



CPPM-T-2012-03

THESE DE DOCTORAT

Towards the first $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ measurements with the LHCb detector

Spécialité : Physique et Sciences de la Matière

Mention : Physique des Particules et Astroparticules

présentée par

Cosme Adrover Pacheco

en vue d'obtenir le grade de docteur de la Aix-Marseille Université

soutenue le 10 setembre 2012 devant le jury composé de

Prof..	Mossadek Talby	Président du jury
Dr.	R. Le Gac	Co-Directeur de thèse
Dr.	G. Mancinelli	Directeur de thèse
Dr.	P. Robbe	Examinateur
Dr.	G. Lanfranchi	Rapporteur
Dr.	U. Langenegger	Rapporteur

Acknowledgements

There are some people I would like to mention first as they counseled me before entering the Ph.D. program: my particle physics teacher Rafel, and my Master's supervisors Lluis and Abelardo. I would also like to thank Steve, now my father in law, for his careful review of my curriculum at that time.

Proceeding chronologically, I need to thank those that trusted in me, and granted me the opportunity to start a research career in the exciting first years of the LHC: my supervisors Renaud and Giampi, and Eric, director of the laboratory.

From the very beginning of my arrival to Marseille I found myself surrounded by nice people always keen to raise me a hand, specially with computing issues, like Stephane, Bassem, Emilie, and my former officemates Garo and Nicolas. Many thanks to all of them.

The structure of the office changed through my PhD and in the end I spent very nice moments with Emilie and Mathieu. It has been a pleasure to have such wonderful colleagues. Merci pour tout les enfants!

The warm conviviality has been constant at CPPM. I would like to thank all the members of this wonderful institute: PhD students, postdocs, researchers, engineers, and administrative staff.

Kind thanks to all LHCb members in Marseille for their continuous support and review of my work. Special thanks to Justine who always gave me good directions and wise advice.

At a broader level, I would like to thank the whole LHCb collaboration for providing me with such a challenging environment, and especially those with whom I worked closely on the golden measurement of the experiment. Many thanks for your support and hard work.

I would like to thank some people that helped me with specific topics of the thesis like Yann who introduced me the BDT and Nazila who helped me with the theoretical interpretation.

It has been an honor for me to have such brilliant and different supervisors. Many thanks to Renaud for his precise and clear advice that guided me in the end of my work. This thesis does not only belong to me, but also to Giampi who always proposed new topics of study and pushed me to never stop working. He never hesitated to sit down in front of my results and review them with a close eye. This document has been carefully reviewed (more than once) by him. It is hard to condense into a few sentences what Giampi has represented for me. Many thanks for everything.

I would like to thank as well the jury of my thesis for their meticulous review of my thesis, and for their instructive questions that broadened my perspective of the analysis.

I have special words of acknowledgement are for my parents and my brother, to whom I am very grateful for all their support, and my friends and family, always interested in what I was doing to make a living, although I am afraid they still do not know what I do and think I am still at school.

Finally, my last words of thanks are for my wife, Ali, who always makes dark moments shine.

Contents

Résumé	1
Introduction	13
1 $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ as a particle physics benchmark	15
1.1 Towards a theory of fundamental interactions	15
1.1.1 The electroweak interaction	16
1.2 Theoretical predictions for $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$	20
1.2.1 Effective field theory for B decays	20
1.2.2 $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ in the framework of EFT	21
1.3 Conclusions	25
2 The LHCb experiment	29
2.1 Introduction	29
2.1.1 LHCb general features	29
2.1.2 Data taking periods	32
2.2 LHCb Tracking System	35
2.2.1 The vertex locator (VELO)	35
2.2.2 The trackers	37
2.2.3 Magnet	37
2.3 Particle identification	37
2.3.1 The RICH detectors	38
2.3.2 Calorimeters	39
2.3.3 Muon system	40
2.3.4 Muon identification	40
2.4 Trigger system	41
2.4.1 Level-0 trigger	42
2.4.2 High Level Trigger	43
2.4.3 Types of event according to the trigger	45
2.5 Conclusions	45
3 Analysis challenges to measure $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$	47
3.1 Analysis overview	47
3.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ background rejection	49
3.2.1 Prompt background	49

3.2.2	B hadron cascade decays	51
3.2.3	Peaking background	51
3.2.4	Combinatoric from semileptonic B decays	51
3.2.5	Other exclusive background	52
3.3	Conclusions	54
4	Validation studies using the Geometrical Likelihood classifier	55
4.1	Mathematical method to combine correlated variables	55
4.2	GL calibration using the low mass resonance decay $D^0 \rightarrow K\pi$	56
4.3	Estimation of the GEC impact on the GL distribution	61
4.4	Geometrical properties of B cascade decays	61
4.5	Conclusions	63
5	Optimization of the discriminant classifier for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$	65
5.1	Extraction of the signal distributions for some variables using data events	65
5.2	Introduction to boosted decision trees	70
5.3	Implementation of a selection BDT classifier (BDTS)	72
5.3.1	Definition of the BDT for the selection	73
5.3.2	Effect of the BDTS selection on combinatorial background	75
5.3.3	Effect of the BDTS selection on normalization channels	75
5.4	BDT classifier as a background discriminant	78
5.5	Further optimization of the discriminant BDT	80
5.6	Optimization of the BDTS-based selection	84
5.7	Further signal data and simulation comparisons	89
5.8	Conclusions	89
6	Extraction of the signal yields using an extended maximum likelihood fit	93
6.1	Overview of the fit procedure	93
6.1.1	Description of the <i>pdfs</i> needed for the analysis	94
6.2	Validation of the fit method using simulated data	100
6.3	Yields from 1 fb^{-1} of data	107
6.4	Conclusions	110
7	The measurement of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$	115
7.1	Normalization of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$	115
7.2	$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ obtained for 1 fb^{-1}	118
7.2.1	Estimate of the systematic uncertainties	119
7.3	Comparison of the measured $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ with the exclusion limits	119
7.3.1	Exclusion limits	122
7.3.2	Comparison of the measured $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ with the exclusion limit .	123
7.4	Implications of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ on NP models	125
7.4.1	Implications on CMSSM	125
7.4.2	Implications on general models	126
7.5	Conclusions	128

Conclusion	129
Bibliography	131

Résumé

Situé en Suisse, le LHC a commencé à faire collisionner des protons à une énergie de 7 TeV au centre de masse en Mars 2010. Il s'agit du plus grand collisionneur de particules du monde avec une circonférence d'environ 27 km. Le but de cette machine est d'aider la communauté scientifique à résoudre les questions liées aux interactions fondamentales entre les particules élémentaires.

La théorie décrivant avec une grande précision ces interactions entre particules est connue comme le Modèle Standard (MS). Malgré que le MS explique presque tous les phénomènes observés en-dessous de 200 GeV, il est considéré par les physiciens comme une théorie valide jusqu'à une énergie d'environ quelques TeV. Et c'est précisément cette gamme d'énergies que le LHC explore.

Un des plus grands problèmes conceptuels qui doit affronter le MS est le problème de la hiérarchie. Afin de réussir à expliquer la masse des particules fondamentales, la brisure spontanée de la symétrie nécessite une nouvelle particule: le boson de Higgs. En tant que particule scalaire, le boson de higgs reçoit des corrections de masse proportionnelles à l'échelle d'énergie valide pour le MS.

Deux procédures sont possibles pour découvrir l'origine de la nouvelle physique au-delà du Modèle Standard: par de mesures directes et indirectes. Dans la première situation, on essaie d'observer directement des nouvelles particules à des niveaux d'énergie jamais envisagés, cependant dans la deuxième on mesure l'effet de nouvelles particules sur des observables expérimentales, telles les rapports de branchement de désintégrations de hadrons beaux. Au LHC deux expériences ont été conçues pour des observations directes: ATLAS et CMS; et une autre expérience designée pour des mesures indirectes: LHCb. Le principal avantage des mesures indirectes par rapport aux directes est la possibilité d'accéder à des masses plus élevées. Un exemple de mesure indirecte est la mesure des rapports de branchement mettant en jeu des changements de saveurs par courant neutre. Ces processus sont interdits au niveau arbre et surviennent, par exemple, au niveau des processus en boucle. Ils sont aussi très intéressants car, dans des processus en boucle hypothétique qu'il y ait de nouvelles particules, elles pourraient intervenir dans le boucle et changer les prédictions attendues dans le modèle standard. La figure 0.1 montre des diagrammes de Feynman pour des transitions $b \rightarrow sl\bar{l}$ ou $b \rightarrow sf\bar{f}$, avec l lepton et f fermion, dans le modèle standard et dans un hypothétique nouveau paradigme.

Les rapports de branchement de $B_s^0 \rightarrow \mu^+ \mu^-$ et $B^0 \rightarrow \mu^+ \mu^-$ ($\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$) sont

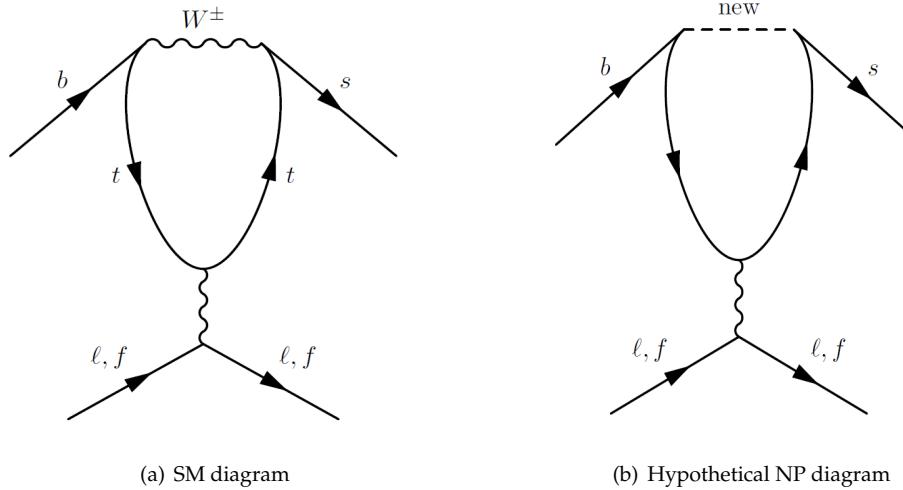


Figure 0.1: Diagrammes pingouins contribuant à $b \rightarrow s l \bar{l}$ avec des particules observées (a) et hypothétiques (b).

prédits avec une bonne précision dans le MS [1]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM} = (3.2 \pm 0.2) \times 10^{-9} \quad (1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{SM} = (1.0 \pm 0.1) \times 10^{-10} \quad (2)$$

D'autres caractéristiques font de l'étude de $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ une voie pour la recherche de nouvelle physique. La première est liée aux possibles déviations prévues dans des nouveaux modèles, telles que la supersymétrie, par rapport aux valeurs attendues dans le MS. La supersymétrie, ou symétrie qui relie les propriétés des fermions avec celles des bosons, est un bon candidat pour étendre le MS, car elle présente une solution au problème de hiérarchie, mais elle prévoit aussi une nouvelle particule stable, qui est un des candidats le plus important de la matière sombre. Elle prévoit aussi l'unification des couplages fort, faible et électromagnétique.

Un autre point d'intérêt est que ces processus n'ont jamais été observés, seulement des limites sur les rapports de branchements ont été établis [2]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 43 \times 10^{-9},$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 76 \times 10^{-10}.$$

Le LHCb est une expérience spécialement conçue pour l'étude des hadrons beaux au LHC. Le LHC (Large Hadron Collider) est un accélérateur installé dans le même tunnel de 26.7 km situé à la frontière franco-suisse qui héberge le LEP dans les années 90. Depuis Mars 2010, le LHC délivre des collisions à 7 TeV (8 TeV depuis 2012) d'énergie au centre de masse.

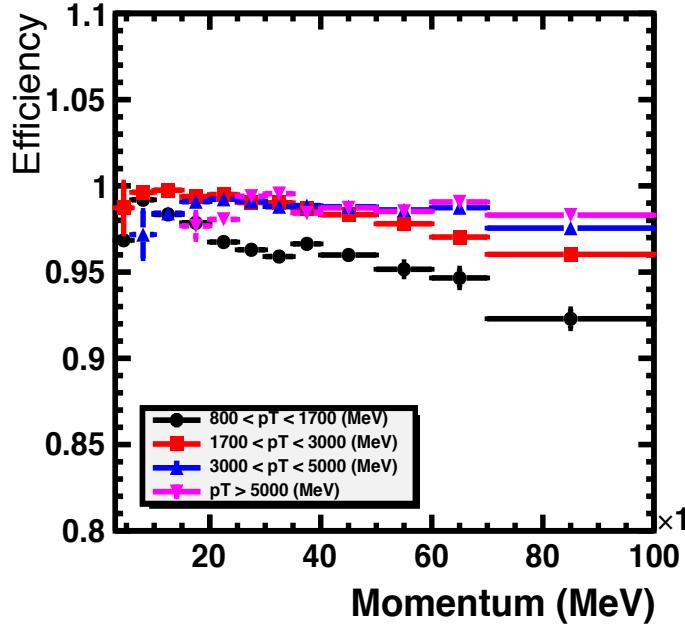


Figure 0.2: Efficacité en la identification des muons en fonction du moment transverse et en différent gammes de moment transverse.

Les deux principales caractéristiques rendant LHCb très intéressant pour la mesure des désintégrations très rares des mesons B avec deux muons dans l'état final sont: une bonne efficacité du système de déclenchement et une bonne efficacité d'identification des muons. La figure 0.2 montre cette efficacité d'identification en fonction de l'impulsion et en quatre intervalles d'impulsion transverse. L'efficacité, pour des impulsions transverses supérieures à $1700 \text{ GeV}/c^2$, est plus grande que 97%.

L'objectif est de mesurer le rapport de brachement de $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ à partir d'une normalisation en utilisant le rapport de brachement de désintégrations connues:

$$\frac{\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(\text{normalization channel})} = \text{Constant} \times \frac{N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}}{N_{\text{normalization channel}}}.$$

Avec cette normalisation on évite l'utilisation de la section efficace totale de production des paires $b\bar{b}$ et la luminosité absolue.

Les critères pour choisir les désintégrations utilisées pour cette normalisation sont basés sur l'erreur de son rapport de brachement, des similarités avec le signal: même efficacité de déclenchement, nombre similaire et type de particules particules. Devant l'impossibilité d'avoir des désintégrations répondant à toutes ces exigences, on choisit trois désintégrations ayant des caractéristiques différentes, afin d'annuler les incertitudes systématiques.

La première étape de l'analyse consiste à sélectionner des événements en essayant de supprimer au maximum le bruit de fond. Ce bruit de fond correspond principalement, à

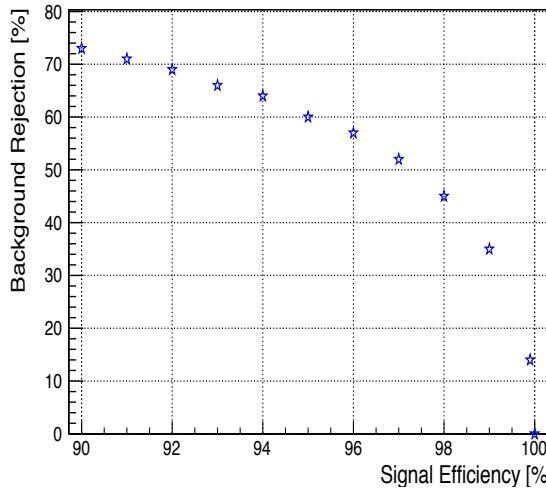


Figure 0.3: Rejet de bruit du fond par rapport à l'efficacité du signal d'une sélection basée sur BDTS calculé avec événements simulés $B_s^0 \rightarrow \mu^+ \mu^-$ en tant que signal et $B_{(s)}^0 \rightarrow h^+ h^-$ bandes latérales de données pour les événements du bruit de fond.

cette étape, à des combinaisons de muons venant du vertex primaire ou la collision a eu lieu.

L'idée est de définir une sélection similaire pour toutes les désintégrations utilisées: signal et normalisation. Une partie de mon travail de thèse a consisté à optimiser une sélection multivariée basée sur les arbres de décision boostés. Cette sélection est définie avec les variables suivants:

- Paramètre d'impact du méson B .
- Significance du paramètre d'impact du méson B .
- Angle entre l'impulsion du méson B et le vecteur défini par le vertex primaire et le vertex de désintégration du méson B .
- La distance minimale d'approche entre les muons.
- La qualité du vertex secondaire
- Le plus petit des paramètres d'impact des muons.

La figure 0.3 montre l'efficacité sur le signal $B_s^0 \rightarrow \mu^+ \mu^-$ contre l'efficacité de réjection du bruit de fond présent dans la sélection $B_{(s)}^0 \rightarrow h^+ h^-$. Les erreurs liées aux deux efficacités sont plus petites que 0.1%. On voit que pour une efficacité du signal de 95% on rejette 60% du bruit de fond. La sélection basée sur ce classificateur est optimisée par rapport aux performances obtenues sur un classificateur supplémentaire.

Pour mieux visualiser l'effet de la sélection sur le bruit de fond, la figure 0.4 présente la masse invariante des candidats $B_{(s)}^0 \rightarrow h^+ h^-$ avant et après deux sélections BDTS. BDTS>0.03 équivaut à une efficacité en signal de 97%, et BDTS>0.03 à 95%.

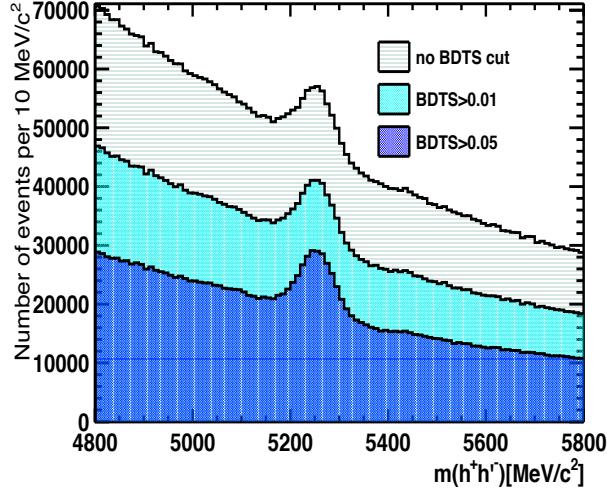


Figure 0.4: Distribution de la masse invariante des événements $B_{(s)}^0 \rightarrow h^+h'$ pour différents selections en BDTS.

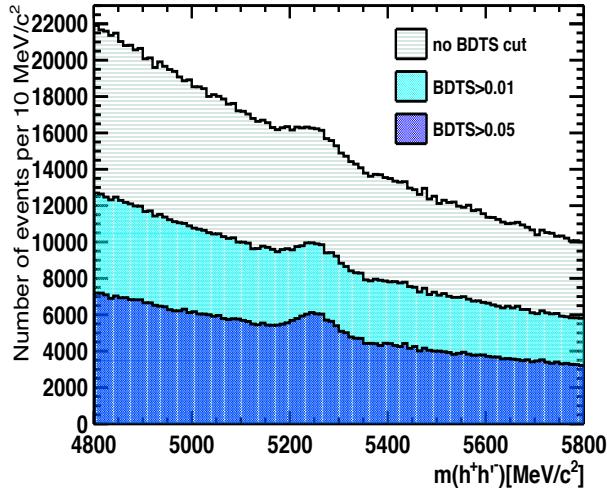


Figure 0.5: Distribution de la masse invariant des événements $B_{(s)}^0 \rightarrow h^+h'$ -TIS pour différents selections en BDTS.

Un autre effet important est la selection en événements surnomés TIS, ou événements qui passent le système de déclenchement alors que les produits de désintégration ne sont pas responsables de cette sélection. Ces événements sont très importants car ils sont utilisés pour estimer la distribution du classificateur final en utilisant des données réelles, c'est-à-dire non-simulées. La figure 0.5 montre la masse invariante des candidates $B_{(s)}^0 \rightarrow h^+h^-$ avant et après deux sélections en BDTS.

Afin d'obtenir $N_{B_s^0 \rightarrow \mu^+\mu^-}$ il faut encore un autre classificateur maximisant la séparation entre le signal et le bruit de fond présent après la sélection. Ce bruit de fond est composé

des combinaisons des muons provenant des différentes désintégrations semi-leptoniques des hadrons B .

En utilisant des événements simulés, on définit un BDT avec 9 variables:

- paramètre d'impact du meson B ;
- temps de vol ou temps propre du méson B ;
- l'isolation du méson B ;
- l'impulsion transverse du méson B ;
- la distance minimale d'approche entre les muons;
- l'impulsion transverse minimale des muons;
- le plus petit des paramètres d'impact en unités de signification des muons;
- l'angle entre l'impulsion du muon chargé positivement, dans le système de référence du candidat B , avec le vecteur perpendiculaire à l'impulsion du B et du faisceau;
- la somme des isolations des muons

La figure 0.6 montre les distributions attendues pour quatre des variables définies. Les courbes bleues correspondent aux distributions du signal $B_s^0 \rightarrow \mu^+ \mu^-$, en rouge le bruit de fond simulé, et les points noirs sont le bruit de fond issues des données. On voit la séparation entre signal et bruit de fond, et aussi un bon accord entre événements simulés et données réélees, pour ce dernier type.

Une partie importante du travail a été dédié à optimiser le BDT. Néanmoins, cette optimisation est limitée par un faible quantité de données simulées de bruit de fond. La figure 0.7 montre les performances, en terme de rejet du bruit de fond pour deux BDT: un avec les variables décrites avant et un autre avec plusieurs. Étant donné que les performances sont similaires, on choisit le classificateur avec neuf variables en raison de sa simplicité.

Un ajustement de la fonction de vraisemblance attendue permet d'obtenir le nombre d'événements du signal. La masse invariante et la sortie du BDT optimisée précédemment entrent dans la définition de cet ajustement.

Soit la fonction de vraisemblance:

$$\mathcal{L} = e^{-\mathcal{N}} \times \frac{\mathcal{N}^N}{N!} \prod p(x_i; \alpha_1, \dots, \alpha_m), \quad (3)$$

où $p(x_i; \alpha_1, \dots, \alpha_m)$ est la fonction de densité de probabilité pdf (par N événements x_1, \dots, x_N en fonction de m paramètres $\alpha_1, \dots, \alpha_m$). \mathcal{N} est le nombre d'événements attendus en supposant statistique Poissonienne. Les paramètres inconnus α_i sont obtenus en minimisant $-2 \log \mathcal{L}$ par rapport à chaque α_i .

$p(x_i; \alpha_1, \dots, \alpha_m)$ a quatre contributions, chacune définie en fonction de sa masse invariante et son BDT:

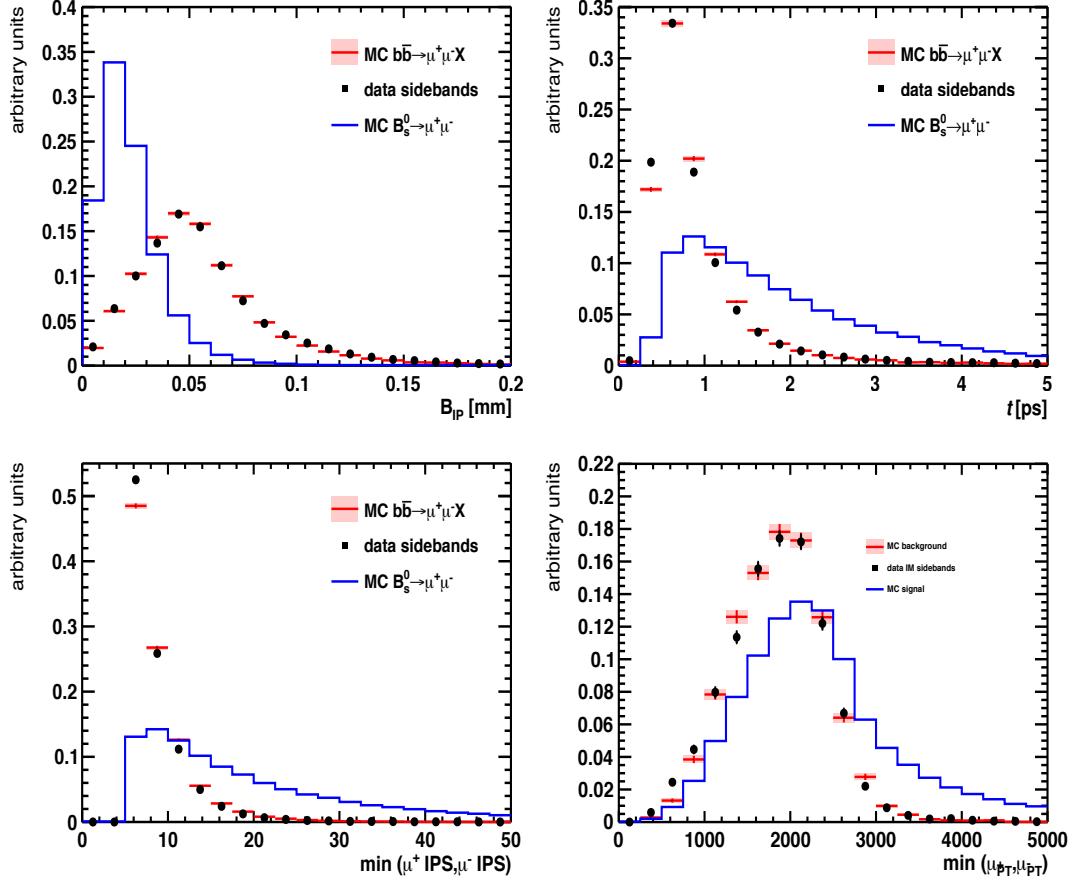


Figure 0.6: Distributions des variables entrant dans le BDT, pour le signal $B_s^0 \rightarrow \mu^+ \mu^-$ (ligne bleue) et le bruit du fond $b\bar{b} \rightarrow \mu\mu X$ MC (rouge avec des incertitudes), et des événements de données à partir des bandes latérales de la distribution de masse invariante. Toutes les distributions sont réalisées avec des événements avec $BDTS > 0.03$.

- $B_s^0 \rightarrow \mu^+ \mu^-$, défini avec une fonction CrystallBall pour la masse avec des paramètres issus des données et le BDT obtenu à travers des événements $B_{(s)}^0 \rightarrow h^+ h^-$ non-biaisés pour le système de déclenchement;
- $B^0 \rightarrow \mu^+ \mu^-$, défini de manière similaire à $B_s^0 \rightarrow \mu^+ \mu^-$ en termes de masse et BDT;
- événements de $B_{(s)}^0 \rightarrow h^+ h^-$ avec double misidentification des hadrons,
- le bruit de fond combinatoire, défini avec une fonction exponentielle obtenue à travers des événements dans les bandes latérales de masse pour la masse et avec une combinaison de trois exponentielles pour le BDT.

Toutes les pdf sont obtenues à la fin avec des données réelles.

Avec $1 fb^{-1}$ de données enregistrées par le LHCb pendant 2011, les résultats obtenus

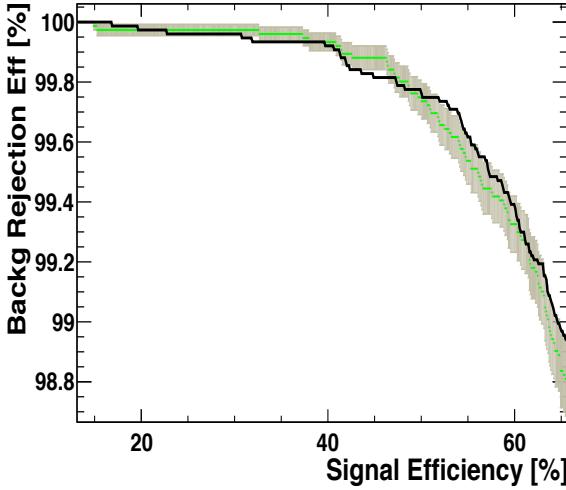


Figure 0.7: Les courbes de rejet de bruit de fond par rapport à l'efficacité du signal pour deux BDT différentes calculés avec des événements après une critère de sélection BDTS >0.05 : la courbe verte avec l'incertitude associé correspond au BDT des 9 variables. La courbe noire représente les performances du BDT optimisé dans cette section avec treize variables.

sont:

$$N_{B_s^0} = 4.5 \left(\begin{array}{l} +5.1 \\ -3.5 \end{array} \right)_{(stat)} \left(\begin{array}{l} +1.7 \\ -2.3 \end{array} \right)_{(syst)} \quad (4)$$

$$N_{B^0} = 3.6 \left(\begin{array}{l} +6.3 \\ -4.5 \end{array} \right)_{(stat)} \left(\begin{array}{l} +3.4 \\ -3.6 \end{array} \right)_{(syst)} \quad (5)$$

où les erreurs statistiques vient de l'ajustement et les incertitudes systématiques sont détaillées dans la suite. La figure 0.8 montre les projections de l'ajustement en BDT et en masse invariante, à l'exclusion des événements avec une sortie du BDT plus petite que 0.25 pour une meilleure visualisation.

Les erreurs systématiques ont différentes sources, mais sont principalement liées à la paramétrisation des *pdfs*. L'évaluation de chaque contribution liée aux différentes composantes de l'ajustement conclut que l'erreur résultant de la paramétrisation du BDT pour le bruit de fond combinatoire est la plus importante.

À l'aide des expériences simulées, on peut estimer la compatibilité de ces résultats avec la valeur prédictive pour le Modèle Standard (10.0), pas prenant en compte compte des erreurs systématiques. La figure 0.9 montre les valeurs obtenues pour le nombre des $B_s^0 \rightarrow \mu^+ \mu^-$ en simulant 10000 expériences. La fraction des événements au-dessus de 0 et inférieure à 10 est de 83% pour une couverture positive.

Les résultats susmentionnés montrent une absence d'excès de signal. Ensuite, des limites d'exclusion sont calculées avec une procédure standard: la méthode CL_s . Avec cette

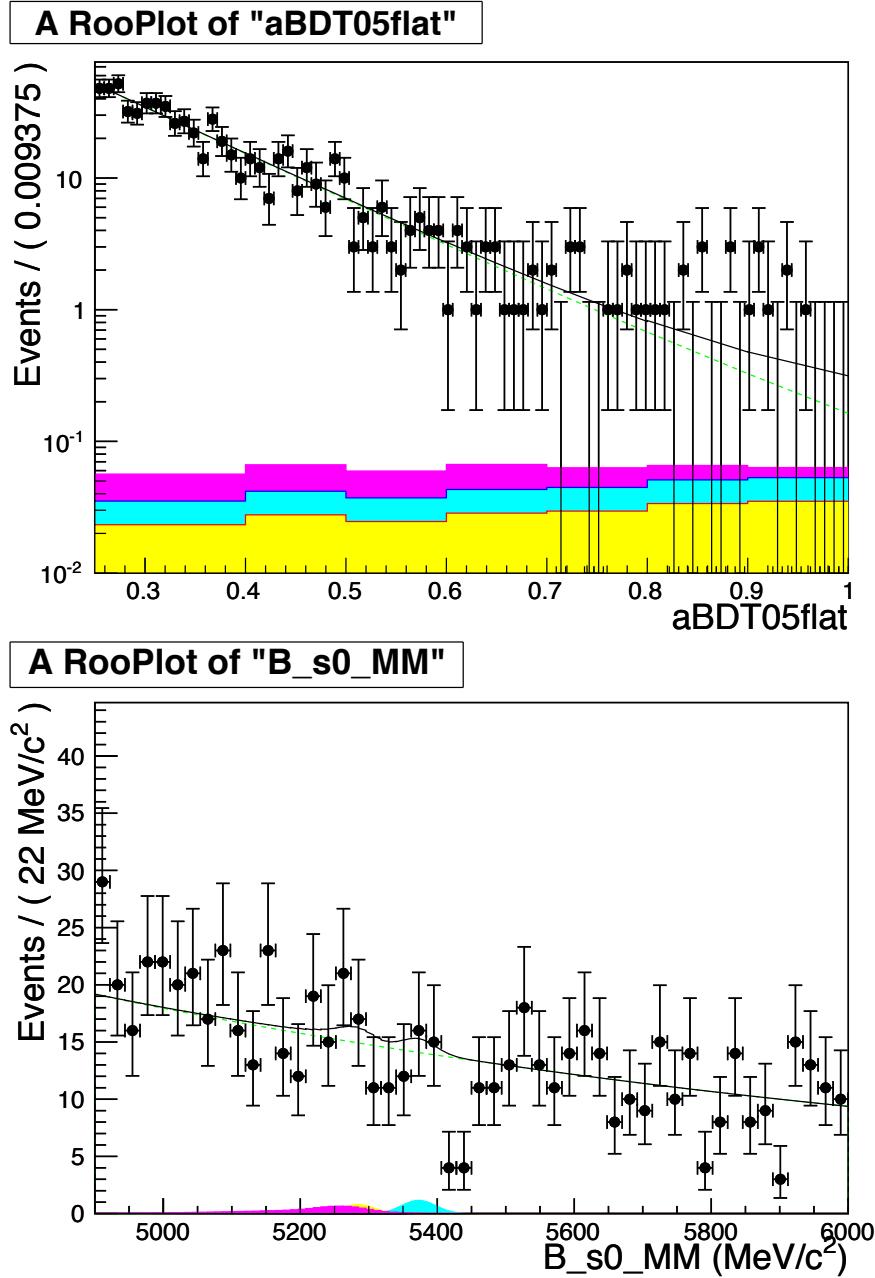


Figure 0.8: Projections de l'ajustement sur les axes du BDT (en haut) et de la masse invariante (en bas), et avec l'exclusion des événements $BDT < 0.25$. Les zones bleues correspondent au $B_s^0 \rightarrow \mu^+ \mu^-$, les zones jaunes au $B^0 \rightarrow \mu^+ \mu^-$ et les zones roses au bruit de fond de $B_{(s)}^0 \rightarrow h^+ h^-$ avec double misidentification.

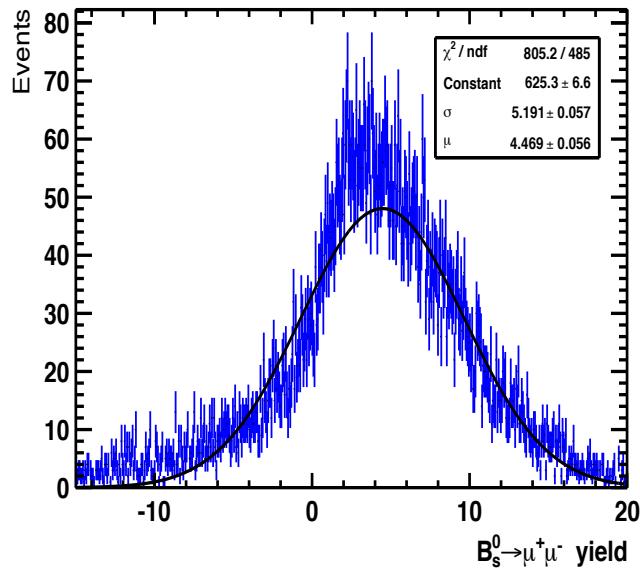


Figure 0.9: Répartition du nombre d’expériences simulées avec un nombre de $B_s^0 \rightarrow \mu^+ \mu^-$ égal à 4.5. Le nombre total de simulations générées est 10000.

méthode, on estime la compatibilité des événements observés et attendus, pour chaque rapport de branchement.

La figure 0.10 montre la distribution de CL_s pour $B_s^0 \rightarrow \mu^+ \mu^-$ (gauche) et $B^0 \rightarrow \mu^+ \mu^-$ (droite). Les valeurs de $\text{CL}_s=0.05$ permettent d’obtenir les limites avec une certitude de 95%: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.5 \times 10^{-9}$ et $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 10.3 \times 10^{-10}$.

Les limites obtenues ont un fort impact sur modèles au-delà du modèle standard.

