$

Taking the $\mathcal{B}$ expected limit at a p-value of 0.05:

- $\mathcal{B}(B_s^0 \to \tau^{\pm}\mu^{\mp}) < 3.0 \cdot 10^{-5}$ at 95%CL,
- $\mathcal{B}(B^0 \to \tau^{\pm}\mu^{\mp}) < 2.0 \cdot 10^{-5}$ at 95%CL.

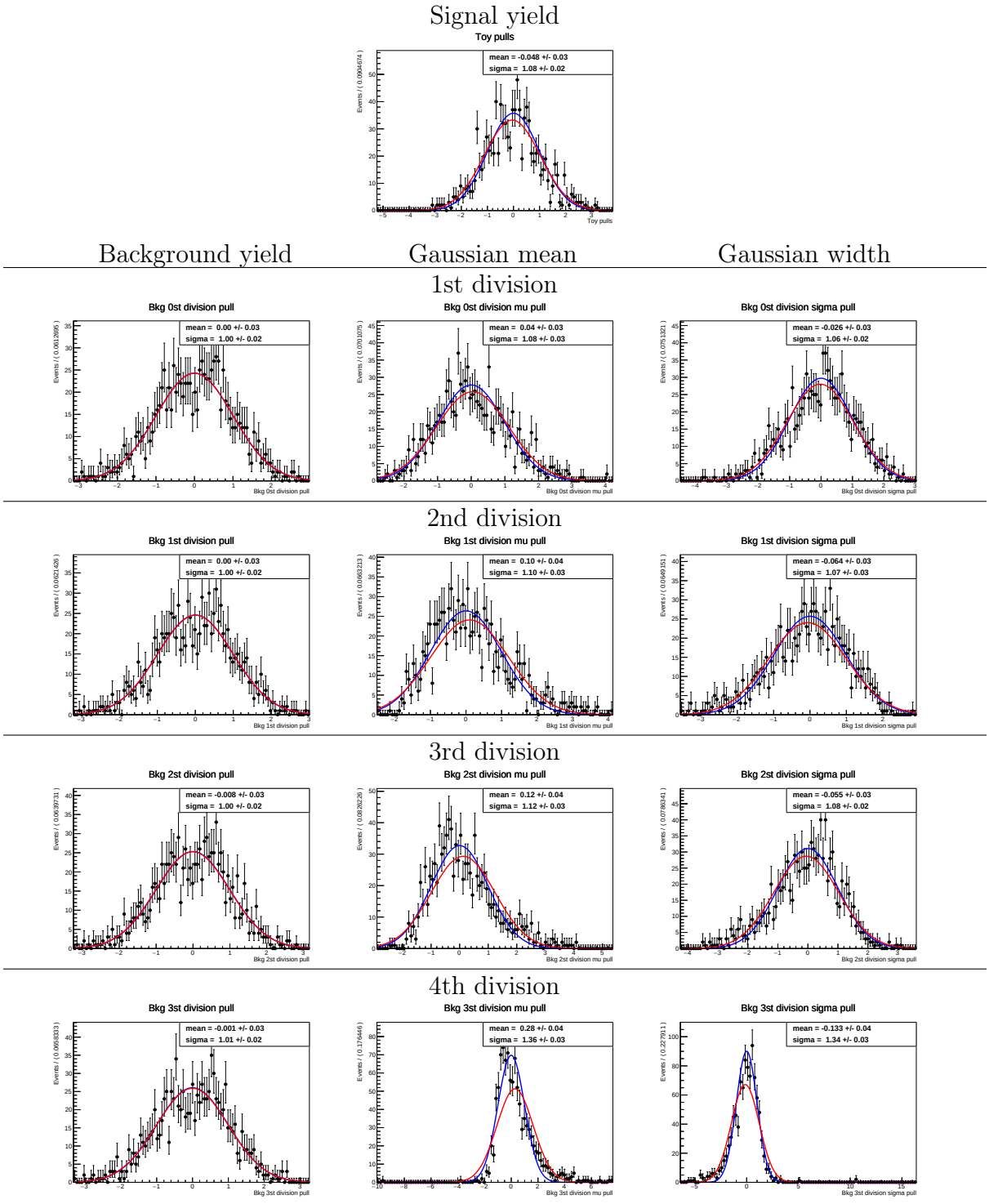

Figure 3.59: $B_s^0 \to \tau^{\pm}\mu^{\mp}$ simultaneous fit free parameters pull distributions for 1000 pseudo-experiments. The red lines correspond to a Gaussian fit to the toy distributions. The blue lines correspond to a N(0,1).

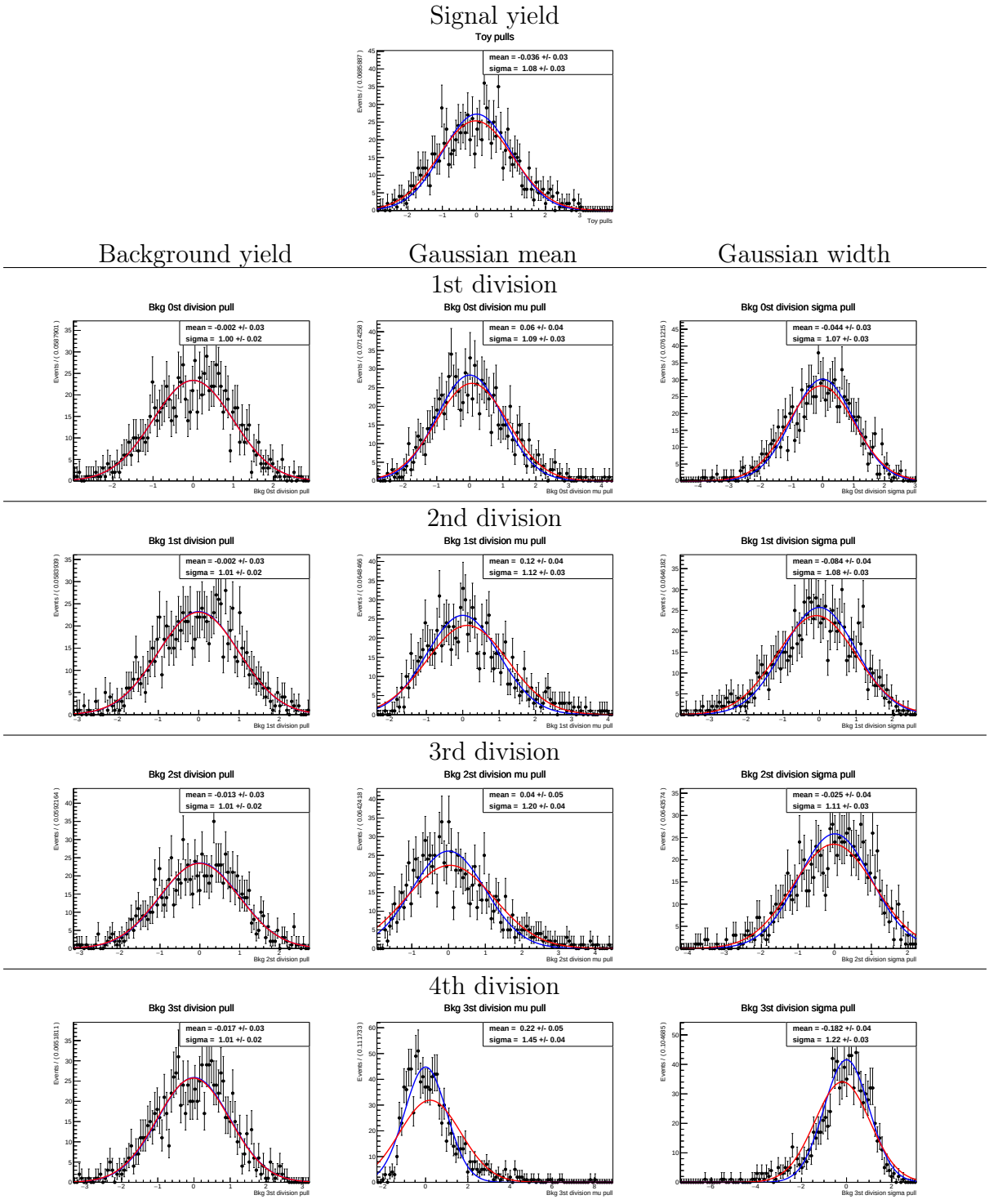

Figure 3.60: $B^0 \to \tau^{\pm}\mu^{\mp}$ simultaneous fit free parameters pull distributions for 1000 pseudo-experiments. The red lines correspond to a Gaussian fit to the toys distribution. The blue lines correspond to a N(0,1).

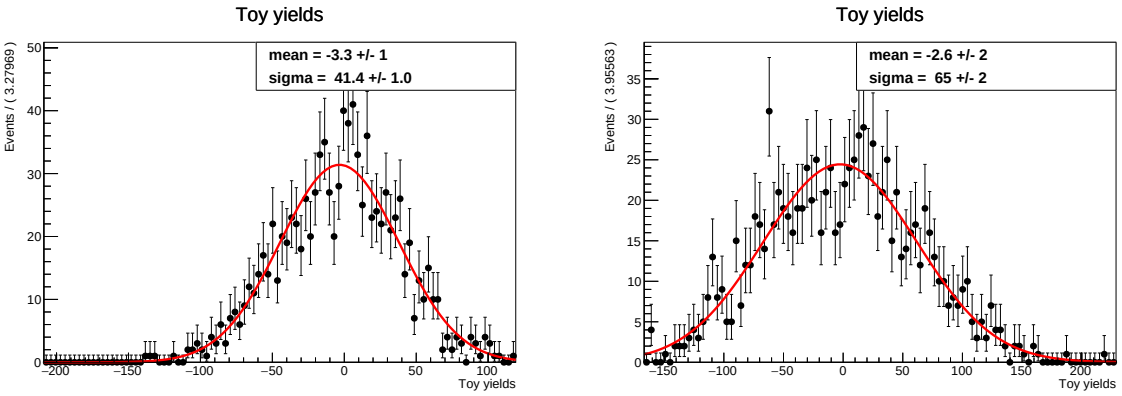

Figure 3.61: $B_s^0 \to \tau^{\pm}\mu^{\mp}$ (left) and $B^0 \to \tau^{\pm}\mu^{\mp}$ (right) fitted signal yield distributions among 1000 toy experiments.

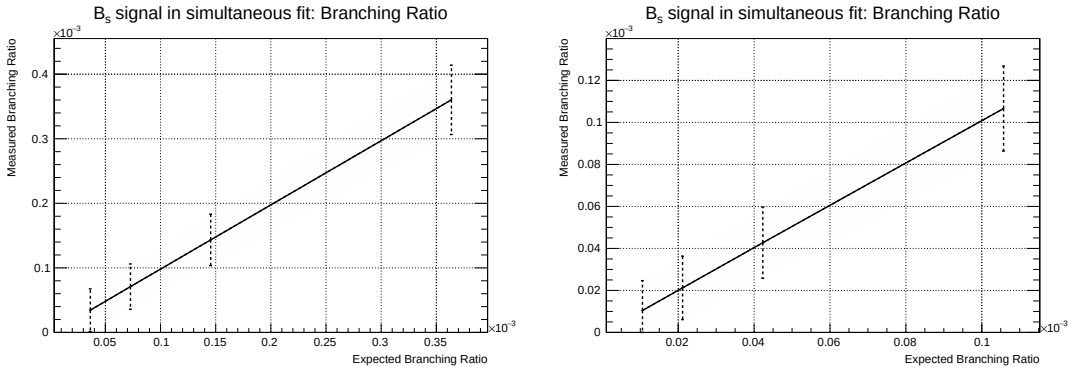

Figure 3.62: Predicted vs computed Branching Ratio ($\mathcal{B}$) using the fitted events on the toy studies including generated signal.

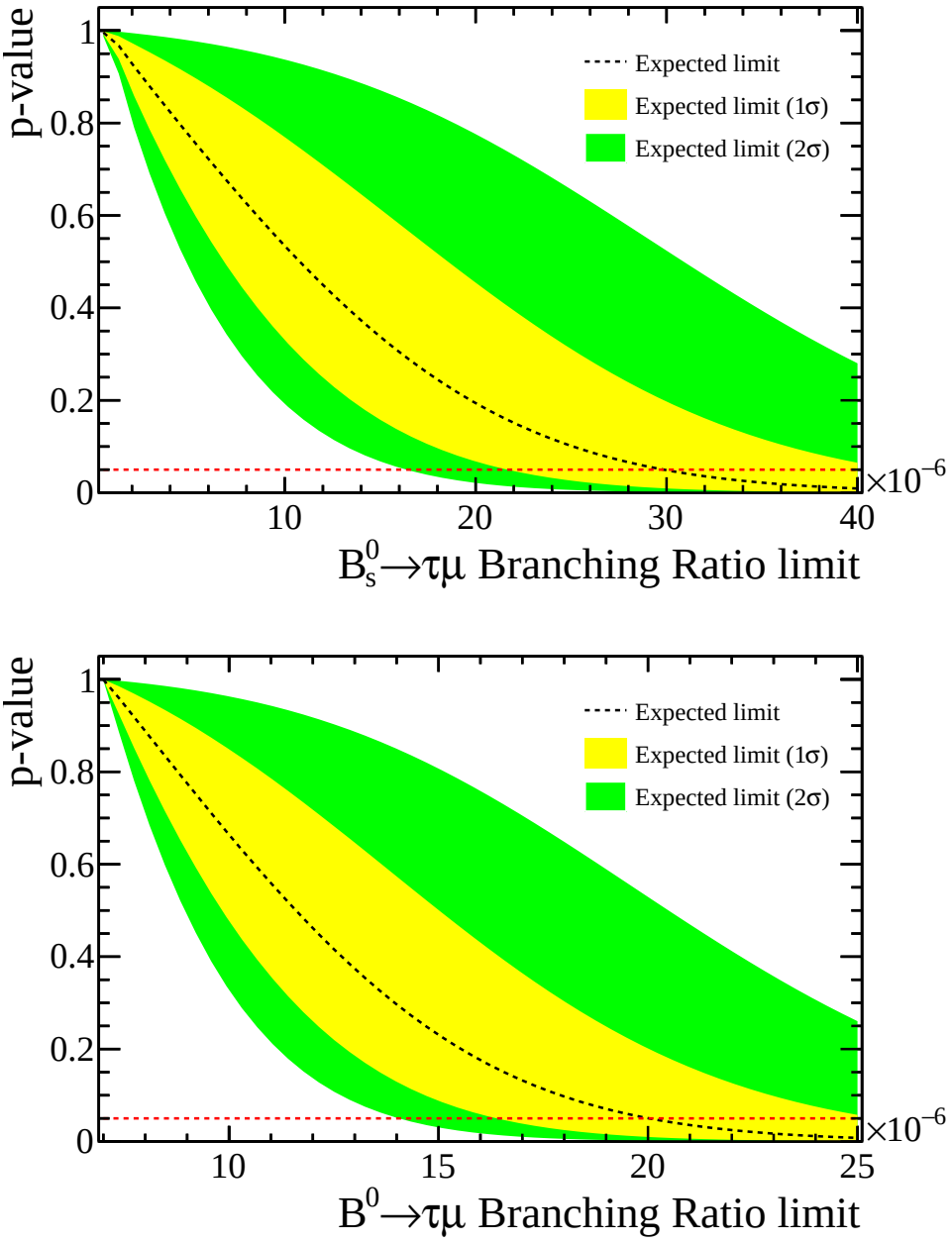

Figure 3.63: $B_s^0 \to \tau^\pm \mu^\mp$ (top) and $B^0 \to \tau^\pm \mu^\mp$ (bottom) exclusion intervals used to evaluate the expected upper limit.

3.9 Post-unblinding strategy

In the not-blinded region, the Same Sign data shape is compatible with the Opposite Sign data. Therefore, the methods presented in section 3.8.2 to extract the signal yields and set limits on the branching ratios are using the Same Sign data as a background proxy. However, once unblinded, the Opposite Sign data may deviate from the expected behaviour. Possible adaptations of the analysis are specified in this section.