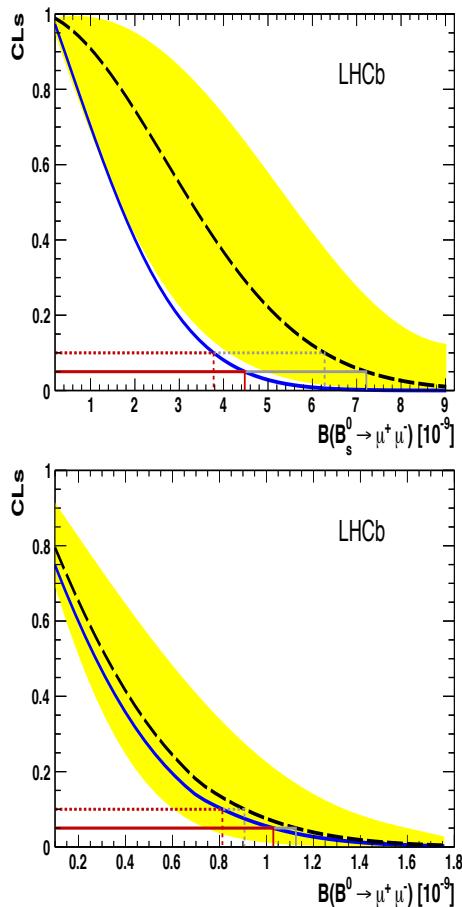


Figure 0.10: CL_s en fonction de $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. La ligne noire en pointillés correspond à la médiane des valeurs du CL_s , tandis que la zone jaune couvre la région de 68% compatibilité avec la valeur moyenne. La ligne bleue correspond à la CL_s observée. Le 90% CL et 95 % sont donnés par l'horizontale en pointillés et solides lignes rouge, respectivement.

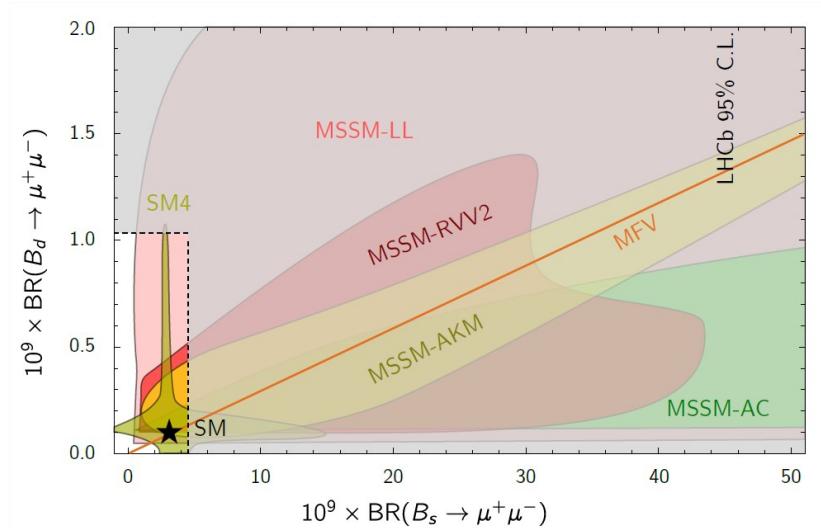


Figure 0.11: $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ vs $\mathcal{B}(B_d^0 \rightarrow \mu^+\mu^-)$ prédicté par l'hypothèse de la violation de saveur minimale (MFV) et par différentes scénarios de nouvelle physique [3]: modèles MSSM avec plusieurs contraintes et la quatrième génération avec un seul boson de Higgs (SM4). Superposé aux prévisions théoriques, la zone ombrée montre les limites d'exclusion obtenues par LHCb [4]. La valeur du MS est marqué avec une étoile.

Introduction

The Large Hadron Collider (LHC) began its activity with the first collisions of protons at a center of mass energy of 7 TeV in March 2010. For the scientific community, this machine represents an opportunity to unveil the still unanswered questions concerning the basic laws governing the fundamental interactions between elementary particles.

The Standard Model (SM) of fundamental interactions properly explains almost all observed phenomena below 200 GeV. Nevertheless, physicists tend to consider the SM an effective theory valid only up to an energy scale around the TeV, which corresponds to the energies that the LHC is exploring.

With the LHC, we expect to reveal new physics with the discovery of particles at masses never before explored. The goal of the LHC general-purpose detectors such as ATLAS and CMS is to observe directly these particles on-shell. An intrinsic drawback of this approach is that, were the new particles masses larger than a few TeV, their discovery would be impossible under the initial running conditions of the LHC at 7 TeV. Nonetheless, having indications of such particles through indirect detection is also feasible. Flavor Changing Neutral Currents (FCNC) are transitions between elementary particles which are prohibited at tree level in the SM, and occur only through loop processes. Several observables, such as branching fractions, shall serve as indications of new particles entering off-shell the FCNC loop processes.

The measurement of the branching fractions of the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ ($\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$), plays a key role in indirect searches for new physics. The small theoretical uncertainty of the SM predictions for these decays together with their clean experimental signature make them interesting probes of physics beyond the SM. In such new scenarios $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ can have different values than the SM predictions.

The experimental state of the art for such branching fractions in March 2010 was restricted to the exclusion limits set by the CDF collaboration with no indications of signal. The difference between the SM prediction and these limits was of about one order of magnitude.

Especially conceived to study b -hadron decays given its forward orientation, the LHCb detector represents the optimal experimental device to discover the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ processes.

One of the main challenges in measuring $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is to drastically reduce the background, as these branching fractions are of order 10^{-9} - 10^{-10} . After describing the theoretical and experimental frameworks of this thesis in chapters 1 and 2, respectively, chapters 3, 4, and 5 aim to describe the sources of background for these rare decays and propose

methods to reduce it. Chapters 6 and 7 detail the extraction of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, as well the implications of such measurements on new physics models.

Chapter 1

$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ as a particle physics benchmark

This chapter identifies the measurement of the branching fractions of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ as benchmark tests of the Standard Model (SM) of fundamental interactions and its extensions. Sec. 1.1 recapitulates the relevant aspects of the SM and remarks the necessity of a more complete theory. Sec. 1.2 describes the methodology used to predict the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ in the SM and beyond (BSM).

1.1 Towards a theory of fundamental interactions

Almost all known phenomena from 1 eV up to almost 200 GeV are well described by the Standard Model. However, experimental evidence and theoretical implications motivate the study of new models extending the SM description of nature.

The SM is a quantum field theory that includes the symmetry groups of strong ($SU(3)_C$) and electroweak interactions ($SU(2)_L \times U(1)_Y$). The sub-indices C , L , and Y refer to color charge, left-handed fields, and hypercharge, respectively. The Lagrangian describing the SM has been developed on the basis of experimental facts and certain invariance principles.

Fermions constitute the basic building blocks of matter (Tabs. 1.1 and 1.2), while the mediators of the interaction between them arise by requiring invariance of the Lagrangian under gauge transformations. These intermediate particles are known as bosons (Tab. 1.3).

Table 1.1: The lepton content of the SM. For each particle it exists a corresponding anti-particle with opposite charge, for charged particles.

Family 1	electron, e^-	electron neutrino, ν_e
Family 2	muon, μ^-	muon neutrino, ν_μ
Family 3	tau, τ^-	tau neutrino, ν_τ
Electric charge	-1	0

Table 1.2: The quark content of the SM. For each particle it exists a corresponding anti-particle with opposite charge.

Family 1	down, d	up, u
Family 2	strange, s	charm, c
Family 3	bottom, b	top, t
Electric charge	-1/3	2/3

Table 1.3: The boson content of the SM.

Interaction	Particle
Electromagnetic	γ
Weak	Z^0 and W^\pm
Strong	gluons

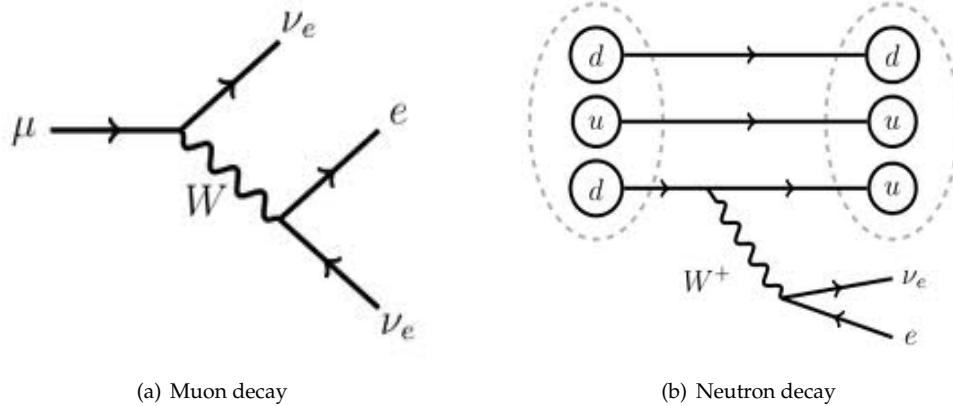


Figure 1.1: Example of charged currents in the SM.

1.1.1 The electroweak interaction

The electroweak (EW) interactions are divided into charged currents (CC) and neutral currents (NC). In the first group, quarks and leptons interact with the W^\pm bosons, as in the decays $\mu^- \rightarrow e^- \bar{\nu}_e \nu_\mu$ and $n \rightarrow p e^- \bar{\nu}_e$ (Fig. 1.1). These processes confirmed that left-handed (right-handed) fermion (anti-fermion) chiralities take part in weak processes. The second group of EW interactions corresponds to interactions of the photon and Z with a fermion and its anti-fermion (Fig. 1.2). Transitions such as $\mu \rightarrow e\gamma$ or $Z \rightarrow e\mu$ have never been observed.

The two fundamental features of the EW interactions are the *mixing of the quark flavor* and the *EW symmetry breaking*.

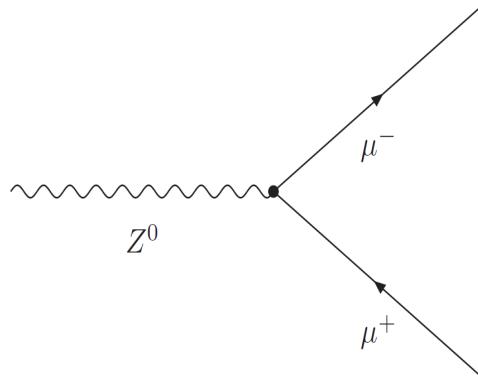


Figure 1.2: Example of a neutral current in the SM corresponding to the decay of a Z boson into a dimuon pair.

Quark mixing

Quark flavor mixing emerges from the observational fact that processes like $K^+(u\bar{s}) \rightarrow \mu\nu_\mu$ exist. Since processes arising from the invariance $SU(2) \times U(1)$ do not mix quarks from different families, it is necessary to postulate that the weak interaction couples to a linear superposition of mass eigenstates. The mixing between families emerges when quark fields are expressed in the basis that diagonalizes the time evolution operator of the theory. This change of basis introduces the Cabibbo-Kobayashi-Maskawa matrix V_{CKM} [5], [6], which depends on three real parameters and one complex phase. This complex phase results in the violation of charge-parity (CP), for which it is the only source within the SM. The CKM matrix is written in the standard parameterization as:

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

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

where $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$, and i, j the family indices from 1 to 3. The values in the diagonal of the matrix are close to 1, $|V_{cd}|$ and $|V_{us}|$ are of the order of 20%, $|V_{cb}|$ and $|V_{ts}|$ of the order of 4%, and $|V_{td}|$ and $|V_{ub}|$ of the order of 1%. These values are obtained via direct measurements of processes like $K \rightarrow l\bar{\nu}\pi$ ($|V_{us}|$) or through indirect measurements such as the oscillation of the B^0 ($|V_{td}|$) or B_s^0 ($|V_{ts}|$) mesons.

The GIM scheme and flavor changing neutral currents

In 1970, Glashow, Iliopoulos and Maiani ensured to forbid transitions that could change flavor but not charge ($d \rightarrow sZ$) [7] and postulated the existence of the quark c . Several

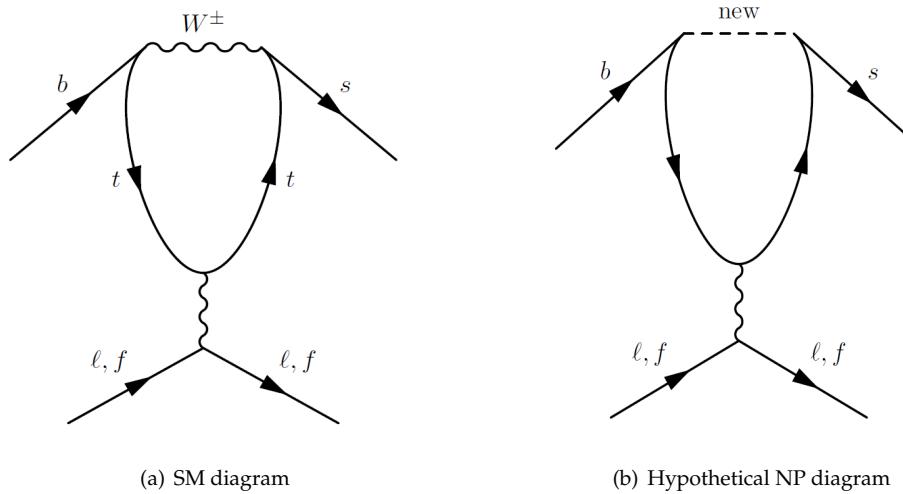


Figure 1.3: Penguin diagrams contributing to $b \rightarrow s l \bar{l}$ involving observed particles (a) and hypothetical particles (b).

reasons supported this hypothesis: the lack of experimental evidence of $K^0 \rightarrow \mu^+ \mu^-$; the unitarity of V_{CKM} that implies that just $d \rightarrow d$ and $s \rightarrow s$ transitions are allowed; the part of the SM Lagrangian corresponding to the Z^0 boson (γ) current is proportional to $Z^\mu \bar{u}_L \gamma_\mu u_L$ ($-A^\mu \bar{u}_L \gamma_\mu u_L$), which lacks a term permitting transitions between quarks of different flavors. As a result, the GIM mechanism forces flavor changing neutral currents amplitudes to be suppressed by $\sum_i V_{bi}^\dagger V_{id}$

Why is it interesting to study FCNC?

FCNC transitions are prohibited at tree level and highly suppressed in the SM. It is possible that, in new physics scenarios, new particles entering a FCNC loop process could alter the amplitude of such decay.

Fig. 1.3 displays the Feynman diagrams of the FCNC transition $b \rightarrow s l \bar{l}$ in the SM and in a possible new physics scenario, where a new particle enters off-shell in the diagram. By measuring observables such as branching fractions we can infer the presence of new particles in FCNC processes.

Furthermore, such processes represent a benchmark for new theories as their parameters must respect the current experimental observables given by such loop processes.

EW symmetry breaking

Experiments have long established that fermions and gauge bosons (excluding the photon and the gluons) are massive. Introducing mass terms in the SM Lagrangian implies that gauge invariance is lost and that $SU(2)_L \times U(1)_Y$ is not a good symmetry of the vacuum, whereas $U(1)_{em}$ is, given that γ is mass-less.

The gauge symmetry is therefore broken by the vacuum, which triggers the Spontaneous Symmetry Breaking (SSB) of the electroweak group to the electromagnetic subgroup:

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}. \quad (1.1)$$

This constitutes the so-called SSB, which enables fermions and bosons to acquire mass. As a result of this mechanism, a new particle emerges: the *Higgs* boson. Until July fourth 2012, this boson remained undiscovered [8] [9] [10]. On this day, the CMS and ATLAS collaborations confirmed a formal discovery of a boson of mass between 125 and 127 GeV/c^2 , consistent with a Higgs boson. Discovering the Higgs was one of the main reasons to build the Large Hadron Collider (LHC).

The puzzles in the SM

Despite the confirmation of many SM predictions, this theory has several limitations and is unable to account for some observational facts.

Dark matter. The content of visible matter in the Universe cannot account for the observed rotation of galaxies [11] [12] in the context of general relativity. Furthermore, studies of the fluctuations of the cosmic microwave background indicate the existence of cold dark matter in the Universe [13] for which there is no SM candidate.

Difference between content of matter and antimatter. There is a large asymmetry between the quantity of matter and antimatter observed in the Universe. Assuming that both were equally created in the initial state of the Universe, a condition such as the violation of the CP symmetry is necessary to account for such observed differences. The problem is that the magnitude of CP violation possible in the SM is not enough to explain them [14] [15].

Gravity. There is not yet a consistent procedure to introduce gravity in the SM.

Large number of free parameters. The SM lacks predictions for particles' masses and mixing patterns which translates into a large number of unpredicted parameters.

The mass hierarchy problem. The mass of a scalar such as the Higgs boson suffers from quantum corrections of its mass due to the physics above a certain scale Λ , $m_{H_{SM}}^2(\text{phys}) \simeq m_{H_{SM}}^2 + \frac{c}{16\pi^2}\Lambda^2$. An upper bound for the physical mass at $130\text{ MeV}/c^2$ due to electroweak precision measurements indicates that Λ should be $\mathcal{O}(\text{TeV})$, in order to avoid a fine tuning of the bare mass term.

Hereafter we describe the main ideas to avoid fine tuning of the bare mass term, which are mostly based on new mechanisms enabling the EWSB.

Assumption 1. There are new degrees of freedom suppressing the diverging term of the scalar mass. This represents the main reasoning of *supersymmetry*, which is a symmetry that relates properties of bosons and fermions [16] [17].

Assumption 2. There are no elementary scalars. Instead, there are composite states or fermions condensates, which introduce new fermions (*technifermions*) and a new interaction arising from a non-abelian gauge symmetry: *technicolor* [18].

Assumption 3. Gravity effects are important at the EW scale, changing the standard four-dimensional space-time structure by adding *extra dimensions* [19] [20].

Assumption 4. The quadratic divergences to the Higgs mass appear at two-loop level. Scenarios including such assumption correspond to *Little Higgs* models [21].

Technicolor models face serious problems with FCNC constraints in order to generate fermion mass hierarchies for the SM fermions. In supersymmetry the squarks must be degenerate in mass in order to avoid FCNC. If supersymmetry exists close to the TeV scale it could solve the hierarchy problem and accomplish the unification of interactions. Nevertheless, the mentioned degeneracy would indicate signs of fine tuning.

The Minimal Flavor Violation (MFV) hypothesis.

The bounds on physics beyond the Standard Model coming from precise flavor tests conclude that new sources of flavor symmetry breaking beyond the SM Yukawa couplings between the Higgs field and a massless fermion field are excluded for new physics at the TeV scale. MFV argues that there are no deviations from the SM in flavor changing processes [22].

In conclusion, taking MFV as valid implies that no other source of flavor symmetry breaking other than the SM Yukawa couplings are accessible at low energies.

1.2 Theoretical predictions for $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$

The branching fraction of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ predicted by the SM are $(3.2 \pm 0.2) \times 10^{-9}$ and $(1.0 \pm 0.1) \times 10^{-10}$ [1]. Secs. 1.2.1 and 1.2.2 introduce the necessary tools to derive such predictions, and those corresponding to new physics scenarios.

1.2.1 Effective field theory for B decays

In B decays, the momentum exchange between the initial and final states is of the order of the mass of the decaying meson (~ 5 GeV). The masses of the intermediate bosons for these processes are $M_{W,Z} \sim 100$ GeV. The key idea behind the Effective Field Theory (EFT) is that physics at a certain low energy scale should decouple from physics at much higher scales, which translates into ruling out the heavy fields.

Therefore, we can construct an effective Hamiltonian that does not contain the W and Z explicitly using the Operator Product Expansion (OPE) approach:

$$H_{\text{eff}} = \sum_i C_i(\mu, Q) \mathcal{O}_i(\mu), \quad (1.2)$$

where C_i are the Wilson coefficients that encode the information from the high energy scale, \mathcal{O}_i the local operators that represent the low-energy effective theory, Q the matching scale

or cutoff and μ the renormalization scale that accounts for the separation of information between high and low energies. Given an initial B meson state $|B\rangle$ and a final state $|X\rangle$, the corresponding contribution to the amplitude can be expressed as:

$$\mathcal{M} \approx \langle |X|H_{\text{eff}}|B\rangle = \sum_i \mathcal{C}_i \langle |X|\mathcal{O}_i|B\rangle. \quad (1.3)$$

We shall see in the next section that calculating the hadronic matrix elements for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ takes a particularly simple form.

1.2.2 $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ in the framework of EFT

The EFT approach is applied to describe the four external fields $b \rightarrow ql^+l^-$ ($q=s,d$) transitions with two quarks of different flavor and two charge-conjugated leptons. These processes are the basis for the analysis of decays such as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, and are represented by the following Hamiltonian [23]:

$$H_{\text{eff}} = \sum_{i=7}^{10} \mathcal{C}_i \mathcal{O}_i + \mathcal{C}_S \mathcal{O}_S + \mathcal{C}_P \mathcal{O}_P + \sum_{i=7}^{10} \mathcal{C}'_i \mathcal{O}'_i + \mathcal{C}'_S \mathcal{O}'_S + \mathcal{C}'_P \mathcal{O}'_P. \quad (1.4)$$

The operators \mathcal{O}_1 to \mathcal{O}_6 are not present in the previous equation as they contribute only in hadronic decays. Also, \mathcal{O}_8 does not contribute as represents the gluonic contribution. \mathcal{O}_7 arises from the photon contribution, while \mathcal{O}_9 and \mathcal{O}_{10} are the semileptonic operators, and \mathcal{O}_S and \mathcal{O}_P the scalar and pseudo-scalar operators:

$$\begin{aligned} \mathcal{O}_7 &= m_b (\bar{q} \sigma_{\mu\nu} P_R b) F^{\mu\nu}, \\ \mathcal{O}_9 &= (\bar{q} \gamma^\mu P_L b) (\bar{l} \gamma_\mu l), \\ \mathcal{O}_{10} &\equiv \mathcal{O}_A = (\bar{q} \gamma^\mu P_L b) (\bar{l} \gamma_\mu \gamma_5 l), \\ \mathcal{O}_S &= m_b (\bar{q} P_R b) (\bar{l} l), \\ \mathcal{O}_P &= m_b (\bar{q} P_R b) (\bar{l} \gamma_5 l). \end{aligned}$$

Exchanging the right-handed projection operator P_R by the left-handed P_L and m_b by m_q allows to describe the primed partners of these operators.

The matrix element for the decay $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ can be expressed as factorization of the leptonic and the hadronic parts:

$$\mathcal{M} \approx \langle l \bar{l} | H_{\text{eff}} | B(p) \rangle = \sum_i \langle l \bar{l} | \mathcal{O}_{\text{lept}}^i | 0 \rangle \langle 0 | \mathcal{O}_{\text{hadr}}^i | B(p) \rangle \quad (1.5)$$

where the sum i represents the scalar, pseudo-scalar, the γ loop and semi-leptonic contributions.

The matrix element:

$$\langle 0 | \bar{q} \gamma_\mu \gamma_5 b | \bar{B}(p) \rangle = i p_\mu f_{B_q}, \quad (1.6)$$

where p_μ is the four-momentum and f_{B_q} the decay factor of the B meson, vanishes when contracted with $\bar{l} \gamma_\mu l$. Furthermore, the matrix element $\langle 0 | \bar{q} \sigma_{\mu\nu} b | \bar{B}(p) \rangle$ also vanishes as it is

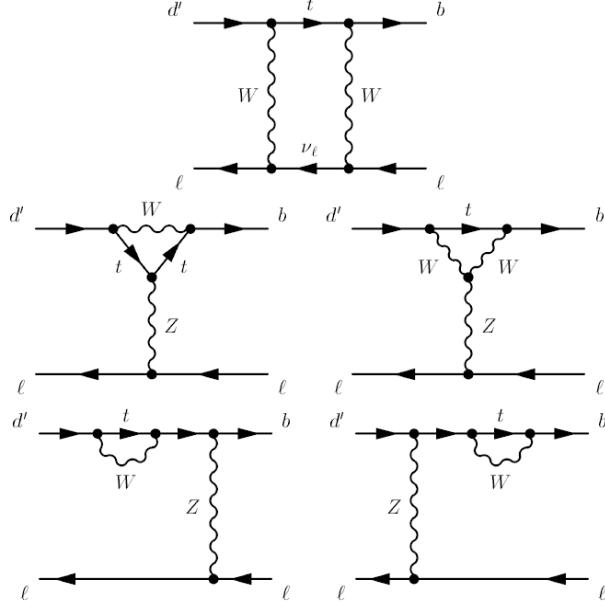


Figure 1.4: SM diagrams contributing to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. The diagrams are analogous for $B_s^0 \rightarrow \mu^+ \mu^-$ exchanging d for s .

not possible to construct an antisymmetric combination made up of p^μ . Consequently, the operators \mathcal{O}_7 and \mathcal{O}_9 do not contribute to the decay $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, which is only governed by the axial part of \mathcal{O}_{10} and the scalar and pseudo-scalar operators.

We quote the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ as obtained in Ref. [23]:

$$\mathcal{B}(\bar{B}_q \rightarrow l^+ l^-) = \frac{G_F^2 \alpha^2 M_{B_q} \tau_{B_q}}{16\pi^3} |V_{tb} V_{tq}^*|^2 \sqrt{1 - \frac{4m_l^2}{M_{B_q}^2}} \left[\left(1 - \frac{4m_l^2}{M_{B_q}^2} \right) |F_S|^2 + |F_P + 2m_l F_A|^2 \right]. \quad (1.7)$$

The corresponding form factors can be represented as:

$$F_{S,P} \equiv i f_{B_q} M_{B_q} \frac{C_{S,P} m_b - C'_{S,P} m_s}{m_b + m_s},$$

$$F_A \equiv i f_{B_q} (C_{10} - C'_{10}).$$

$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ within the SM

Fig. 1.4 shows the main diagrams contributing to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ in the SM associated to C_{10} as the scalar and pseudo-scalar are suppressed by a factor $(m_l m_{b,q} / M_W^2)$. These branching fractions can be obtained computing the Wilson coefficients related with the high-energy regime of the processes.

Furthermore, there is an intrinsic source of uncertainty due to the experimental knowledge of the decay constant ($f_{B_s} = 238.8 \pm 9.5 \text{ MeV}$ [24]). A method proposed in Ref. [1]

translates the dependence on the decay constant into the meson mass difference, available from experimental observations ($\Delta M_{Bs} = 17.77 \pm 0.12 \text{ ps}^{-1}$ [25]). Using this approach the SM predictions have an uncertainty of 6% for $B_s^0 \rightarrow \mu^+ \mu^-$ and 10% for $B^0 \rightarrow \mu^+ \mu^-$:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{SM} = (3.2 \pm 0.2) \times 10^{-9} \quad (1.8)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)_{SM} = (1.0 \pm 0.1) \times 10^{-10} \quad (1.9)$$

The aforementioned branching fractions are implied through this thesis when mentioning a SM prediction. Nevertheless, improvements in lattice calculations predicting $f_{Bs} = 227.7 \pm 6.2 \text{ MeV}$, imply slight differences on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ predictions [26]. Furthermore, as suggested in Ref. [27], the experimental observables are time independent measurements that, to be compared with the theoretical predictions, these should be enlarged by a 9%.

Contributions to $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ from NP beyond the SM

The decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ are sensitive probes of new physics given the smallness of their branching fraction uncertainty. Potential contributions to the branching fraction of these decays arise from the modification of the Wilson coefficients.

Two Higgs Doublet Model 2HDM [28] is an extension of the SM which contains a larger Higgs sector: two Higgs doublets both with a vacuum expectation value vev different from zero. An important parameter of this framework is the ratio of both vev $\tan \beta \equiv V_b/V_a$. Moreover, it contains two neutral scalars H^0 and h^0 , a neutral pseudoscalar A^0 , two charged scalars H^\pm , and three would-be Goldstone bosons G^0 and G^\pm . Fig. 1.5 shows the relevant diagrams contributing to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in 2HDM.

In 2HDM-II, down-type quarks and charged leptons couple to one Higgs doublet, and up-type quarks and neutrinos couple to the other doublet. In this model, both C_S and C_P receive contributions proportional to $\tan^2 \beta$. Fig 1.6 shows the branching fraction prediction as a function of the parameter M_{H^+} and different $\tan \beta$ values [29]. Values of M_{H^+} below 295 GeV are excluded by the measurement of $b \rightarrow s\gamma$ [30]. In spite of these constraints, enhancements of the branching fraction are possible, especially for low values of M_{H^+} and large $\tan \beta$.

2HDM with MFV Scenarios in which the b -quark Yukawa coupling becomes $\mathcal{O}(1)$ (large $\tan \beta$) present enhancements with respect to the SM expectations for observables such as $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ [31]. Fig. 1.7 shows the correlation of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ in models with Higgs-mediated FCNC respecting the MFV hypothesis ($\Gamma(B_s^0 \rightarrow \mu^+ \mu^-)/\Gamma(B^0 \rightarrow \mu^+ \mu^-) \approx |V_{ts}/V_{td}|^2$) and reflects that an enhancement of both branching fractions respecting the MFV hypothesis would be a clear hint of MFV at large $\tan \beta$.

MSSM. The minimal supersymmetric extension of the Standard Model (MSSM) [32] is constructed by adding the corresponding partners (super-particles) of the 2HDM particle

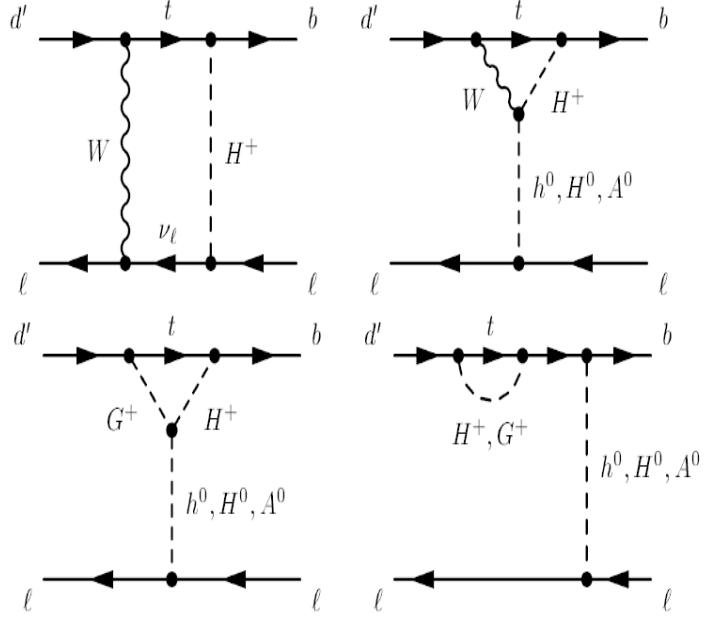


Figure 1.5: 2HDM diagrams contributing to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. The diagrams are analogous for $B_s^0 \rightarrow \mu^+ \mu^-$ exchanging d for s [23].

content. There is no experimental evidence of superparticles, therefore supersymmetry (SUSY) is supposed to be spontaneously broken. Not requiring SUSY breaking explicitly translates into more than one hundred parameters in the theory. Nevertheless, this number is usually constrained by assuming universality within the parameters.

MSSM contains the same diagrams as in the SM and 2HDM-II, although diagrams with loop exchange of the particles with their super-partners (Fig. 1.8) are also present in the computation of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ within MSSM.

New contributions on the Wilson coefficients appear in the MSSM model, which affect mainly C_A . As a result, a significant enhancement of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is possible at large values of $\tan\beta$.

In MSSM, as in 2HDM, $C_{S,P}$ and $C'_{S,P}$ have comparable size which leads to their suppression. As a result, values of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ smaller than the SM are expected in minimal supersymmetric scenarios.

$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is computed in the context of two of such scenarios through a likelihood built according to current electroweak precision measurements, B physics data, cosmological observations and recent LHC results [33].

Fig. 1.9 shows that the most probable value of the likelihood gives a $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ prediction slightly larger than the SM in the constrained MSSM (CMSSM), which assumes universality for the mass of the scalar fields, the gaugino masses, and the coupling of the

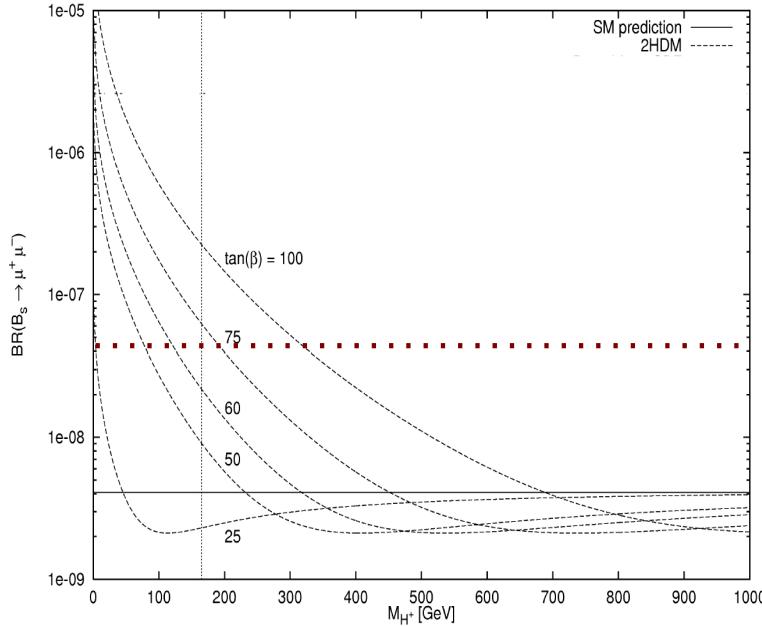


Figure 1.6: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ in 2HDM as a function of the parameter M_{H^+} . The horizontal dotted line represents the experimental limit on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ from Ref. [2].

triple scalar vertices [34]. Including the mass of the Higgs $\simeq 125$ GeV does not alter remarkably the minimum of the $\Delta\chi^2$ distribution if $(g - 2)_\mu$ is included.

Fig. 1.10 shows that in MSSM with non-universal Higgs masses (MSSM-NUHM1) departures from the SM prediction are compatible with the empirical input of the likelihood. This model inherits all the universality assumptions of the CMSSM except those related with the Higgs masses [35]. As a result, two new free parameters arise: the modulus and the mass of the pseudoscalar Higgs M_A . The best fit value reflects that $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ in NUHM1 is significantly larger than the SM prediction. As in the previous case, including $M_h \simeq 125$ GeV does not alter remarkably the distribution of $\Delta\chi^2$ if $(g - 2)_\mu$ is included.

1.3 Conclusions

This chapter described the theoretical framework of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, emphasizing the role to be played by this measurement in constraining the new paradigm of fundamental interactions between elementary particles.

The accurate theoretical predictions in the SM, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.2 \pm 0.2) \times 10^{-9}$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.0 \pm 0.1) \times 10^{-10}$, represent an opportunity to search for contributions arising from new physics beyond the SM.

Large enhancements from the SM prediction can be possible in several new physics models, especially for MSSM-NUHM, where the most probable value for $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ is about 60% larger than the SM prediction.