First, the way to handle any bias on the fitted signal yield is described in section 3.9.1. Then, potential gain from background shapes being compatible between the four BDT bins is discussed in section 3.9.2. Finally, the strategy to extract the limit on the branching ratios of the seek after signal is developed in section 3.9.3.

3.9.1 Accounting for a potential bias in the fit

The B_s^0 signal yields obtained in 3.8.3.3 when fitting 1000 pseudo-experiments without signal exhibits a small intrinsic bias of $\Delta N_{sig} = -3.3 \pm 1.0$. This corresponds to $\sim$8% of the statistical error ($\delta N_{sig} = 41$) on the fitted yield. As shown on Figure 3.64, the bias remains constant when signal is injected in the data generation.

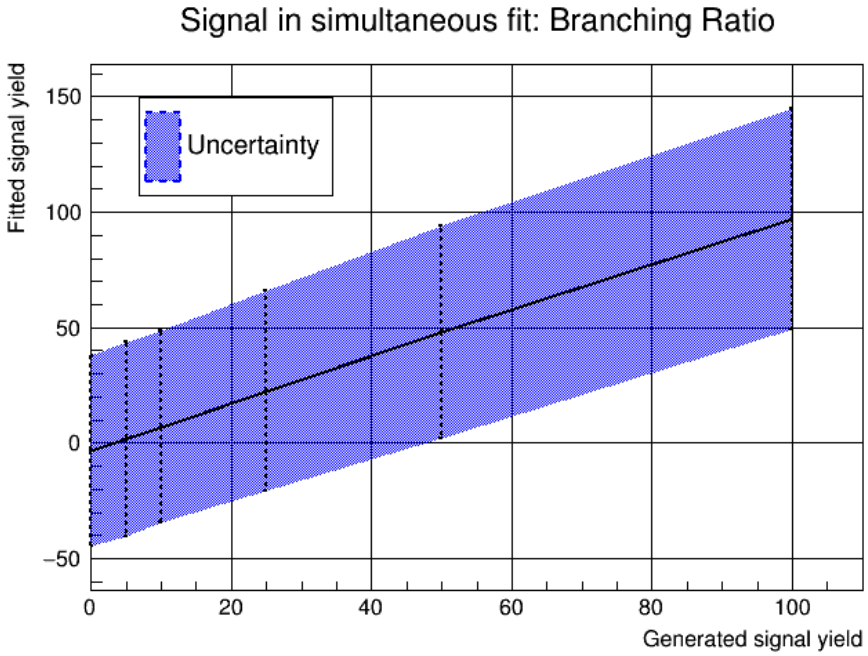

Figure 3.64: Fitted signal events -vs- Generated signal events (1000 toys).

In order to assure a correct coverage of the CLs method in presence of such a bias, the PDF is adapted:

$$PDF^{tot} = \sum_{i}^{\text{BDT bins}} \left((\mathcal{B}\alpha^{-1} + BIAS)\epsilon_i^{sig} Hyp_i^{sig} + n_i^{bkg} Gaus_i^{bkg}(\mu, \sigma) \right), \tag{3.36}$$

where $BIAS$ is a gaussian constrained parameter centred on the observed bias (ΔN_{sig}) with a sigma of 1, corresponding to the uncertainty on ΔN_{sig}. When performing the fit

sensibility studies accounting for the bias using the Same Sign data shape, the obtained signal yield and the signal yield pull are the ones shown in Figure 3.65. It can be seen that the bias has been suppressed. The obtained expected limit is increased by ∼1% when the bias is accounted for.

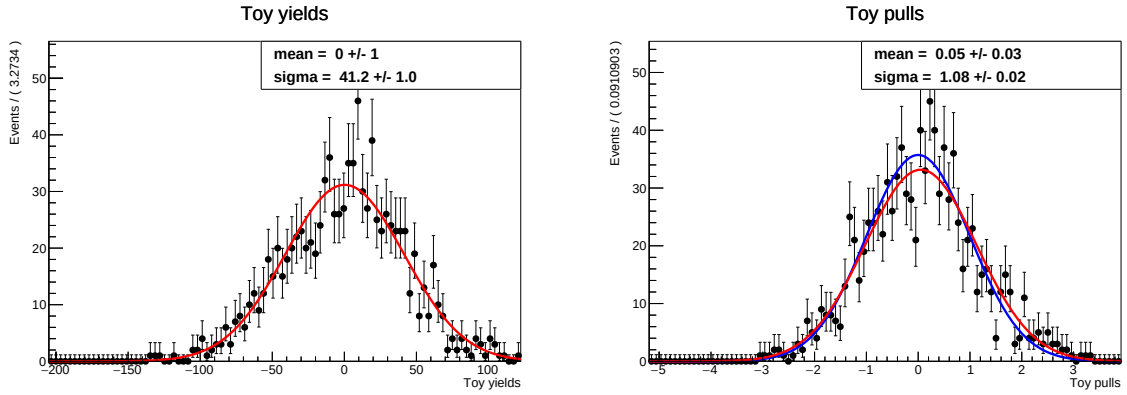

Figure 3.65: $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ signal yield (left) and the signal yield pull (right) for 1000 pseudo-experiments when accounting for the fit bias in the fit model.

3.9.2 Potential improvements

Nothing guaranties that the background shape in each BDT bin will be compatible in the unblinded data. However, given the compatibility of the Same Sign data shapes in each BDT bin, a fit with Gaussian background parameters shared among the BDT bins has been explored. When the background is generated with the same shape in each bin, the bias on the signal yield central value (see Figure 3.66) as well as its uncertainty ($\delta N_{sig} = 32$) is reduced. However, until the data are unblinded, we don't know whether the background shape in each bin are similar enough to allow using this strategy.

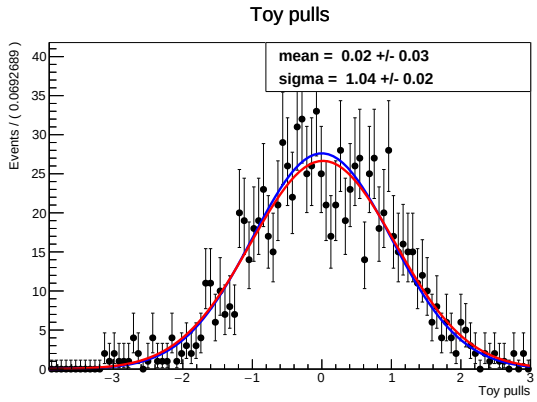

Figure 3.66: Signal yield pulls when the same gaussian background shape is used in all bins and when the parameters of the Gaussian fit to the background is shared among all BDT bins.

3.9.3 Post-unblinding strategy

The strategy to extract the limit with the unblinded data is described below. The dead end possibilities are in red.

1) Fit the unblinded data with the bkg+signal model.

   - Reduce the number of BDT bins if the fit has a bad χ^2, the fitted signal is 3σ lower than 0 or the fit is not convergent.
   - If with 2 BDT bins the fit is still incorrect, cut on the last BDT variable and perform a single unbinned fit. The cut should be optimized following these 3 premises: maximizing the fit sensitivity, that the current model holds and avoiding biases.
   - If the last strategy fails the data should be re-blinded and the fit model should be necessarily changed.
   - Go to next step if the fit is fully convergent with a good χ^2 and the number of fitted signal events is greater than or compatible within 3σ with 0.

2a) In case a signal yield compatible with 0 within 3σ is found:

   i. Extract the unblinded data shape (with the background only model) and the correct background yield per BDT bin.
      - If the data shapes are compatible among the different BDT bins, the sharing of background PDFs parameters will be explored in order to increase the fit sensitivity.
   ii. Perform the toys study as depicted in the Analysis note and estimate the bias of the fit using the toys.
      - In case that the effect of the fit bias becomes dramatic and it can not be overcome, the data should be re-blinded and the fit model should be necessarily changed.
   iii. Extract the limit with the systematic associated to the possible bias.

2b) In case a signal yield greater than 0 with a significance of at least 3σ is found, a branching ratio will be set. In that case, the 2D model (fit at the same time of B_s and B_d signals) strategy should be fully explored and optimized to extract the B_s and B^0 branching ratios simultaneously (see Appendix C).

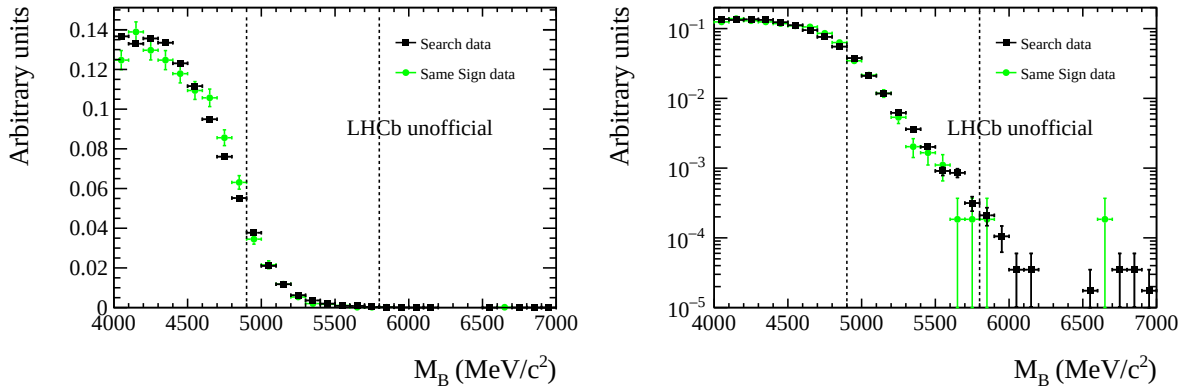

Figure 3.67: Data M_B distribution once the complete offline selection have been applied. The Opposite Sign data is shown in black and the Same Sign data in green. The two plots show the same data, the left one with a linear axis and the right one with a logarithmic axis.

3.10 Unblinded data

The M_B shapes of the Same Sign and the unblinded Opposite Sign data after the complete selection are shown in Figure 3.67. The following qualitative facts are observed:

- good agreement in the blinded region between the Same Sign and the Opposite Sign data,
- no clear peak is observed in the signal region.

In the fitted region, the unblinded Opposite Sign data contains 17746 events which is as expected ten times more than the 1804 Same Sign candidates.

3.10.1 Using the model based on the Same Sign data shape

The model described in Section 3.8.2.3 is based on the Same Sign data shape and consists of an Hypatia distribution to describe the signal and a Gaussian to describe the background.

The simultaneous fit to the unblinded Opposite Sign data in four BDT bins are shown on Figure 3.68. The resulting fit presents a bad χ^2_{ndof} in the first BDT bin. The situation do not improve when the number of bins is changed.

In addition, when the fit is performed independently in each bin, the fitted signal yields do not behave at all like the expected signal contribution that should be equally distributed among the bins. Instead, the fitted signal yields is distributed as a background component with ∼63%, ∼20%, ∼10% and ∼7% of the total fitted signal yield from the first to the last bins. This clearly indicates the presence of a non gaussian tail at high mass in the background distribution that could not be observed with the limited statistics of the Same Sign data.

Therefore, following the post-unblinding strategy, the data in the three most sensitive bins are re-blinded and a new background model is tuned on the first BDT bin of the Opposite Sign data which is largely dominated by background.

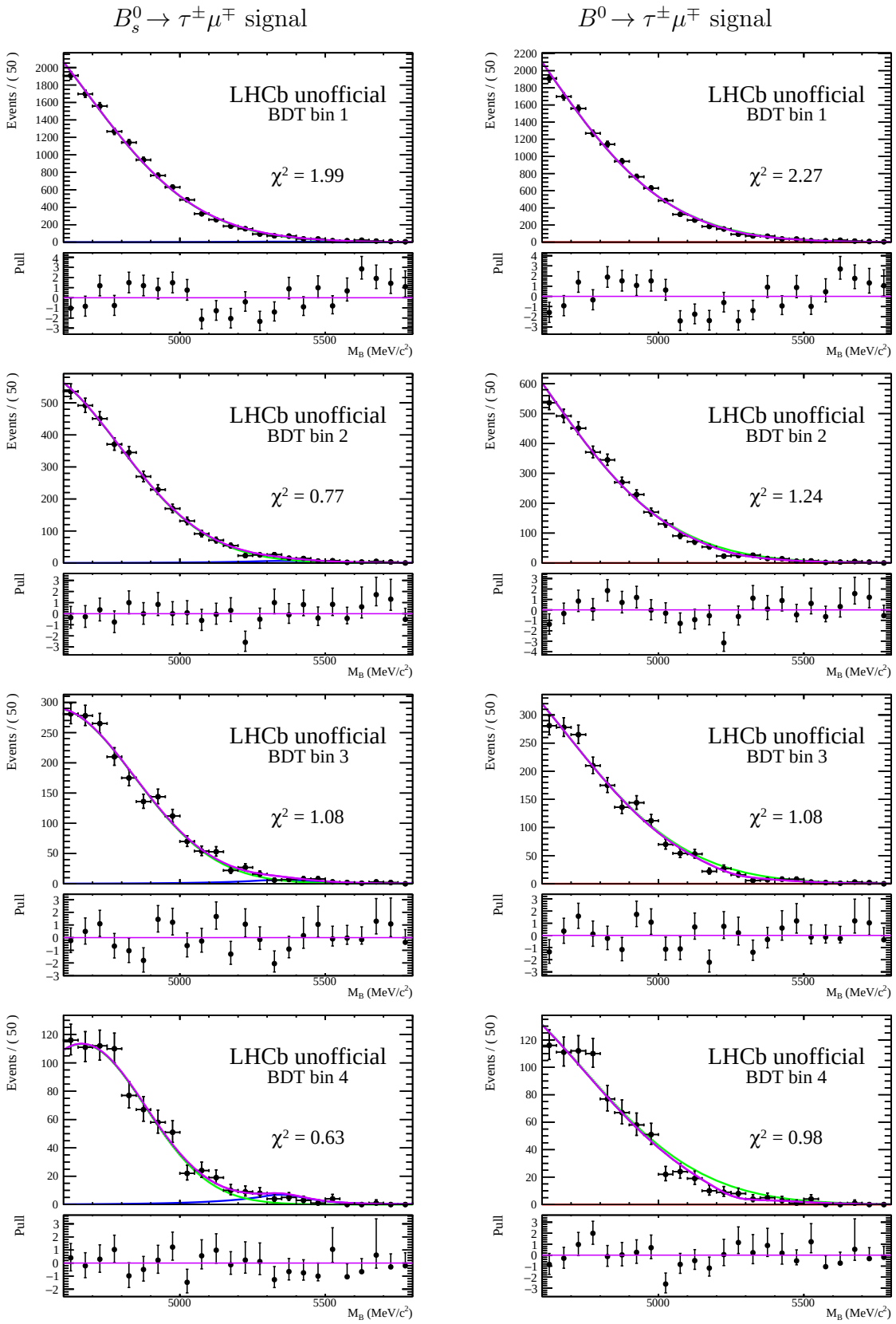

Figure 3.68: $B_s^0 \to \tau^\pm\mu^\mp$ (left) and $B^0 \to \tau^\pm\mu^\mp$ (right) Opposite Sign data fits with the signal + Gaussian background model described in Section 3.8.2.3. The total PDF is shown in purple, the signal in blue for B_s^0 (red for B^0) and the background in green.

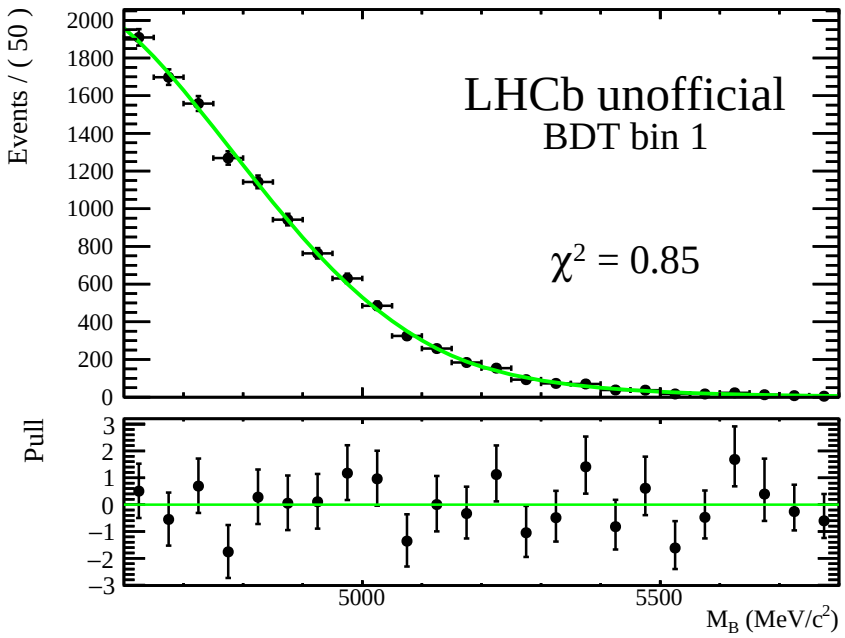

Figure 3.69: Crystal Ball background only fit to the the first BDT bin of the Opposite Sign data.

3.10.2 Fit model generalization

3.10.2.1 Background only model from the mass distribution in the first BDT bin

A Crystal Ball (CB) distribution [66] which naturally extends a Gaussian distribution with an exponential tail at high mass fits nicely to the mass distribution in the first BDT bin, which is expected to be dominated by background, as shown on Figure 3.69. The CB distribution will therefore be used to describe the background in all bins instead of the Gaussian shape.

3.10.2.2 Simultaneous fit with the signal and the new background model

The full model describing the signal and background components in each bin is then the following:

$$PDF^{tot} = \sum_{i}^{\text{BDT bins}} \left(\mathbf{N^{sig}} \epsilon_i^{sig} Hyp_i^{sig} + n_i^{bkg} CB_i^{bkg}(\boldsymbol{\mu}, \sigma_i, \boldsymbol{\alpha}, \boldsymbol{\eta}) \right) \tag{3.37}$$

where:

- $\mathbf{N^{sig}}$: total signal yield (free). Common parameter between the fits.
- n_i^{bkg}: background yield in each BDT bin (free).
- ϵ_i^{sig}: signal efficiency per BDT bin taken from MC. Floating with Gaussian constraints. The width of the Gaussian constraint corresponds to the total uncertainty of the per bin signal efficiency reported in Table 3.31.

- Hyp_i^{sig}: Hypatia PDF describing the signal shape with the parameters fixed from MC fit. The hypathia width is allowed to vary with Gaussian constraints.
- $CB_i^{bkg}(\boldsymbol{\mu}, \sigma_i, \boldsymbol{\alpha}, \boldsymbol{\eta})$: Crystal Ball PDF describing the background shape (parameters free).

The parameters in red are left free during the fit process and the bold parameters are shared among the BDT bins. The tail parameters and the mean of the Crystal Ball are shared among all bins as the mass distributions in each bin are compatible. This reduces the number of free parameters and thus allows to get a good fit stability. However, the width of the CB in each bin is left free in the fit to account for potential small remaining differences in the background shapes in each bin. In summary, the simultaneous fit in 4 BDT bins fit contains:

- 12 free parameters,
- and 8 constrained parameters (with Gaussian constraints).

The Opposite Sign data fit is shown in Figure 3.70, leading to the following number of signal events observed:

- $\mathrm{N}^{\mathrm{sig}}_{B_s^0 \to \tau^\pm \mu^\mp} = -18 \pm 38$,
- $\mathrm{N}^{\mathrm{sig}}_{B^0 \to \tau^\pm \mu^\mp} = -63 \pm 57$.

corresponding to a 0.3 and 1.1σ downward fluctuations respectively. Therefore, no significant signal excess is observed. The fit parameters are shown in Table 3.35.

3.10.2.3 Fit validation

Pseudo-experiment studies (generated as a background only distribution) are performed as depicted in Section 3.8.3, by substituting the mentioned Gaussian background PDF by a Crystal Ball background PDF, thus checking the stability of the model in Equation 3.37. The rate of convergence of this model is of ∼93%.

Attempts to free more Crystal Ball parameters in the fit were performed yielding to a smaller convergence rate (∼30% when all parameters are freed) and larger bias. The model with shared mean and tail parameters of the background Crystal Ball is therefore used to extract the branching ratio limits. The number of bins is kept at 4 as increasing the number of BDT bins does not provide a significant improvement of the fit sensitivity.

Being the fit absent of a significant bias in the B_s^0 channel, the B^0 case presents a bias on the signal yield corresponding to ∼10% of the signal yield uncertainty, which is overcome with the strategy described in Section 3.9.1; this information is shown in Figure 3.71, containing the signal yield and pull distributions among the pseudo-experiments.

3.10.3 Branching ratio limits extraction

The model used to extract the limits is the one depicted in Section 3.10.2.2. The asymptotic CLs method (described in Section 3.8.4) is used to extract upper limits on the branching fractions. The expected upper limit (Figure 3.72) is obtained using the following scan parameter ranges:

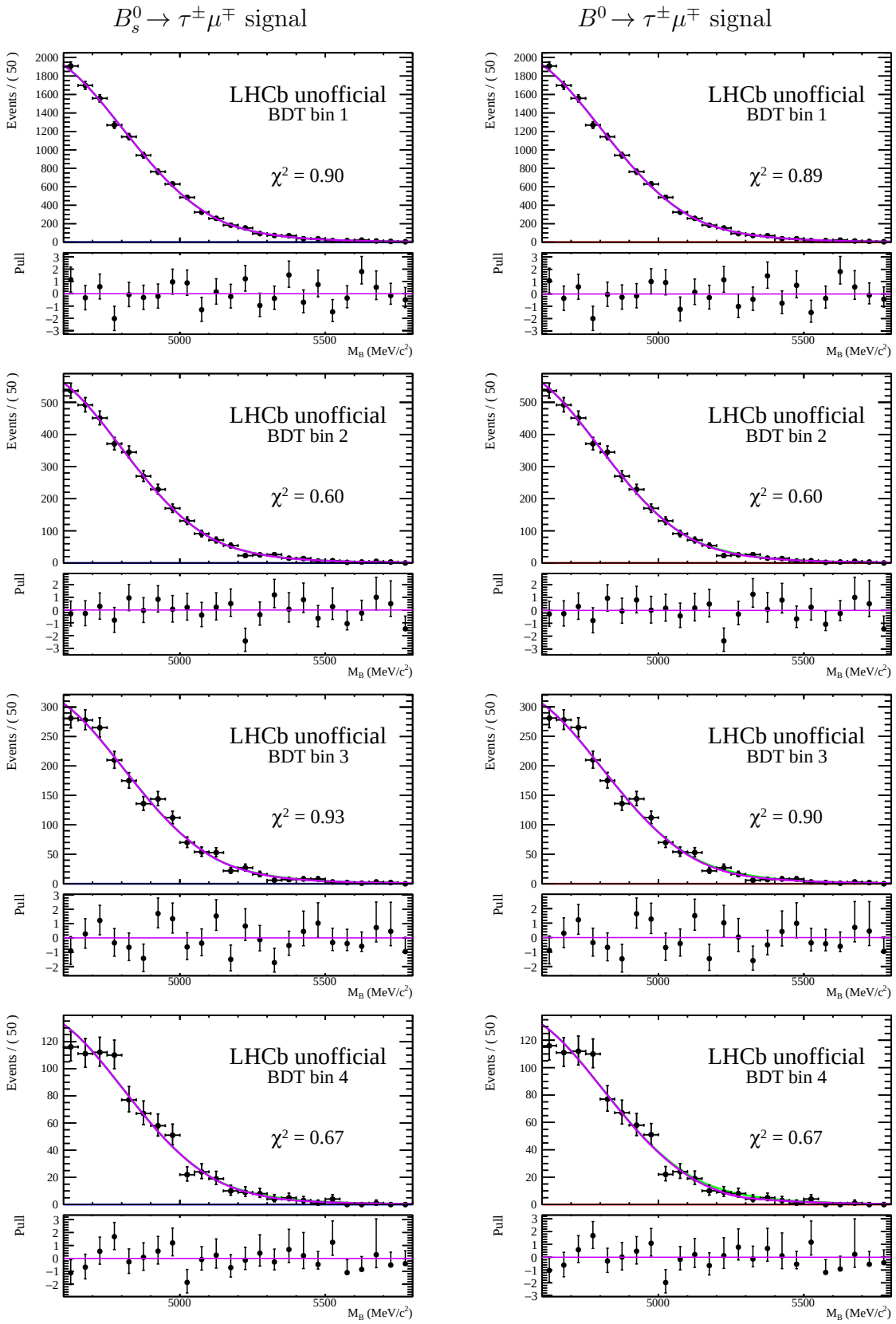

Figure 3.70: $B_s^0 \to \tau^\pm\mu^\mp$ (left) and $B^0 \to \tau^\pm\mu^\mp$ (right) Opposite Sign data fits with the signal+background model described in Section 3.10.2.2. The total PDF is shown in purple, the signal in blue for B_s^0 (red for B^0) and the background in green.

		B_s^0	B^0
Division	Parameter	Value	Value
-	N_{sig}	-18 ± 38	-63 ± 57
	CB μ	4484 ± 17	4486 ± 28
	CB α	-1.9 ± 0.2	-1.9 ± 0.2
	CB η	15 ± 13	24 ± 27
1	CB σ	313 ± 14	312 ± 12
	n_{bkg}	11716 ± 108	11728 ± 109
2	CB σ	307 ± 8	307 ± 13
	n_{bkg}	3344 ± 59	3355 ± 59
3	CB σ	316 ± 9	317 ± 14
	n_{bkg}	1887 ± 44	1898 ± 47
4	CB σ	315 ± 12	319 ± 17
	n_{bkg}	816 ± 30	826 ± 31

Table 3.35: $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$ Opposite Sign data fits parameter values with the signal+background model described in Section 3.10.2.2.

Mode	Limit	$90\%CL$	$95\%CL$
$B_s^0 \to \tau^\pm\mu^\mp$	Observed	$2.5 \cdot 10^{-5}$	$3.0 \cdot 10^{-5}$
	Expected	$3.0 \cdot 10^{-5}$	$3.6 \cdot 10^{-5}$
$B^0 \to \tau^\pm\mu^\mp$	Observed	$1.0 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$
	Expected	$1.3 \cdot 10^{-5}$	$1.6 \cdot 10^{-5}$

Table 3.36: Upper limits for the lepton flavor violating decays $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$.

- $B_s^0 \to \tau^\pm\mu^\mp$: 50 scan points in range $[5.0 \cdot 10^{-8}, 4.6 \cdot 10^{-5}]$
- $B^0 \to \tau^\pm\mu^\mp$: 50 scan points in range $[8.0 \cdot 10^{-8}, 2.1 \cdot 10^{-5}]$

The resulting $\mathcal{B}$ observed and expected limits are shown in Table 3.36.

The limits for the B_s^0 light and heavy mass-eigenstate are computed in Appendix D.1) and no significant differences are found.

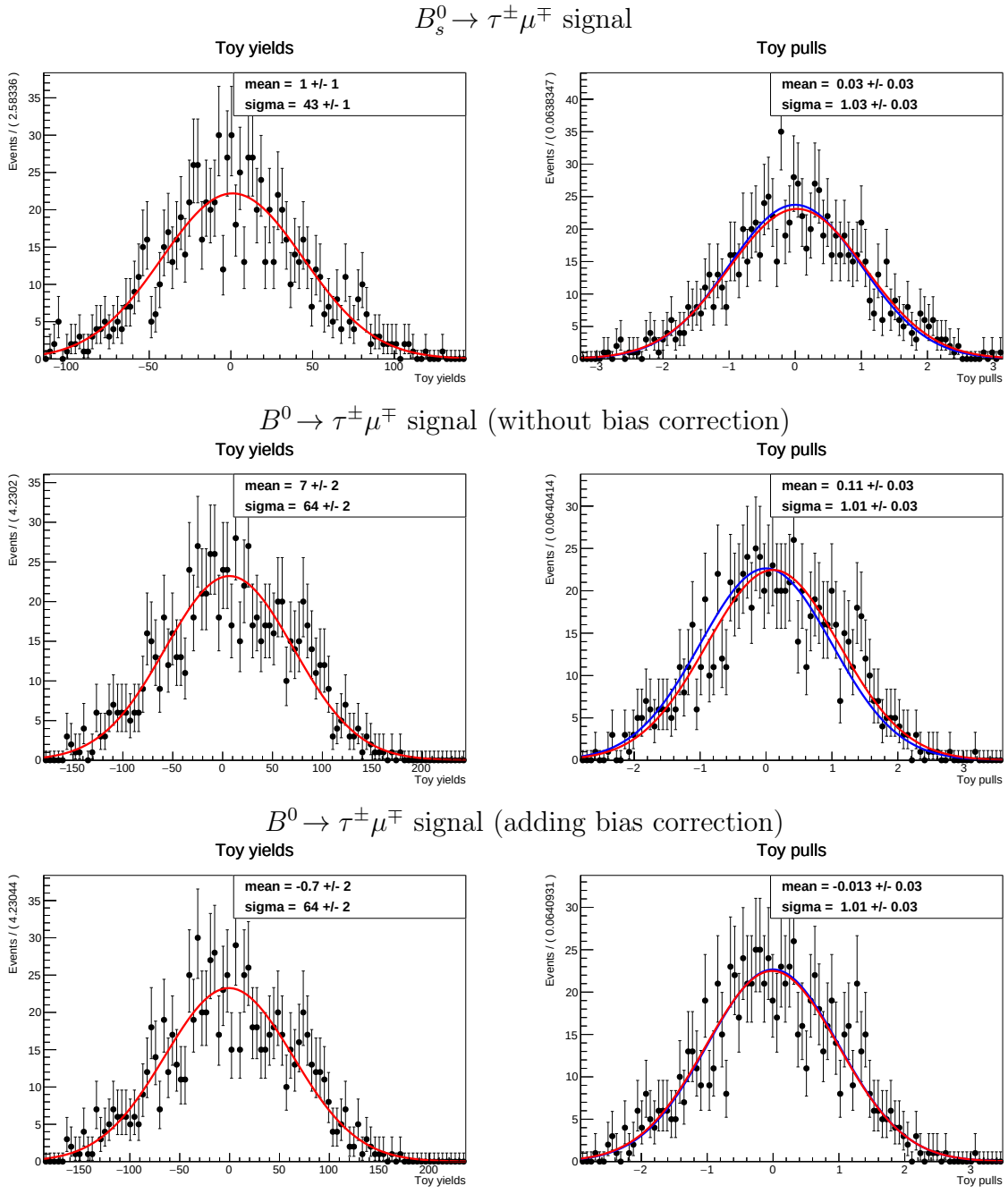

Figure 3.71: Signal yield (left) and signal yield pull (right) distributions among 1000 background only pseudo-experiments. The Gaussian fitted distributions are shown in red and the blue line represent a reference N(0,1).