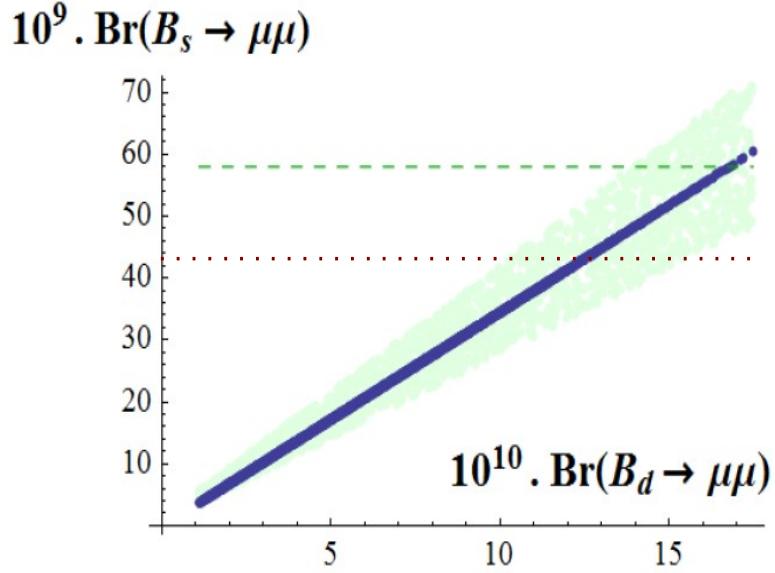


Figure 1.7: Correlation between $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ in Higgs-mediated FCNC respecting the MFV hypothesis. The central value (red line) corresponds to $\Gamma(B_s^0 \rightarrow \mu^+ \mu^-)/\Gamma(B_d^0 \rightarrow \mu^+ \mu^-) \approx |V_{ts}/V_{td}|^2$, while the green points account for the uncertainties in $|V_{ts}|$ and $|V_{td}|$. The horizontal dotted line represents the experimental limit on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ from Ref. [2].

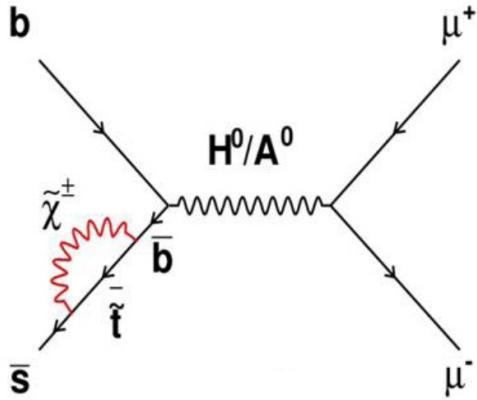


Figure 1.8: Feynman diagram contribution to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ within MSSM with superpartners entering in the loop.

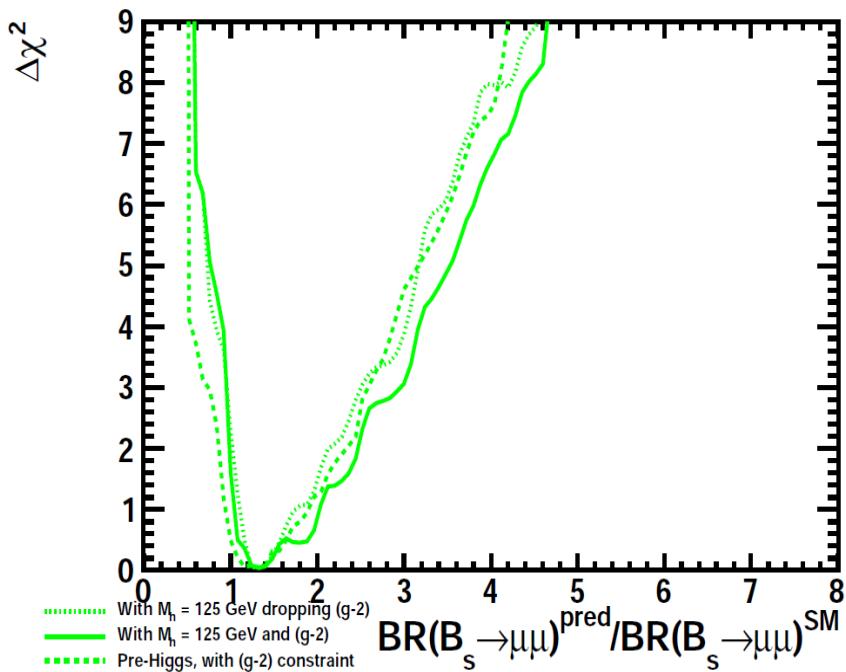


Figure 1.9: Likelihood functions for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ in the CMSSM. The solid line assumes a Higgs mass $\simeq 125$ GeV and includes $(g-2)_\mu$, and the dotted line not including $(g-2)_\mu$. The dashed line does not include M_h in the computation.

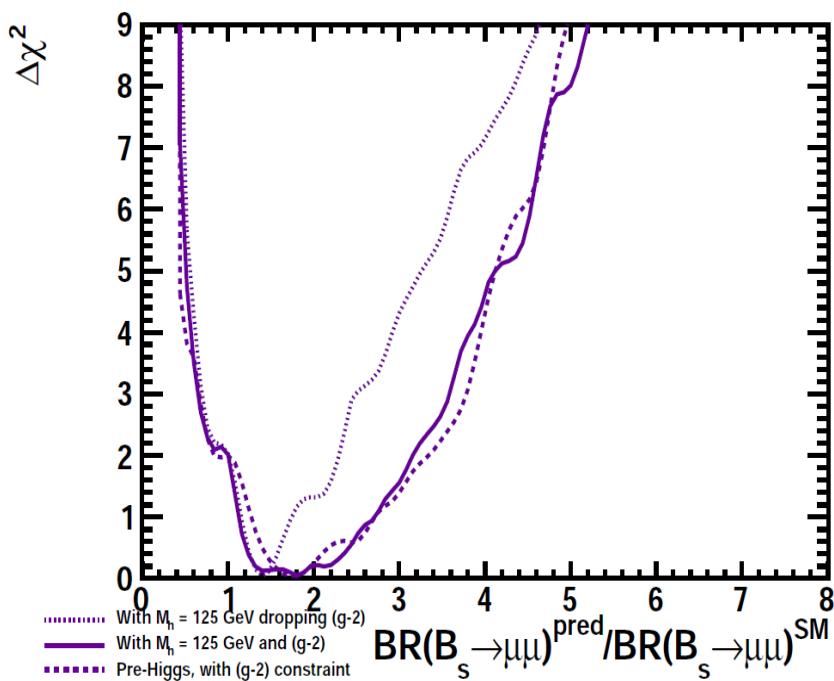


Figure 1.10: Likelihood functions for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ in the NUHM. The solid line assumes a Higgs mass $\simeq 125$ GeV and includes $(g-2)_\mu$, and the dotted line not including $(g-2)_\mu$. The dashed line does not include M_h in the computation.

Chapter 2

The LHCb experiment

The B physics era started with the discovery of the b -quark by a fixed target experiment (FNAL) in 1977. Then, most of the experimental outcome arose from B factories: CLEO and ARGUS collaborations discovered the $b \rightarrow u + W$ transitions in 1990; Belle and BaBar validated the Kobayashi-Maskawa mechanism as major source of CP violation, and ended their activity with $\sim 1.1 \text{ ab}^{-1}$ of integrated luminosity and $\sim 1.3 \times 10^9 B\bar{B}$ pairs.

Despite the arduous environment, experiments at the Tevatron hadron collider showed its competitiveness with B factories, benefiting from a large $b\bar{b}$ cross-section, and studying the unknown B_s^0 system.

This chapter describes the LHCb experiment [36]. Sec. 2.1 summarizes the experimental conditions of the analysis presented in this thesis. Secs. 2.2 and 2.3 contain a brief description of each LHCb sub-detector, focusing on their purpose and performances. Sec. 2.4 discusses the LHCb trigger system with special attention to the aspects relevant for the measurement of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$.

2.1 Introduction

The Large Hadron Collider (LHC) [37] is a two ring superconducting accelerator, where the protons travel in opposite directions. It was installed in the same tunnel of 26,7 km unearthened for the LEP collider in the 80's, and it was designed to collide protons at an energy of 14 TeV at the center of mass. Nevertheless, the experience acquired with the first tests performed in 2008 showed that new bus bars were necessary in order to avoid magnet quenches at such energies.

LHC is delivering data until the end of 2012 at an energy of 7 TeV (8 TeV since the beginning of 2012) at the center of mass since March 2010, before a two-year technical stop. Fig. 2.1 illustrates the acceleration chain allowing to reach such energies.

The colliding bunches at a rate of 40 MHz have $\sim 10^{11}$ protons, and the data taking scheme during the beam lifetime is known as a fill.

2.1.1 LHCb general features

LHCb is an experiment conceived for heavy flavor physics studies at the LHC. Its main goal is to search for indirect evidence of new physics in CP violation and in rare decays of

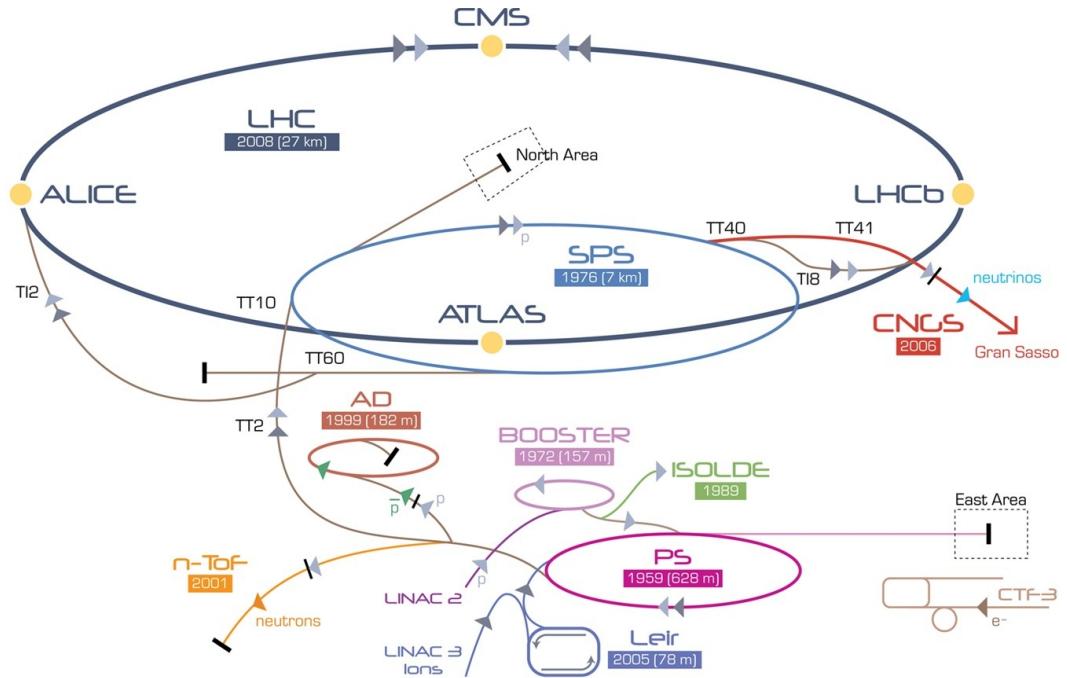


Figure 2.1: The CERN acceleration complex. Protons are first accelerated with the LINAC up to an energy of 50 MeV, then injected in the PS where they reach 26 GeV, and subsequently in the SPS. In the last step, the protons enter in the LHC with an energy of 450 GeV.

beauty and charm hadrons.

To be able to exploit the large number of b -hadrons produced at the LHC (Fig. 2.2), an efficient and flexible trigger is used to cope with the coarse hadronic environment.

b -hadrons arise from $b\bar{b}$ quark pairs mainly produced through hard QCD interactions (see Fig. 2.3) which hadronize into a large family of hadrons: B^+ , B^0 , B_s^0 , B_c^+ , Λ_b . The expected angular distribution of the produced $b\bar{b}$ pair is highly adjacent to the beam axis (see. Fig. 2.4), which justifies the choice of the LHCb detector layout: a single-arm forward spectrometer (see Fig. 2.5) build as a compromise between budget and acceptance.

A dipole magnet providing an integrated field of 4 Tm allows for momentum measurement of charged particles. The tracking system consists of a vertex locator, called VELO, situated close to the interaction point and including a pile-up veto counter. It is followed by a silicon micro-strip detector and tracking stations, downstream and upstream the magnet, respectively. Ring Imaging Cherenkov counters (RICH) allow an excellent hadron identification in the momentum range from 2 to 100 GeV/c. The calorimeter system providing reconstruction of photons, π^0 and identification of electrons comprises a Scintillator Pad Detector (SPD), a Preshower (PS), and an electromagnetic (ECAL) and hadronic (HCAL) calorimeters. The muon system providing muon identification and contributing to the trigger comprises five muon chambers.

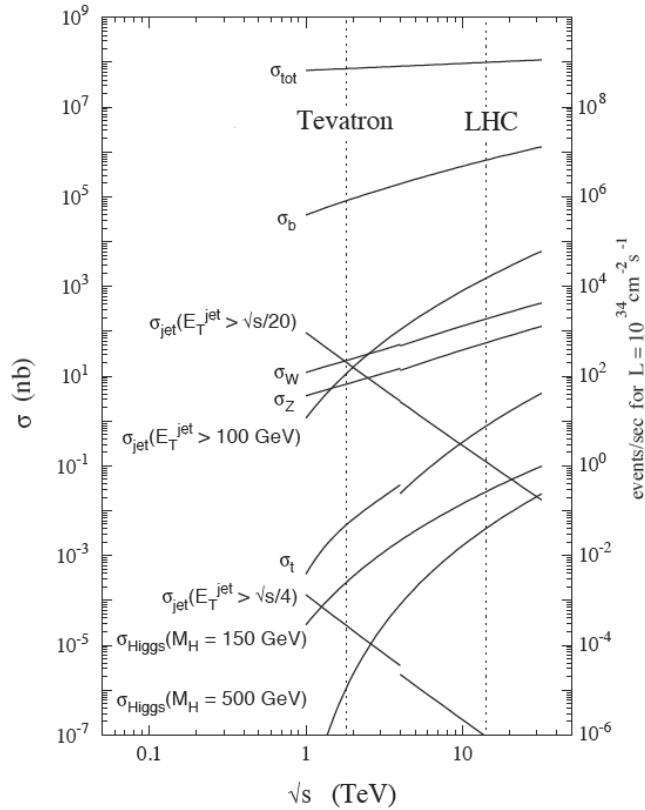


Figure 2.2: Production cross-section and number of events per second produced as a function of energy of the collision at the center of mass. Almost three times more $b\bar{b}$ pairs are produced at 7 TeV (first LHC run) compared to 2 TeV (Tevatron's highest energy) for a given instantaneous luminosity.

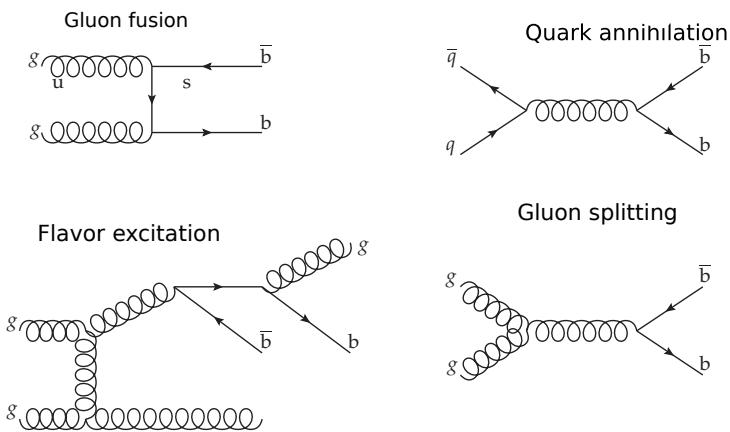


Figure 2.3: Main b-quark production mechanisms. Flavor excitation (47%) and gluon splitting (28%) [38] are the dominant processes at LHC.

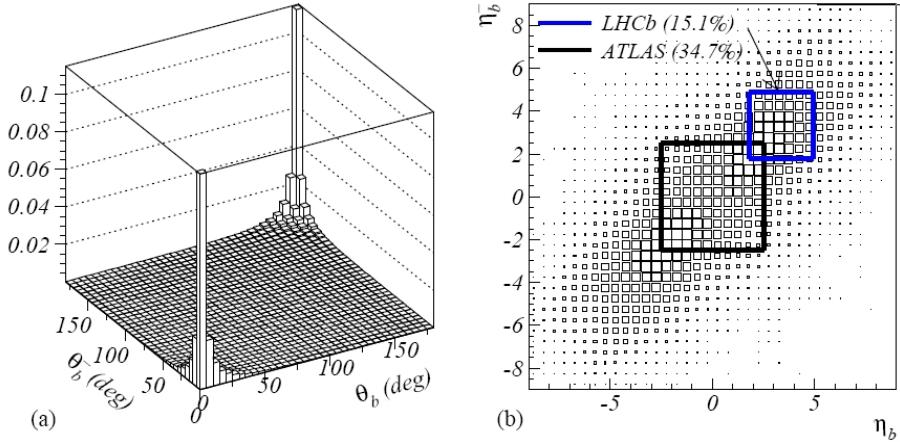


Figure 2.4: (a) : Correlation of the polar angles of B mesons produced in proton-proton collisions at $\sqrt{s} = 14\text{ TeV}$ [38]. (b) : as in (a) but in pseudo-rapidity. The plot shows the complementarity of the Atlas and LHCb experiments for the detection of $b\bar{b}$ pairs. ATLAS covers 34.7% of the produced $b\bar{b}$ pairs, while LHCb covers 15.1%.

2.1.2 Data taking periods

At the end of 2009, LHCb recorded the first pp -collisions at the injection energy of the LHC, $\sqrt{s} = 0.9\text{ TeV}$.

In 2010, the energy of the LHC beams ramped up to half their nominal value delivering collisions at $\sqrt{s} = 7\text{ TeV}$. These running conditions started with very low instantaneous luminosities (smaller than $10^{28}\text{ cm}^{-2}\text{s}^{-1}$) and almost no pile-up, to reach $10^{30}\text{ cm}^{-2}\text{s}^{-1}$ with relatively large pile-up, up to 3 collisions per pp bunch crossing. Fig. 2.6 shows the delivered luminosity as a function of time that peaks at the end of the 2010 data run.

Although the pile-up had been more than 4 times above the nominal value, the luminosity was two orders of magnitude below the nominal LHCb luminosity ($2 \times 10^{32}\text{ cm}^{-2}\text{s}^{-1}$) due to the low number of pp bunches in the LHC. Nevertheless, the physics output was not compromised.

During 2011, the number of bunches in the machine continuously increased until about half of the machine was filled. This larger number of bunches allowed to reduce the pile-up over the year while running at luminosities close to twice the nominal LHCb luminosity. A particular feature of the 2011 data taking is the luminosity leveling: adjusting the displacement of the two colliding beams in order to obtain constant luminosity during the normal lifetime of the beams, as shown in Fig. 2.7. This figure displays the instantaneous luminosity as a function of time delivered to each of the four LHC experiments. Fig. 2.8 shows an steady increase of the integrated luminosity as a function of the LHC fill number.

The studies presented in Chap. 4 use data taken during 2010. The rest of the studies, including the detector performances described in the following sections, are based solely on 2011 data.

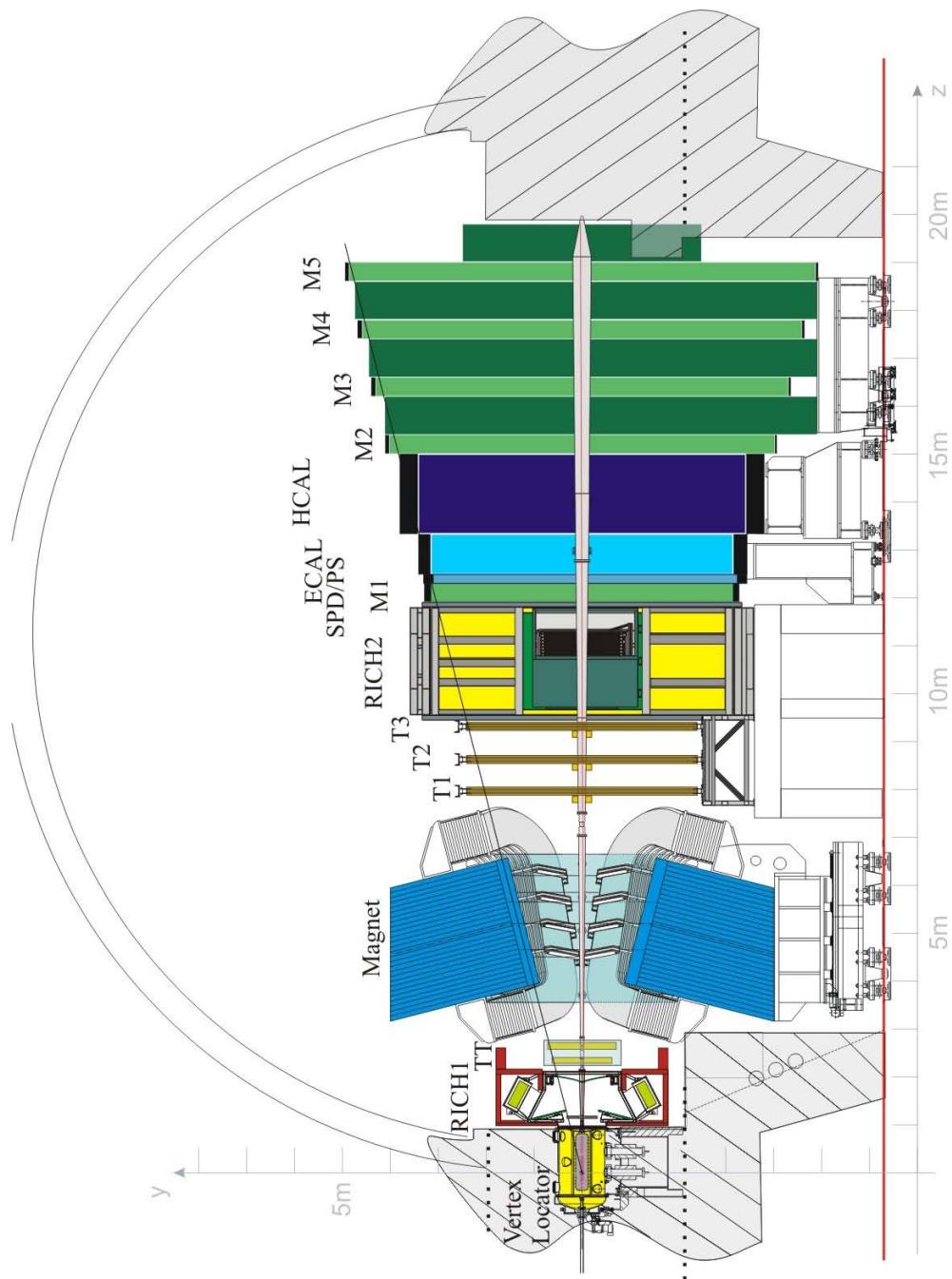


Figure 2.5: General view of the LHCb experiment. It covers an acceptance defined by $0.01 < \theta_{xz} < 0.25$ rad and $0.01 < \theta_{yz} < 0.30$, where θ_{xz} and θ_{yz} refer to the angles with respect to xz and yz planes respectively. The collisions take place at the origin of the z horizontal axis, inside the Vertex Locator.

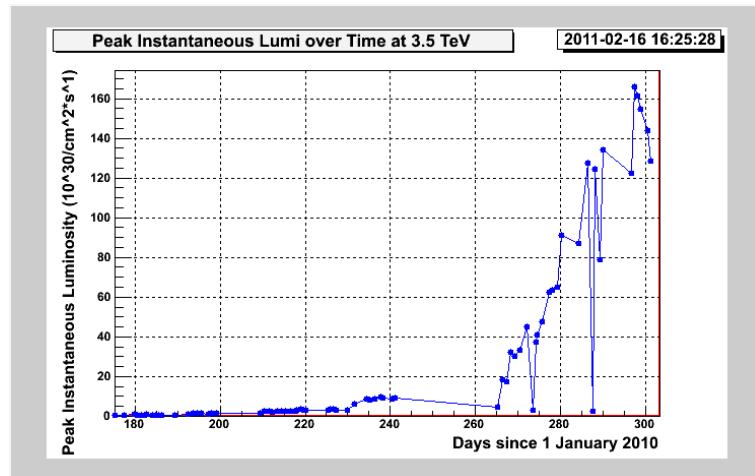


Figure 2.6: 2010 delivered luminosity as a function of time.

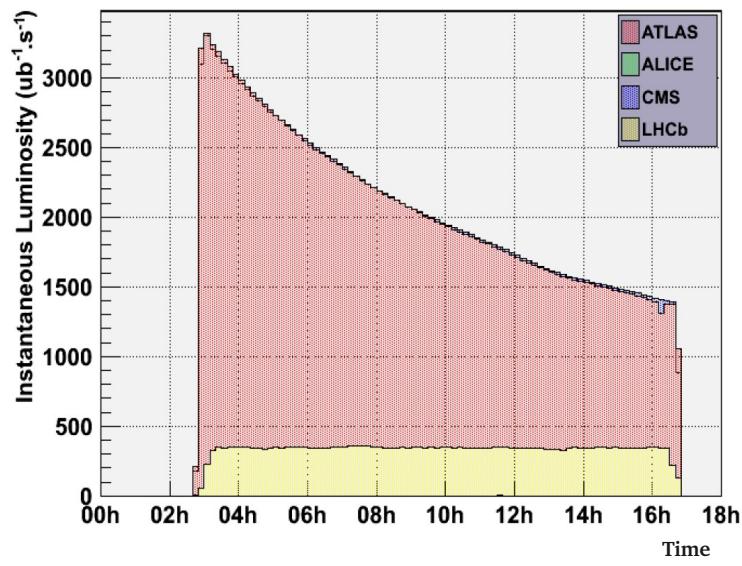


Figure 2.7: Instantaneous luminosity as a function of time delivered to the four LHC experiments during 2011 data taking: ALICE, ATLAS, CMS, and LHCb.

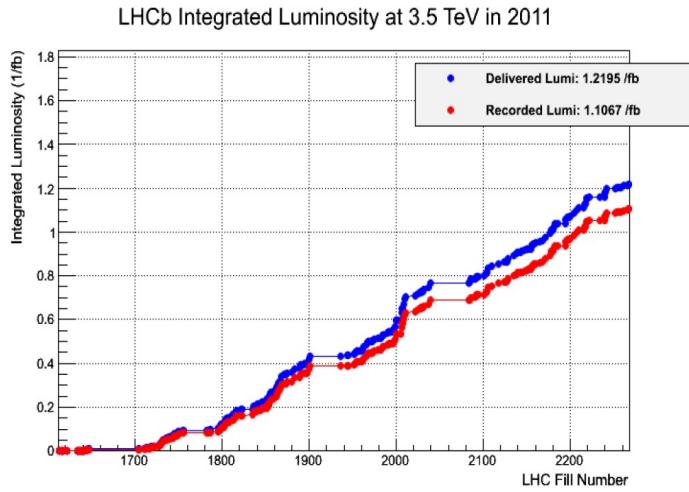


Figure 2.8: LHCb integrated luminosity through 2011 as a function of LHC fill number.

2.2 LHCb Tracking System

The B meson lifetimes are about 1.5 ps. This, together with the fact that their momenta at the LHC is much larger than their mass, results in an average separation of the order of 1 cm between decay and production vertices. Therefore, an accurate vertexing is crucial to distinguish the primary from the decay vertices to be able to identify B hadron decays. Furthermore, good momentum resolution is needed in order to obtain good mass resolutions.

The key elements of the LHCb tracking system are:

2.2.1 The vertex locator (VELO)

The VELO (Fig. 2.9) provides information on the trajectory of charged particles in the region closest to the interaction point, which is used to determine the primary and secondary vertices and as input to the track reconstruction. It consists of two-half detectors, each made out of silicon modules along the beam direction providing both radial and azimuthal information. These two halves are retractable to be as close as 8.2 mm to the beam pipe during data taking. In addition, each half contains two Pile-Up veto stations composed of a sensor designed to reject events with more than one interaction by the trigger system.

The primary vertex resolution has been estimated to be of about $12 \mu\text{m}$ in x and y , and $65 \mu\text{m}$ in z for 30-35 track vertices, which is the average number of tracks per primary vertex (see Fig. 2.10 (a)). Another important feature for a B -physics experiment is the impact parameter resolution which, for the 2011 data sample, we found to be $13.2 + 24.7/p_T \mu\text{m}$ (see Fig. 2.10 (b)).

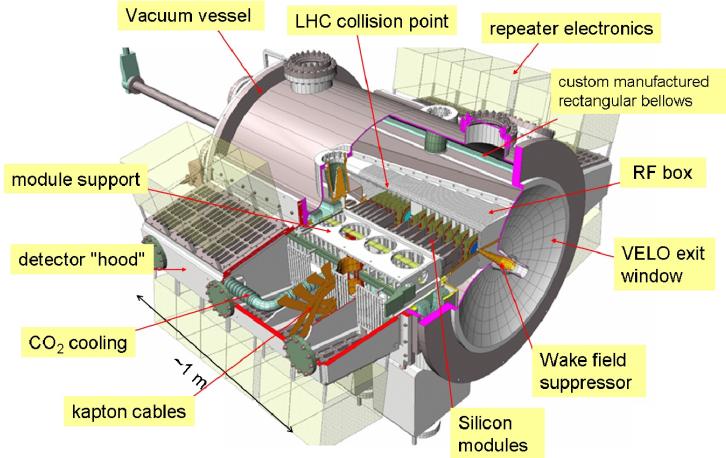


Figure 2.9: Vertex locator VELO layout.

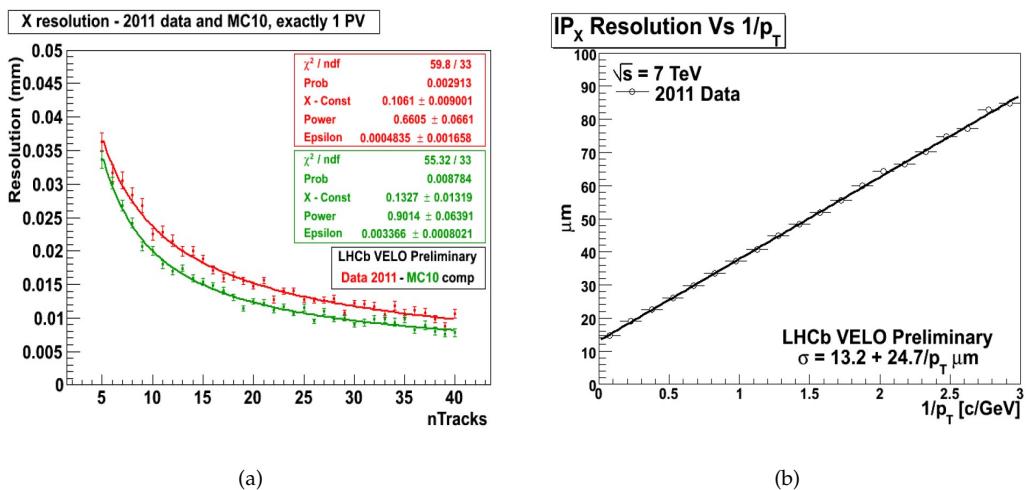


Figure 2.10: Primary vertex resolution in the transverse plane as function of the number of tracks (a) and impact parameter resolution in the transverse plane as function of $1/p_T$ (b).

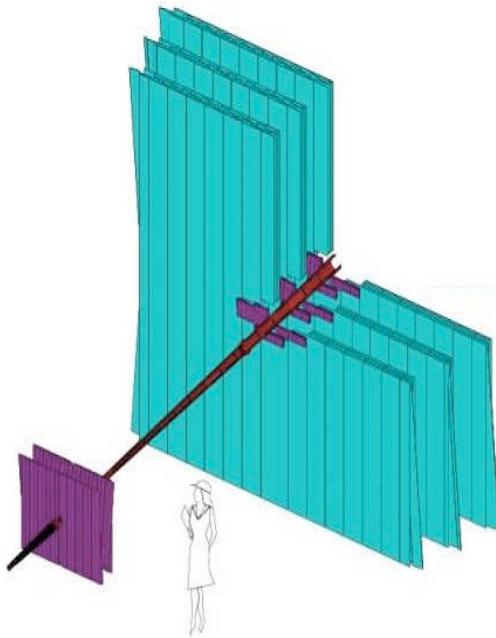


Figure 2.11: Overview of the tracking stations, with the beam pipe represented for reference. TT and IT are displayed in purple, while OT is shown in blue.

2.2.2 The trackers

The trackers include the tracker turicensis (TT), the inner-tracker (IT) and the outer tracker (OT). These sub-detectors, based of silicon sensors layers, are situated upstream and downstream of the magnet as displayed in Fig. 2.11 and provide a momentum resolution of 0.3 to 0.5%.

2.2.3 Magnet

Indirectly part of the tracking system, the magnet is used to establish the tracks' momenta. Its magnetic field corresponds to 4 Tm and its polarity is regularly reversed in order to record data in both orientations to evaluate systematic effects.

2.3 Particle identification

Important B physics measurements contain photons, electrons, muons, pions or kaons in the final state which need to be properly identified. For instance, the determination of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ relies not just on the muon identification but also requires the separations between kaons and pions to extract the invariant mass resolution and mean of B_s^0 and B^0 (see Sec. 6). At the LHCb, the identification of particles is achieved by combining the information furnished by the following sub-detectors:

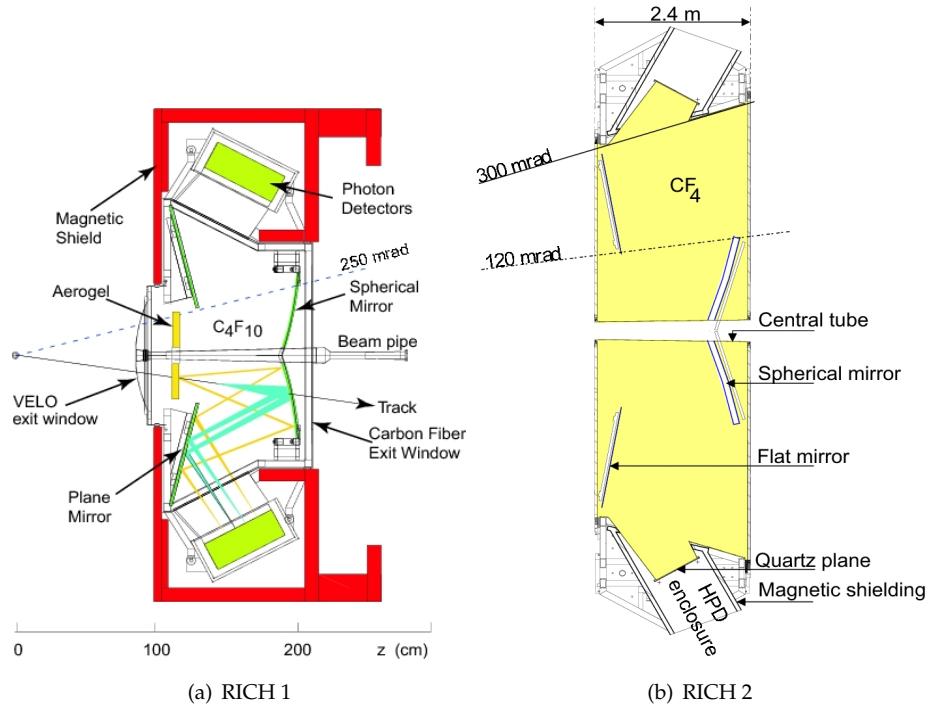


Figure 2.12: Lateral view of RICH 1 and view from below of RICH 2.

2.3.1 The RICH detectors

The Ring Imaging Cherenkov detectors (RICH) allow to measure particles' velocities via the rings of Cherenkov light produced by their passage through certain radiators such as silicon aerogel or gaseous C_4F_{10} . Two of these detectors, RICH1 and RICH2 (Fig. 2.12), perform an identification in different ranges of momentum: below 60 GeV for the RICH1 and from 15 GeV up to 100 GeV for the RICH2. Both detectors achieve an angular resolutions of 1.62 and 0.62 mrad with respect to the xz and yz planes.