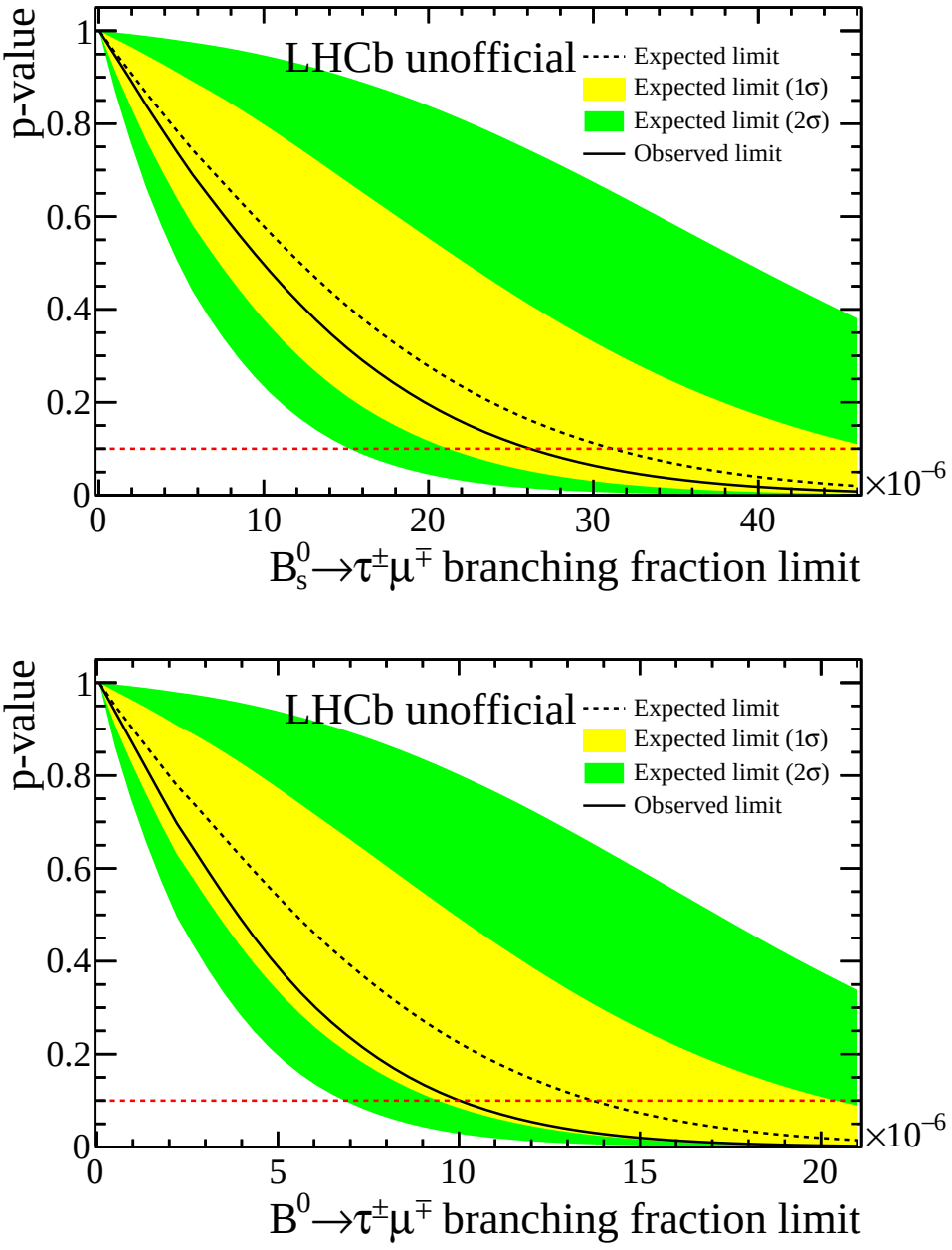

Figure 3.72: $B_s^0 \to \tau^\pm\mu^\mp$ (top) and $B^0 \to \tau^\pm\mu^\mp$ (bottom) exclusion intervals used to evaluate the expected upper limit.

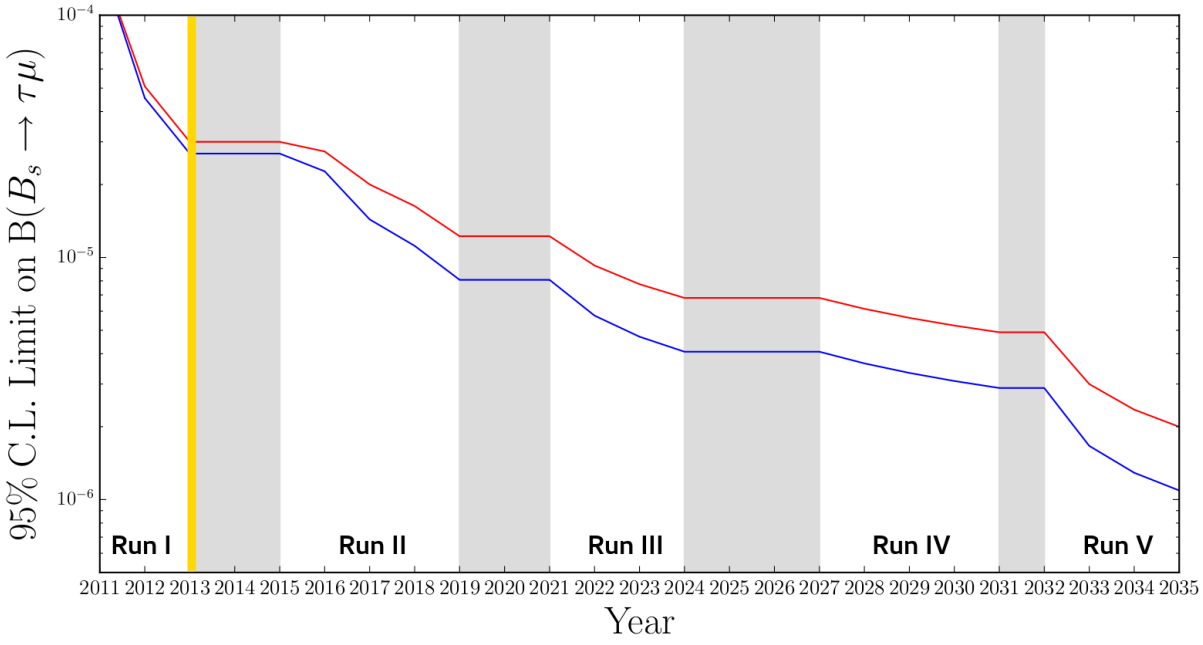

Figure 3.73: $B_s^0 \to \tau^\pm\mu^\mp$ expected branching fraction limit with respect the total LHCb recorded data for the coming years. The red line assumes the same signal efficiency as the Run I analysis and the blue line assumes also certain additional improvement conditions detailed in Section 3.11.

3.11 Prospects

Figure 3.73 shows the extrapolated evolution of the $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp)$ expected limit for the planned data taking periods of LHCb, from Run I (years 2011-2012) to Run V (2032-2035).

The expected limit will improve continuously once more data is available. However, aside from the increase in statistics, the following assumptions on additional improvements have been used to perform a more reliable extrapolation:

- 25% additional contribution from $B_s^0 \to \tau^\pm(\to \pi^\pm\pi^\mp\pi^\pm\pi^0\nu)\mu^\mp$ from Runs II (partial optimization of the future selection for this channel).
- 20% of additional improvement in Run III trigger upgrade.
- 15% of additional improvement in Run V from reconstruction (magnet chambers).

From the extrapolation is expected that the $\mathcal{B}(B_s^0 \to \tau^\pm\mu^\mp)$ limit reaches $\sim 10^{-6}$ by the end of the LHC, contributing this way to the discard of more BSM models predicting Lepton Flavor Violation.

3.12 Summary and conclusion

The search for the $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$ decays presented in this document is performed with the LHCb run 1 data. In this search, the τ lepton is reconstructed in the 3-prong $\tau^\pm \to \pi^\pm\pi^\mp\pi^\pm\nu$ channel. A custom mass computation taking into account the unmeasured momentum of the neutrino which peaks around at the measured B mass

value is used to extract the signal yield via an unbinned simultaneous maximum likelihood fit in bins of the output of a BDT discriminating signal and background. Several selection procedures, some of which using BDT's, have been devised to drastically reduce the background level, allowing such fit to be performed. The efficiency of the selection is estimated using simulation and some data-driven techniques to account for effects not well reproduced in simulation.

No significant excesses are observed for the two modes and the obtained branching ratio expected (observed) limits are: $\mathcal{B}\left(B_s^0 \to \tau^{\pm}\mu^{\mp}\right) < 3.0(2.5) \cdot 10^{-5}$ and $\mathcal{B}\left(B^0 \to \tau^{\pm}\mu^{\mp}\right) < 1.3(1.0) \cdot 10^{-5}$ at 90%CL. These results represent the best upper limits to date being the first measurement for the B_s^0 mode.

With further improvements in the selection process, e.g. increasing the contribution of the $\tau^{\pm} \to \pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^0\nu$ mode on the signal efficiency, and accounting for the additional improvements of LHCb trigger and reconstruction, the limits are expected to improve by one order of magnitude by the end of the LHC.

Appendix A

Analytic reconstruction of the decay kinematics

The B candidate decay vertex position (V_B) and the neutrino 4-momentum (P_ν) are expressed in terms of the measured quantities :

- Primary Vertex coordinates (PV).
- τ decay vertex coordinates (V_τ).
- 3 π system 4-momentum ($P_{3\pi}$).
- μ 4-momentum (P_μ).

The decay plane of the B two-body decay is defined by the μ direction ($\hat{n}_\mu$), given by the μ momentum, and the B flight direction ($\hat{n}_B$), given by the PV and the V_B positions (Figure 3.9). The orthonormal reference frame ($\hat{\imath}\hat{\jmath}\hat{k}$) is defined such that $\hat{\imath}$ is along the μ direction ($\hat{\imath} = \hat{n}_\mu$), $\hat{\jmath}$ lies in the decay plane and $\hat{k}$ is orthogonal to the decay plane.

The following conditions have to be fulfilled:

- Only the neutrino and the 3π momentum $\hat{k}$ components are outside the decay plane.

$$p_{3\pi}^{\hat{k}} = p_\nu^{\hat{k}} \tag{A.1}$$

- V_B along the μ direction. Being U_μ the coordinates of a given point in the μ track:

$$\begin{aligned} \vec{P}_\mu \wedge (U_\mu - V_B) &= 0 \\ p_\mu^{\hat{\imath}} \left(V_B^{\hat{\jmath}} - U_\mu^{\hat{\jmath}} \right) &= 0 \end{aligned} \tag{A.2}$$

- B momentum in the direction given by the PV and V_B when the neutrino is added:

$$\begin{aligned} \vec{P}_B \wedge (V_B - PV) &= 0 \\ \left(p_{3\pi}^{\hat{\jmath}} + p_\nu^{\hat{\jmath}} \right) \left(V_B^{\hat{\imath}} - PV^{\hat{\imath}} \right) - \left(p_{3\pi}^{\hat{\imath}} + p_\mu^{\hat{\imath}} + p_\nu^{\hat{\imath}} \right) \left(V_B^{\hat{\jmath}} - PV^{\hat{\jmath}} \right) &= 0 \end{aligned} \tag{A.3}$$

- τ momentum in the direction given by the V_B and V_τ when the neutrino is added:

$$\begin{aligned} \vec{P}_\tau \wedge (V_\tau - V_B) = 0 \\ \left(p_{3\pi}^{\hat{i}} + p_\nu^{\hat{i}}\right)\left(V_\tau^{\hat{j}} - V_B^{\hat{j}}\right) - \left(p_{3\pi}^{\hat{j}} + p_\nu^{\hat{j}}\right)\left(V_\tau^{\hat{i}} - V_B^{\hat{i}}\right) = 0 \end{aligned} \tag{A.4}$$

- 4-momentum conservation in the τ decay vertex.

$$\begin{aligned} P_\tau^\alpha - P_{3\pi}^\alpha - P_\nu^\alpha = 0 \\ \frac{1}{2}\left(M_\tau^2 - M_{3\pi}^2\right) - p_{3\pi}^{\hat{k},2} + p_{3\pi}^{\hat{i}} p_\nu^{\hat{i}} + p_{3\pi}^{\hat{j}} p_\nu^{\hat{j}} \\ -\sqrt{p_{3\pi}^{\hat{i},2} + p_{3\pi}^{\hat{j},2} + p_{3\pi}^{\hat{k},2} + M_{3\pi}^2}\sqrt{p_\nu^{\hat{i},2} + p_\nu^{\hat{j},2} + p_{3\pi}^{\hat{k},2}} = 0 \end{aligned} \tag{A.5}$$

With this system of equations the process is kinetically closed and the expression of the V_B and the P_ν are found. As the 4-momentum conservation in the τ decay vertex is a second order equation, the P_ν has a two-fold ambiguity. Finally, the B candidate invariant mass can be computed using the conservation of the 4-momentum in the decay chain:

$$\begin{aligned} P_B^\alpha =& P_\mu^\alpha + P_{3\pi}^\alpha + P_\nu^\alpha \\ M_B^2 =& \left(\sqrt{p_\mu^{\hat{i},2} + p_\mu^{\hat{j},2} + p_{3\pi}^{\hat{k},2} + M_\mu^2} + \sqrt{p_{3\pi}^{\hat{i},2} + p_{3\pi}^{\hat{j},2} + p_{3\pi}^{\hat{k},2} + M_{3\pi}^2} + \sqrt{p_\nu^{\hat{i},2} + p_\nu^{\hat{j},2} + p_{3\pi}^{\hat{k},2}}\right)^2 \\ & - \left(p_\mu^{\hat{i}} + p_{3\pi}^{\hat{i}} + p_\nu^{\hat{i}}\right)^2 - \left(p_\mu^{\hat{j}} + p_{3\pi}^{\hat{j}} + p_\nu^{\hat{j}}\right)^2 \end{aligned} \tag{A.6}$$

Two solutions exist for the B invariant mass depending of which P_ν is used: $B\oplus$ and $B\ominus$. The correct behavior of the analytic solution can be seen by applying the reconstruction process to the B_s^0 signal Monte-Carlo truth, which corresponds to the simulation of the unique decay $B_s^0 \to \tau(\pi\pi\pi\nu)\mu$ (Figure A.1). It can be seen that the invariant B masses are close to a Dirac delta in the B_s^0 mass value with some radiative tails. However, there are 4% of the cases where there are no physical solutions to the reconstruction, in case of negative discriminant.

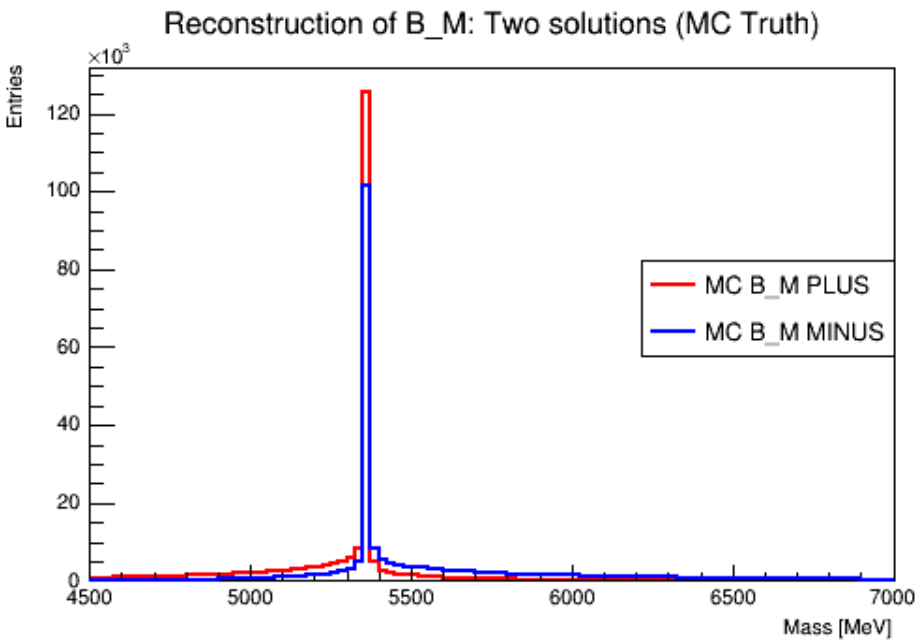

Figure A.1: Analytic reconstruction applied to the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ Monte-Carlo truth sample.

Appendix B

Variables data-MC agreement

2011 and 2012 data-MC agreement for the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ selection and last BDT is shown in Figures B.1, B.2, B.4 and B.4. The $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ signal MC sample and the data sample with the implemented offline selection described in Section 3.3 are used to check the data-MC agreement.

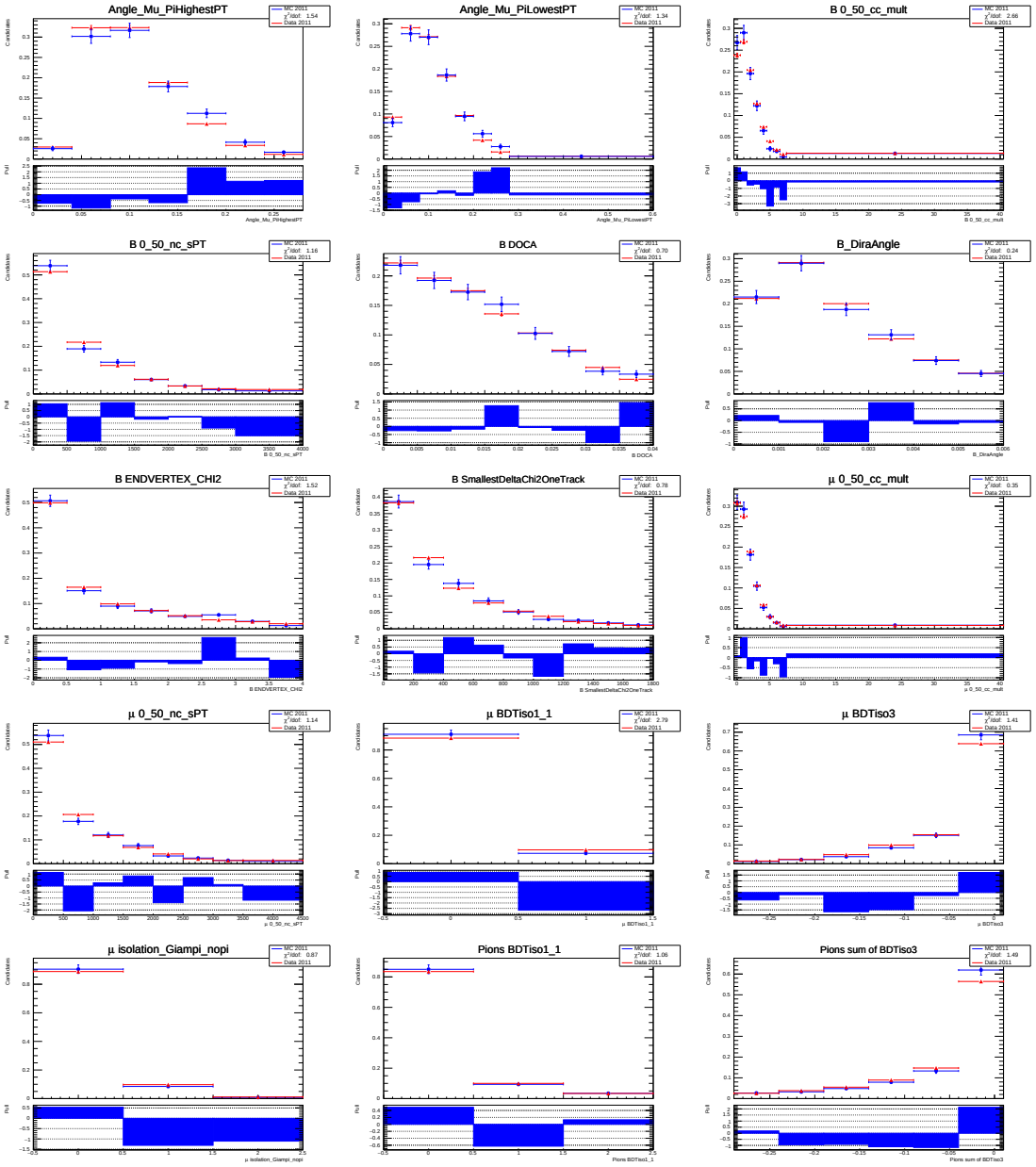

Figure B.1: (Part 1) 2011 Data-MC agreement for the variables used in $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection without weights applied.

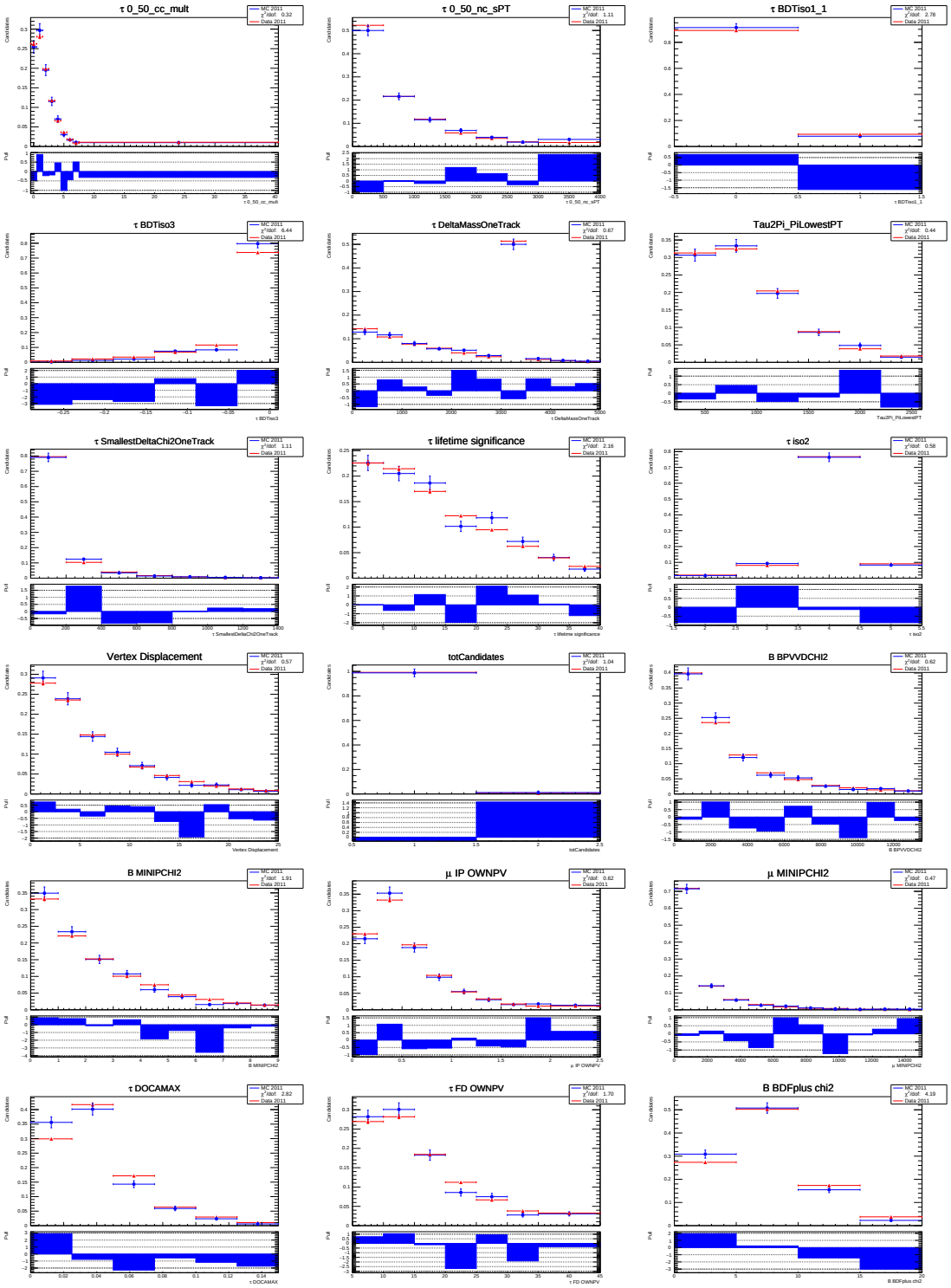

Figure B.2: (Part 2) 2011 Data-MC agreement for the variables used in $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection without weights applied.

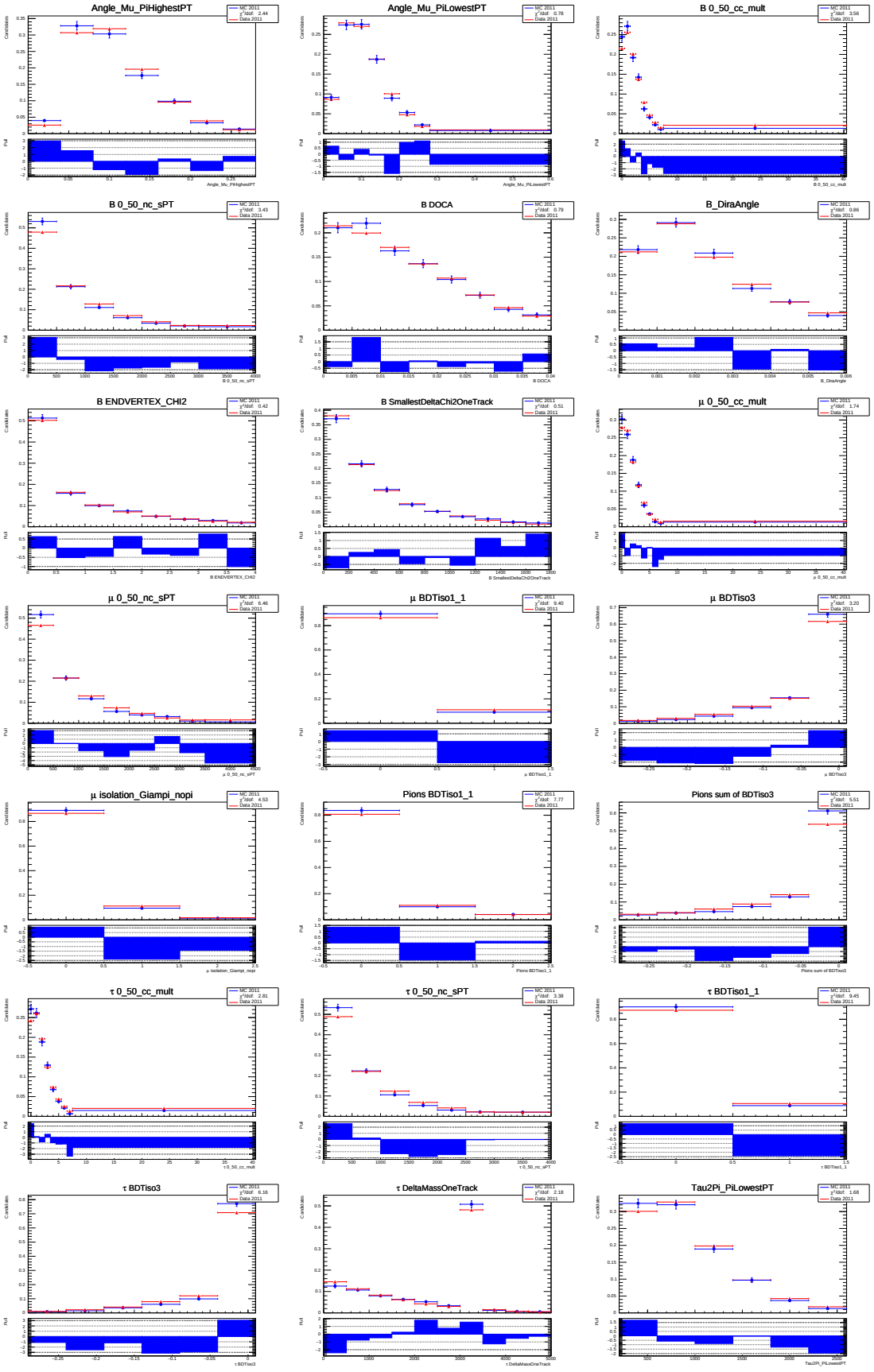

Figure B.3: (Part 1) 2012 Data-MC agreement for the variables used in $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection without weights applied.

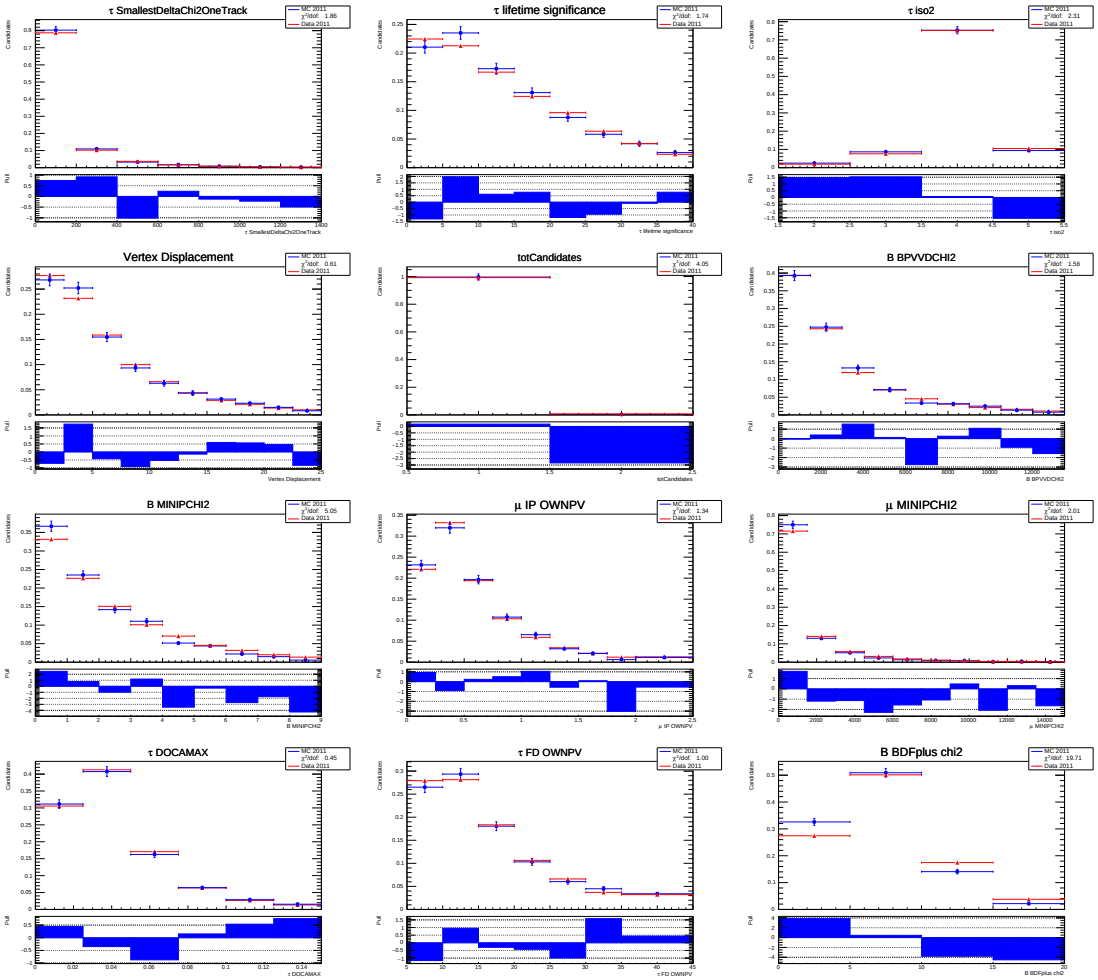

Figure B.4: (Part 2) 2012 Data-MC agreement for the variables used in $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection without weights applied.

Appendix C

2D simultaneous fit exploration

Due to the limited separation of the $B^0 \to \tau^{\pm}\mu^{\mp}$ and $B_s^0 \to \tau^{\pm}\mu^{\mp}$ signals, a 1D fit has been used to extract the signal yield and compute the expected upper $\mathcal{B}$ limits as described in section 3.8. The 1D fit consists in performing the data fit twice: a first time assuming all observed signal events as produced by the B^0 decay, and a second time considering all observed signal events as if they were coming from a B_s^0 decay. The 1D fit is the standard strategy where the data does not contain signal as it provides reliable results without complicating uselessly the signal yield extraction strategy.