Hadron separation The information furnished by the RICH detectors is used to separate between pions, kaons, and protons. The particle identification is performed by an algorithm based on a log-likelihood approach matching observed patterns to what expected from the reconstructed tracks under a given set of particle hypothesis. The method, referred as global pattern-recognition, consists in maximize the likelihood by varying the particle hypothesis of each track. The log-likelihood difference (ΔLL) between a kaon hypothesis and the pion hypothesis is given by:

$$\Delta LL(K - \pi) = \ln \left(\frac{\mathcal{L}_K}{\mathcal{L}_\pi} \right)$$

Fig. 2.13 shows the efficiency and the fake rate of π/K identification over the entire momentum range. The mis-identification of pions into kaons is $\sim 3\%$ over the momentum range requiring $\Delta LL(K - \pi) > 5$.

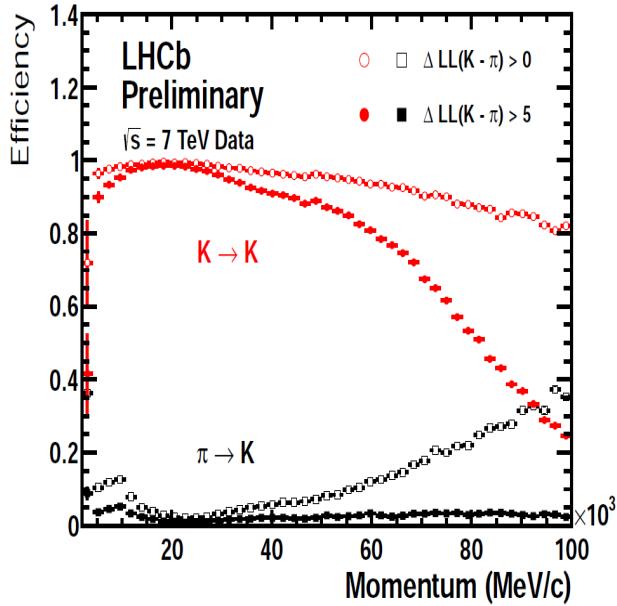


Figure 2.13: Efficiency and fake rate identification of the Ring Imaging Cherenkov detectors.

2.3.2 Calorimeters

The deposited energy in the calorimeters allows to estimate the energy of charged and neutral particles and provide information for their identification. Together with the momentum measurement through the tracking system it allows to reject background from low-energy interactions. Four calorimeters comprise the calorimetry of the LHCb detector, here ordered according to their z coordinate:

- Scintillator Pad Detector (SPD). Situated in front of the calorimeters, it distinguishes between photons and electrons. It consists of a single layer of scintillator cells which provide a binary outcome depending on the deposited energy.
- Pre-shower (PS) detector. A layer of scintillator cells measures the energy of the particle shower originated through a 15 mm thick lead barrier (2.5 radiation lengths) downstream the PS. It allows to distinguish photons and electrons from hadrons.
- Electromagnetic calorimeter (ECAL). It consists of stacks of alternating lead-scintillator plates that represent 25 radiation lengths. In combination with the PS, it provides an energy measurement with a resolution of $\sigma_E/E = 9\%/\sqrt{E} \otimes 0.8\%$, with E in GeV.
- Hadronic calorimeter (HCAL). It comprises iron and scintillator plates arranged parallel to the yz plane (5.6 interaction lengths). The resolution in the estimation of the shower energy is $\sigma_E/E = 69\%/\sqrt{E} \otimes 9\%$, with E in GeV.

The information recorded by the calorimeters is used during the trigger process.

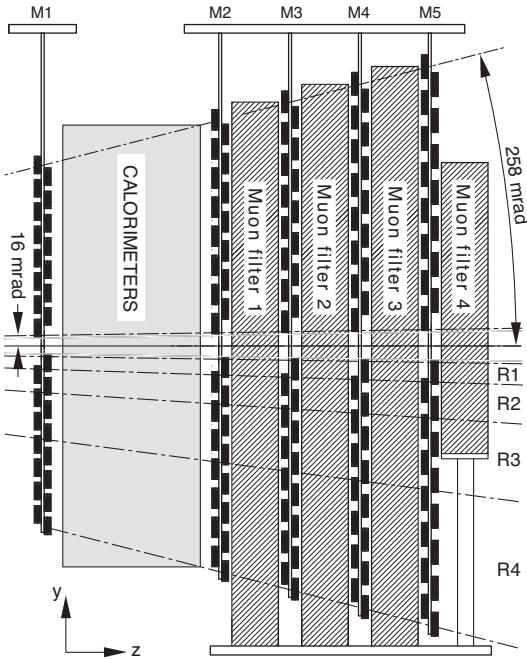


Figure 2.14: Lateral view of the muon system.

2.3.3 Muon system

The muon system aims to identify muons. This sub-detector consists in five stations placed along the z axis (M1 to M5) and it plays an important role in the trigger system (next section) during data taking. Moreover, the offline muon identification is crucial as muons are present in many key channels studied in LHCb as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $B_s^0 \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) \phi (\rightarrow K^+ K^-)$, $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, among others.

The muon stations are furnished with 1368 multi-wired proportional chambers (MWPC) divided in four regions of different granularities, at the exception of M1 which used gas electron multiplier technology (GEM), to handle the high flux of particles. Fig. 2.14 shows a lateral view of the muon system. It gathers 25 926 logical channels that are used by the trigger system.

2.3.4 Muon identification

Properly identifying muons is crucial for the measurements of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$. Hits in the muon stations which match the direction of a track reconstructed in the tracking system constitute a muon candidate. For each muon candidate, likelihoods for muon and non-muon hypothesis are computed. This test is based on the average distance of the closest hits in the muon station to the extrapolation from the track stations.

The performance of the muon identification is estimated using muons from $J/\psi \rightarrow \mu^+ \mu^-$, protons from $\Lambda \rightarrow p\pi$, and kaons and pions from $D^0 \rightarrow K\pi$ decays. Fig. 2.15 shows the efficiency of the muon candidate selection (*IsMuon*) as a function of the track momentum in different p_T ranges. This efficiency is above 97% for tracks with $p_T > 1700 \text{ MeV}/c$

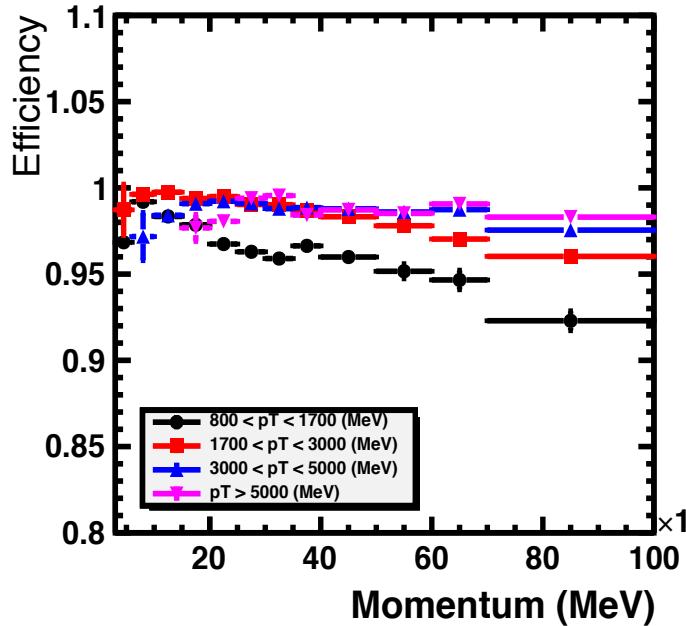


Figure 2.15: Efficiency of the muon candidate selection as a function of momentum for different ranges of transverse momentum.

and above 92% for $800 < p_T < 1700$ MeV/c in the interval $2 < p < 100$ GeV/c. Fig. 2.16 shows the probabilities of incorrect identification of kaons, pions, and protons as muons, being below 5% in all cases.

2.4 Trigger system

The LHCb trigger system is conceived to select B and D decays with hadrons and leptons in the final state. It is flexible and redundant, and very efficient for muons. These features are relevant for the analysis described below.

Reducing the event rate at LHC down to 3 kHz allowed by the storage resources, constitutes the main goal of the trigger. Dedicated electronics processing the information from the Velo, calorimeters, and muon stations, allow to reduce the rate down to 1 MHz, at which the detector can be read out. This initial level which reduces the initial rate down to 1 MHz is called Level-0 (L0) trigger. Consecutively, the High Level Trigger (HLT), a software trigger running in a computer farm of thousands of CPU nodes, performs the remaining reduction of the event rate down to 3 kHz, by accessing the information of the entire detector. Fig 2.17 displays a synopsis of the LHCb trigger.

The trigger information is embedded into a *trigger configuration key* (TCK), an hexadecimal key that identifies the set of trigger decisions, algorithms and selections applied. Different configurations allow to fully exploit the trigger infrastructure available for any given data taking condition.

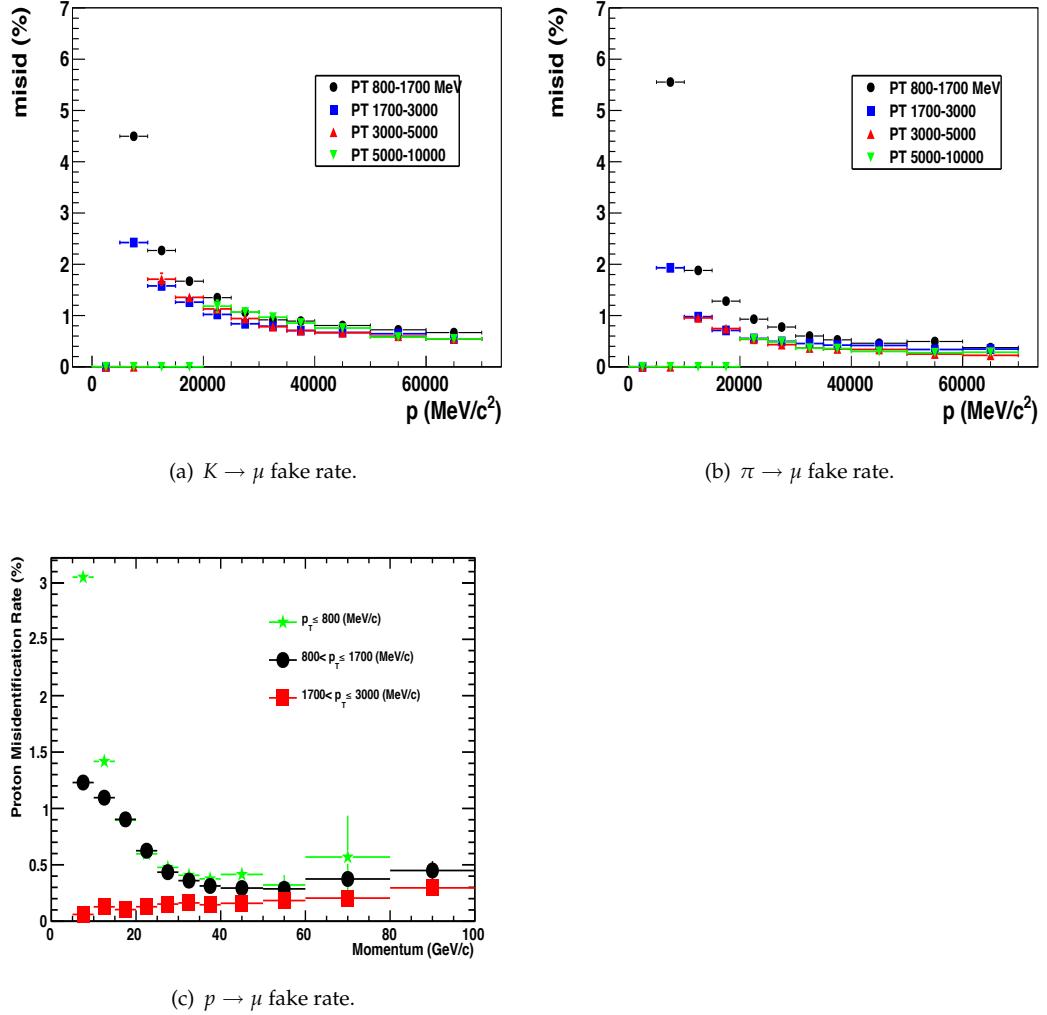


Figure 2.16: Probability of incorrect identifications as muons of several hadrons as a function of momentum, in different p_T ranges.

In nominal conditions, the required number of primary interactions per crossing should be 0.6, while this number peaked to 2 at the end of 2010 and was kept almost constant around 1 during 2011 data taking. These relatively high pile-up conditions translate into more particles in the event which imply not only a degradation of the processing time, but also a reduction of the rejection performances of the trigger system. A selection based on event quantities, called *global event cuts* (GEC) intend to suppress events with high-multiplicity requiring a maximum number of hits in the SPD, VELO, IT, and OT, on top of the other trigger requirements. Solely the SPD hits requirements changed during 2010 data taking.

2.4.1 Level-0 trigger

The L0 requirements are based on several detector measurements:

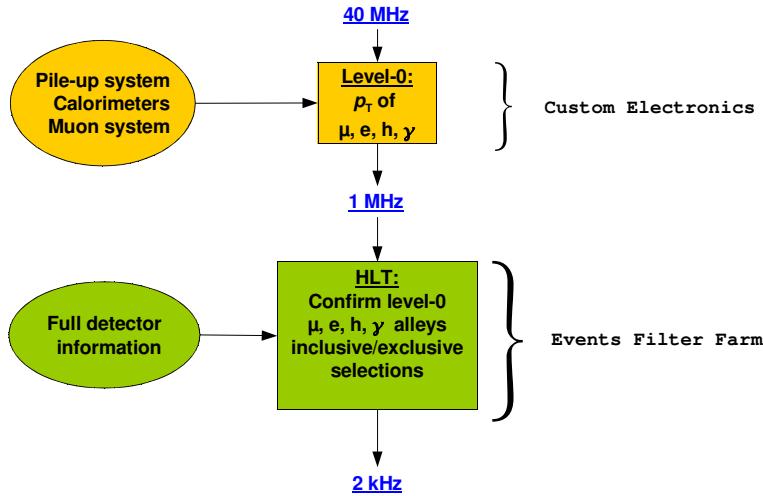


Figure 2.17: LHCb trigger synopsis.

- Transverse energy of the candidates in the calorimeters. With the information provided by the SPD and the PS, each candidate is classified as either a photon, an electron or a hadron.
- Transverse momentum of the candidates in the muon system. Muon candidates require hits in all muon stations.
- Number of primary interactions estimated by the Pile-Up system.
- Multiplicity of charged tracks measured by the SPD.

The L0 requirements (high p_T photons, electrons or hadrons, and single or double high p_T muons) that fulfill the presence of a B decays in an event for the 2010 and 2011 data are reported on Tab. 2.1.

2.4.2 High Level Trigger

HLT algorithms use the information of the full detector in successive steps. In the first step, HLT1, only the information from the VELO and the tracking stations helps to refine the L0 candidates. In the second step, HLT2, the full reconstruction of the event is performed, which enables to either accept or reject events according to several selection requirements.

The HLT1 main decisions, reported in Tab. 2.2, are:

- *Hlt1TrackMuon*, requires events triggered by the L0 on muons with a muon candidate satisfying p_T and impact parameter significance (IPS) requirements.
- *Hlt1SingleMuonNoIP*, same as *Hlt1TrackMuon* without IPS requirement.
- *Hlt1DiMuonNoIP*, requires events triggered by the L0 on muons with a pair of muons satisfying selections based on p_T and their combination, but not requiring any displacement with respect to the primary vertex.

Table 2.1: L0 requirements for four different TCK categories used during 2010 (first three raws) and 2011 (bottom raw). Each TCK is classified according to its L0- μ , L0-dip μ , and L0-hadron criteria: muon candidate or candidates p_T (GeV/c) and the maximum number of SPD hits (nSPD).

TCK	L0- μ p_T / nSPD	L0-dip μ p_{T1} / p_{T2} / nSPD	L0-hadron p_T (GeV/c) / nSPD
xx0030 ($\sim 2 \text{ pb}^{-1}$)	1.4 / 900	0.56 / 0.48 / 900	2.6 / 900
xx002A ($\sim 14 \text{ pb}^{-1}$)	1.4 / 900	0.56 / 0.48 / 900	3.6 / 900
xx002C ($\sim 12 \text{ pb}^{-1}$)	1.4 / 900	0.56 / 0.48 / 900	3.6 / 450
xx0032 -xx0038 ($\sim 1 \text{ fb}^{-1}$)	1.5 / 600	$\sqrt{p_{T1} \times p_{T2}} = 1.3$ / 900	3.5 / 600

Table 2.2: HLT1 requirements for four different TCK categories used during 2010 (first three raws) and 2011 (bottom raw). Each TCK is classified according to its HLT1 decisions: muon candidate p_T (GeV/c), its impact parameter IP (mm), and its impact parameter significance IPS, and the combination of transverse momentum of both muon candidates $m_{\mu\mu}$. (a) stands for *Hlt1SingleMuonNoIP* and (b) for *Hlt1DiMuonNoIP*.

TCK	<i>Hlt1TrackMuon</i> p_T	(a) $m_{\mu\mu}$	(b) p_T / IP / IPS	<i>Hlt1TrackAllL0</i> p_T / IP / IPS
xx0030	800 / 0.11 / 5	1.8	2.5	1450 / 0.11 / $\sqrt{50}$
xx002A				
zz002C				
xx0032	1850 / 0.1 / 1.5	1.8	0.5	1850 / 0.1 / 16
-xx0038				

- *Hlt1TrackAllL0*, requires a track in the event satisfying criteria based on p_T and IPS, regardless of the L0 decision.

The HLT2 lines that select channels with J/ψ in the final state require two muons with invariant mass of $120 \text{ MeV}/c^2$ around the J/ψ mass. Similar requirements, although a restriction of invariant mass above $4.7 \text{ GeV}/c^2$, select $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ events. To select two-body b-hadron decays the HLT2 trigger requires two tracks in the invariant mass with good reconstructed secondary vertex, satisfying IP and p_T selection criteria.

Many other HLT2 lines are used but they are not relevant for this study.

2.4.3 Types of event according to the trigger

The events can be classified according to the information required by the trigger, which records all the necessary information for such classification. The main difference in the trigger selection is introduced when the signal products were responsible or not for in the triggering decision. Three main categories are defined according to the trigger selection:

- *trigger on signal* (TOS) or events that would be triggered only requiring the selection criteria on the studied signal B decay products;
- *trigger independent of signal* (TIS) or events that would be triggered even if the signal B and its decay products were not present in the event. TIS events constitute a sample of trigger unbiased events and are commonly used, for instance, to estimate trigger efficiencies;
- *trigger on both* (TOB) or events that do not fit in the other categories. Both the signal and the rest of the event are needed to trigger these events.

2.5 Conclusions

This chapter presents the LHCb detector and summarizes the key features that make it the optimal experiment to search for rare B meson decays with muons in the final state. Among this features we distinguish an optimal trigger for B decays, together with a good muon identification efficiency; a good invariant mass resolution allowing to reduce background; and good primary and secondary vertex reconstruction, crucial in selecting B decays.

Chapter 3

Analysis challenges to measure $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$

The main challenge in determining $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is to disentangle the signal from the background. We shall describe the background sources in Sec. 3.1. Kinematic and geometrical informations allow to discriminate between signal and background.

The idea behind this chapter is twofold: first to familiarize the reader with the analysis strategy adopted to obtain $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (Sec. 3.1) and then to focus on the characterization and removal of the background (Sec. 3.2).

3.1 Analysis overview

Fig. 3.1 displays the typical dimuon invariant mass spectrum for L0- μ triggered events. This figure shows that in the region close to the B^0 and B_s^0 masses, 5280 MeV and 5365 MeV, respectively, there is a large amount of events. We recall from the first chapter that the values of the branching fractions of $B_s^0 \rightarrow \mu^+ \mu^-$ or $B^0 \rightarrow \mu^+ \mu^-$ are very small. Therefore, the number of signal events embedded in this sample is negligible. The expected number of events in the 2010 is 0.7 $B_s^0 \rightarrow \mu^+ \mu^-$ and 0.08 $B^0 \rightarrow \mu^+ \mu^-$, while 10.0 and 1.2 are expected for the 2011 data taking. In conclusion, understanding and exploiting the properties of signal candidates that differ from those of background events represents the most important step of our analysis. Consequently we require an efficient suppression of such background while retaining most of signal as well as most events from control channels, which are processes used to determine certain quantities without relying on simulation. For example, $B^0 \rightarrow K\pi$ decays are used to extract from data the mean value of the B^0 mass (see Chap. 6.1.1), and $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$, $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$, and $B^0 \rightarrow K\pi$ decays are used in the yield normalization (see Chap. 7). $B_{(s)}^0 \rightarrow h^+ h^-$ decays are used to estimate in data the output of the discriminant multivariate classifier between signal and background (see Chap. 5).

The trigger, already described in Sec. 2.4, constitutes the initial part of this efficient selection. We stress here the importance of keeping the selection as similar as possible between different channels, in order to avoid biases on the measured quantities.

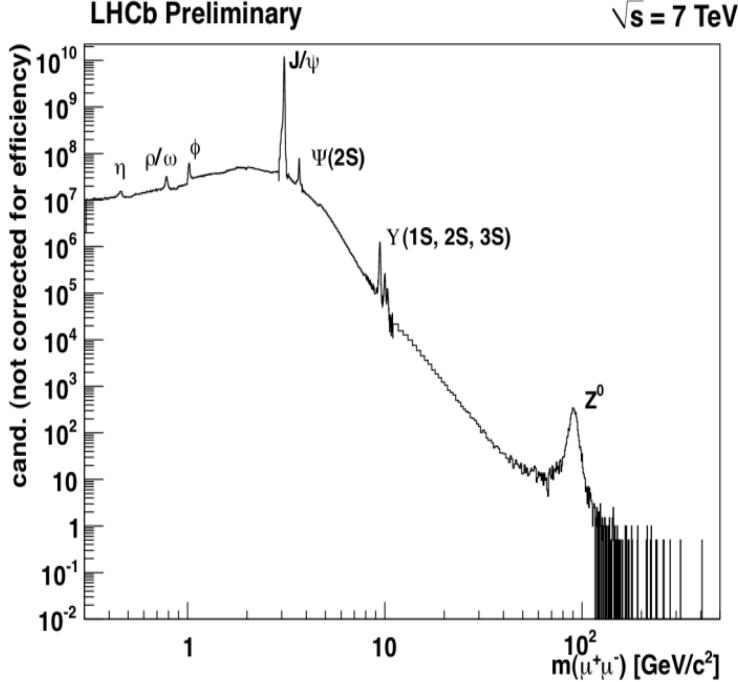


Figure 3.1: Dimuon mass spectrum.

A further offline selection suppresses more background, reducing the data set to manageable levels. This selection is detailed in the next section.

After this selection we are still left with residual background. The separation between signal and this residual background is accomplished with a multivariate classifier which combines several kinematic and geometrical properties of the event into a single output. This output can be converted into a probability for an event of being signal or background.

During the two first years of analysis, three different classifiers have been used to discriminate between signal and background: a combination of decorrelated likelihoods, used for the 2010 data [39], and two boosted decision tree classifiers [40] [4]. In order to use all data available and to avoid biases, these multivariate classifiers are trained using simulated events.

We obtain the number of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ from our data sample with a fit of a two-component (invariant mass of the muon pair and the multivariate classifier) functional. Given the integrated luminosity (L_{int}) and the total $b\bar{b}$ cross-section ($\sigma_{b\bar{b}}$), we translate the number of signal events ($N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$), extracted from our data sample, into a branching fraction:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \propto \frac{1}{L_{int} \times \sigma_{b\bar{b}}} \times N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}.$$

However, the normalization of the number of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ to the number of events for known decay allows to withdraw, from the previous formula, the terms L_{int} and $\sigma_{b\bar{b}}$, which

are not precisely measured:

$$\frac{\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(\text{norm})} = \text{Constant} \times \frac{N_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}}{N_{\text{norm}}}$$

We devote Sec. 7.1 to review the parameters entering in *Constant*.

In the absence of signal evidence, we set limits on $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ (see Sec. 7.3). The procedure followed in this case is:

- The two-dimensional plane defined by the multivariate classifier and the invariant mass is divided in bins.
- For each bin, the expected number of background events is estimated. Then, for a given branching fraction hypothesis, the compatibility between the observed and expected number of events is computed with the CL_s method [41].

We perform a so-called ‘blind analysis’. The events with invariant mass around (± 60 MeV/c 2) the B_s^0 and B^0 masses are excluded from the analysis until all the necessary quantities are ready.

3.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ background rejection

This section describes the main sources of background and the methods used for their removal.

3.2.1 Prompt background

Muons coming directly from the primary vertex (PV) or collision vertex, are the main constituents of what we call prompt background.

The B mesons produced at 7 TeV collisions travel, in average, about one centimeter before decaying. This relatively large decay length is due to the boost of the produced $b\bar{b}$ pair. This means that the tracks originated in a B decay are displaced from the PV. The decay vertex is called secondary vertex (SV).

The impact parameter (IP) of a track with respect to a certain primary vertex, is the distance between the track and the vertex at the point of closest approach. Hence, requiring a relatively large IP of the tracks with respect to the PV allows the removal of most prompt background. For illustration, Fig. 3.2 represents a $B_s^0 \rightarrow \mu^+ \mu^-$ candidate. This graphic shows how the projections of the muon tracks are displaced from the PV.

Removing the prompt background and other unphysical events constitutes the first step in our analysis. After the online selection performed at the trigger level, we reduce offline the data size (summarized in Tabs 3.1, 3.2) by enforcing well reconstructed tracks ($\text{track-}\chi^2/n.d.f$), with a minimum transverse momentum (p_T) and a minimum impact parameter χ^2 ($\text{IP}\chi^2$). In order to remove unphysical events, tracks are selected up to a maximum momentum (p), maximum transverse momentum, and by imposing the proper time of the

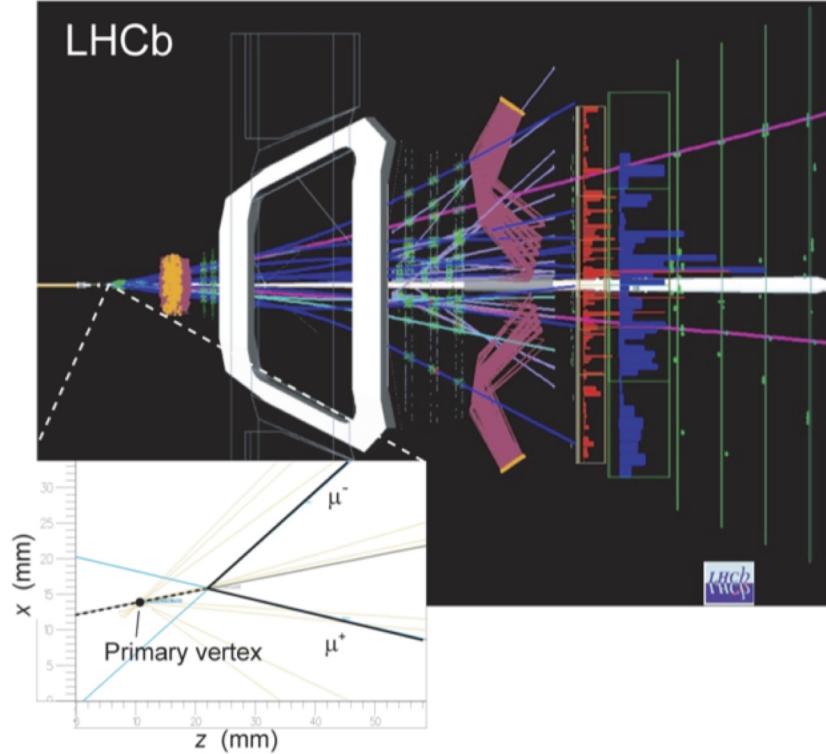


Figure 3.2: Event display of a $B_s^0 \rightarrow \mu^+ \mu^-$ candidate at the LHCb detector. The region close to the collision point is zoomed on the bottom left of the image.

B candidate (t) to be lower than nine times the lifetime of the B_s^0 . We select tracks with an invariant mass within a mass window around the B_s^0 mass. The distance of closest approach between the tracks (DOCA) and the χ^2 of the SV fit allow to suppress tracks that do not make a good vertex. The separation between primary and secondary vertices χ^2 (VDS) allows to select well displaced decay vertices. Finally, by requiring a transverse momentum of the B candidate higher than 500 MeV/c we reduce the exclusive dimuon production ($pp \rightarrow p\mu^+\mu^-p$). An important fraction of such processes have a dimuon invariant mass which is in the B_s^0/B^0 search region. Moreover, these events can have a large separation between primary and secondary vertices. The efficiency for signal of this selection criteria is $\sim 50\%$.

The criteria to select signal (and $B_{(s)}^0 \rightarrow h^+h'^-$) or $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$ and $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-)$ differ only by the selection of the extra kaon or kaons. DOCA and vertex χ^2 correspond to the J/ψ candidate for $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$ and $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-)$.

In Chap. 5 we present a more efficient selection method developed through this thesis work.

3.2.2 B hadron cascade decays

Another source of background are $B \rightarrow D(\rightarrow \mu X)\mu X$ decays. In these processes, a B meson decays semileptonically into a D meson, a muon and other decay products. Thereupon, the D meson decays semileptonically into another muon of opposite charge with respect to that of the previous muon. The geometrical features of these processes, so called cascade decays, can be similar to those of a $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay. For instance, the muon tracks are displaced from the primary vertex. Due to the unreconstructed decay products, the invariant mass of the dimuon pair of these decays is lower than the B_s^0 or B^0 masses. Nevertheless, we need to remove this background to properly extrapolate, using the sidebands of the invariant mass distribution, the amount of expected background in the signal region (see Sec. 6.1.1). We present a detailed classification of these processes in Sec. 4.4.

3.2.3 Peaking background

$B_{(s)}^0 \rightarrow h^+ h'^-$ decays in which each hadron decays into a muon represents the most dangerous type of peaking background. In order to reduce this background, we require tight particle identification requirements.

The invariant mass and the expected yields in our data sample of such processes are described in Chap. 6.

3.2.4 Combinatoric from semileptonic B decays

After applying the selection criteria described in the previous sections we are mostly left with another kind of background: randomly selected pairs of muons coming from different semileptonic B decays.

Table 3.1: Selection criteria for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h'^-$ channels.

	Variable	Requirement
		$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h'^-$
μ / h	track χ^2/ndf	< 4
	IP χ^2	> 25
	p_T	> 0.25 and < 40 GeV/c
	p	< 500 GeV/c
$B_{(s)}^0$	$ M_{hh} - M(B_{(s)}) $	< 600 MeV/c
	DOCA	< 0.3 mm
	vertex χ^2	< 9
	VDS	> 15
	IP χ^2	< 25
	t	$< 9 \cdot \tau(B_s^0)$
	p_T	> 500 MeV/c2

Table 3.2: Selection criteria for $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$ and $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$ normalization channels.

$B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$			$B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$		
	Variable	Requirement		Variable	Requirement
μ	track χ^2/ndf	<4	μ	track χ^2/ndf	<4
	IP χ^2	>25		IP χ^2	>25
J/ψ	DOCA	<0.3 mm		DOCA	<0.3 mm
	vertex χ^2	<9		vertex χ^2	<9
	Δm	<60 Mev		Δm	<60 Mev
	VDS	>15		VDS	>15
K^\pm	track χ^2/ndf	<5	K^\pm	track χ^2/ndf	<5
	IP χ^2	>25		IP $\chi^2 [K^+, K^-]$	[> 4, > 4]
			ϕ	Δm	<10 Mev
				IP χ^2	> 25
B^+	IP χ^2	<25	B_s^0	IP χ^2	<25
	Δm	<100 Mev		Δm	<100 Mev
	vertex χ^2	< 45		vertex χ^2	<75
	p_T	> 500 MeV/c2			

Monte Carlo methods are used to generate a sample of events that have two muons and a b -quark within the 400 mrad acceptance. We refer to this sample in the following as $b\bar{b} \rightarrow \mu\mu X$, where X stands for any other decay product. The total number of $b\bar{b} \rightarrow \mu\mu X$ events corresponds to $\sim 570 \text{ pb}^{-1}$ of integrated luminosity, assuming $\sigma_{b\bar{b}} = 288 \pm 4 \pm 48 \mu\text{b}$ [42] measured by LHCb.

Fig. 3.3 shows the distributions of several variables after applying the selection described in Sec. 3.2.1. We denote the impact parameter significance ($\frac{IP}{\sigma(IP)}$) as IPS. For each variable, we show the distributions of signal $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ background, both obtained from simulation, and data sidebands excluding the region [5000 MeV - 5418 MeV], to avoid the presence of signal and most of the peaking background. In Fig. 3.3, the variables on the top (left) are more discriminant than those placed on the bottom (right). Data sidebands and $b\bar{b} \rightarrow \mu\mu X$ simulation are in agreement with the exception of the p_T of the B candidate. Even though we optimize and train the multivariate classifier (see Chap. 5) using simulation, the final distribution of this classifier for background is extracted from data events in the invariant mass sidebands (see Sec. 6.1.1).

We devote the next two chapters to describe the multivariate techniques that combine some of these variables to improve the separation between signal and background.

3.2.5 Other exclusive background

We do not discuss in detail other sources of background, such as $B_c^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\mu^+\nu$, $B^+ \rightarrow \pi^+ \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^- \gamma$, which represent other types of peaking background.

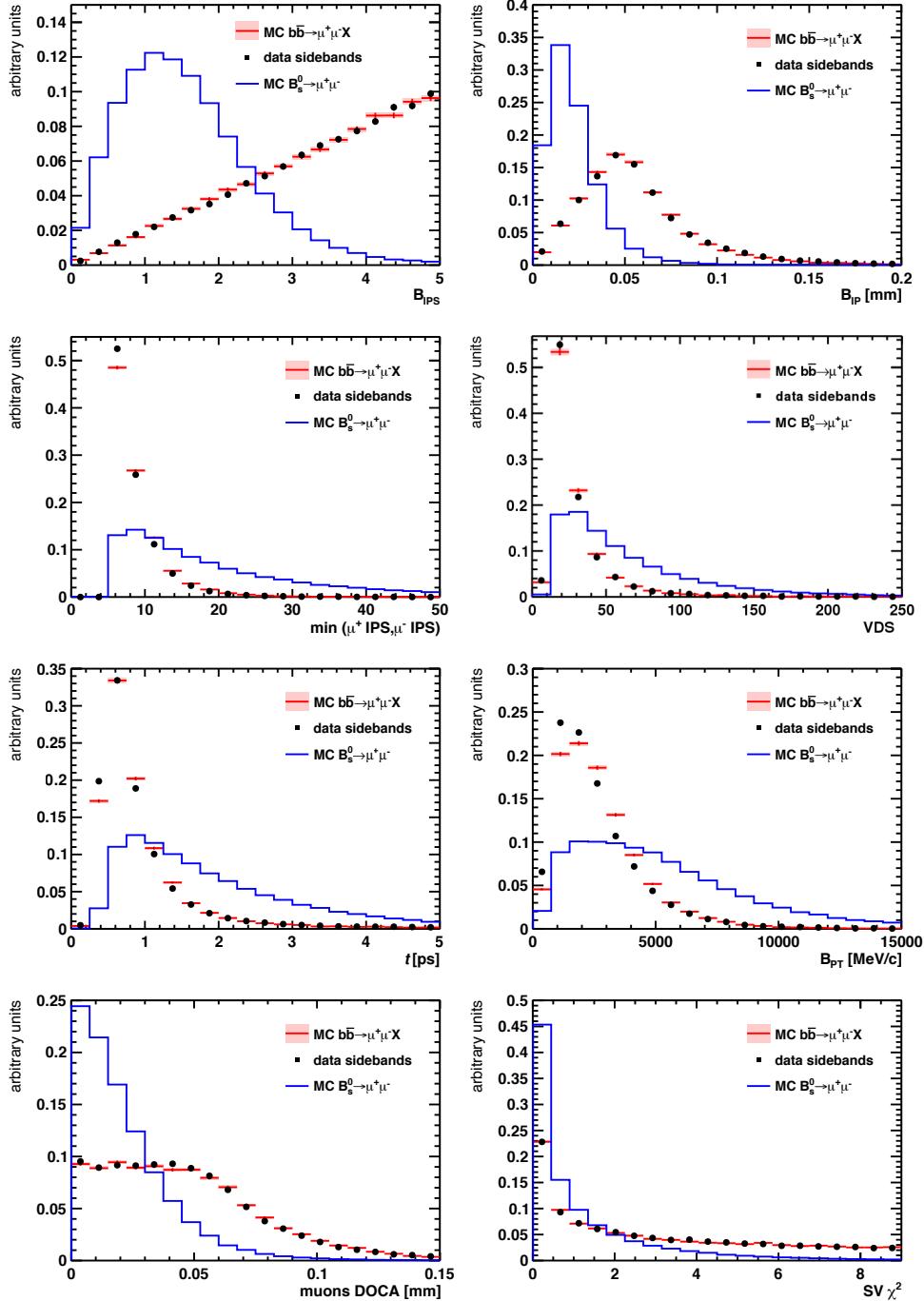


Figure 3.3: Distributions, for signal $B_s^0 \rightarrow \mu^+ \mu^-$ MC (blue line) and background $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC (red segments), of several geometric and kinematic Al variables. Overlaid, data sidebands with IsMuon requirement (black points).