A 2D fit consists in fitting the B^0 and B_s^0 signal yield. The used model must describe the background and the two signal shapes in a unique total Probability Density Function (PDF). Studies of the use of a 2D simultaneous fit in the $B_{(s)}^0 \to \tau^{\pm}\mu^{\mp}$ search have been carried out using the following model in the simultaneous fit:

$$PDF^{tot} = \sum_{i}^{\text{BDT bins}} \left(\mathbf{N}_{\mathbf{B_s}}^{\mathbf{sig}} \epsilon_{i,B_s}^{sig} Hyp_{i,B_s}^{sig} + \mathbf{N}_{\mathbf{B_d}}^{\mathbf{sig}} \epsilon_{i,B_d}^{sig} Hyp_{i,B_d}^{sig} + n_i^{bkg} Gaus_i^{bkg}(\mu_i, \sigma_i) \right),$$

where

- $\mathbf{N}_{\mathbf{B_{(s)}}}^{\mathbf{sig}}$: Total $B_{(s)}$ signal yield. Common parameters in the simultaneous fit.
- n_i^{bkg}: Background yield in each BDT bin. These parameters is left free in the fit.
- $\epsilon_{i,B_{(s)}}^{sig}$: Fixed from MC with Gaussian constraints accounting for the statistical uncertainty. Computed in 3.8.1.3.
- $Hyp_{i,B_{(s)}}^{sig}$: Hypathia PDF accounting for the signal. They are fixed from MC. The widths are allowed to vary within Gaussian constraints.
- $Gaus_i^{bkg}(\mu_i, \sigma_i)$: Gaussian PDF accounting for the background. Free parameters.

In order to check the fit stability and bias, studies using 1000 pseudo-experiments following the Same Sign data Gaussian shape have been carried out in the way described in Section 3.8.3. The last BDT output is divided in 4 bins and the number of background events are generated from the expected number of unblinded events in the fit region: 14000.

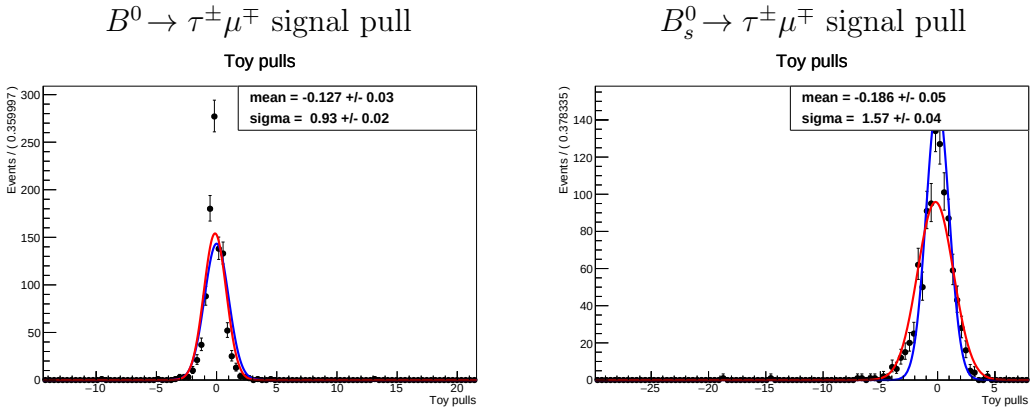

Figure C.1: Signal yield pull obtained using a 2D simultaneous fit model to background only pseudo-experiments. $B^0 \to \tau^\pm\mu^\mp$ (left) $B_s^0 \to \tau^\pm\mu^\mp$ (right).

C.1 Absence of signal

Generating the pseudo-experiments as background only distributions, the corresponding signals pull distributions, computed as in Equation 3.33, are shown in Figure C.1. Using a Gaussian fit to the observed signal yields distributions among the pseudo-experiments, the results are:

- for $B^0 \to \tau^\pm\mu^\mp$ $N_{sig}^{obs} = -24 \pm 102$ events and
- for $B_s^0 \to \tau^\pm\mu^\mp$ $N_{sig}^{obs} = -3 \pm 42$ events.

The fit strategy works as more than 99% of the fits are convergent. However, from the signal yield pulls it can be observed that the fit is biased. No solution has been found to correct for this bias. Therefore, for 0 signal, no benefits are expected from 2D fit strategy with respect to the 1D fit strategy.

C.2 Presence of signal

In this case a significant amount of signal has been injected to the pseudo-experiments: 500 B^0 and 500 B_s^0 signal events. A given pseudo-experiment is illustrated in Figure C.2.

The corresponding signal yield pull distributions are shown in Figure C.3. Using a Gaussian fit to the observed signal yields distributions among the pseudo-experiments, the results are:

- for $B^0 \to \tau^\pm\mu^\mp$ $N_{sig}^{obs} = 468 \pm 450$ events and
- for $B_s^0 \to \tau^\pm\mu^\mp$ $N_{sig}^{obs} = 469 \pm 92$ events.

In this case the fit model can not disentangle if the signal events come from B^0 or B_s^0 decays, as it is reflected in the B^0 case signal yield uncertainty. The fit strategy lacks information on the correlation between the 2 signals.

If in future updates of the measurements a significant signal excess is found, the strategy of a 2D simultaneous fit will need to be completed and fully explored, as it will be the only way of setting unambiguously the two Branching Ratios.

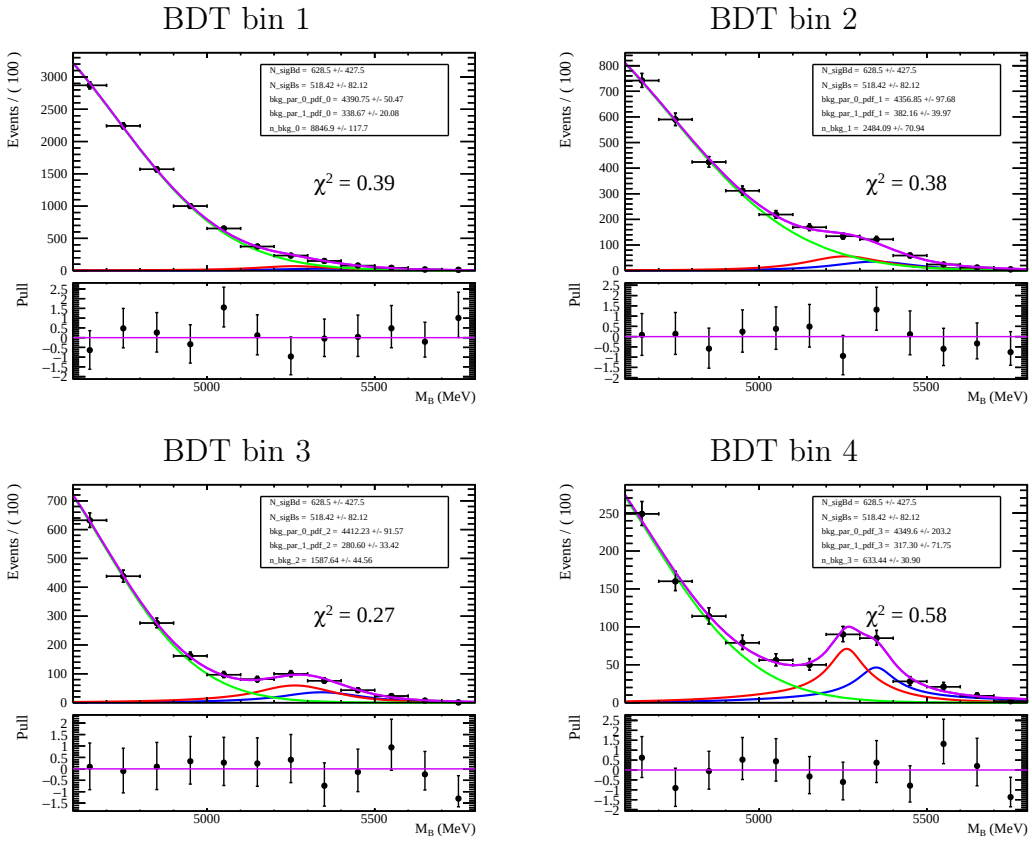

Figure C.2: Pseudo-experiment generated with 14000 background events and 500 events for each signal. It is fitted with the 2D simultaneous fit model.

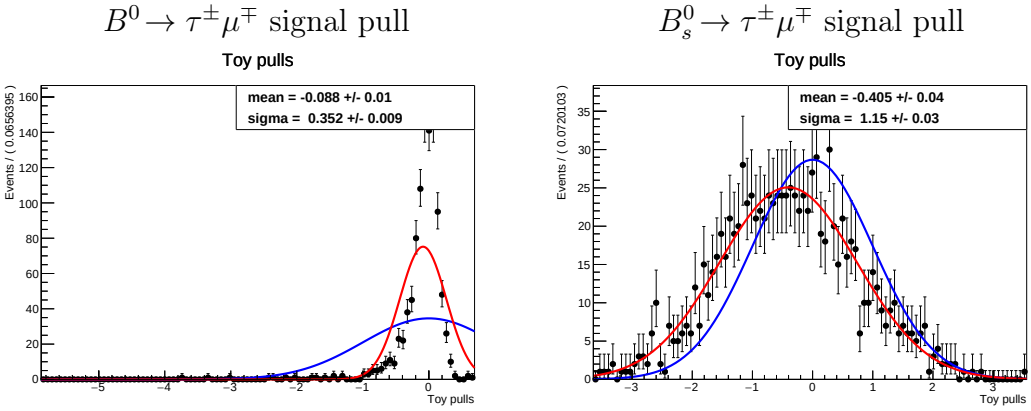

Figure C.3: Signal yield pull obtained using a 2D simultaneous fit model to pseudo-experiments containing background and signals. $B^0 \to \tau^{\pm}\mu^{\mp}$ (left) $B_s^0 \to \tau^{\pm}\mu^{\mp}$ (right).

Appendix D

Mass resolution evolution

The M_B reconstruction method, described in Section 3.4, is mainly sensitive to the decay vertices position resolution. This fact can be used to extrapolate the signal Hypathia width with respect to the resolution of the LHCb tracking system, in view of future upgrades.

In order to check how the Hypatia shape is affected by the vertex measurement precision, new sets of events are created from the the $B_s^0 \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulated samples where the B and τ decay vertex true positions are smeared simultaneously. Different samples are produced where the smearing is done with different widths ranging from 10% to 100% of the resolution estimated in the original sample. The M_B reconstruction method is applied to these modified MC samples and the fit procedure is repeated with all parameters fixed to their nominal values but the widths.

The range of the possible variation in the simulated vertex resolution is inferred using the normalization channel $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ MC and data samples. The angle between the D momentum and the direction given by the line formed by the B and D decay vertexes is compared in the data and MC samples. The relative difference between the width of the distributions of this angle in data and MC is about 6% (Figure D.1). This is used as an upper bound to the possible variation of the resolution on the B and τ decay vertex positions in the $B_s^0 \to \tau^{\pm}\mu^{\mp}$ simulated samples.

The relation between the vertex position resolution and the Hypatia width is shown in Figure D.2 and exhibits an almost linear correlation between the two quantities. It can be seen that an improvement of a factor 2 of the current resolution will lead to a separation of the B_s^0 and B^0 signals of 0.8σ, almost a factor 2 better than nowadays.

D.1 Limits on the branching ratio of the heavy and light eigenstates

Due to flavour mixing, the B_s^0 and B^0 systems have each two mass eigenstates, and per mass eigenstate a priori different lifetimes. The time dependent decay rate is the sum of the decreasing exponential contributions of both light and heavy eigenstates and the time integrated branching ratio can be written as:

$$\mathcal{B}\left(B_{(s)} \to \tau^{\pm}\mu^{\mp}\right) = \frac{1 + \mathcal{A}^{\Delta\Gamma}_{B^0_{(s)}\to\tau^{\pm}\mu^{\mp}} y_{(s)}}{1 - y^2_{(s)}} \mathcal{B}\left(B_{(s)} \to \tau^{\pm}\mu^{\mp}\right)_{t=0}, \tag{D.1}$$

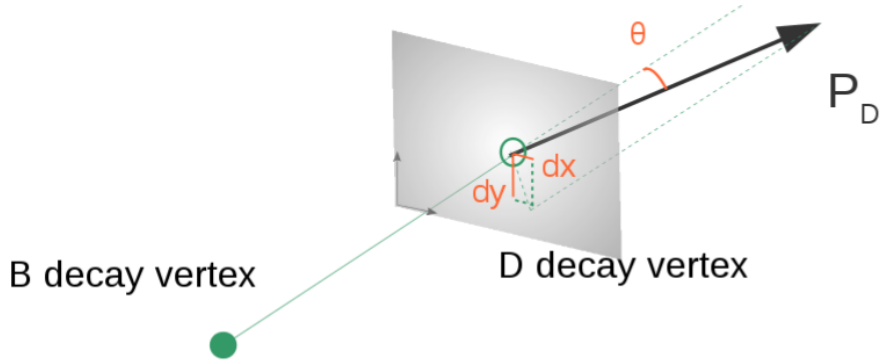

Figure D.1: Current vertex resolution extracted from the normalization channel $B^0 \to D^- \pi^+$.

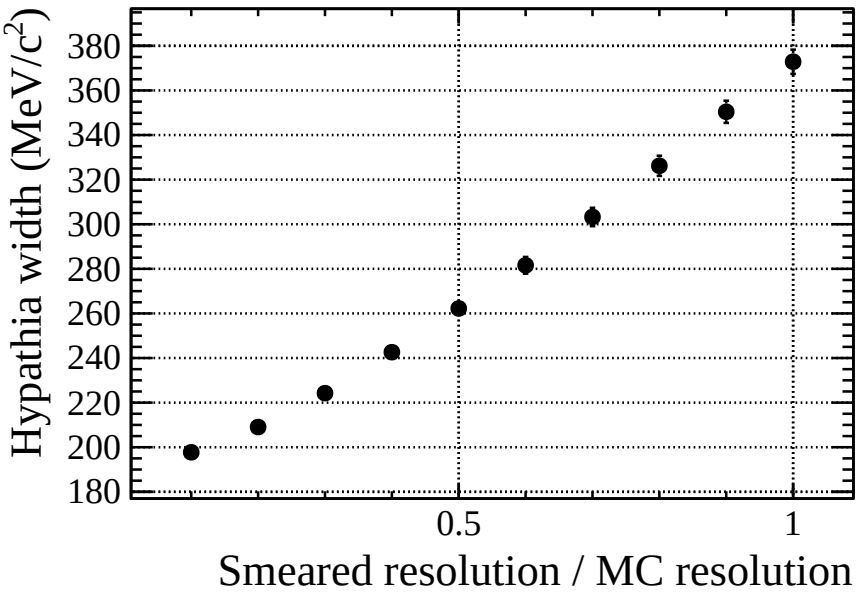

Figure D.2: Signal Hypatia with with respect to the vertexes resolution

where $\mathcal{A}^{\Delta\Gamma}_{B^0_{(s)}\to\tau^\pm\mu^\mp}$ and $y_{(s)}$ are defined as:

$$y_{(s)} = \frac{\Gamma^{(s)}_L - \Gamma^{(s)}_H}{\Gamma^{(s)}_L + \Gamma^{(s)}_H}, \tag{D.2}$$

$$\mathcal{A}^{\Delta\Gamma}_{B^0_{(s)}\to\tau^\pm\mu^\mp} = \frac{\Gamma^{(s)}_{H\to\tau^\pm\mu^\mp} - \Gamma^{(s)}_{L\to\tau^\pm\mu^\mp}}{\Gamma^{(s)}_{H\to\tau^\pm\mu^\mp} + \Gamma^{(s)}_{L\to\tau^\pm\mu^\mp}} \tag{D.3}$$

with $\Gamma^{(s)}_L$ and $\Gamma^{(s)}_H$ the total widths of the $B_{(s)}$ light and heavy states, and $\Gamma^{(s)}_{L\to\tau^\pm\mu^\mp}$ and $\Gamma^{(s)}_{H\to\tau^\pm\mu^\mp}$ their partial decay widths to $\tau^\pm\mu^\mp$.

$\mathcal{A}^{\Delta\Gamma}$	$\tau_{\text{effective}}$ (ps)
1.0	1.61407
-1.0	1.42561

Table D.1: B_s^0 effective lifetimes for different values of $\mathcal{A}^{\Delta}$.

As the possible new physics behind the LFV decay $B^0_{(s)} \to \tau^{\pm}\mu^{\mp}$ is not known, the branching fraction would be affected via $\mathcal{A}^{\Delta\Gamma}_{B^0_{(s)}\to\tau^{\pm}\mu^{\mp}}$ in a non trivial way if new physics entered differently the $B_{(s)}$ light and heavy decay amplitudes. For B^0, the width difference between the two mass eigenstates is negligible, giving $y \simeq 0$ and the branching fraction is not sensitive to the mixing and decay interplay. Therefore, only the B_s^0 case is studied as it has non-zero measured lifetime difference ($y_s = 0.062 \pm 0.006$ [13]). The two extreme cases $\mathcal{A}^{\Delta\Gamma} = -1$ and $\mathcal{A}^{\Delta\Gamma} = 1$ are considered, corresponding to the cases where only the light or heavy eigenstate contributes.

The decay rates are characterised by an effective lifetime, computed as follows:

$$\tau_{\text{effective}} = \frac{2\tau_{B_s^0} y_s \mathcal{A}^{\Delta\Gamma}_{B_s^0\to\tau^{\pm}\mu^{\mp}} + (1+y_s^2)\,\tau_{B_s^0}}{(1+y_s^2) + \mathcal{A}^{\Delta\Gamma}_{B_s^0\to\tau^{\pm}\mu^{\mp}} y_{B_s^0}\,(1-y_s^2)} \tag{D.4}$$

where $\tau_{B_s^0} = 1.511 \pm 0.014$ ps is the B_s^0 lifetime. The values of the different effective lifetimes for the two relevant cases are reported in Table D.1.

To extract the limit on the branching ratios of the $B_{s,L} \to \tau^{\pm}\mu^{\mp}$ and $B_{s,H} \to \tau^{\pm}\mu^{\mp}$ decays, the CLs method is used with modified efficiencies per BDT bins. The efficiency correction factors for the selection (including stripping) and the last BDT output distribution are computed by re-weighting the $B_s^0 \to \tau^{\pm}\mu^{\mp}$ MC signal in each of the 4 BDT bins. Each MC event is reweighted using:

$$\omega = e^{-t\left(\frac{1}{\tau_{\text{effective}}} - \frac{1}{\tau_{B_s^0}}\right)}, \tag{D.5}$$

where t is the true B lifetime. The results are shown in Table D.2.

As the B decay time is not used explicitly in the selection nor in the final BDT, the correction factors are very close to one and the effects on the expected (observed) limits are very small:

- $\mathcal{B}\left(B_{s,L} \to \tau^{\pm}\mu^{\mp}\right) < 2.99(2.51) \cdot 10^{-5}$ at 95%CL,
- $\mathcal{B}\left(B_{s,H} \to \tau^{\pm}\mu^{\mp}\right) < 2.92(2.45) \cdot 10^{-5}$ at 95%CL.

Year	$\mathcal{A}^{\Delta\Gamma}$	BDT bin	ϵ correction factor
2011	+1	1	1.0009
		2	1.0034
		3	1.0073
		4	1.0093
	-1	1	0.9959
		2	0.9938
		3	0.9903
		4	0.9882
2012	+1	1	1.0026
		2	1.0031
		3	1.0071
		4	1.0114
	-1	1	0.9945
		2	0.9942
		3	0.9905
		4	0.9863

Table D.2: Efficiency correction factors of the last BDT bins for different values of $\mathcal{A}^{\Delta\Gamma}$.

Appendix E

Run 1 and Run 2 variables distribution comparison

The shape of the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection are compared between Run 1 and Run 2.
Figures E.1 and E.2 contains the comparison of the variables shapes between the 2012 and 2016 $B^0_s \to \tau^{\pm}\mu^{\mp}$ signal samples. Both samples do not contain the PID requirements applied in the stripping line.
Figures E.3 and E.4 contains the comparison of the variables shapes between Run1 (2011 and 2011) and Run2 (2015 and 2016) of the same sign data.
Each plot contains a $\chi^2/ndof$ value for the corresponding variable (a). It is computed the following way for non-empty bins:

$$\chi^2/ndof = \frac{1}{N_{bins}} \sum_{i}^{bins} \frac{\left(a_i^{Run1} - a_i^{Run2}\right)^2}{\left(\delta a_i^{Run1} + \delta a_i^{Run2}\right)^2}. \tag{E.1}$$

On same sign data, the uncertainty (δa) on the variable bins is low due to the high statistics. Therefore the $\chi^2/ndof$ values tend to be high even for small shape differences. In general, all variables show a good agreement between Run 1 and Run 2. The variables showing the greatest discrepancy are the ones related to the secondary vertex properties. Furthermore, it can be seen that in the same sign data, the B invariant mass variable shape is different. The shape difference is reduced drastically during the offline selection process, as it can be seen in Section 3.6, where the B mass shapes are presented after the complete selection.

The analysis with the Run 2 data can be performed following the same strategy than Run 1, given that no obvious showstoppers are seen. Special care should be payed to the efficiency corrections and systematic uncertainties which will consume the most of the working time.

The Run 1 and Run 2 data can not be merged together as the data taking conditions are very different. However, in order to take advantage of the higher statistics, a merge between the Run 1 and Run 2 data at the likelihood level during the fit strategy could be explored.

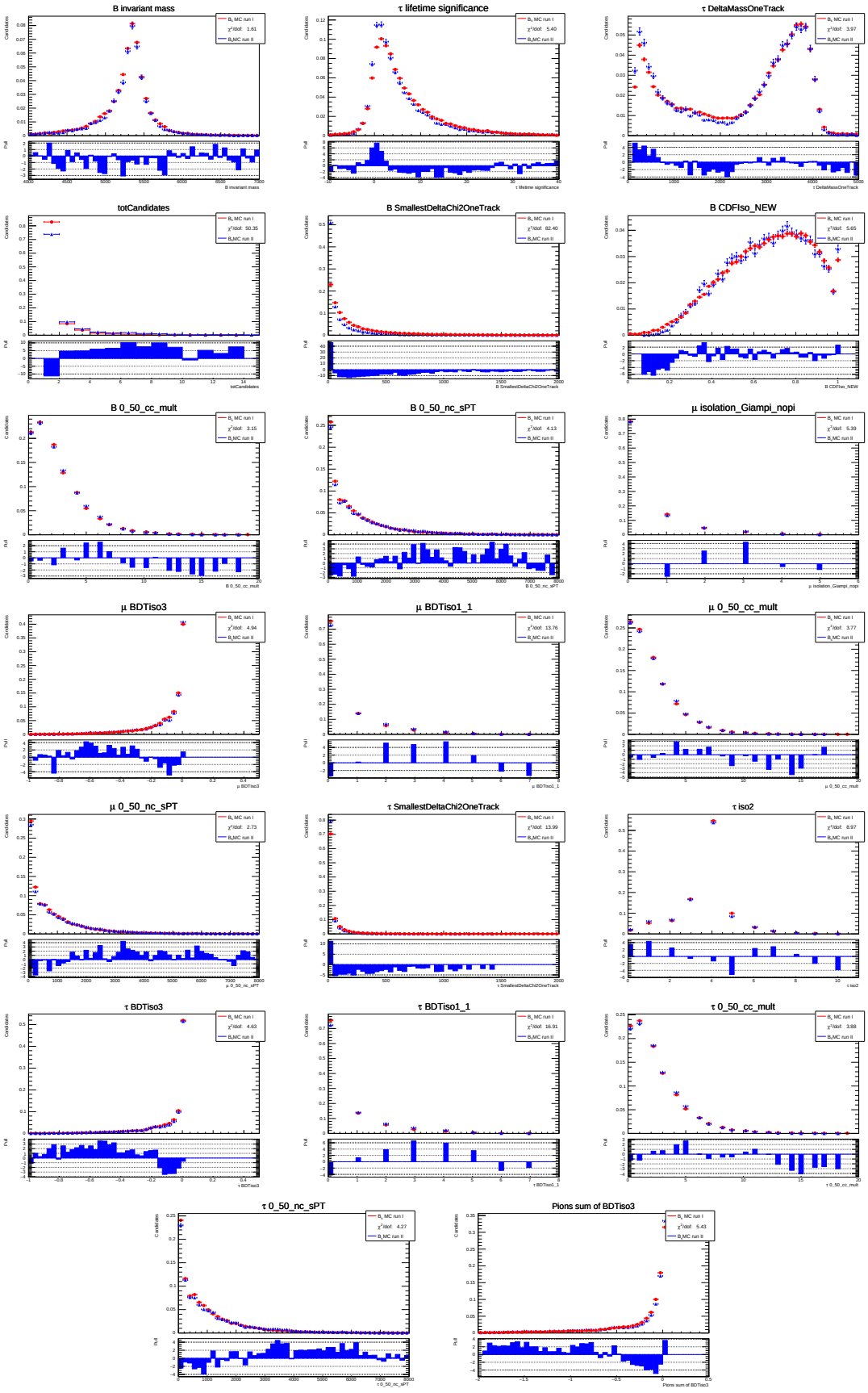

Figure E.1: (Part 1) Run1 and Run2 distribution comparisons for the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection on the 2012 and 2016 $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulation samples.

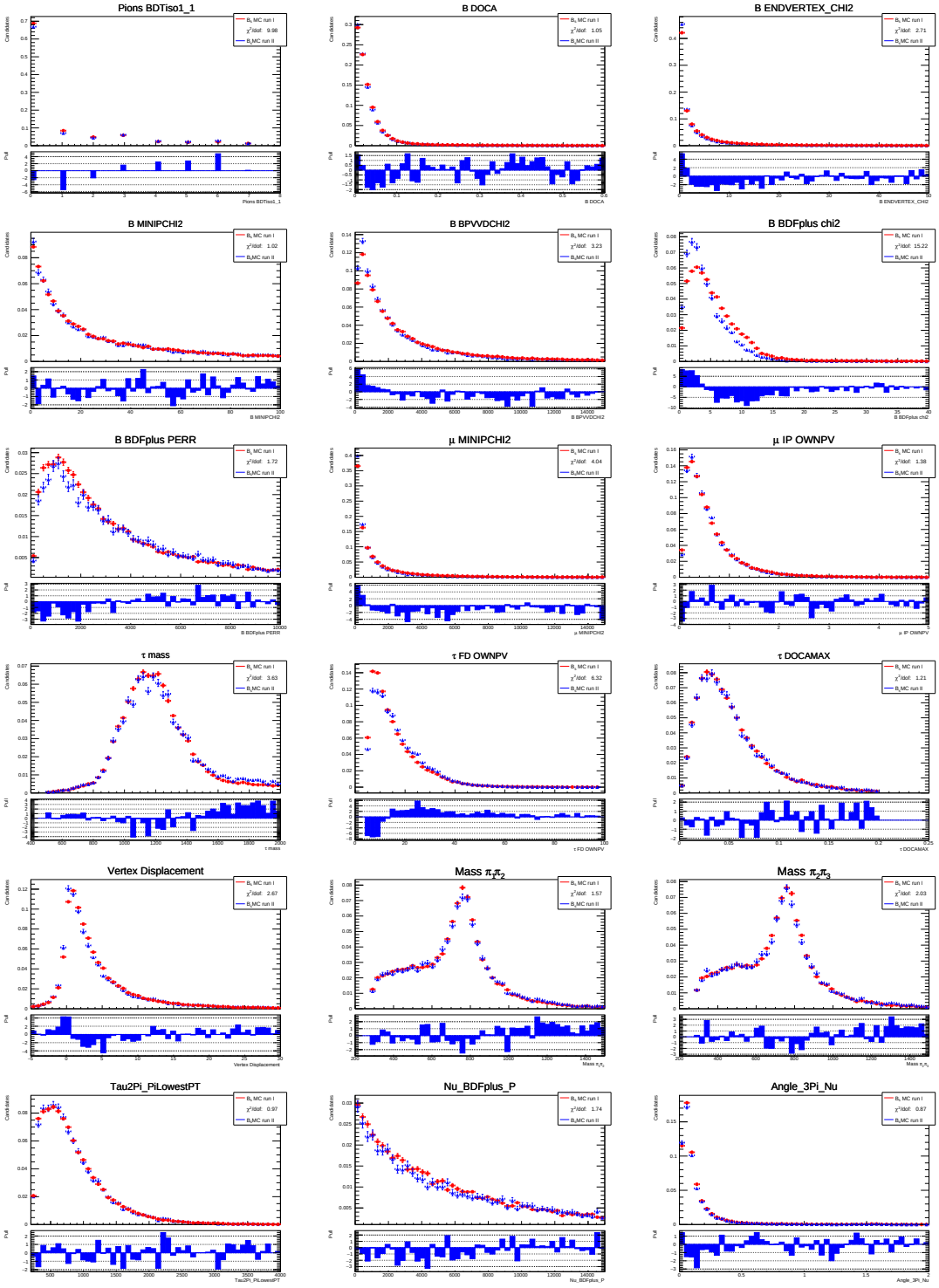

Figure E.2: (Part 1) Run1 and Run2 distribution comparisons for the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection on the 2012 and 2016 $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ simulation samples.

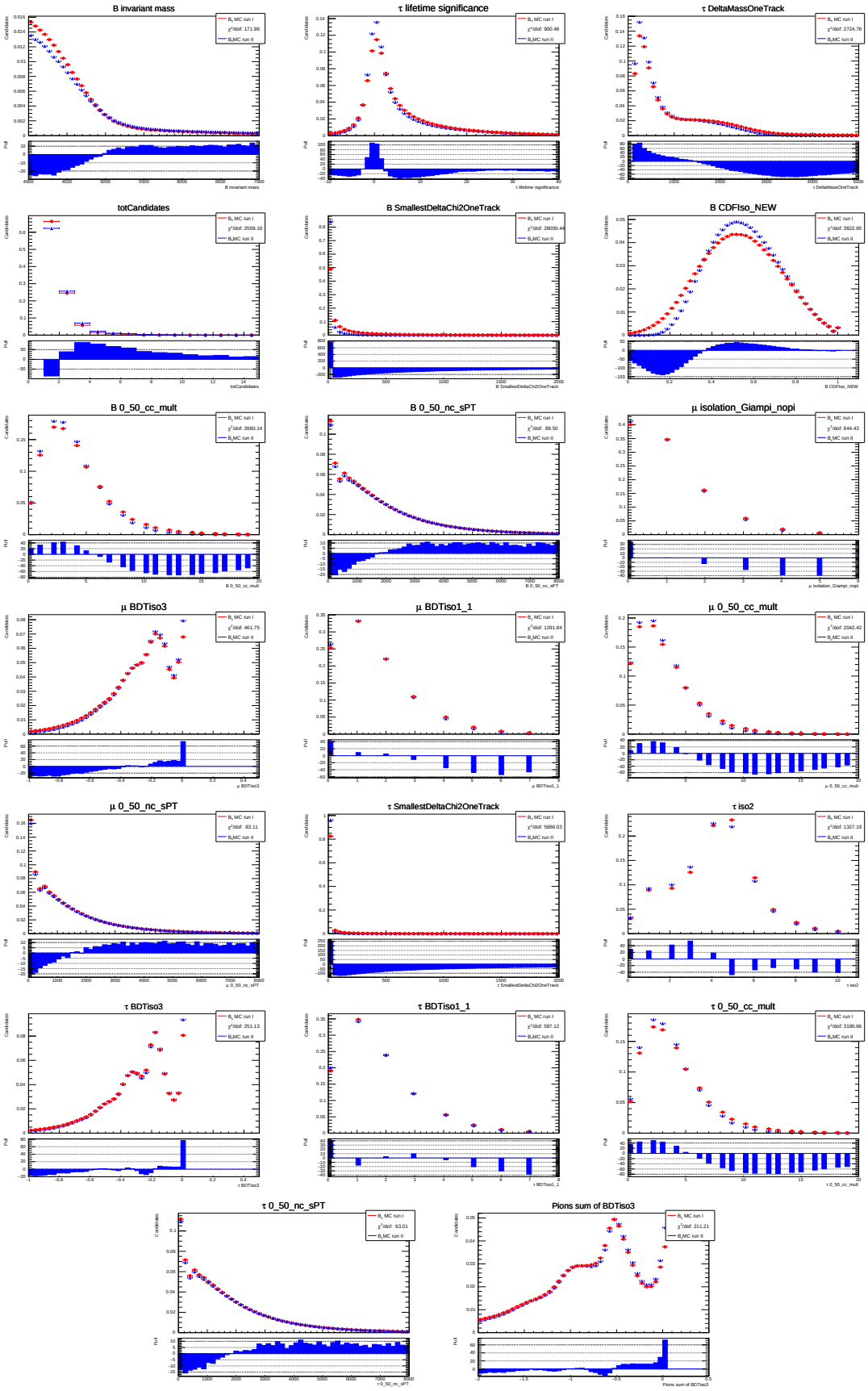

Figure E.3: (Part 1) Run1 and Run2 distribution comparisons for the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection on the Run1 and Run2 same sign data samples.