Nevertheless, as Ref. [43] shows, their impact on the final conclusions of this document is negligible. Using simulation it has been estimated that the number of the first type of events per fb^{-1} is of 0.123 ± 0.073 in the B^0 region and 0.259 ± 0.141 in the B_s^0 region. For the second case, the expected number of events is 2.87 in the sensitive region of the analysis. The expected number of events of the third type, assuming a $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^- \gamma) = 3.2 \times 10^{-8}$, is smaller than 0.058 and 0.096 candidates in the B_s^0 and B^0 regions.

3.3 Conclusions

We have described some of the main steps of the analysis. Then we classified the different sources of background for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and presented some of the procedures adopted to reduce them. The separation between signal and combinatorial background from semileptonic B decays is accomplished with multivariate techniques, for which we refer the reader to Chaps 4 and 5. Furthermore, we shall review the cascade background in the next chapter.

Chapter 4

Validation studies using the Geometrical Likelihood classifier

The Geometrical Likelihood (GL) method combines several variables, related with the geometry of an event, into one single variable. In the analyses presented in Ref. [39], the GL was used to separate the signal from the combinatorial background. However, Chap. 5 presents a different multivariate technique that improves the GL's capability in discriminating the background from the signal. We dedicate this chapter to summarize the studies performed during 2010 using the GL. These results are used in the analysis presented in this thesis.

Sec. 4.1 introduces the method used to combine correlated variables, while Sec. 4.2 summarizes the method to extract the GL distribution from data. We devote Sec. 4.3 to study the effect of the global event cuts (GEC) (see Sec. 2) on the GL distribution. Finally, Sec. 4.4 describes the cascade background according to its characteristic GL.

4.1 Mathematical method to combine correlated variables

To exploit the background discrimination power of several variables it is useful to combine them. In this section we describe the method applied to combine correlated information.

Given a sample of events, N variables are transformed into Gaussian distributed variables (centered at zero and width equal to one) to easily compute probabilities and treat the correlations. Then, the new set of variables is rotated in order to be linearly independent. The new set of rotated variables is transformed into Gaussian distributed variables.

The procedure described is applied to a signal and a background samples as the correlations between the variables can be different. Thereupon, for each event we compute $\chi_s^2 = \sum s_i^2$ and $\chi_b^2 = \sum b_i^2$, where s_i and b_i represent the distance to the mean value of the ith Gaussian variable, for signal and background, respectively. The discriminant variable is defined as $\Delta\chi^2 = \chi_b^2 - \chi_s^2$.

Finally, this variable is transformed to be uniformly distributed between 0 and 1 for signal. Due to this transformation, the background peaks at zero.

Table 4.1: List of selection criteria for $D^0 \rightarrow K\pi$ decays.

$D^0 IP\chi^2 < 30$	$D^0 FD\chi^2 > 40$
$DOCA < 0.07 mm$	$DIRA > 0.999$
$\chi_{SV}^2 < 10$	$D^0 P > 5000 MeV$
$D P_T > 1500 MeV$	$K P_T > 1100 MeV$
$K P > 5000 MeV$	$DaughTrack\chi^2 < 3$
$K DLL(K - \pi) > 10$	

The variables included in the definition of the GL are: the IP and the decay time of the B candidate, the DOCA, the minimum IPS and the isolations of both muon candidates. The isolation of a muon candidate corresponds to the number of tracks displaced from the PV, that make a good vertex with the muon: the angle between the candidate and the track is lower than 0.27 rad, their DOCA is lower than 130 μm and they make a vertex whose distances to the primary and secondary vertices remain between $0.5 cm < d_{PV} < 4 cm$ and $-0.15 cm < d_{SV} < 30 cm$ in the z direction, respectively. The other muon candidate in the event is not considered in the computation. Hence, the isolation of the muons from $B_s^0 \rightarrow \mu^+ \mu^-$ or $B^0 \rightarrow \mu^+ \mu^-$ decays peaks at zero, while it has larger values for background (see Fig. 5.16).

4.2 GL calibration using the low mass resonance decay $D^0 \rightarrow K\pi$

In this section we derive the expected GL distribution for data $D^0 \rightarrow K\pi$ TIS events (see Sec. 2.4), to compare it with what obtained from MC $D^0 \rightarrow K\pi$ events. Ideally, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ distribution is calibrated with $B_{(s)}^0 \rightarrow h^+ h^-$ events triggered by hadronic triggered lines. Being these different from the muon lines, there is a bias due to the trigger different for both types of decays. Using TIS events allows to avoid this problem, at the price of a very low efficiency. At the beginning of the data taking period, the TIS $B_{(s)}^0 \rightarrow h^+ h^-$ statistics was extremely low that rendered $D^0 \rightarrow K\pi$ events an interesting decay to study the GL performances. Although having a much larger sample, these decays have different geometrical and kinematic properties than $B_s^0 \rightarrow \mu^+ \mu^-$ due to the lower mass and decay time of the D meson.

Samples of $D^0 \rightarrow K\pi$

The data sample that we employ corresponds to about $20 pb^{-1}$ of data collected during 2010. To select $D^0 \rightarrow K\pi$ decays we apply the requirements listed in Tab. 4.1. The direction angle (DIRA) is the angle between the reconstructed momentum of the meson candidate (D^0 in this case) and the vector defined by the PV and the SV. We require particle identification of the kaon.

We apply the same selection to simulated events. After the selection, we require the events to be TIS and triggered with TCK x0030 which is then emulated on the simulated

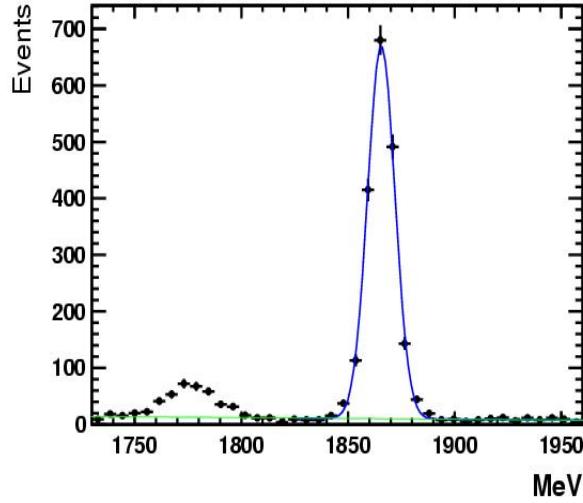


Figure 4.1: Invariant mass distribution for MC $D^0 \rightarrow K\pi$ triggered unbiased events selected using the requirements listed in Tab. 4.1.

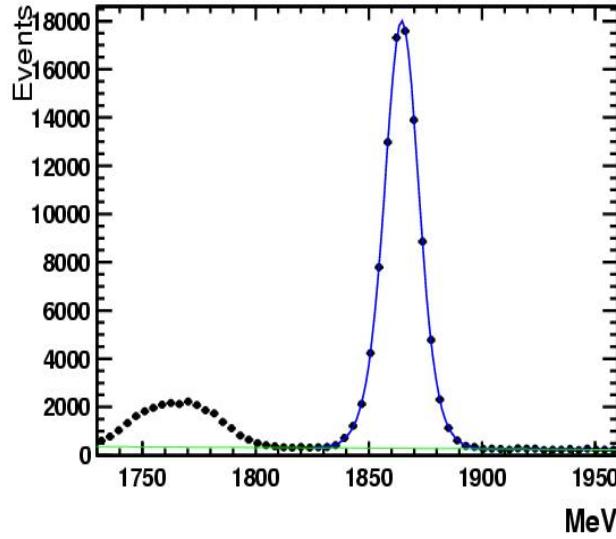


Figure 4.2: Invariant mass distribution for data $D^0 \rightarrow K\pi$ triggered unbiased events selected using the requirements listed in Tab. 4.1.

sample. Figs. 4.1 and Fig. 4.2 show the invariant mass distributions obtained for simulated and data events after this selection.

By fitting the invariant mass we estimate that the yield of signal events is 91300 ± 400 for the data sample.

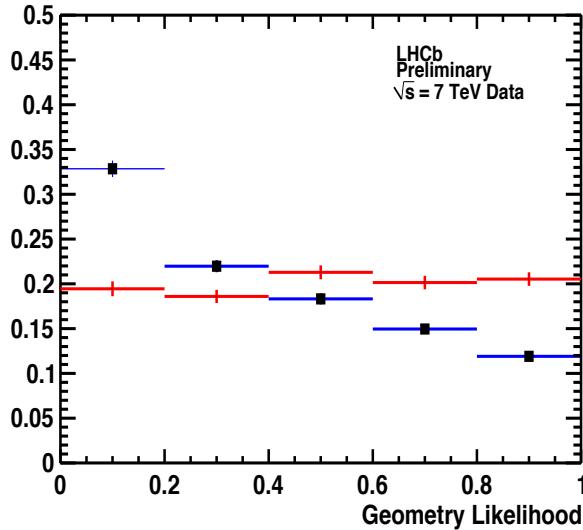


Figure 4.3: Signal GL distributions for simulation (red) and data (blue with black squares) obtained from $D^0 \rightarrow K\pi$ triggered unbiased decays.

Method to extract the GL distributions

The training is the procedure in which we compute the GL variable using a signal and a background samples.

The signal training sample corresponds to truth matched MC events, while as background sample we employ data from the higher sideband of the invariant mass distribution [1900-1960] MeV/ c^2 . We are forced to use the higher sideband in order to avoid the D physical background that can mimic the signal properties of $D^0 \rightarrow K\pi$. This type of background consists mainly of $D^0 \rightarrow K^+K^-$ with one of the kaons reconstructed as a pion. Consequently, these events present a lower invariant mass than $D^0 \rightarrow K\pi$. In Figs. 4.1 and 4.2, the D physical background are the bump in the lower sideband of the peak.

In order to extract the GL distribution for signal we perform a background subtraction method. First, the GL distribution is computed for a signal plus background sample (from events in the mass peak region), and then subtracted the background component (obtained from events in the invariant mass sidebands). To estimate the number of background events in the signal plus background sample, we fit the invariant mass distribution in the higher sideband range [1900-1960] MeV/ c^2 . The model used for this fit corresponds to a linear function (dotted green line in Figs. 4.1 and 4.2).

Fig. 4.3 shows the GL distributions which are flat for MC, hence no bias is present after background subtraction, while for data a downward slope as a function of GL is visible. Fig. 4.4 shows that the distributions obtained for background peak at zero as expected.

The reason for this discrepancy in signal is mainly due to the differences between data and simulation of two variables: the impact parameter of the D and the degree of isolation (see Fig. 4.6). To prove this we scale the IP distribution in MC in order to correspond with

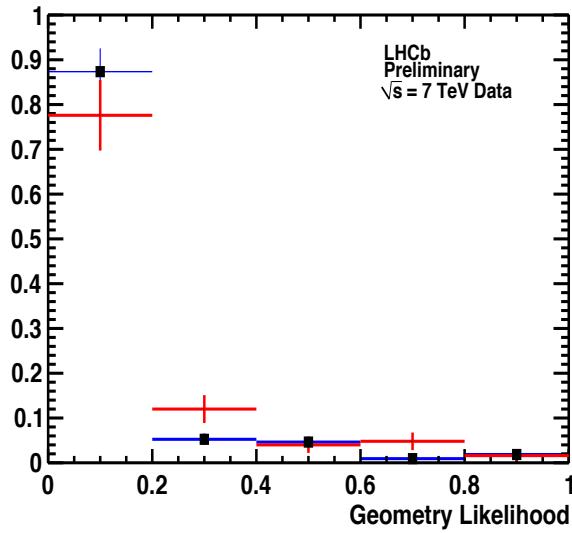


Figure 4.4: Background GL distributions for simulation (red) and data (blue with black squares) obtained from $D^0 \rightarrow K\pi$ triggered unbiased decays.

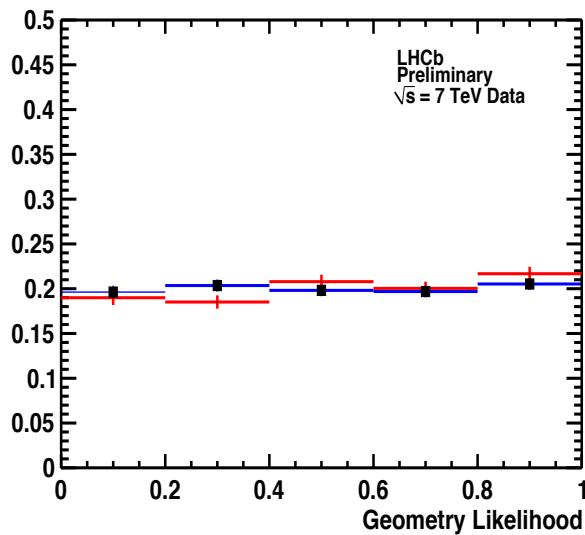


Figure 4.5: Signal GL distributions for simulation (red) and data (blue with black squares) obtained from $D^0 \rightarrow K\pi$ triggered unbiased decays after scaling the impact parameter of the D and removing the isolation.

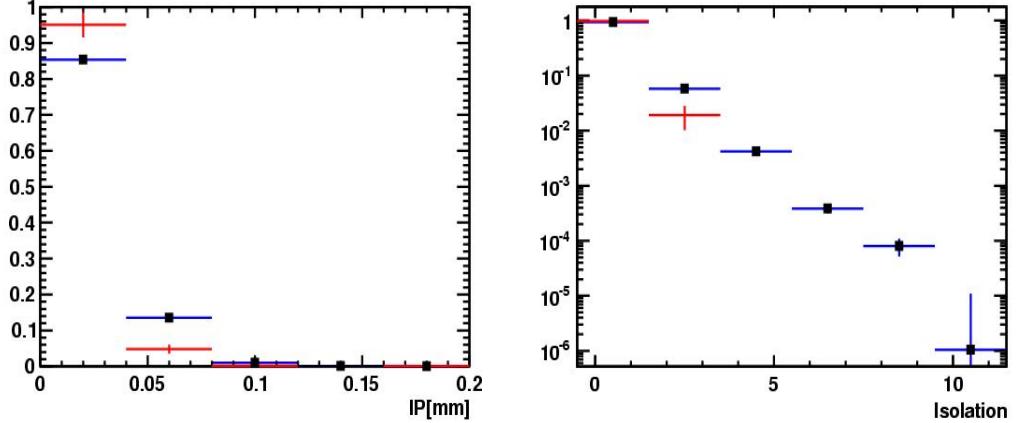


Figure 4.6: Data (blue, black squares) and simulation (red) comparisons for the distributions of impact parameter of the D^0 (left) and the isolation of the tracks (right). The plots are obtained using $D^0 \rightarrow K\pi$ TIS decays.

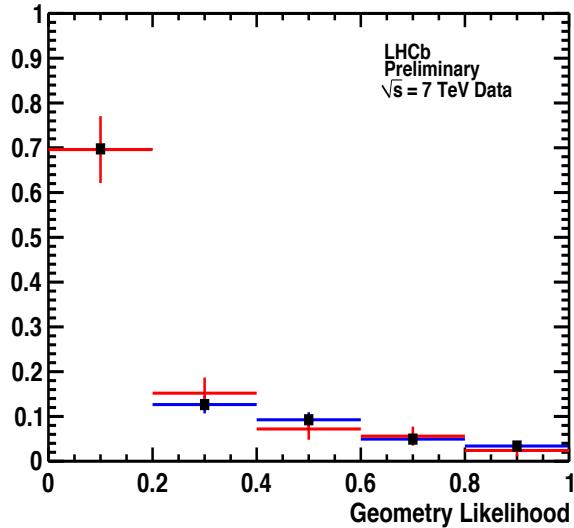


Figure 4.7: Background GL distributions for simulation (red) and data (blue with black squares) obtained from $D^0 \rightarrow K\pi$ triggered unbiased decays after scaling the impact parameter of the D and removing the isolation.

data, and remove the isolation from the input variables. Fig. 4.5 shows that the distributions for MC and data agree for signal, and Fig. 4.7 shows that the agreement between data and MC distributions for background improves after the aforementioned changes.

This study gave an indication of the GL shape expected for $B_s^0 \rightarrow \mu^+ \mu^-$ events extracted from $B_{(s)}^0 \rightarrow h^+ h^-$ TIS events. The discrepancy in the impact parameter between data and MC observed in $D^0 \rightarrow K\pi$ decays, and expected on two-body B decays, lead to the introduction of the smearing technique (see Sec. 5.1). The isolation variable in simulation does

Table 4.2: L0 requirements according to the TCK.

TCK	SPD hits
x002A	900
x002C	450

not properly describe data for $D^0 \rightarrow K\pi$ decays, nevertheless this variable was conceived for B decays. Sec. 5.1 shows a good agreement between data and simulation in B decays.

4.3 Estimation of the GEC impact on the GL distribution

In this section we study the effect of the global event cuts on the GL distribution.

During the 2010 data taking, the number of collisions per crossing was approximately two, while in nominal conditions it should be 0.6. In this data taking conditions, the L0 rate was too high. The adopted solution to reduce this rate relied on applying GEC to select events with lower multiplicity. Among these requirements, the maximum number of SPD hits at the L0 decision changed during the data taking. We aim to study the influence of the SPD requirement on the geometrical properties of the event by directly comparing the GL distributions of events triggered in different conditions.

The following study employs two samples triggered by different TCK configurations, which differ only by the number of SPD hits required: x002A and x002C (see Tab. 4.2)

Fig. 4.8 shows the GL distributions obtained for the two different TCKs. These distributions are obtained following the same procedure of background subtraction presented in the previous section. Their agreement reflects the negligible effect of the GEC. Furthermore, as the SPD requirements are modified just for the L0-hadron decision, we attempt also to check whether the hadronic TIS lines introduce any quantifiable bias on the GL distributions. Thus, we split both samples, one with events selected with TCK=x002A and the other with TCK=x002C, in events that are TIS for muonic and for hadronic lines. Fig. 4.9 shows the results obtained for TCK=x002A and Fig. 4.10 for TCK=x002C: in both the two distributions are in agreement which translates in lack of bias originating from SPD requirements on events that are TIS for hadron lines.

The outcome of this study allows to use of the whole data sample without any compulsory separation into samples where different GEC were applied.

4.4 Geometrical properties of B cascade decays

In Sec. 3.2.2 we described the B cascade background. Here we aim to study the geometrical properties of this type of events.

We use a sample of simulated events with at least one b -quark in the event (inclusive $b\bar{b}$ MC), which corresponds to about 130 pb^{-1} of integrated luminosity. Then, using MC truth matched events we require one of the muons to be a decay product of a B meson and

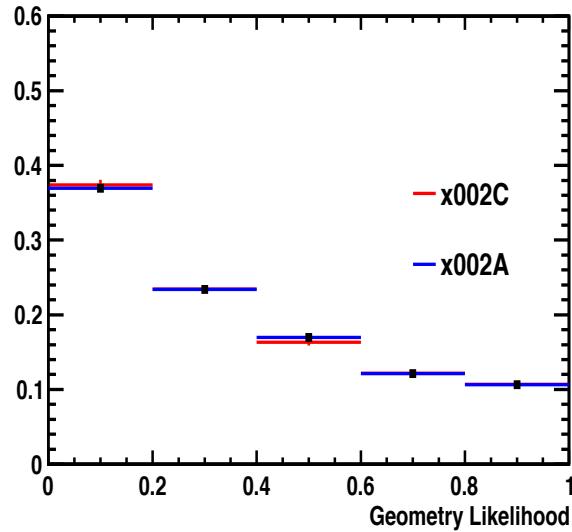


Figure 4.8: Signal GL distributions obtained from $D^0 \rightarrow K\pi$ triggered unbiased decays for events triggered with the configuration TCK=x002C (red) and TCK=x002A (blue with black squares).

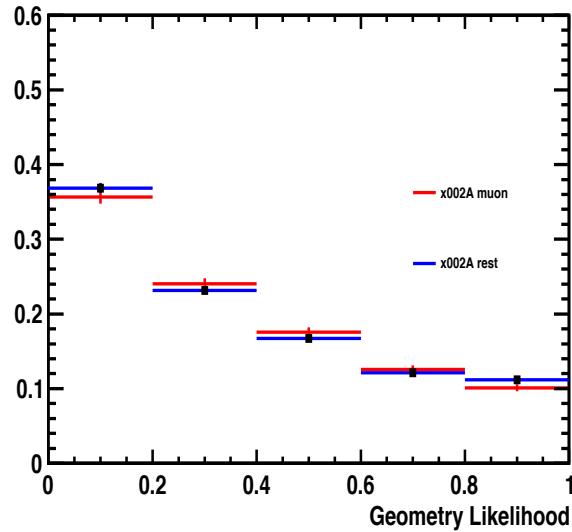


Figure 4.9: Signal GL distributions obtained from $D^0 \rightarrow K\pi$ events triggered with the configuration TCK=x002A that are triggered unbiased for muonic lines (red) and for the rest of lines (blue with black squares).

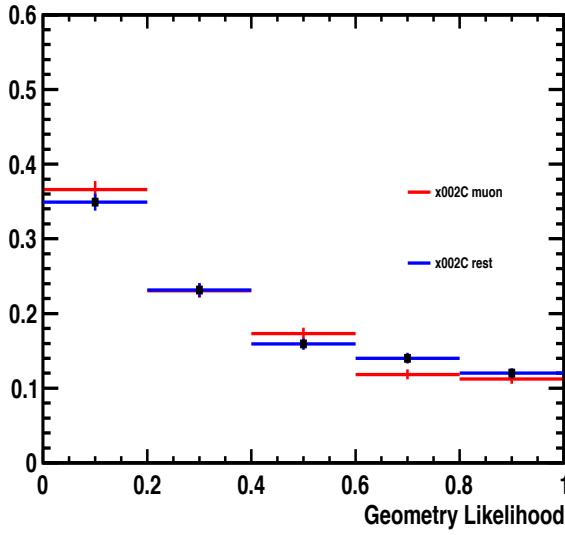


Figure 4.10: Signal GL distributions obtained from $D^0 \rightarrow K\pi$ events triggered with the configuration TCK=x002C that are triggered unbiased for muonic lines (red) and for the rest of lines (blue with black squares).

another muon to be the decay product of a D meson from the same B . Fig 4.11 shows the invariant mass and the GL (for GL values above 0.1) for this type of events. In the GL plot we display as well the distribution for the $b\bar{b}$ MC sample. In spite of the small number of events it is clear that most of the B cascade background has an invariant mass lower than 4900 MeV. Furthermore, the GL distribution of such events has signal-like behavior, as the events spread along the x axis. From a total yield of 19 $B \rightarrow D(\rightarrow \mu X)\mu X$ events, 10 populate the GL region above 0.1. The fraction of $B \rightarrow D(\rightarrow \mu X)\mu X$ events in this region is much larger than the 6% computed for $b\bar{b}$ background.

In order to remove most of this background, we require the dimuon invariant mass to be higher than $4900 \text{ MeV}/c^2$ in the next steps of the analysis.

4.5 Conclusions

This chapter describes the method used to separate signal from combinatorial background for the analysis published in Ref. [39]. The classifier employed in this publication includes a non-geometrical discriminant variable, the p_T of the B candidate. In the next chapter we shall describe the improvements achieved to increase the rejection power of such classifier.

Using the low mass resonance decay $D^0 \rightarrow K\pi$, we have found a discrepancy in the impact parameter between data and MC that lead to the introduction of a track smearing technique (see Sec. 5.1). Furthermore, we have characterized the cascade decays in terms of their geometrical properties and established a method to suppress them. Finally, we have used the GL classifier to study the possible effects of the GEC applied at the L0 trigger level. The outcome of these studies proved a lack of bias on the geometrical properties of signal

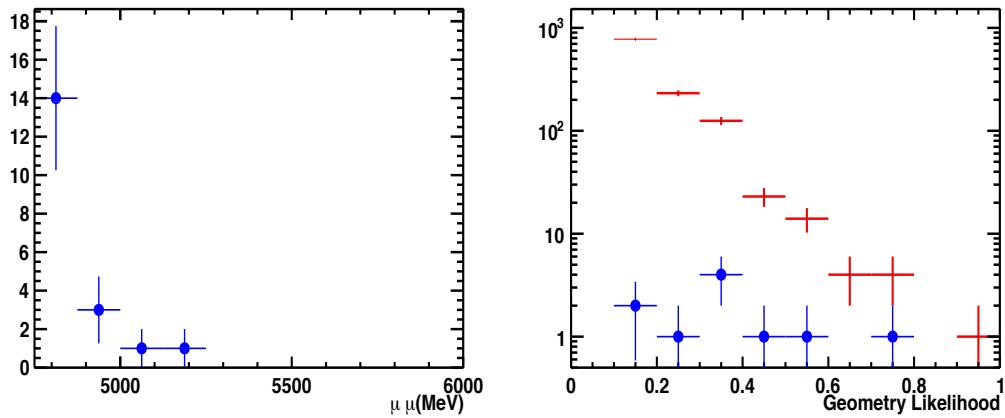


Figure 4.11: Invariant mass of the dimuon pairs from $B \rightarrow D(\rightarrow \mu X)\mu X$ decays (left). On the GL distribution (right) the blue circles represent $B \rightarrow D(\rightarrow \mu X)\mu X$ decays while in red we show $b\bar{b}$ background distribution.

events due to different SPD requirements at the L0 level.

Chapter 5

Optimization of the discriminant classifier for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$

This chapter presents a new multivariate classifier: the boosted decision trees (BDT). Firstly, Sec. 5.1 shows a comparison of data and simulation signal distributions for some of the variables used to define the multivariate classifiers presented in this thesis. Sec. 5.2 introduces the idea behind decision trees. Then, in Sec. 5.3, we describe the implementation of this type of classifier in the selection process of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$. We devote Secs. 5.4 and 5.5 to identify the discriminant classifier used to separate signal from combinatorial background. We shall compare the performances of the GL with those obtained with BDT. Finally, Sec. 5.6 details the final step of the optimization and presents the final classifier for the analysis.

5.1 Extraction of the signal distributions for some variables using data events

As shown in the last chapter, the combinatorial background in data and the simulated $b\bar{b} \rightarrow \mu\mu X$ background are in agreement. All the multivariate classifiers described in this thesis are defined using simulated events. In order to guarantee that the performances of such classifiers hold after calibrating them with data, we verify whether the simulated signal also describes the data signal.

Ideally we should compare simulated $B_s^0 \rightarrow \mu^+ \mu^-$ with $B_{(s)}^0 \rightarrow h^+ h^-$ TIS data events, as the hadronic trigger biases the distributions. Nevertheless, this approach is not feasible given the small available statistics for data TIS events. In order to compare data and MC selected by the same trigger lines, we use $B^0 \rightarrow K\pi$ simulated decays as a proxy of $B_s^0 \rightarrow \mu^+ \mu^-$, and assume that the geometrical and kinematic properties of such decays are compatible. The distributions for each variable are extracted using a fit of the inclusive $B_{(s)}^0 \rightarrow h^+ h^-$ data sample. Then, each distribution is compared with simulated $B^0 \rightarrow K\pi$ decays. Both data and MC events are required to be triggered by the same TCK (x0032) and to have a selection based on a classifier described in Sec. 5.3.

The $B_{(s)}^0 \rightarrow h^+ h^-$ data has a pion-pion mass hypothesis and the model used to fit the invariant mass distribution is described in Ref. [43]. Here we highlight the main features of

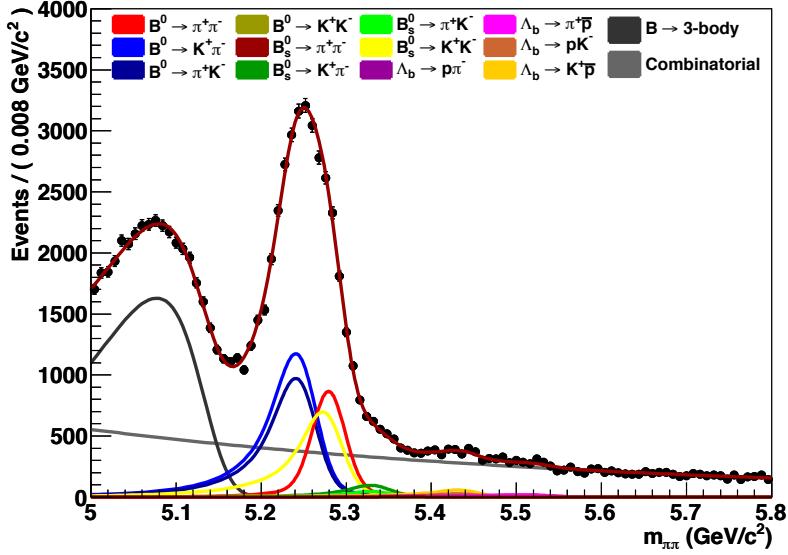


Figure 5.1: Invariant mass distribution of $B_{(s)}^0 \rightarrow h^+h'^-$ simulated events in the $\pi\pi$ mass hypothesis extracted from a b inclusive sample.

such a model:

- The signal $B_{(s)}^0 \rightarrow h^+h'^-$ component is described by means of a Crystal Ball (CB) function. This parameterization is broadly used in HEP and comprises a core Gaussian portion that describes the detector response and a power-law low-end tail that accounts for the presence of radiative decays. The Crystal Ball function notes as:

$$\mathcal{S}(x; \mu, \sigma, \alpha, n) = \begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & \frac{x-\mu}{\sigma} > -\alpha \\ f(x; \mu, \sigma, \alpha, n)^{-\alpha}, & \frac{x-\mu}{\sigma} \leq -\alpha, \end{cases}$$

where $f(x; \mu, \sigma, \alpha, n) = \left(\frac{n}{|\alpha|}\right)^n e^{-\frac{|x|}{2}} \left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\mu}{\sigma}\right)$. The parameters α and n are extracted from the simulation, while the mean and the resolution of the Gaussian are extracted from data. This CB function has no physical meaning as the sample we intend to fit is a superposition of 4 exclusive modes with a wrong mass hypothesis. Nevertheless, it is a good approximation as shown in Fig. 5.1 for simulated events.

- We parameterize Λ_b decays, which peak at a higher mass than the B mesons, with a Gaussian fixing its yield with respect to the B meson yield. The fraction of produced B mesons with respect to Λ_b baryons is 0.9469, according to the PDG.
- One of the background components is the physical background which consists of partially reconstructed three-body B meson decays, like $B^0 \rightarrow \rho^\pm \pi^\mp$. The reconstructed $\pi^+\pi^-$ invariant mass fall in the lower region of the $B_{(s)}^0 \rightarrow h^+h'^-$ invariant mass distribution due to the missing π^0 . The model used to describe such background, a modified Argus function, is described in Ref. [43].

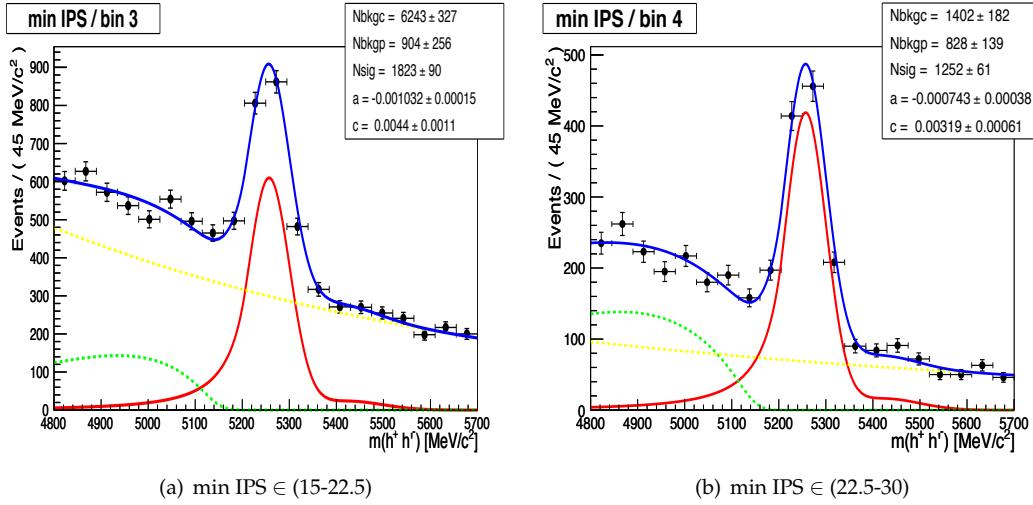


Figure 5.2: $B_{(s)}^0 \rightarrow h^+ h^-$ events with smallest IPS of the muons (min IPS) [15-22.5] (a) and [22.5-30] (b) fit with the model described in Sec. 5.1

- Finally, the last component of the fit describes the combinatorial background for $B_{(s)}^0 \rightarrow h^+ h^-$ decays, that we model with an exponential function.

As an example, we detail the extraction of the distribution of the smallest impact parameter significance of the tracks (min IPS): the $B_{(s)}^0 \rightarrow h^+ h'^-$ sample is divided into 10 different sub-samples according to an equally spaced binned min IPS distribution between 0 and 75. Then, each sub-sample is fit using the model previously described. Fig. 5.2 displays the result of the fit for the third and fourth bin. The parameters describing the signal are fixed when fitting the distributions in each bin. In the upper right box we identify the obtained yields for the combinatorial background (N_{bkgc}), the physical background (N_{bkgp}) and the signal plus Λ decays (N_{sig}). These components correspond to the dotted yellow, dotted blue and red distributions respectively. We show the rest of fits performed for the other bins of min IPS in Fig. 5.3.

For each one of the bins we extract the number of $B_{(s)}^0 \rightarrow h^+ h'^-$ decays and compare it with the yields of $B^0 \rightarrow K\pi$ events in each bin. Fig. 5.4 shows the agreement between data and simulation for min IPS. These simulated events have been modified using the smearing technique that we review in the following. The uncertainties displayed in this figure are the uncertainties associated with the fit method.

Smearing of the track parameters

The smearing technique aims to account for observed differences between data and simulation impact parameter resolution by manipulating the track parameters so that the IP resolutions in x and y in simulation correspond with data.