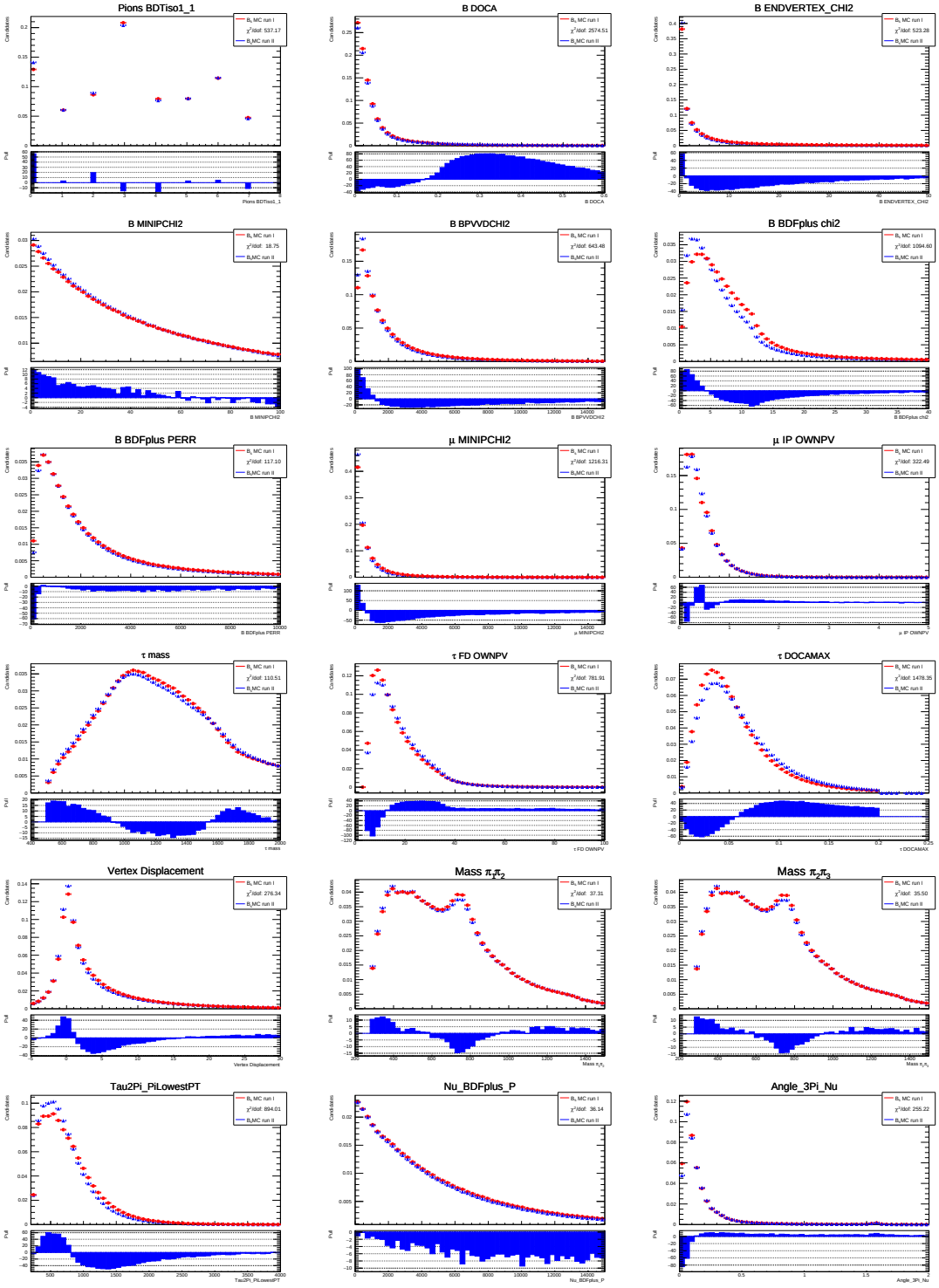

Figure E.4: (Part 1) Run1 and Run2 distribution comparisons for the variables used in the $B^0_{(s)} \to \tau^{\pm}(\to \pi^{\pm}\pi^{\mp}\pi^{\pm}\nu)\mu^{\mp}$ offline selection on the on the Run1 and Run2 same sign data samples.

Bibliography

[1] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv:1703.05747.

[2] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[3] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **113** (2014) 151601, arXiv:1406.6482.

[4] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016*, arXiv:1612.07233, updated results and plots available at http://www.slac.stanford.edu/xorg/hflav/.

[5] D. Boubaa, A. Datta, M. Duraisamy, S. Khalil, *Predictions for $B \to \tau\overline{\mu} + \mu\overline{\tau}$*, arXiv:1211.5168.

[6] D. Becirevic, N. Kosnik, O. Sumensari, R. Zukanovich Funchal, *Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive $b \to sl_1l_2$ modes*, arXiv:1608.07583.

[7] A. D. Smirnov, *Vector leptoquark mass limits and branching ratios of $K^0_l, B^0, B_s \to l^+_i l^-_j$ decays with account of fermion mixing in leptoquark currents*, arXiv:1801.02895.

[8] A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski, J. Rosiek, *Lepton-flavour violating B decays in generic Z' models*, arXiv:1504.07928.

[9] A. Crivellin, *B Decays and Lepton Flavour (Universality) Violation*, arXiv:1505.01527.

[10] D. Becirevic, O. Sumensari, R. Zukanovich Funchal, *Lepton Flavour Violation in Exclusive $b \to s$ Decays*, arXiv:1602.00881.

[11] Babar collaboration, *Searches for the decays $B^0 \to l^+\tau^-$ and $B^+ \to l^+\nu$ (l=e,μ) using hadronic tag reconstruction*, arXiv:0801.0697.

[12] The LHCb collaboration, A. A. et al, *The LHCb detector at the LHC*, Journal of Instrumentation **3** (2008), no. 08 S08005.

[13] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001.

[14] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[15] F. Dulat, B. Mistberger, *Limit setting procedures and theoretical uncertainties in Higgs boson searches*, arXiv:1204.3851.

[16] D. V. Peskin, Michael E and Schroeder, *An introduction to quantum field theory*, Westview, Boulder, CO, 1995. Includes exercises.

[17] S Glashow, J Iliopoulos, L Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) .

[18] Falk, Adam F. , *The CKM matrix and the heavy quark expansion*, in *Flavor physics for the millennium. Proceedings, Theoretical Advanced Study Institute in elementary particle physics, TASI 2000, Boulder, USA, June 4-30, 2000*, pp. 379–427, 2000. arXiv:hep-ph/0007339.

[19] T. Blake, G. Lanfranchi, D. M. Straub, *Rare B decays as tasts of the Standard Model*, arXiv:1606.00916.

[20] A. Ali, B. Pecjak et al, *Towards $B \to V\gamma$ decays at NNLO in SCET*, Eur. Phys. **55** (2008) 577.

[21] C. Bobeth *et al.*, *$B_{s,d} \to l^+l^-$ in the standard model with reduced theoretical uncertainty*, Phys. Rev. Lett. **112** (2014) 101801.

[22] LHCb, CMS, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68, arXiv:1411.4413.

[23] LHCb Collaboration, A. et al. *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801.

[24] CDF Collaboration, A. et al. *Search for the decays $B_{(s)}^0 \to e^+\mu^-$ and $B_{(s)}^0 \to e^+e^-$ in cdf run ii*, Phys. Rev. Lett. **102** (2009) 201801.

[25] D. Bečirević, S. Fajfer, N. Košnik, and O. Sumensari, *Leptoquark model to explain the b-physics anomalies, R_K and R_D*, Phys. Rev. D **94** (2016) 115021.

[26] LHCb Collaboration, *Test of lepton universality with $B^0 \to K^{0*}l^+l^-$ decays*, arXiv:1705.05802.

[27] L.-S. Geng *et al.*, *Towards the discovery of new physics with lepton-universality ratios of $b \to s\ell\ell$ decays*, Phys. Rev. D **96** (2017) 093006.

[28] W. Altmannshofer, P. S. Bhupal Dev, and A. Soni, *$R_{D^{(*)}}$ anomaly: A possible hint for natural supersymmetry with r-parity violation*, Phys. Rev. D **96** (2017) 095010.

[29] CDF Collaboration, A. et al. *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016*, Eur. Phys. J. C (2017) **77** (2017) 805.

[30] D Greljo, A Isidori, et al. , *B physics anomalies: a guide to combined explanations*, J. High Energ. Phys. **44** .

[31] M. Bordone, C. Cornella, J. Fuentes-Martin, G. Isidori, *Low-energy signatures of the PS^3 model: from B-physics anomalies to LFV*, arXiv:1805.09328.

[32] MEG II Collaboration, *The design of the MEG II experiment*, arXiv:1801.04688.

[33] L. Bartoszek et al. , *Mu2e Technical Design Report*, arXiv:1501.05241.

[34] COMET, B. E. Krikler, *An Overview of the COMET Experiment and its Recent Progress*, arXiv:1512.08564.

[35] A. Alekou *et al.*, *Accelerator system for the PRISM based muon to electron conversion experiment*, in *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013. arXiv:1310.0804.

[36] Belle, D. Liventsev, *Recent results from the Belle experiment*, arXiv:1608.04500.

[37] Belle II, B. Wang, *The Belle II Experiment and SuperKEKB Upgrade*, arXiv:1511.09434.

[38] NA62 collaboration, *The Beam and detector of the NA62 experiment at CERN*, arXiv:1703.08501.

[39] MEG Collaboration, *Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment*, Eur. Phys. **76** (2016) 434.

[40] U. Bellgardt *et al.*, *Search for the decay $\mu^+ \to e^+e^+e^-$*, Nuclear Physics B **299** (1988), no. 1 1 .

[41] A search for $\mu - e$conversion in muonic gold, *W. Bertl, R. Engfer, E.A. Hermes, G. KurzT. Kozlowski, J. KuthG. Otter, F. Rosenbaum, N.M. Ryskulov, A. van der Schaaf, P. Wintz, I. Zychor, The SINDRUM II Collaboration*, Eur. Phys. **47** (2016) 337.

[42] SINDRUM II Collaboration, *Search for μ^- to e^- conversion*, Journal of Physics G: Nuclear and Particle Physics **17** (1991), no. S S47.

[43] BABAR Collaboration, e. a. Aubert, B.K̃aryotakis, *Searches for lepton flavor violation in the decays $\tau^\pm \to e^\pm\gamma$ and $\tau^\pm \to \mu^\pm\gamma$*, Phys. Rev. Lett. **104** (2010) 021802.

[44] K. Hayasaka, K. Inami et. al. , *Search for lepton-flavor-violating τ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs*, Physics Letters B **687** (2010), no. 2 139 .

[45] C. Patrignani and P. D. Group, *Review of particle physics*, Chinese Physics C **40** (2016), no. 10 100001.

[46] KTeV Collaboration, e. a. E.Ãbouzaid, *Search for lepton-flavor-violating decays of the neutral kaon*, Phys. Rev. Lett. **100** (2008) 131803.

[47] BNL E871 Collaboration, D. A. et al. *New limit on muon and electron lepton number violation from $k_l^0 \to \mu^\pm e^\mp$ decay*, Phys. Rev. Lett. **81** (1998) 5734.

[48] NA62 Collaboration, *Forbidden Kaon and Pion Decays in NA62*, arXiv:1306.3361.

[49] ATLAS Collaboration, G. Aad, and B. Abbott et. al. , *Search for the lepton flavor violating decay $z \to e\mu$ in pp collisions at $\sqrt{s} = 8$ TeV with the atlas detector*, Phys. Rev. D **90** (2014) 072010.

[50] CMS collaboration, *Search for lepton-flavour-violating decays of the higgs boson*, Physics Letters B **749** (2015) 337 .

[51] ATLAS collaboration, A. et al. *Search for lepton-flavour-violating $h \to \tau\mu$ decays of the higgs boson with the atlas detector*, Phys. Rev. Lett. **211** (2015) .

[52] LHCb collaboration, *Search for the lepton-flavour violating decay $d^0 \to e^\pm\mu^\mp$*, Physics Letters B **754** (2016) 167 .

[53] LHCb collaboration, B. A. e. a. R.Ãaij, *Search for the lepton-flavour violating decays $b^0_{(s)} \to e^\pm\mu^\mp$*, Journal of High Energy Physics (2018) .

[54] LHCb Collaboration, e. a. R.Ãaij, *Search for lepton number violating decays $B^+ \to \pi^-\mu^+\mu^+$ and $B^+ \to K^-\mu^+\mu^+$*, Phys. Rev. Lett. **108** (2012) 101601.

[55] BaBar collaboration, *The BABAR Detector: Upgrades, Operation and Performance*, arXiv:1305.3560.

[56] D0 collaboration, *The Upgraded D0 Detector*, arXiv:0507191.

[57] CDF collaboration, *The CDF Experiment at the Tevatron - the First Two Years of Run II*, arXiv:0306046.

[58] LHCb Collaboration, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions. Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Tech. Rep. CERN-EP-2016-201. LHCB-PAPER-2016-031. CERN-EP-2016-201. 5, Dec, 2016.

[59] E. Norrbin and T. Sjostrand, *Production and hadronization of heavy quarks*, Eur. Phys. J. C (2000) **17** (2007) 137.

[60] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[61] R. L. Eubank, *A Kalman Filter Primer*, Chapman and Hall/CRC, 2006.

[62] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[63] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to pythia 8.1*, Computer Physics Communications **178** (2008), no. 11 852 .

[64] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[65] M. G. Pia, T. Basaglia, Z. W. Bell, P. V. Dressendorfer, *Geant4 in Scientific Literature*, arXiv:0912.0360.

[66] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[67] I. M. Nugent, T. Przedzinski, P. Roig, O. Shekhovtsova, Z. Was, *Resonance Chiral Lagrangian Currents and Experimental Data for* $\tau \to \pi^-\pi^-\pi^+\nu_\tau$, arXiv:1310.1053.

[68] F. James and M. Winkler, *MINUIT User's Guide*, .

[69] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[70] LHCb collaboration, *Study of Muon Isolation in the* $B_s^0 \to \mu^+\mu^-$ *channel*, LHCb-INT-2010-011.

[71] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[72] LHCb tracking group, *Lhcb tracking efficiencies*, URL: https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbDetectorAlignment.

[73] A. Powell *et al.*, *Particle identification at LHCb*, PoS **ICHEP2010** (2010) 020, LHCb-PROC-2011-008.

[74] LHCb collaboration, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039.

[75] Z. Was, J. Zaremba, *Study of variants for Monte Carlo generators of* $\tau \to 3\pi\nu$ *models*, arXiv:1508.06424.

[76] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, Erratum ibid. **119** (2017) 169901, arXiv:1612.05140.

[77] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001, and 2017 update.

[78] LHCb collaboration, *Updated average* f_s/f_d *b-hadron production fraction ratio for* 7 TeV *pp collisions*, LHCb-CONF-2013-011.

[79] D. Martinez Santos, F. Dupertuis, *Mass distributions marginalized over per-event errors*, arXiv:1312.5000.

[80] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) .