Fig. 5.5 shows the IP distribution obtained with the procedure described in the previous section compared with $B^0 \rightarrow K\pi$ simulated events with no smearing applied (a); and after

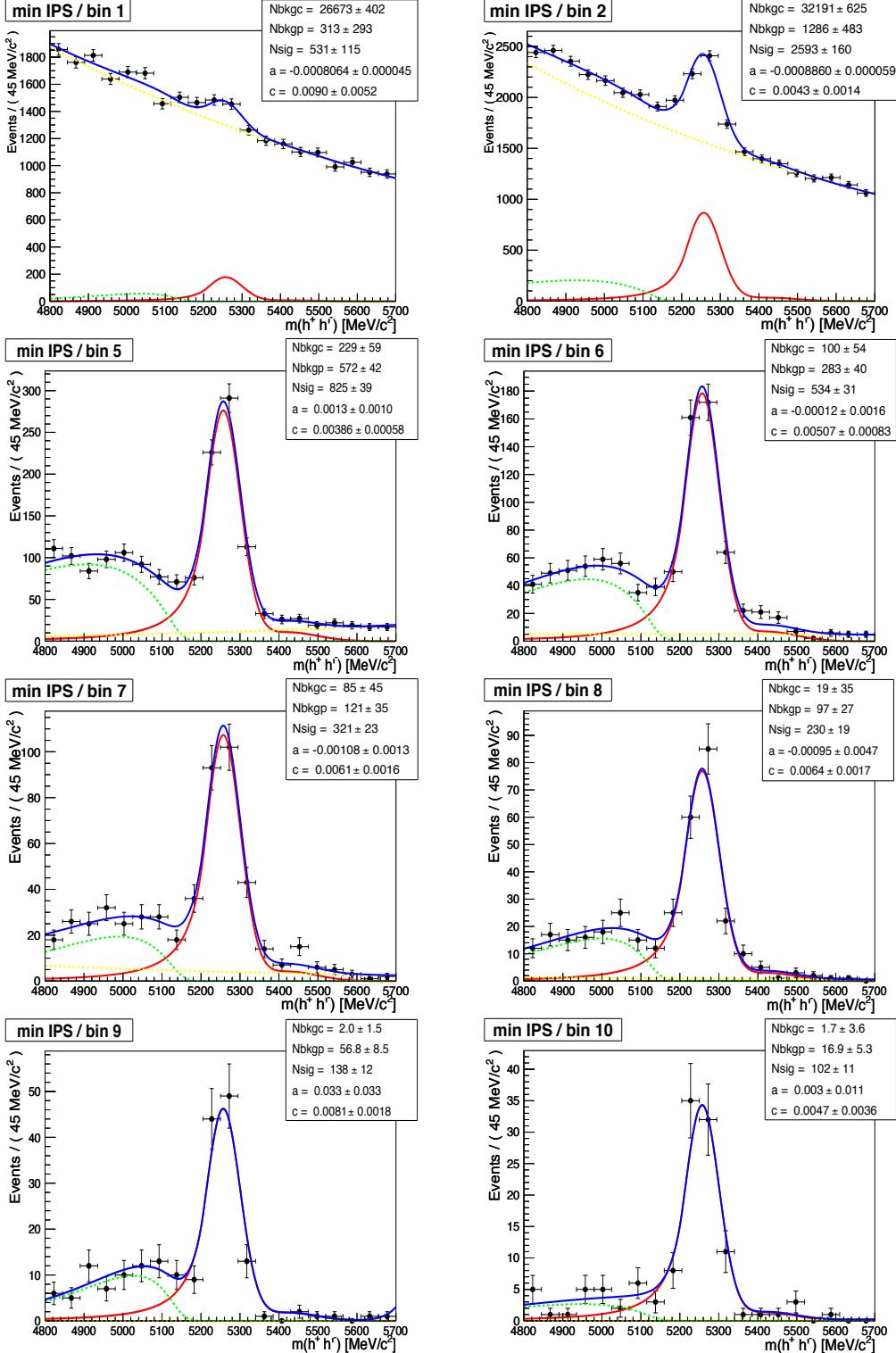


Figure 5.3: $B_{(s)}^0 \rightarrow h^+ h^-$ events, in bins of min IPS of the muons, fit with the model described in Sec. 5.1

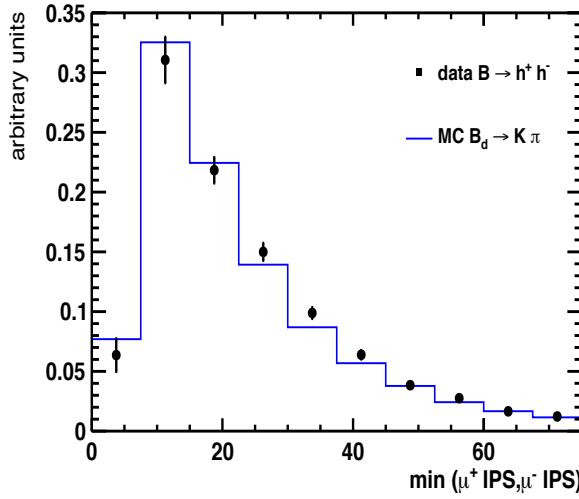


Figure 5.4: A comparison of the min IP of the tracks for $B^0 \rightarrow K\pi$ simulated events (blue line) and data $B_{(s)}^0 \rightarrow h^+h^-$ events (black points).

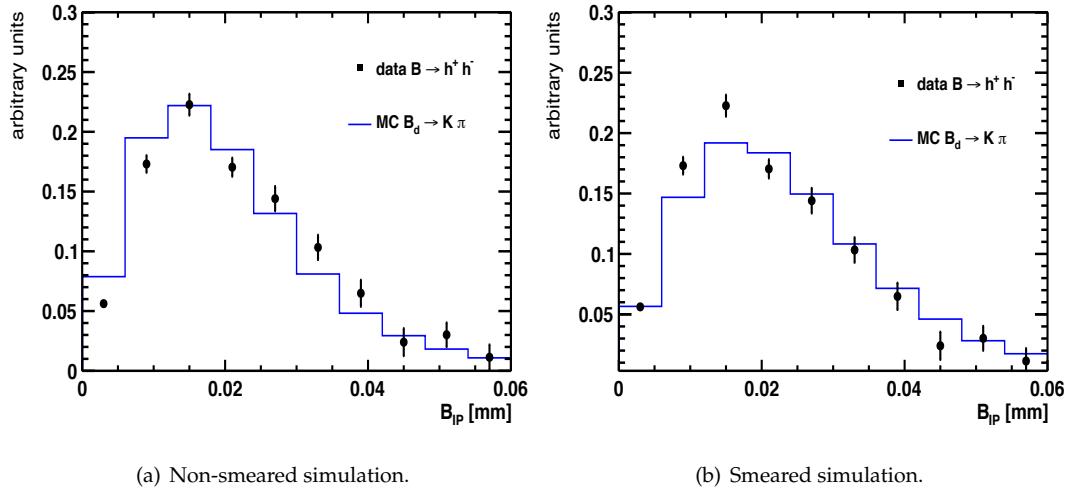


Figure 5.5: Comparisons of data and simulation before the smearing of the track parameters (left) and after (right).

applying the smearing (b). The agreement between distributions is more visible in (b), reflecting the good performance of the smearing procedure.

The same approach is used to extract the rest of the fits of some other variables used to define the classifier. Figs. 5.6 shows that data and simulation signals are in relatively good agreement. The uncertainties in all figures are the errors of the signal yield obtained with the fit.

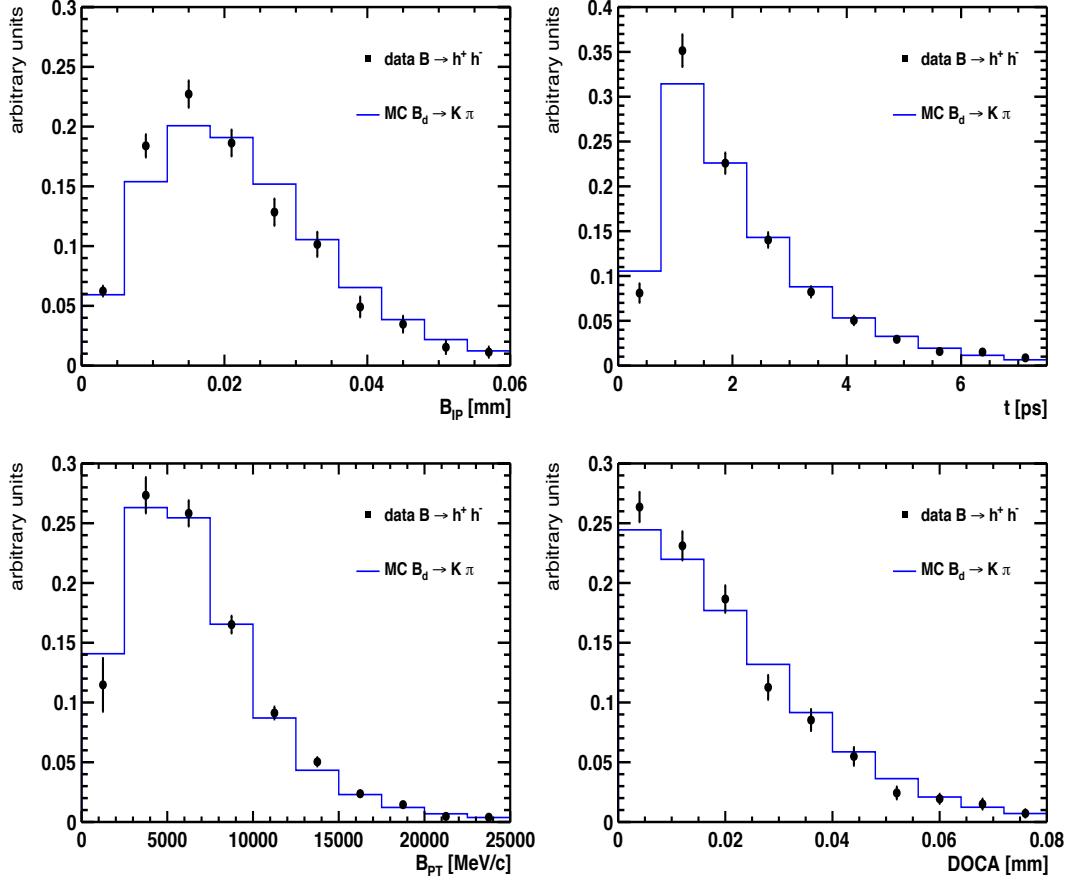


Figure 5.6: Comparisons of four variables that define classifiers presented in this document for $B^0 \rightarrow K\pi$ simulated events (blue line) and data $B_{(s)}^0 \rightarrow h^+ h'^-$ events (black points).

5.2 Introduction to boosted decision trees

A decision tree is a machine learning technique developed in the context of data mining and pattern recognition [44]. Its basic principle consists in extending a cut-based analysis into a multivariate technique by consecutively analyzing a sample of events that fail a certain hypothesis. The concept behind decision trees is not to reject events that miscarry a criterion but instead to explore other criteria that may properly classify such events.

The analysis presented in this thesis is based on binary trees in which each node or data sample recursively splits into two branches: enriched in signal or in background. The main steps of a decision tree algorithm are:

1. Consider a sample composed of signal and background. Fig. 5.7 represents an example of tree in which the starting sample contains 35 signal and 65 background events.
2. Start with a set of n variables to classify the events. In our example we use α, β, γ .
3. Choose the most discriminating variable between signal and background (α in Fig. 5.7) and set a certain splitting criterion. This requirement separates the original node into

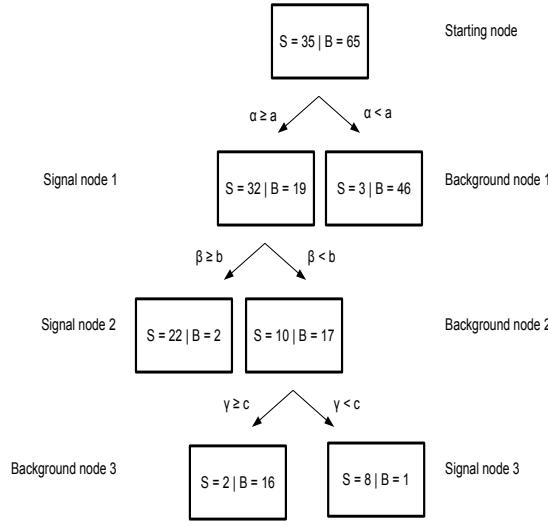


Figure 5.7: A decision tree diagram example.

two nodes: one whose events fulfill $\alpha < a$ and contains a large number of background events (background node 1); and another in which the events have $\alpha \geq a$. We denote this last node as ‘signal node 1’ as it contains a large number of signal events.

4. If the signal purity (number of signal events divided by the total events) in a node can be improved, split previous node again. Based on the variable that discriminates better the signal from the background left in the node, two new nodes are generated through a splitting of signal node 1. Background node 1 remains as it is since no further division leads to an increase of signal purity.
5. Once the tree is created (training process), the output of the tree can be evaluated for another set of events (testing process). The output for such set of events will depend on how its variables behave in the tree.

Furthermore, the algorithm can impose several constraints on the tree’s characteristics. For instance, to assure a statistical significance of the purity measurement in each node, we can require a minimum number of events in each leaf (node in a tree). Also, to avoid large computing time one can require a maximum number of layers. This last parameter is strongly related with the concept of boosting described later.

A drawback inherent to the decision tree classifier is that a small change in the training sample can lead to large differences in the tree structure. Moreover, the output of the tree is discrete as it depends on the purities of signal and background in each node. This demands an increase of the tree size which reduces the significance of the purity measurement. Furthermore, decision trees have an inherent problem related with the number of misclassified events. These are background events misclassified in the final signal nodes

and signal events misclassified in the background nodes. For instance, in Fig. 5.7, signal node 2 contains two misclassified events, while background node 1 contains three misclassified signal events.

For this reasons, the idea of boosting emerged to improve the performances of decision trees (and other types of classifiers). It relies on the combination of different decision trees into a new one, more stable and with improved performances in terms of significance.

The boosting technique used for the work performed in this document, AdaBoost, uses the misclassification rate in order to assign a certain weight to a given tree. The misclassified events in the tree T_k are weighted with a factor e^{α_k} in the next tree T_{k+1} . Consequently, the factor $\alpha_k = \beta \ln \frac{1-\epsilon_k}{\epsilon_k}$ forces the algorithm to concentrate harder on the misclassified events. β is the boosting parameter and ϵ_k the misclassification rate.

The final output of the boosted decision tree corresponds to a weighted average of the single responses of each of the trees (N_{trees}):

$$T(i) = \frac{1}{\sum_{k=1}^{N_{trees}} \alpha_k} \sum_{k=1}^{N_{trees}} \alpha_k T_k(i). \quad (5.1)$$

All the boosted decision trees classifiers reported in this thesis have been constructed using the TMVA package [45].

5.3 Implementation of a selection BDT classifier (BDTS)

Sec. 3.1 mentions that $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h'^-$ decays have the same selection criteria. We do not require particle identification to the tracks, collecting $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B_{(s)}^0 \rightarrow h^+ h'^-$ candidates at the same time. This translates into a large event data sample as the background for $B_{(s)}^0 \rightarrow h^+ h'^-$ is abundant.

We shall illustrate this by requiring the same selection as described in 3.2 to a sample of about 800 pb^{-1} . Fig. 5.8 shows the $B_{(s)}^0 \rightarrow h^+ h'^-$ distributions for all the events and those that are trigger unbiased (TIS). Using these types of events is crucial for the calibration of the final discriminant classifier. The previous figure reflects the large amount of background in both cases and the negligible signal over background ratio for TIS events in the region of invariant mass of $[5100-5400] \text{ MeV}/c^2$.

In this section we review the strategy to accomplish an efficient reduction of the $B_{(s)}^0 \rightarrow h^+ h'^-$ background using the boosted decision trees technique.

The definition of the BDT for the selection starts with an optimization of the variables used, following two steps: the training process, in which the output of the BDT is generated, and the testing process, in which its performances are quantified. Both steps rely on the utilization of different samples: $B_s^0 \rightarrow \mu^+ \mu^-$ as signal and 6% of the events in the sidebands of the $B_{(s)}^0 \rightarrow h^+ h'^-$ invariant mass distributions as background. The ranges that define these sidebands are $[4800-5000]$ and $[5500-6000] \text{ MeV}/c$ to avoid $B_{(s)}^0 \rightarrow h^+ h'^-$ decays.

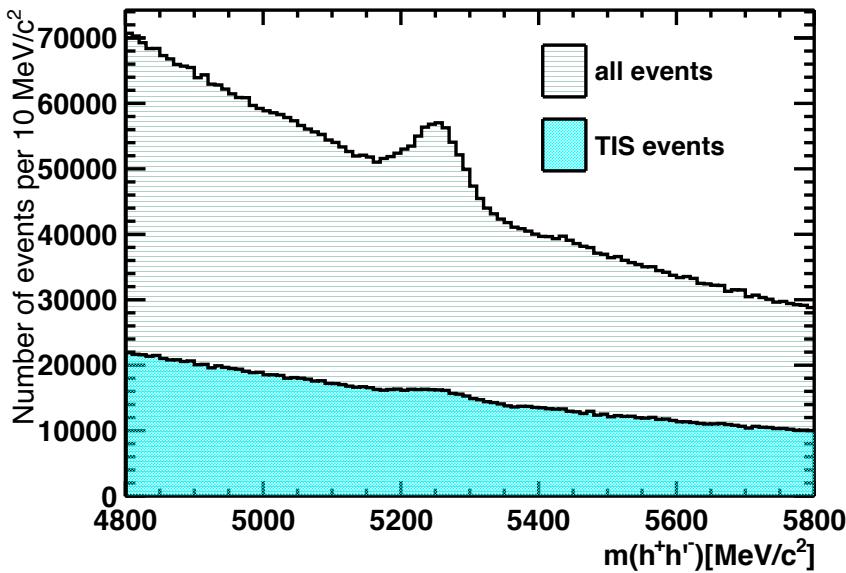


Figure 5.8: Invariant mass distribution of $B_{(s)}^0 \rightarrow h^+h'^-$ events obtained with about 800 pb^{-1} .

5.3.1 Definition of the BDT for the selection

The optimization process of the BDT starts considering nine variables, commonly used in many offline selections: the separation between primary and secondary vertices (VD) and its significance (VDS), B IP, B IPS, DOCA, the direction angle (dira), $\chi^2(\text{SV})$, min IPS and min IP.

Firstly, the background rejection versus the signal efficiency curve is computed using the set of nine variables. Then, one variable at the time is removed from the original set and for each new set we compute the background rejection versus signal efficiency. At this point, we compare the receiver operating characteristic (ROC) curves obtained with each set of variables. Being interested in keeping a high signal efficiency, we choose the set giving the best performance for signal efficiencies above 90%. Only as pedagogical purposes, Fig. 5.9 shows the ROC curves of four different BDT trained with different sets of variables. In this figure, the black and the red curves show the best performances above 90% signal efficiency. For this example, we imagine that the black curve was obtained with all the original variables, and the red curve with all the variables removing min IPS. Hence, as the performances are very similar, and the latter curve has less variables we would keep the variables that define this BDT as a starting set for a further step.

The process of removing a variable at a time is repeated until a decrease in performance observed between the new set and the original one is larger than 5-6%. At the end of the optimization, the final set of variables is: BIP, BIPS, DOCA, DIRA, χ^2 (SV), minIP.

Fig. 5.10 represents the receiver operating characteristic (ROC) curve for the defined BDT, hereafter called BDTS, which shows the signal efficiency versus the background rejec-

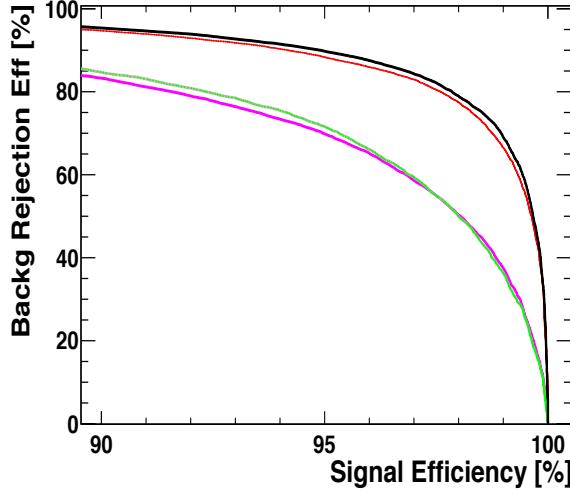


Figure 5.9: Background rejection versus signal efficiency for four hypothetical BDT computed with different input variables.

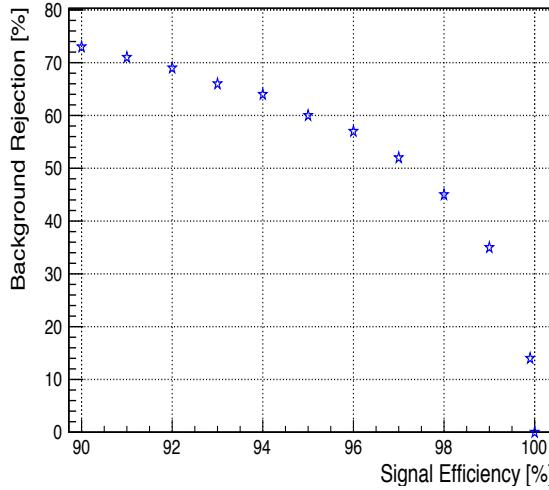


Figure 5.10: Background rejection versus signal efficiency of a BDTS-based selection computed with simulated $B_s^0 \rightarrow \mu^+ \mu^-$ as signal and $B_{(s)}^0 \rightarrow h^+ h^-$ data sidebands as background events.

tion. The uncertainties in both background rejection and signal efficiency are smaller than 0.1%.

In order to stress the discriminant power of this classifier we compare it with a different approach consisting in a tighter selection than in Sec. 3.2.1. This tighter selection, developed initially in our working group, has the following requirements: DOCA<0.08, DIRA>0.9997, min IPS>6 and B IP<0.08, which translated in 50% of background rejection

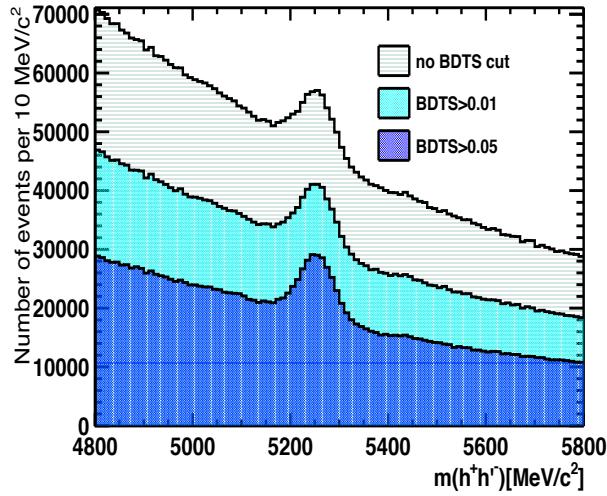


Figure 5.11: $B_{(s)}^0 \rightarrow h^+h^-$ invariant mass distributions with different BDTS requirements.

at a signal efficiency of 92%. The BDTS rejects 40% more background for the same signal efficiency.

As the GL, the BDTS output for signal $B_s^0 \rightarrow \mu^+\mu^-$ is constructed to be uniformly distributed between 0 and 1 and peaking at 0 for background. Fig. 5.11 shows the effect of different BDTS requirements on the $B_{(s)}^0 \rightarrow h^+h^-$ sample mentioned at the beginning of this section. We can see how the signal over background ratio in the mass window improves from 0.08 for no requirement to 0.19 for BDTS>0.05, which corresponds to a signal efficiency of 95% for $B_s^0 \rightarrow \mu^+\mu^-$. As shown in Fig. 5.10 the background rejection that corresponds to this selection criterion is of about 60% Another important aspect, shown in Fig. 5.12, is that the ratio signal over background is highly enhanced for the TIS sample.

5.3.2 Effect of the BDTS selection on combinatorial background

Fig. 5.13 presents the distributions for smeared (see Sec. 5.1) signal $B_s^0 \rightarrow \mu^+\mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ simulated events, compared with data sidebands (with Ismuon requirement). The signal distribution for signal is not flat, as the training is performed with non-smeared MC. The fact that the $B_s^0 \rightarrow \mu^+\mu^-$ smeared events peak more at zero than the non-smeared, flat by construction, reflects that the former are more background-like than the latter sample. Data and simulation backgrounds distributions are in excellent agreement.

The BDTS $b\bar{b} \rightarrow \mu\mu X$ background rejection, visible in Fig. 5.14, is 51.9 ± 0.3 % and 76.4 ± 0.2 % for BDTS>0.01 and BDTS>0.05, respectively. Fig. 5.15 demonstrates that the correlation between the BDTS output and the $\mu\mu$ invariant mass is negligible.

5.3.3 Effect of the BDTS selection on normalization channels

Using smeared simulated events, the efficiencies computed for different BDTS selection requirements on $B_s^0 \rightarrow \mu^+\mu^-$, $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$, $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-)\phi(\rightarrow K^+K^-)$,

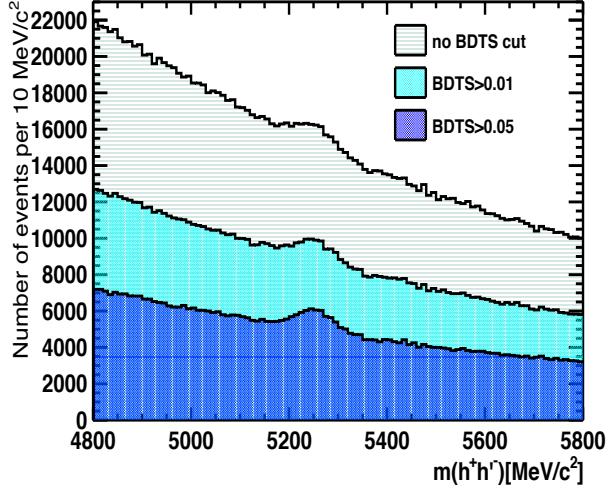


Figure 5.12: $B_{(s)}^0 \rightarrow h^+ h'^-$ TIS invariant mass distributions with different BDTS requirements.

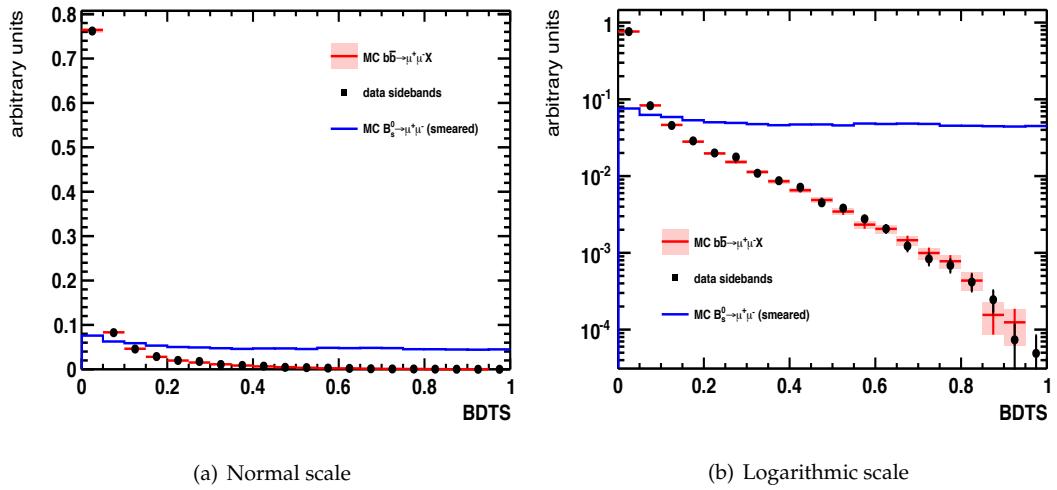


Figure 5.13: BDTS distributions for smeared $B_s^0 \rightarrow \mu^+ \mu^-$ and $b\bar{b} \rightarrow \mu\mu X$ simulated events, and data events from the invariant mass sidebands.

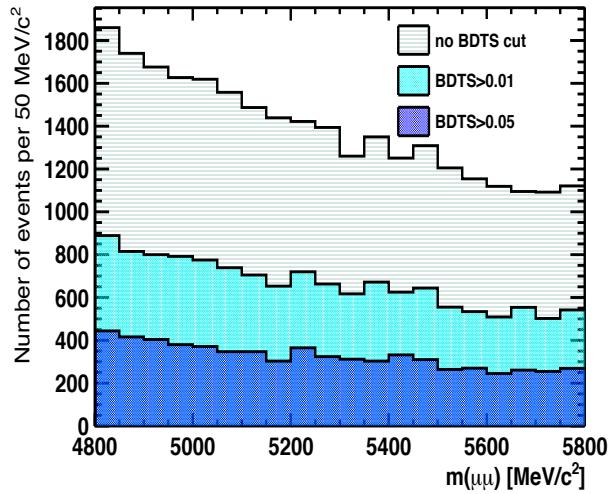


Figure 5.14: $b\bar{b} \rightarrow \mu\mu X$ invariant mass with different BDTS selection requirements.

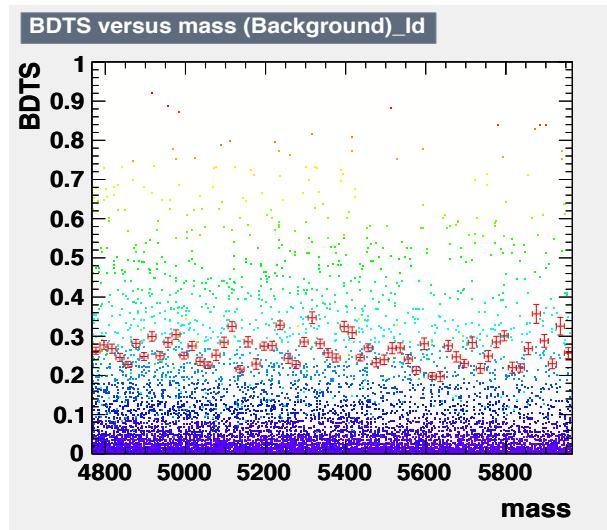


Figure 5.15: BDTS output versus the invariant mass for simulated $b\bar{b} \rightarrow \mu\mu X$ events. The red crosses are the average profile between both variables.

Table 5.1: Efficiencies for different BDTS selection requirements for $B_s^0 \rightarrow \mu^+ \mu^-$ (a), $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$ (b) and $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$ (c) smeared Monte Carlo samples. The efficiency ratios between $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$, $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$, $B^0 \rightarrow K\pi$ (d) and $B_s^0 \rightarrow \mu^+ \mu^-$ are reported in the third, fifth and seventh row respectively, with their uncertainties quoted in parenthesis.

BDTS requirement	0.01	0.03	0.05	0.08	0.1
$\epsilon_{(a)}$	0.9828	0.9501	0.9210	0.8815	0.8570
$\epsilon_{(b)}$ ratio	0.9834	0.9516	0.9228	0.8833	0.8587
$\epsilon_{(c)}$ ratio	1.0006(5)	1.0010(6)	1.002(1)	1.002(1)	1.002(1)
$\epsilon_{(d)}$ ratio	0.9830	0.9487	0.9170	0.8750	0.8495
	1.0001(5)	0.9985(8)	0.996(1)	0.993(1)	0.991(1)
	0.9825	0.9497	0.9200	0.8794	0.8538
	0.9996(4)	0.9995(7)	0.9989(9)	0.998(1)	0.996(1)

and $B^0 \rightarrow K\pi$ are summarized in Tab. 5.1. This table shows that the ratios of efficiencies between each normalization channel and the $B_s^0 \rightarrow \mu^+ \mu^-$ signal ($\epsilon_{channel}/\epsilon_{signal}$) are equal to one within 0.4% for any BDTS requirement.

5.4 BDT classifier as a background discriminant

In the previous chapter we mentioned that the GL method was adopted to discriminate signal from background for the analysis presented in Ref. [39]. This section describes the definition of a new classifier, based on a BDT framework, adopted to improve the rejection of the combinatorial background.

The following discriminant variables, in addition to those used to define the published GL, are used to optimize a BDT classifier:

- ‘B isolation’, the isolation of the B candidate (CDF definition [46]):

$$I_{CDF} = \frac{p_T(B)}{p_T(B) + \sum_{\text{tracks}} p_T(\text{tracks})}$$

where $p_T(B)$ is the B transverse momentum and $\sum_{\text{tracks}} p_T(\text{tracks})$ is the sum of transverse momentum of tracks, excluding the muon candidates, satisfying $\sqrt{\delta\eta^2 + \delta\phi^2}$, with $\delta\eta$ and $\delta\phi$ denoting the difference in pseudorapidity and in ϕ coordinate between the track and the B candidate, respectively;

- ‘ $\min(p_{p_T}^+, p_{p_T}^-)$ ’, the smallest of the muon candidates’ p_T ;
- ‘polarization angle’, the cosine of the angle between the muon momentum in the B candidate rest frame and the vector perpendicular to the B candidate momentum and the beam axis.

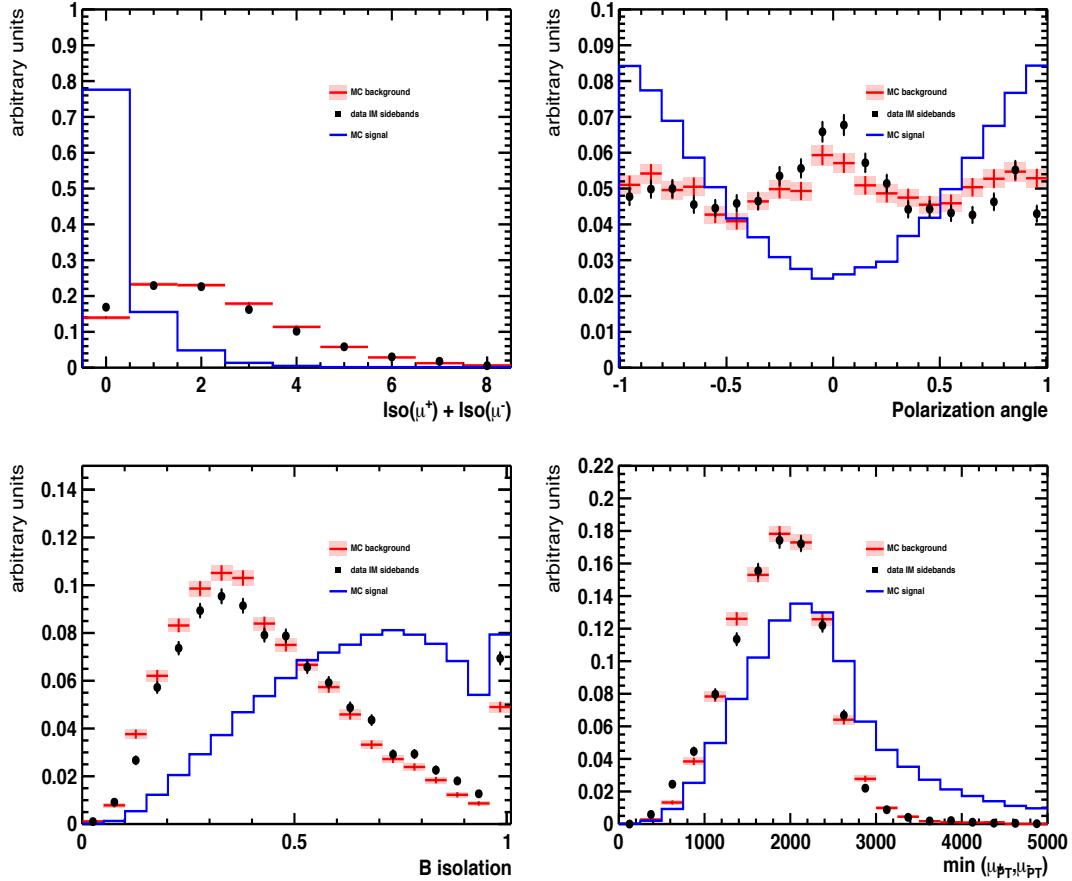


Figure 5.16: Distributions of variables entering the BDT, for signal $B_s^0 \rightarrow \mu^+ \mu^-$ (blue line) and background $b\bar{b} \rightarrow \mu\mu \text{XMC}$ (red with uncertainties), and data events from the sidebands of the invariant mass distribution. All plots are performed with events requiring $\text{BDTS} > 0.03$.

Fig. 5.16 displays these three variables, and the sum of the isolations of the muons. Once again the simulation describes quite well the combinatorial background.

The green curve in Fig. 5.17 shows the background rejection in terms of signal efficiency computed for the following set of six variables: the IP, the p_T and the decay time of the B candidate, the smallest IPS, the sum of the isolations and the DOCA of the muon candidates.

In order to understand what is the effect of introducing these new variables (B isolation, the smallest of the p_T of the muons and the ‘polarization angle’), we first add one of them at a time and compute the ROC curves for these three new sets of variables, displayed in blue, black and red, respectively, in Fig. 5.17. Every new variable improves the performances of the first set. The largest improvement comes from adding the smallest of the p_T of the muons.

In a second step, we add two variables at the time to the original set, while the final step corresponds to adding all three. Fig. 5.18 shows four ROC curves: one corresponds to

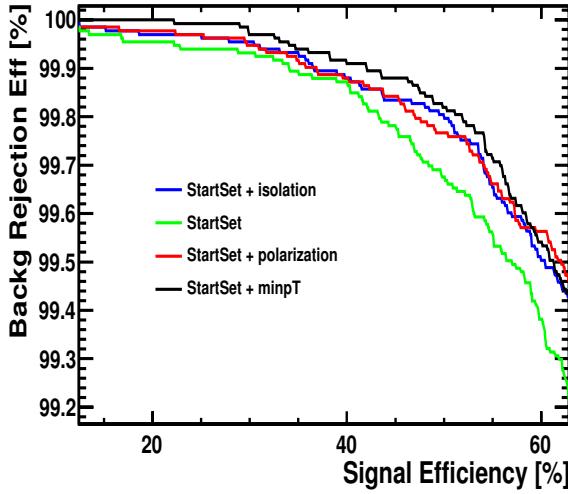


Figure 5.17: ROC curves for BDT defined with different input variables. StartSet stands for the starting set of variables: the IP, the p_T and the decay time of the B candidate, the smallest IPS, the sum of the isolations and the DOCA of the muon candidates.

the set with all variables, one to the set when just ' $\min(\mu_{p_T}^+, \mu_{p_T}^-)$ ' is added to the starting set and two curves that represent the best sets when adding two variables. The errors on the curves, not reported in the figures, are smaller than $\sim 0.04\%$. The samples used for the training and testing contained about 9500 background events and 84000 signal events, respectively.

The criterion to decide the most suitable set of variables is the maximum background rejection at a signal efficiency at 50%, the best set being the one containing all variables. The background in this region of signal efficiency, known as sensitive region, is highly suppressed.

The BDT classifier optimized in the previous paragraph increases the discriminating power of the GL, as defined in Ref. [39], as shown in Fig 5.19. The uncertainty associated to the background rejection for the GL (correlated with the uncertainty for the BDT) is of 0.15% for the low signal efficiency plot. This uncertainty is larger than the one reported for Figs. 5.17 and 5.18 as the background samples used in this case are smaller: about 1000 events for training and testing respectively. The reason to use such a reduced sample is because the GL was previously computed for an old sample with lower statistics. Therefore, in order to compare performances in similar conditions, we use the same number of events to train and test the BDT.

5.5 Further optimization of the discriminant BDT

Aiming to improve the rejection of background for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, new discriminant variables were under study [47] after the publication of Ref. [39]. An important feature desired in such variables is a low correlation with the former variables, which translates into bring-

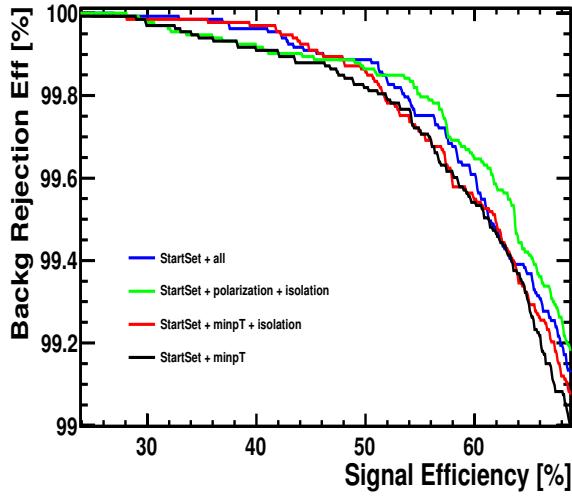


Figure 5.18: ROC curves for BDT defined with different input variables. StartSet stands for the starting set of variables: the IP, the p_T and the decay time of the B candidate, the smallest IPS, the sum of the isolations and the DOCA of the muon candidates.

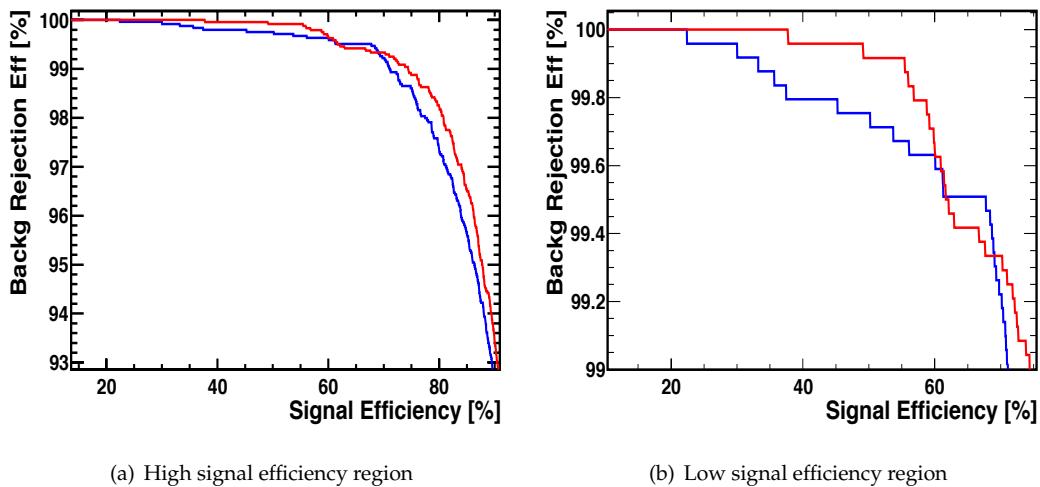


Figure 5.19: ROC curves for the GL [39] (blue line) and the BDT with nine variables defined in this section (red line).

ing new information to the classifier. This new information arises from the different angular distributions between the muon candidates, and between a muon candidate and the momentum of the other B in the event, expected for signal and background processes. We give a brief description of four variables (shown in Fig. 5.20) for which we study the possibility of defining a new BDT with:

- ‘other B angle’: the angle between the B candidate’s momentum and the trust momentum. The trust momentum is defined as the sum of momenta of all long tracks, excluding those associated to primary vertices different the one associated to the B candidate and tracks from long lived particles. If the lack of tracks impedes the definition of the trust momentum this variable is set to 1 in order to simplify a selection criterion separating signal and background;
- ‘ $|\Delta\eta|$ ’: the absolute value of the difference of pseudo-rapidity of the muon candidates;
- ‘angle wrt p_{tr} (boosted RF)’: the angle between the direction of the positive muon candidate in the rest frame of the B candidate and the trust momentum in the same frame. If the lack of tracks impedes the definition of the trust momentum this variable is set to $\pi/2$ in order to simplify a selection criterion separating signal and background;
- ‘abs(ϕ)’: absolute value of the difference of the spherical ϕ coordinate of the muon candidates.

The process of optimization is similar as for the BDT for the selection. With the idea of estimating the combination of variables that has the highest rejection of background power, we start computing the background rejection versus the signal efficiency for a BDT defined with a set of seventeen variables: the nine variables that define the previous BDT, all the variables that entered in the optimization of BDTS and the aforementioned four new variables.

Then, we compute the background rejections removing one variable at a time and compare them with the performances of the original set. If the background rejection of one of the new sets is similar as the original one, we start a new iteration from this last set of variables, removing again one variable at a time. After having removed several variables, if in the next step we observe a large reduction of the performances when trying to remove another variable, we proceed adding one by one those variables already discarded. The reason to do this relies on the fact that some variables might improve the performances, or not, depending on the other variables entering the BDT, due to correlations between them.

After the optimization the best set of variables is: the decay time, IP, isolation, direction angle, VDS and p_T of the B candidate, the smallest p_T DOCA, and isolation of the muon candidates, and the ‘polarization angle’ and the χ^2 (SV).

Fig. 5.21 shows the ROC curves corresponding to the BDT trained with nine variables and the optimized BDT with 13 variables. These curves were obtained using all the statistics available: $\sim 32,000$ background and 160,000 signal events. This was possible with the following procedure: split both $B_s^0 \rightarrow \mu^+ \mu^-$ signal and $b\bar{b} \rightarrow \mu\mu X$ background samples

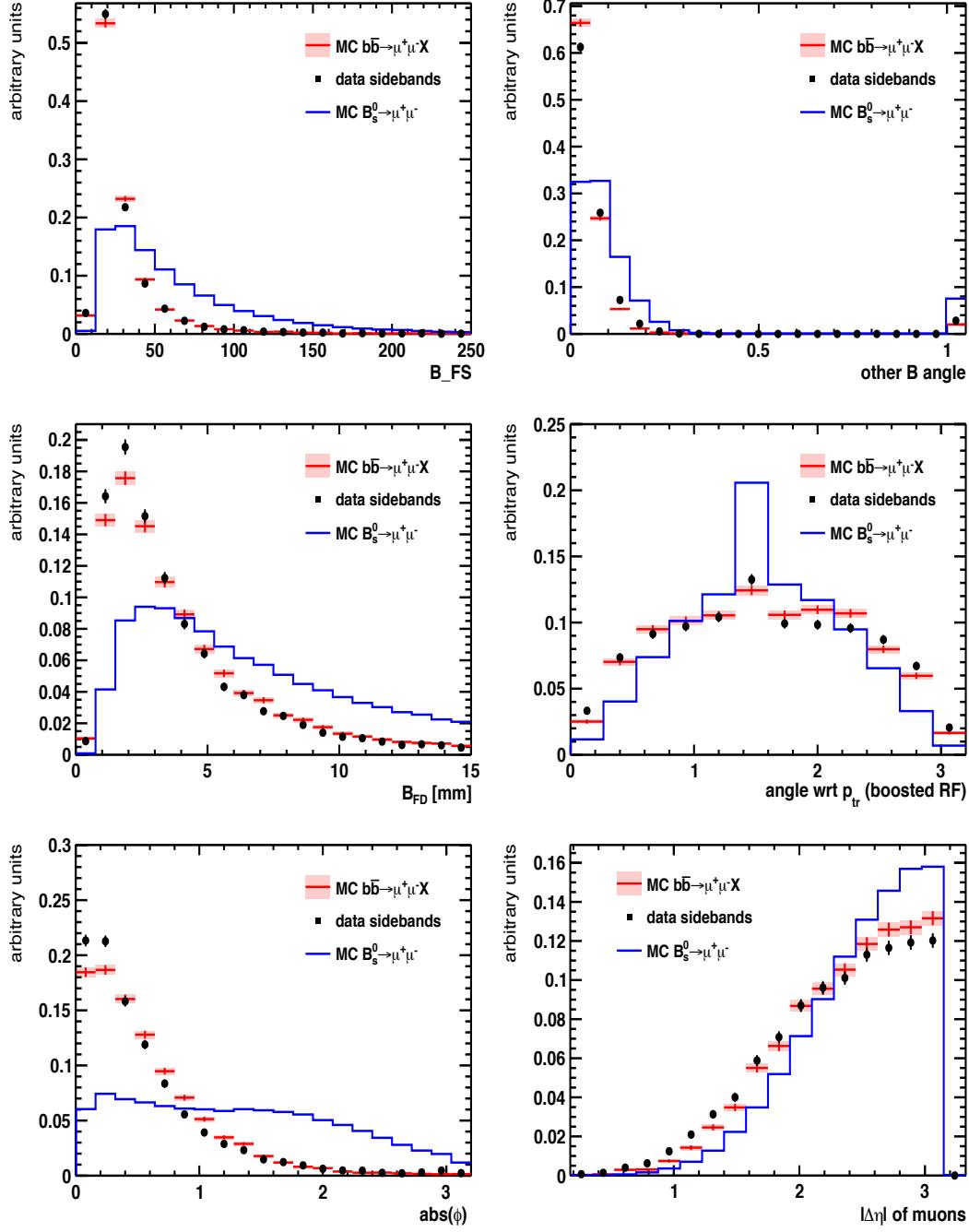


Figure 5.20: Variables included in the optimization of the BDT classifier. All the samples have a $\text{BDTS} > 0.05$ requirement.

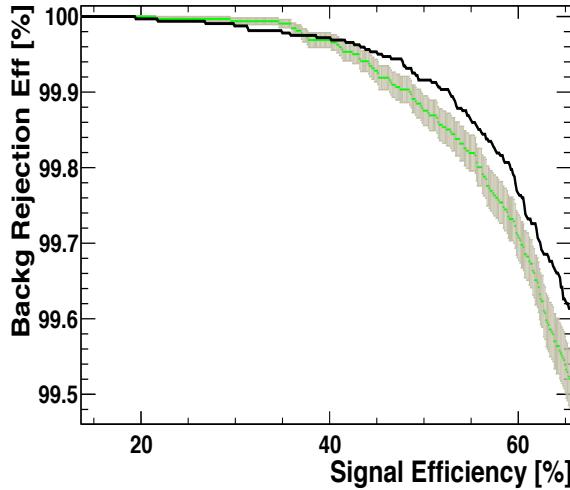


Figure 5.21: ROC curves for two different BDT: the green curve with the associated uncertainty corresponds to the BDT defined in the previous section which has nine variables. The black curve represents the performances of the BDT optimized in this section with thirteen variables.

into two, namely subsample signal A and subsample signal B, and subsample background A and sub-sample background B. Then, the sub-samples A are used to train a BDT that classifies the sub-samples B. On the other hand, the sub-samples B are used to train another BDT that classifies the sub-samples A. Finally, the curves shown in Fig. 5.21 are computed considering together sub-samples A and B.

The aforementioned procedure allows to optimize the variables entering the BDT. The final classifier is trained with all the available simulated events, and then calibrated on data.

In conclusion, the optimized BDT rejects $(33 \pm 7)\%$ more background for a signal efficiency of 50%.

5.6 Optimization of the BDTS-based selection

In principle, the BDTS was conceived to exploit multivariate techniques in order to reduce the background of two-body B decays, keeping a high efficiency in signal and control channels. In this section we propose a study of the performances of various BDT classifiers trained with different samples in terms of BDTS requirements. The BDTS mostly rejects background at low values of BDT (see Fig. 5.22).

We use a $b\bar{b} \rightarrow \mu\mu X$ simulated background sample of about 32,000 events. Requiring BDTS>0.1 reduces the yield to about 5,000 events, as the background rejection is 85%. A tighter criterion would not be appropriate as we need to split the sample into two: one for training and one for testing.

We display in Fig. 5.23 the ROC curves obtained after applying five different BDTS selections. Both training and testing samples are reduced by the same BDTS prerequisite. The

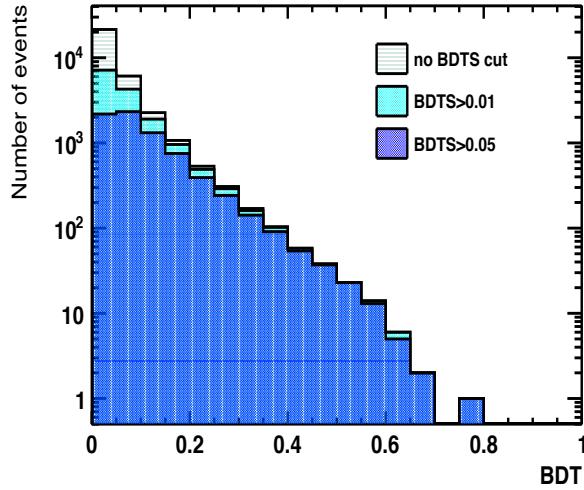


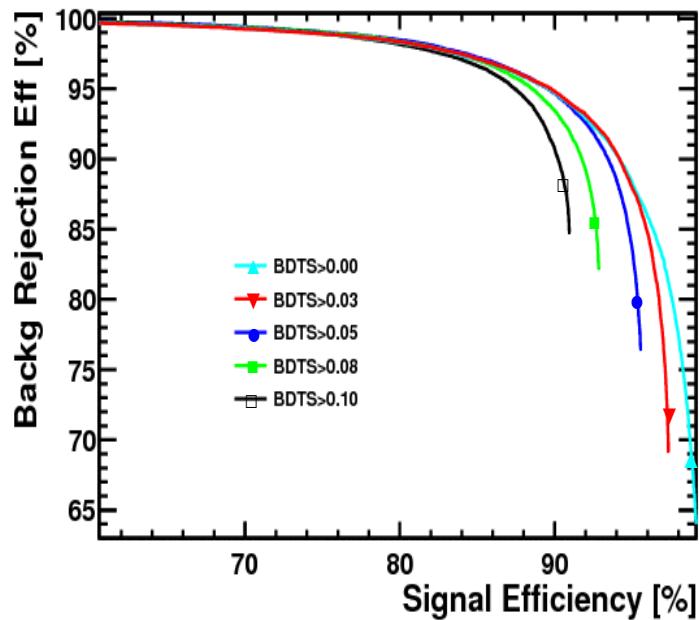
Figure 5.22: BDT distribution for $b\bar{b} \rightarrow \mu\mu X$ events with different BDTS requirements.

plot on the top shows the high signal efficiency region and the bottom one the low signal efficiency. The computation of the signal and background efficiencies of all the curves is performed considering that the total number of events is the same as when not requiring any BDTS selection. This allows a fair comparison of the background rejection in the low efficiency region. As a result, we can see how each curve starts at the background rejection point corresponding to the background rejection of the BDTS selection.

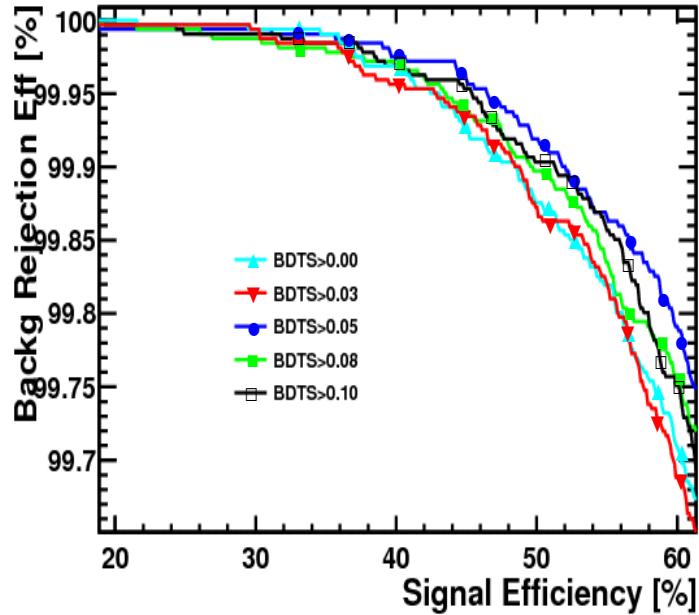
In conclusion, the BDT trained with $\text{BDTS}>0.05$ rejects more background in the region of signal efficiency below 50%. In this region, the improvement with respect to the curve with no BDTS requirement is of about 40%. The uncertainty associated to the background rejection for this last curve at signal efficiency of 50% is of 0.02%. Therefore, considering the errors correlated, the curve with the $\text{BDTS}>0.05$ requirement is 2.5σ above the curve with no BDTS selection, which indicates a significant improvement.

The fact that the performance of the BDT diminish for higher BDTS requirements (i.e. 0.08 and 0.1) than $\text{BDTS}>0.05$ is due to the fact that these requirements reduce the $b\bar{b} \rightarrow \mu\mu X$ background yield below 50% signal efficiency, hence there is not enough statistics for the training process.

After requiring $\text{BDTS}>0.05$, we perform a new study of the different BDT (with nine and thirteen variables). Fig. 5.25 shows the compatibility of the background rejection of both classifiers. The reason that the BDT with 9 variables improves in terms of rejection after the BDTS requirement, but not the BDT with 13 is due to the fact that some of variables defining this last classifier are the same as those already used to define the BDTS. Fig. 5.24 shows four variables shared by both BDTS and the BDT described in this section. The plots on the left are obtained with no BDTS selection, while for those on the right a $\text{BDTS}>0.05$ is enforced. It is clear that the rejection power of the variables after the BDTS requirement is much lower than before the selection.



(a) High signal efficiency region



(b) Low signal efficiency region

Figure 5.23: ROC for different BDT (nine variables) after several BDTS selections.

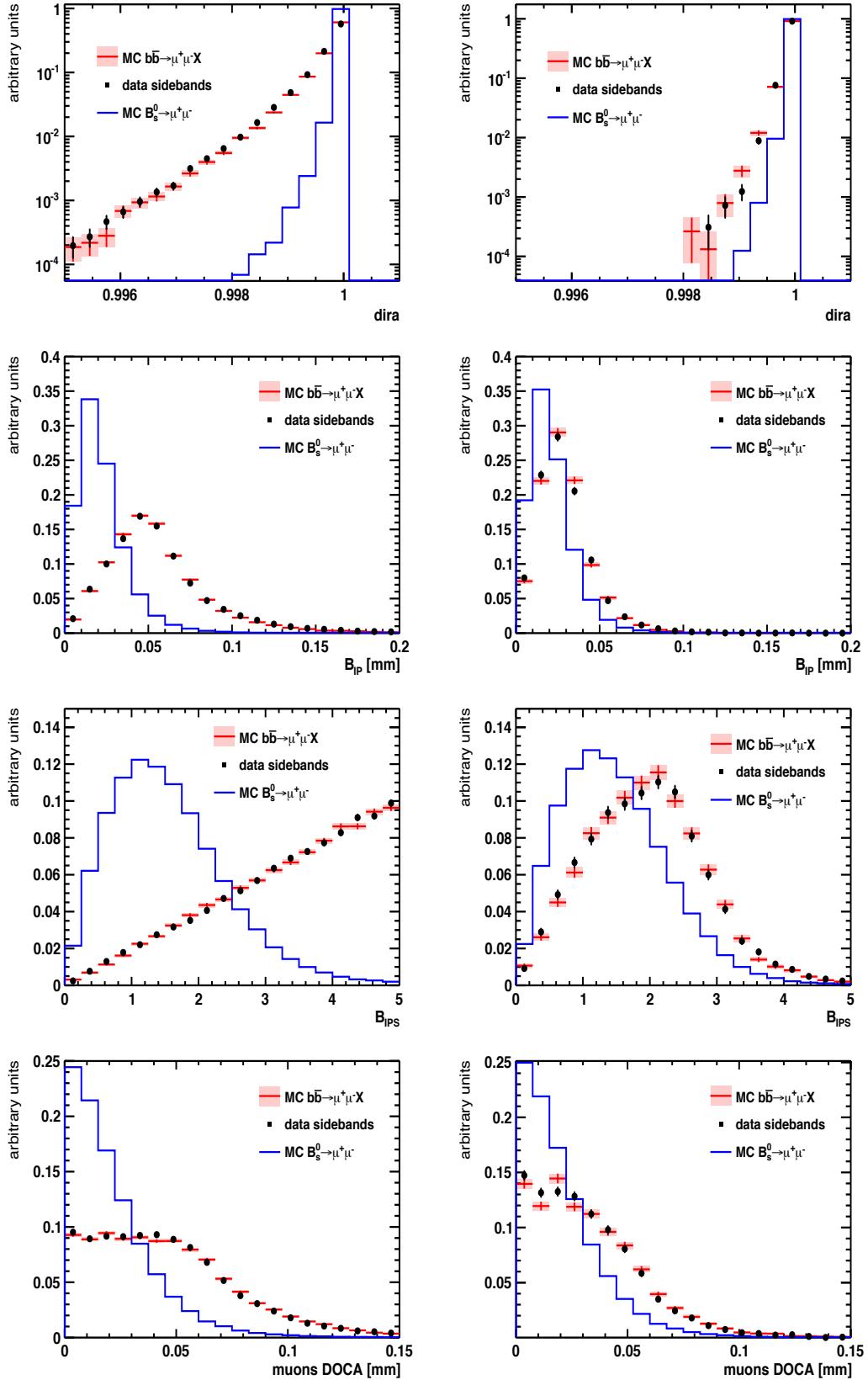


Figure 5.24: Variables shared by BDTS and the optimized BDT with nine variables. The plots on the left have no BDTS selection, while those on the right are required to have $BDTS > 0.05$.

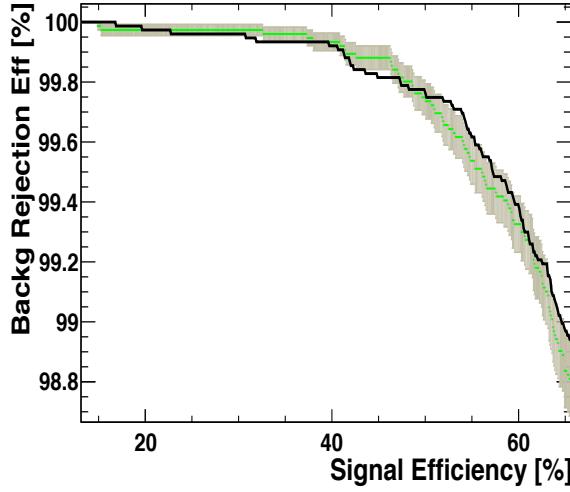


Figure 5.25: ROC curves for two different BDT computed with events after a BDTS>0.05 selection requirement: the green curve with the associated uncertainty corresponds to the BDT (9 variables). The black curve shows the performances of the BDT optimized in this section with thirteen variables.

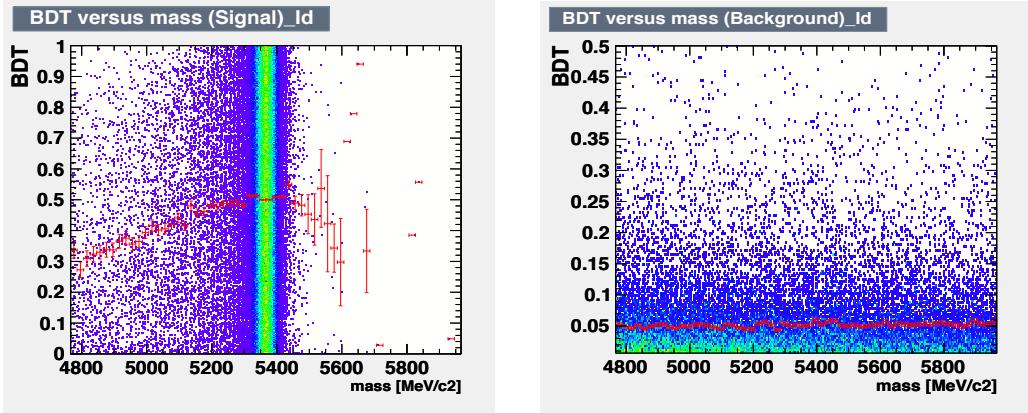


Figure 5.26: Scatter plots of invariant mass and BDT for signal $B_s^0 \rightarrow \mu^+ \mu^-$ (left) and background $b\bar{b} \rightarrow \mu\mu X$ (right) simulated events. The red points are the average profile between both variables.

BDT and invariant mass correlation

In Chap. 6 we perform a two-dimensional unbinned likelihood fit in invariant mass and BDT. In order to allow the product of their probability functions, mass and BDT need to be uncorrelated.

Fig. 5.26 (left) reflects a small correlation of 6% between the invariant mass and the BDT for $B_s^0 \rightarrow \mu^+ \mu^-$ simulated events. This correlation is due to the radiative decays in the lower edge of the mass peak. The correlation in $b\bar{b} \rightarrow \mu\mu X$ simulated events (right plot) is negligible.

Optimization of the BDT parameters

The classifier optimized in the previous section has a good background power. In this section we introduce another possibility to improve the classifier based on the optimization of the parameters that define such BDT.

We recall that the lack of a large $b\bar{b} \rightarrow \mu\mu X$ sample limits a further significant improvement of the classifier as it is almost 100% efficient in rejecting the background in the low signal efficiency region.

Four parameters, presented in Sec. 5.2, define the BDT: the number of trees used (N_{trees}), the boosting parameter (β), the minimum number of events in each node (N_{events}^{min}), and the maximum number of layers of each tree (N_{depth}^{max}). The BDT presented in the previous sections have: $N_{trees} = 250$, $\beta = 1$, $N_{events}^{min} = 400$, and $N_{depth}^{max} = 3$. The idea behind the following study is to search for the combination of these parameters that more efficiently discriminates the background in the sensitive region.

We perform a scan of the parameter space ($N_{trees} \in (100, 250, 400, 550)$; $\beta \in (0.8, 0.9, 1, 1.1)$; $N_{events}^{min} \in (200, 400, 600)$; $N_{depth}^{max} \in (2, 3, 4)$) and compute a figure of merit (FOM) that indicates the rejection power of each classifier. This FOM is the integral over the ROC curve below a signal efficiency of 60%, weighting more the region below 50%: $FOM \equiv 0.75 \times \int_{sig.eff.50\%} ROC + 0.25 \times \int_{sig.eff.60\%} ROC$. Although the main goal is to suppress the background below 50% signal efficiency, we account for an improvement a bit above this value searching for more stable trend in the ROC curve.

Fig. 5.27 shows the improvement in performance of the classifier varying the parameters. The new optimized BDT has: $N_{trees} = 400$, $\beta = 1$, $N_{events}^{min} = 400$, and $N_{depth}^{max} = 3$. Despite this visible improvement, these new settings were not used to define the final classifier as all the necessary parallel studies such as the calibration of the BDT with data (see Sec. 6.1.1) or the computation of the normalization factor (see Sec. 7.1) were finalized with the aforementioned settings.

In conclusion, the classifier that we use for the rest of the analysis corresponds to a BDT trained with the aforementioned nine variables, original parameters ($N_{trees} = 250$, $\beta = 1$, $N_{events}^{min} = 400$, and $N_{depth}^{max} = 3$), and after requiring $BDTS > 0.05$ in our samples.

5.7 Further signal data and simulation comparisons

Fig. 5.28 shows a comparison between data and simulation for signal for each of the other eight variables that define the BDT classifier. As for those comparisons already presented in 5.1, data and simulation are in relatively good agreement.

5.8 Conclusions

This chapter describes the boosted decision trees (BDT) classifier. This technique, used as a selection criterion, allows to reduce the background of two-body B decays by 60%, while keeping an efficiency of 92% in smeared $B_s^0 \rightarrow \mu^+\mu^-$ signal events. The reduction of $b\bar{b} \rightarrow \mu\mu X$ simulated background is of $(76.4 \pm 0.2)\%$ for the same signal efficiency. Furthermore,

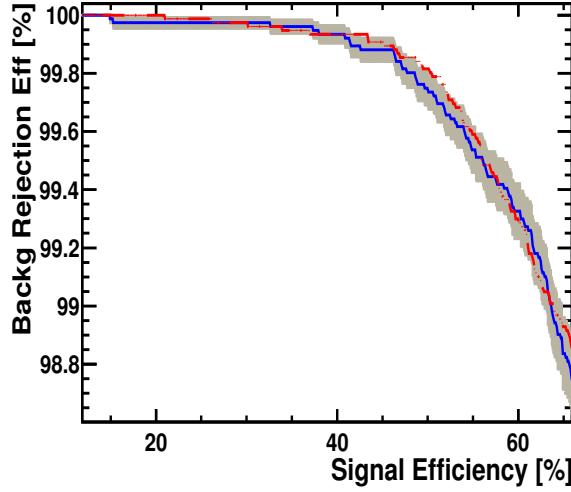


Figure 5.27: ROC for different BDT with same nine input variables and different number of trees: 250 trees (blue with grey error band) and 400 trees (dashed red).

the signal efficiency for such selection requirement in the sensitive region ($\text{BDT} > 0.5$) is 99% according to simulation.

A second BDT classifier is optimized to separate signal from background in the low signal efficiency region. According to simulation, this BDT allows to suppress a further 99.9% of background after the selection process, removing half of the remaining signal events.

Finally, we present a method to extract the signal distributions from data. The obtained data distributions for the variables used to define the BDT classifiers are in good agreement with simulation.

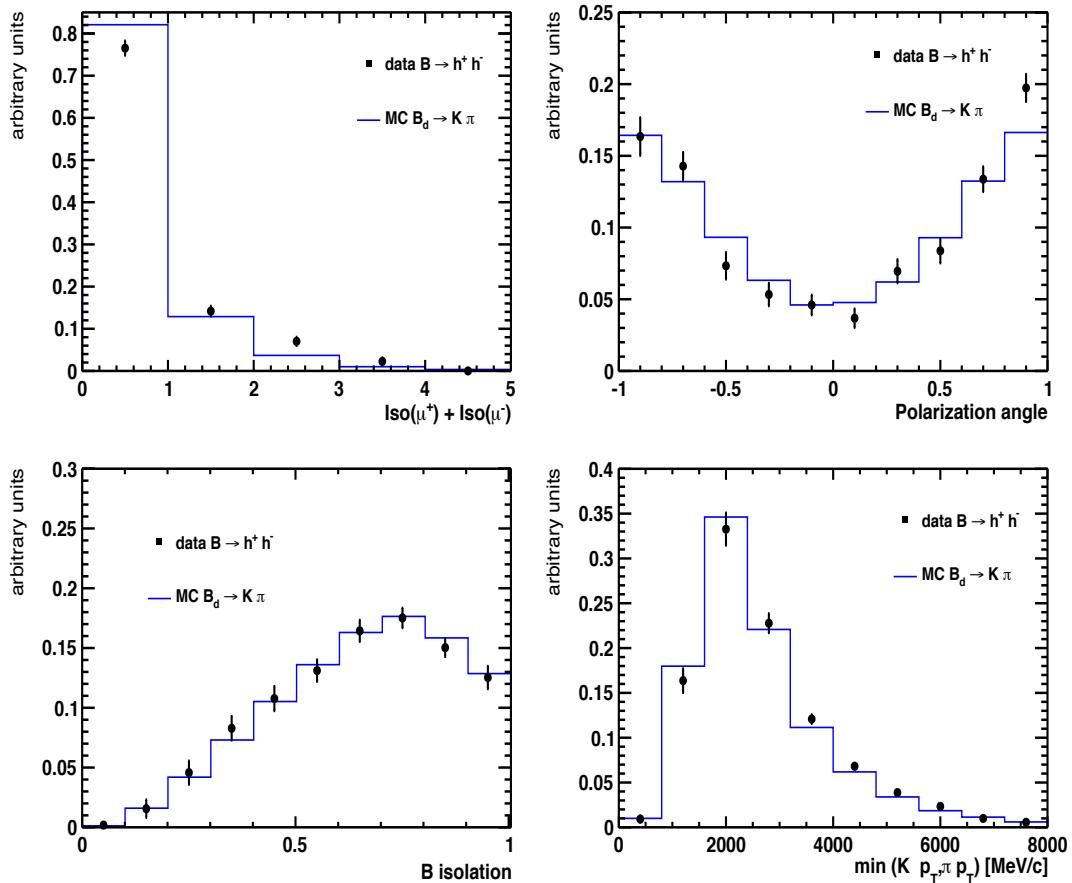


Figure 5.28: Comparisons of several variables that define the BDT classifier for $B^0 \rightarrow K\pi$ simulated events (blue line) and data $B^0_{(s)} \rightarrow h^+ h'^-$ events (black points).

Chapter 6

Extraction of the signal yields using an extended maximum likelihood fit

The technique we use to estimate the signal yields from our data sample is an extended maximum likelihood fit in two dimensions: the invariant mass and BDT output (see Chap. 5).

Sec. 6.1 briefly describes the extended maximum likelihood fit procedure. Then, Sec. 6.2 presents the method followed to extract the yields and to estimate the uncertainties and Sec. 6.3 summarizes the results obtained.

6.1 Overview of the fit procedure

The extended maximum likelihood (EML) fit method [48] assumes the knowledge of the probability density function (*pdf*) for N data events x_1, \dots, x_N in terms of a set of m parameters $\alpha_1, \dots, \alpha_m$: $p(x_i; \alpha_1, \dots, \alpha_m)$. If \mathcal{N} is the expected number of events, the observed number N is given by Poisson statistics, and the likelihood, \mathcal{L} , for this particular data values is:

$$\mathcal{L} = e^{-\mathcal{N}} \times \frac{\mathcal{N}^N}{N!} \prod p(x_i; \alpha_1, \dots, \alpha_m). \quad (6.1)$$

The factor multiplying the product in the extended likelihood is the Poisson term. The unknown parameters are obtained by minimizing $-2 \log \mathcal{L}$ with respect to all α_i .

The general idea behind EML fits is to extend the normalization condition not constraining it to be normalized to 1. The new probability density function does not just describe the shape of the expected distribution but also the size of the sample.

The $p(x_i; \alpha_1, \dots, \alpha_m)$ that represents our domain of interest is the sum of four contributions, each one depending on the mass (M) and the BDT:

$$p = \frac{N_s \mathcal{S}(M, BDT) + N_d \mathcal{D}(M, BDT) + N_p \mathcal{P}(M, BDT) + N_c \mathcal{C}(M, BDT)}{N}, \quad (6.2)$$

where $N = N_s + N_d + N_p + N_c$ is the sum of the number of $B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, peaking background, and combinatorial background events, while $\mathcal{S}(M, BDT)$, $\mathcal{D}(M, BDT)$, $\mathcal{P}(M, BDT)$ and $\mathcal{C}(M, BDT)$ are their respective *pdfs*. Each of these *pdfs* factorizes into a term that depends on the invariant mass and another that depends on the BDT, assuming no correlation between these two variables. For instance $\mathcal{S}(M, BDT) = \mathcal{S}_1(M) \times \mathcal{S}_2(BDT)$.

6.1.1 Description of the *pdfs* needed for the analysis

We report the methods used to obtain the parameters defining each *pdf* for the four types of events and both components (invariant mass and BDT). These methods have the advantage to rely mainly on data.

The reader shall find in Ref. [43] a more detailed description of the procedure to extract such *pdfs*.

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ invariant mass *pdf*

The model used to describe the invariant mass *pdf* for signal is a Crystall Ball function (see 5.1).

The exclusive $B^0 \rightarrow K^+ \pi^-$, $B^0 \rightarrow \pi^+ K^-$, $B^0 \rightarrow \pi^+ \pi^-$ and $B_s^0 \rightarrow K^+ K^-$ decays allow to derive from data the mean of the masses for B_s^0 and B^0 . The four exclusive B decays come from the $B_{(s)}^0 \rightarrow h^+ h^-$ TIS sample with four different mass hypotheses and particle identification requirements. For each exclusive decay, a fit of its invariant mass distribution allows to extract the mean values of the B^0 and B_s^0 masses:

$$m(B^0) = 5284.63 \pm 0.20_{\text{(stat)}} \pm 0.27_{\text{(syst)}} \text{MeV}/c^2,$$

$$m(B_s^0) = 5372.96 \pm 0.32_{\text{(stat)}} \pm 0.28_{\text{(syst)}} \text{MeV}/c^2.$$

The quoted $m(B^0)$ corresponds to a combination of the values obtained for the three B^0 channels. Both measured $m(B_s^0)$ and $m(B^0)$ are about 0.1% above the PDG mass values and the systematic uncertainties take into account variations of particle identification and BDTS requirements. These shifts of the measured masses with respect to the PDG values do not spoil the final results as they are taken into account in the analysis.

The resolution of the Gaussian core of the Crystal Ball is extracted by interpolating the values of the resolutions obtained for the five different dimuon resonances (J/ψ , $\Psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$). The interpolation by a power-law function (power=1.37±0.08) has been validated up to the Z^0 mass within 0.75σ and the systematic uncertainties estimated by fitting with a linear function. The resolutions obtained with this method are:

$$\sigma_{B^0} = 24.3 \pm 0.3_{\text{(stat)}} \pm 0.7_{\text{(syst)}} \text{MeV}/c^2,$$

$$\sigma_{B_s^0} = 24.8 \pm 0.3_{\text{(stat)}} \pm 0.6_{\text{(syst)}} \text{MeV}/c^2.$$

The transition point α and the exponent n are obtained by fitting the $B_s^0 \rightarrow \mu^+ \mu^-$ MC distribution requiring the resolution to match what measured on data:

$$\alpha = 2.094 \pm 0.024_{\text{(stat)}} \pm 0.015_{\text{(syst)}},$$

$$n = 0.993 \pm 0.009_{\text{(stat)}} \pm 0.012_{\text{(syst)}}.$$

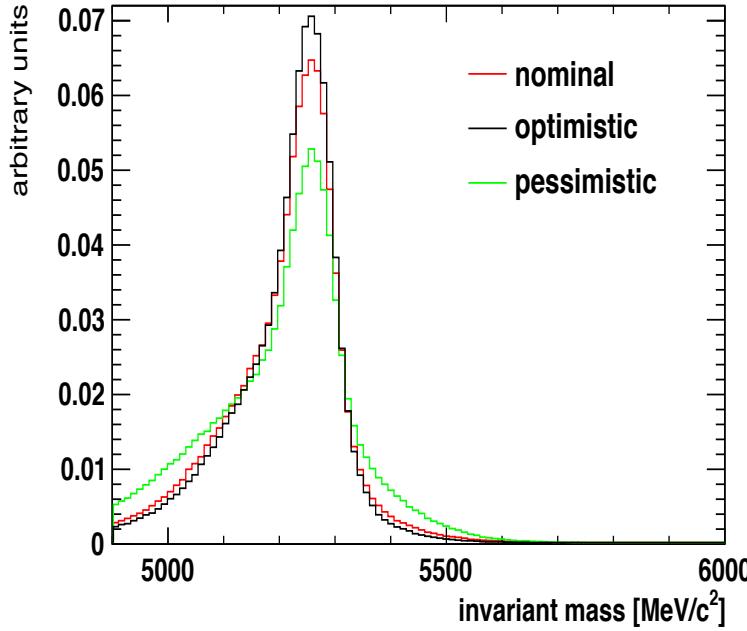


Figure 6.1: Invariant mass *pdf* for $B_{(s)}^0 \rightarrow h^+h'^-$ decays in which both hadrons decay in flight, which constitutes the main peaking background.

Peaking background invariant mass *pdf*

The invariant mass line-shape of $B_{(s)}^0 \rightarrow h^+h'^-$ decays in which both hadrons decay in flight is obtained from $B_{(s)}^0 \rightarrow h^+h'^-$ decays correcting the momentum of both hadrons by a factor that accounts for the loss of momentum due to their decay. This factor is extracted from simulation. Fig 6.1 shows the *pdf* for the invariant mass of the peaking background after such condition (nominal). Two different scenarios are considered: assuming no kink at all (optimistic) and accounting for a kink as large as the one present in the decay in flight of kaons (pessimistic).

The number of peaking background events in the whole invariant mass range is evaluated to be 5.4 ± 0.7 events for our data sample [43]. This number is obtained taking into account the double mis-identification rate ($\epsilon(hh \rightarrow \mu\mu)$) obtained by convoluting the kaon and pion fake rates, measured on data, with the momentum and p_T spectrums of the two hadrons of $B_{(s)}^0 \rightarrow h^+h'^-$ simulated events. The fraction of peaking background events that are in the B_s^0 and B^0 mass windows are $(8.8^{+3.0}_{-2.1})\%$ and $(48.0^{+20}_{-8})\%$, respectively.

Combinatorial background invariant mass *pdf*

To extract the invariant mass *pdf* for the combinatorial background we fit with an exponential function the $\mu\mu$ invariant mass distribution in the ranges [4900-5000] and [5418-5950] MeV/c^2 . The lower edge of the first range excludes most of the cascade background studied in Ref. 4.4, while the higher edge avoids misidentified $B_{(s)}^0 \rightarrow h^+h'^-$ decays. Fig. 6.2

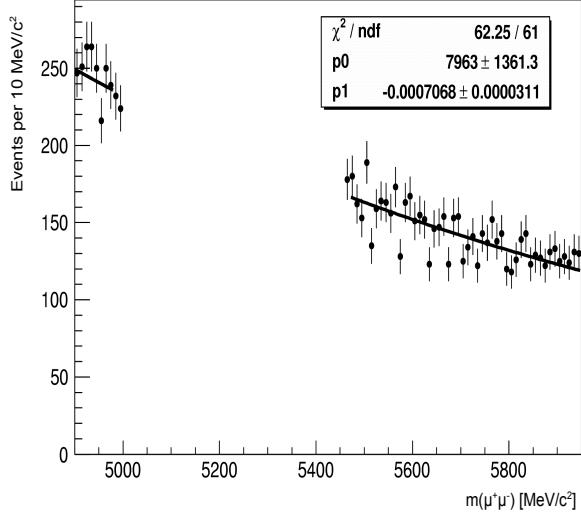


Figure 6.2: Invariant mass distribution of $\mu\mu$ events in the ranges [4900-5000] and [5418-5950] MeV/c^2 fit with an exponential function.

shows the invariant mass fit.

$$B_{(s)}^0 \rightarrow \mu^+\mu^- \text{ BDT } pdf$$

The BDT *pdf* for signal is obtained via data $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events. This sample is divided in different bins in BDT in which the invariant mass distribution is fit to extract the signal yield.

In an analysis presented with a reduced dataset [40], the number of $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events in the first BDT bin could not be determined directly because of a too low S/B ratio. Instead, we did a subtraction to the total yield the events in the rest of the bins. After introducing the BDT-based selection, the yield of TIS $B_{(s)}^0 \rightarrow h^+h'^-$ can be independently fit in all bins.

Fig. 6.3 shows the BDT *pdfs* for $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decays, in which each bins is obtained from the $B_{(s)}^0 \rightarrow h^+h'^-$ -TIS yields corrected with a factor to take into account the muon trigger bias on $B_s^0 \rightarrow \mu^+\mu^-$.

Peaking background BDT *pdf*

The distribution of the BDT of the peaking background is also obtained from the number of $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events in each BDT bin as for the signal BDT. In this case, the correction factor applied to the number of TIS events in each bin takes into account the double misidentification rate $\epsilon(hh \rightarrow \mu\mu)$ in the different BDT bins.

Fig. 6.4 shows the BDT peaking background distribution. The higher value of the first bin indicates that these processes are not perfectly signal-like due to the decays in flight.

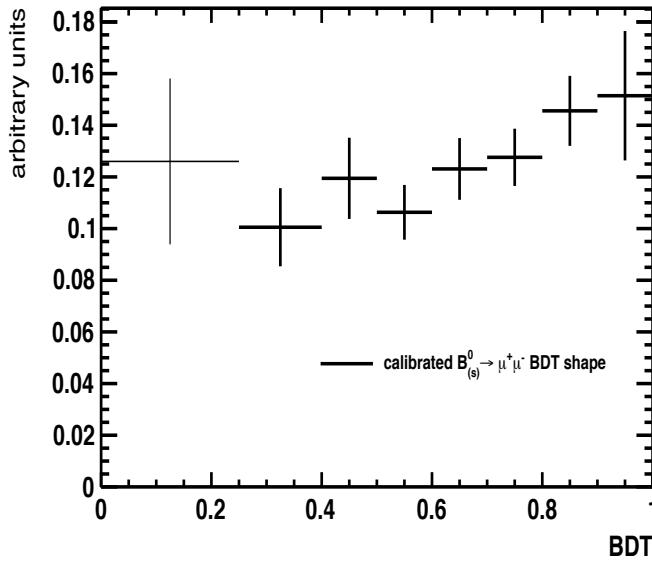


Figure 6.3: $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BDT *pdf* extracted from data $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events. The bins are normalized to their size.

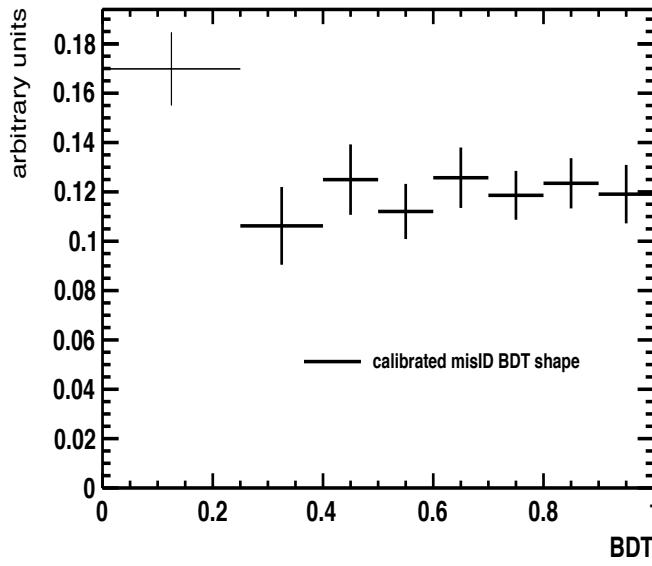


Figure 6.4: Peaking background BDT *pdf* extracted from data $B_{(s)}^0 \rightarrow h^+h'^-$ TIS events in which both hadrons decay in flight after correcting by the double mis-identification rate in each BDT bin. The bins are normalized to their size.

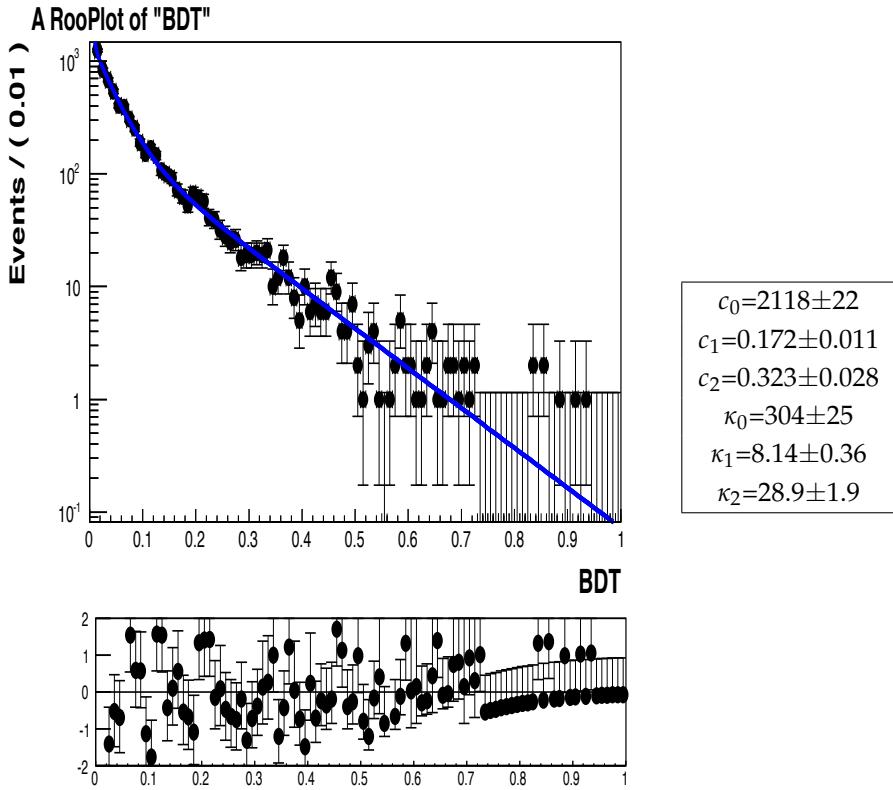


Figure 6.5: BDT distribution for candidates with invariant mass in the ranges [4900-5000] and [5418-5950] MeV/c^2 fit using a sum of three exponentials. The pulls of the fit are displayed at the bottom.

Combinatorial background BDT *pdf*

To extract the *pdf* for the BDT of the combinatorial background we fit the BDT distribution of events in the sidebands of the invariant mass distribution. The model used to fit the distribution is a sum of three exponentials:

$$f(\text{background}; \text{BDT}) = c_0 e^{-\kappa_0 x} + c_1 e^{-\kappa_1 x} + c_2 e^{-\kappa_2 x}.$$

Fig. 6.5 shows the result of the fit, with the pulls ($\frac{N_{\text{bin}(i)} - N_{\text{fit}(\text{bin}(i))}}{\sigma N_{\text{bin}(i)}}$) displayed on the bottom of the figure and the parameters of the fit on the right box. This figure reflects an overall good agreement between model and data, although it does not properly succeed fitting the first bin, whose pull value, 9, falls out of the range of the plot. For this reason, we will cross-check the results with another fit excluding the first bin in BDT. As shown in Fig. 6.6, excluding the region of $\text{BDT} \in (0-0.25)$ most of the pull values remain between -1 and 1.

The model with the exponential functions (in all BDT range) is used as a baseline to describe the BDT background, as it allows a well defined treatment of the systematic uncertainties by fluctuating the parameters with a Gaussian constraint.

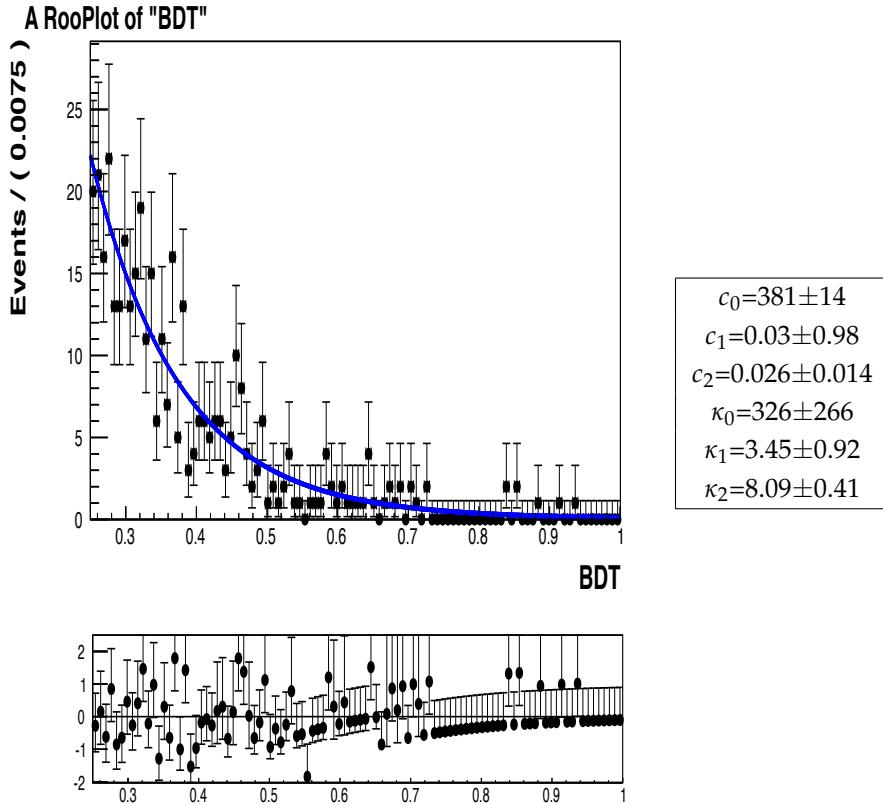


Figure 6.6: BDT distribution for candidates with invariant mass in the ranges [4900-5000] and [5418-5950] MeV/c^2 , excluding $\text{BDT} < 0.25$ events, fit using a sum of three exponentials. The pulls of the fit are displayed at the bottom.

We use two other models to describe the background BDT, to account for different sources of systematic uncertainty: a binned *pdf* with the same binning as the BDT for the signal and a kernel estimator [49] which provides an unbinned and non-parametric estimate of the *pdf* from a given data sample. We use the `RooKeysPdf` implementation in `RooFit` to obtain the kernel estimator from our data. Fig. 6.7 displays the fit performed using this last model which has a pull distribution mostly between -1 and 1. In this case, the first bin is properly fit.

Invariant mass correlation with BDT

The fit performed to extract the final yields uses two variables that are a priori uncorrelated (see right plot in Fig. 5.26): the invariant mass and the BDT. We study the possibility of a correlation of the exponent of the fit of the combinatorial background mass and the BDT by dividing the sample in different BDT ranges. The invariant mass distributions in each bin are fit with an exponential function, and the obtained exponents are displayed in Fig. 6.8 and reported in Tab. 6.1. All exponents agree within errors with the exponent obtained fitting the whole sample. At high BDT the errors increase due to the lack of statistics.

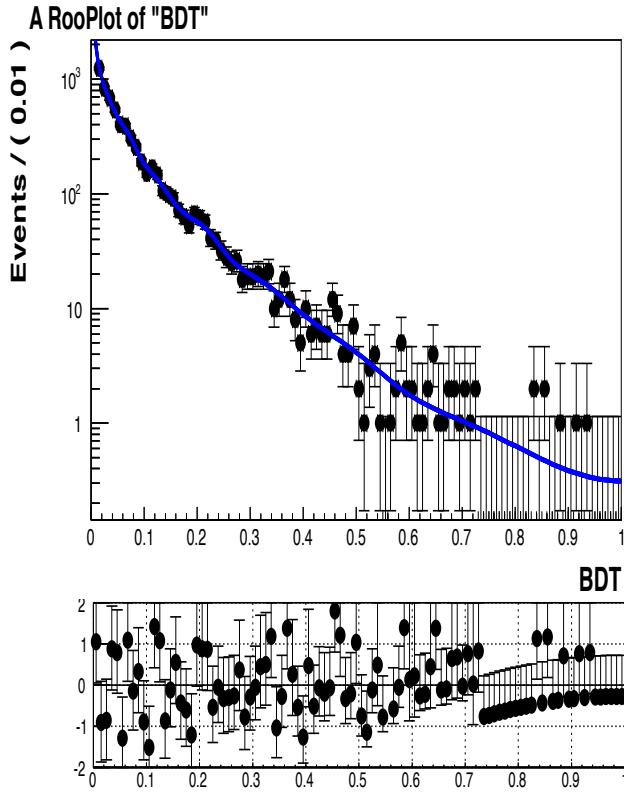


Figure 6.7: BDT distribution for candidates with invariant mass in the ranges [4900-5000] and [5418-5950] MeV/ c^2 fit using a kernel estimator. The pulls of the fit are displayed at the bottom.

6.2 Validation of the fit method using simulated data

Considering a $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ as predicted by the SM, the expected number of events in the data sample after the selection process is of about 10.0 $B_s^0 \rightarrow \mu^+ \mu^-$ and 1.2 $B^0 \rightarrow \mu^+ \mu^-$. To quantify the sensitivity of our method and to verify possible biases in the estimation of the uncertainties we fit a generated data sample. This sample is the result of merging data events in the sidebands of the invariant mass with events generated in the range [5000-5418] MeV/ c^2 of mass according to their *pdf*. The number of generated events in this mass range correspond to: 10 $B_s^0 \rightarrow \mu^+ \mu^-$, 0 $B^0 \rightarrow \mu^+ \mu^-$, 5 peaking background and 8371 combinatorial background. This background yield is estimated through the integration of the exponential function that models the background invariant mass distribution in the mass range [5000-5418] MeV/ c^2 .

Before analyzing the previous sample, we perform a toy study based on simulated events to check for biases in the uncertainty estimates.

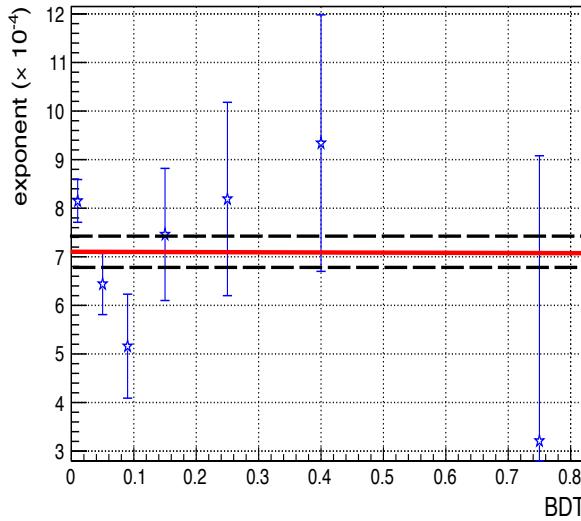


Figure 6.8: Values of the exponents extracted by fitting the invariant mass distribution in the ranges [4900-5000] and [5418-5950] MeV/c^2 with an exponential function, as a function of the BDT range. The horizontal red line shows the value obtained fitting the whole range, being the dashed black lines its uncertainty.

Table 6.1: Values of the exponents extracted by fitting the invariant mass distribution in the ranges [4900-5000] and [5418-5950] MeV/c^2 with an exponential function, as a function of the BDT range.

BDT range	- exponent ($\times 10^{-4}$)
0.000-0.025	8.2 ± 0.4
0.025-0.070	6.3 ± 0.6
0.070-0.125	5.4 ± 1.0
0.125-0.200	5.2 ± 1.3
0.200-0.300	6.5 ± 1.7
0.300-0.500	8.9 ± 2.2
0.500-1.000	13.4 ± 6.7
0.000-1.000	7.1 ± 0.3
0.250-1.000	9.8 ± 1.7

MC toy validation

We simulate 1000 data samples generating a number of events corresponding to the previously described sample: 10 $B_s^0 \rightarrow \mu^+ \mu^-$, 0 $B^0 \rightarrow \mu^+ \mu^-$, 5 peaking background, and 17235 (this number accounts for the events in the whole invariant mass range) combinatorial background.

Each sample is fit and the yields and the uncertainties for each type of events extracted.

Fig. 6.9 shows the distribution of fit yields, uncertainties, and pull values ($\frac{N_{fit} - N_{generated}}{\sigma_{N_{fit}}}$), for each type of event.

- **Yields.** The $B_s^0 \rightarrow \mu^+ \mu^-$ and background yields are well described by Gaussian distributions. On the other hand, the $B^0 \rightarrow \mu^+ \mu^-$ yield distribution is non-Gaussian and it has a tail towards negative values. This tail is caused by an overestimation of one of the other components, mainly the combinatorial background, due to an statistical fluctuation that forces the $B^0 \rightarrow \mu^+ \mu^-$ yield to large negative values in order to recover the stability of the fit.
- **Uncertainties.** A Gaussian model describes properly the $B_s^0 \rightarrow \mu^+ \mu^-$ and the background uncertainties, but not the $B^0 \rightarrow \mu^+ \mu^-$ uncertainty.
- **Pulls.** The $B_s^0 \rightarrow \mu^+ \mu^-$ and background pulls are Gaussian. The plots shown in Fig. 6.9 reflect a 5% overestimation of the statistical uncertainty of the B_s^0 yield that is compatible with zero. Hence, no correction is applied to the final uncertainty. The distribution of $B^0 \rightarrow \mu^+ \mu^-$ pulls is clearly non-Gaussian, and mostly between -2 and 2.

Generated data validation

We test several fit strategies before unblinding the data and choose the one giving the smallest relative error in measuring the $B_s^0 \rightarrow \mu^+ \mu^-$ yield of the simulated data while giving an unbiased result. These approaches consist in fixing or letting free the parameters defining each *pdf* and requiring or not a positive $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields. In all cases, the parameters defining the signal BDT, and the peaking background BDT and mass *pdfs* are fixed.

On one hand, not fixing the parameters results into a larger statistical uncertainty, on the other hand, no systematic uncertainty needs to be accounted for in this case, which can result into smaller total uncertainty. For this reason we test all the possibilities.

Tabs. 6.2 and 6.3 show the results obtained by fixing and letting free the parameters, and using different background BDT models, respectively. On both tables we report the results obtained fitting the BDT distributions with three exponentials including or not the first BDT bin. Furthermore, we apply positive prior requirements on the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields.

The results obtained using the background model with three exponentials (first rows in Tabs. 6.2 and 6.3) are consistent with each other at 7%, reflecting unbiased results. The results obtained using the other three approaches (no first BDT bin, kernel estimator and binned BDT background distribution) are consistent with the baseline method of using three exponentials at 10%, which also indicates the robustness of the fit method.

In the next paragraph we describe the method used to estimate the systematic uncertainty.

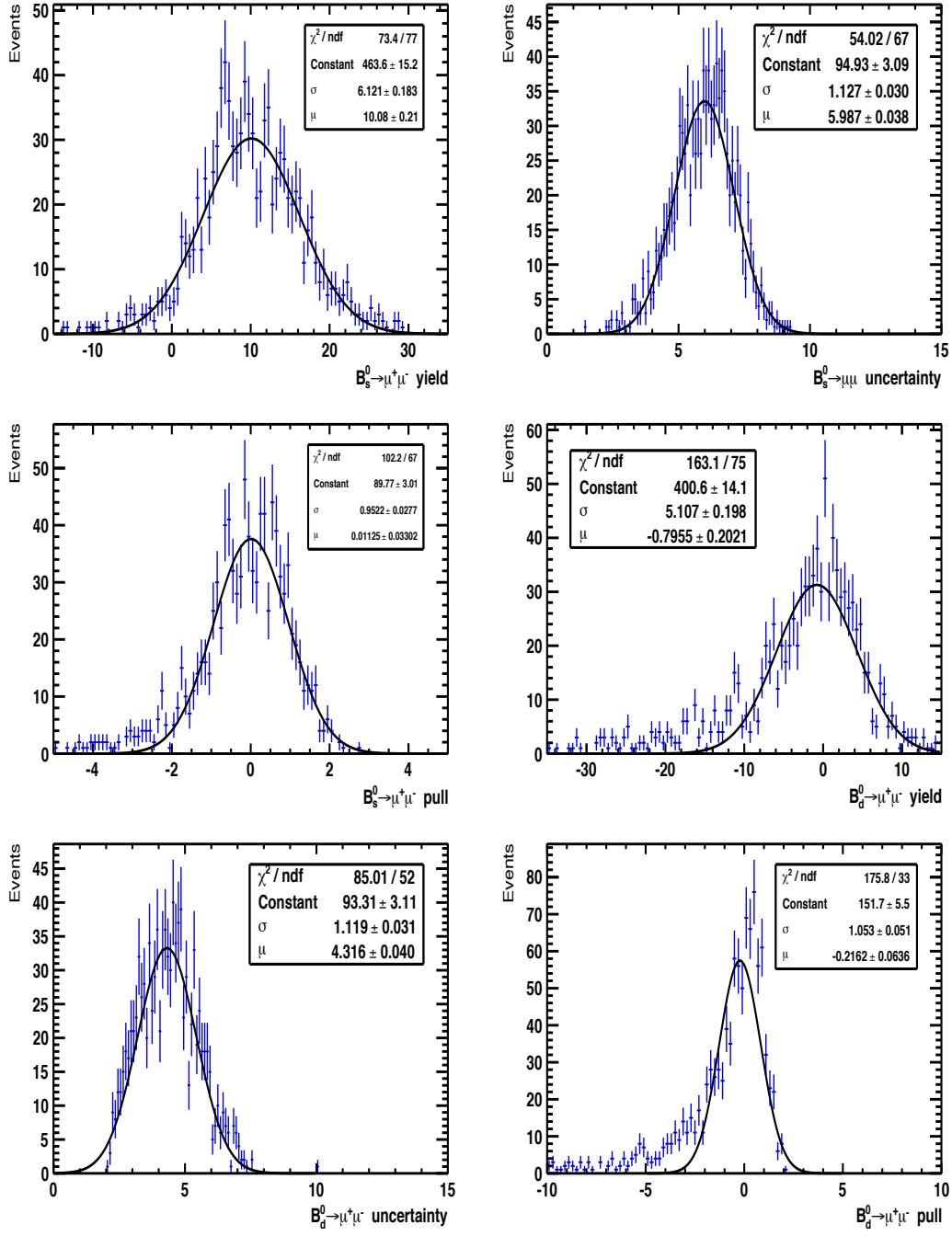


Figure 6.9: Distributions of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ yields, their uncertainties and their pulls. The total number of toys generated is 1000.

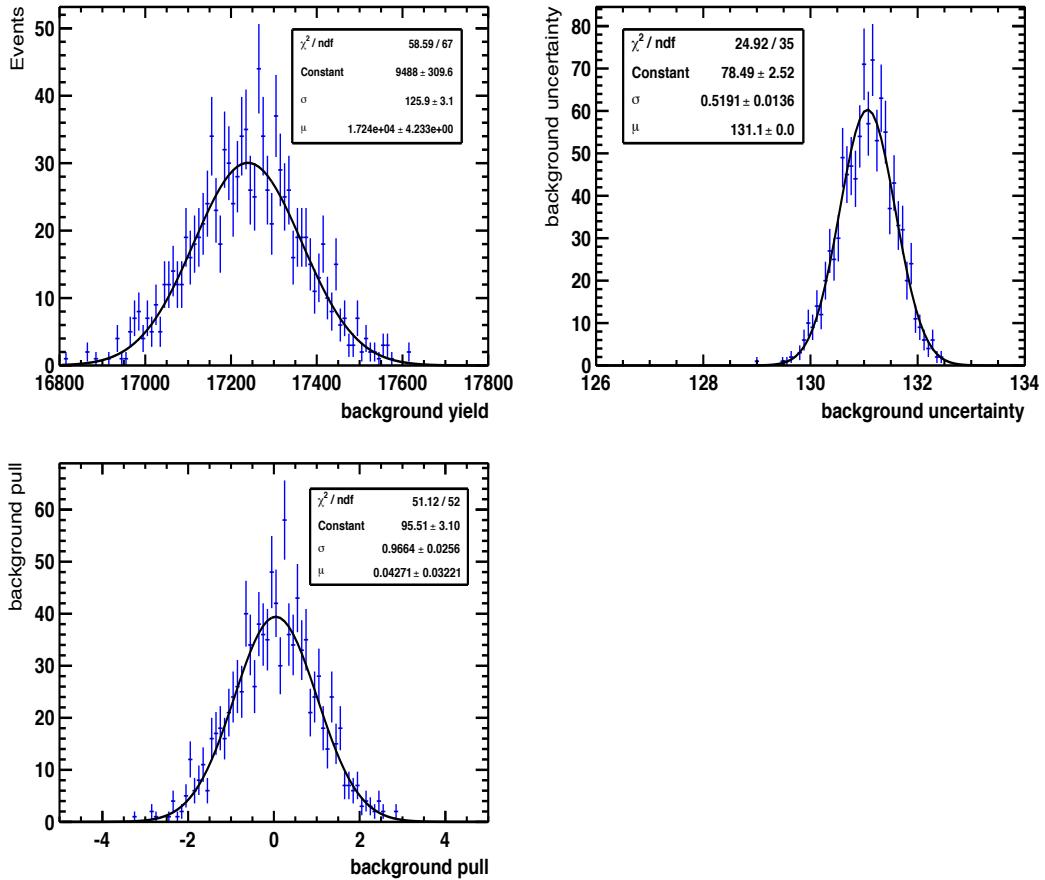


Figure 6.10: Distributions background yields, uncertainties and pulls. The total number of toys generated is 1000.

Table 6.2: $B_s^0 \rightarrow \mu^+ \mu^-$ yields obtained fitting the simulated data fixing the nuisance parameters to their mean values using different BDT background model.

Model	Positive prior	No prior
3 exponentials	$12.4^{+6.9}_{-5.4}$	$13.3^{+7.1}_{-5.7}$
3 exp. (no 1 st bin)	$11.3^{+6.8}_{-5.5}$	$12.1^{+7.1}_{-5.6}$
Rookeyspdf	$9.5^{+7.0}_{-5.6}$	$10.6^{+7.3}_{-5.9}$
Binned histogram	$10.4^{+7.1}_{-5.6}$	$11.4^{+7.4}_{-6.0}$

Table 6.3: $B_s^0 \rightarrow \mu^+ \mu^-$ yields obtained fitting the simulated data letting free the nuisance parameters of the invariant mass signal and background, and the BDT background using different BDT background model.

Model	Positive prior	No prior
3 exponentials	$13.1^{+7.2}_{-5.6}$	$13.3^{+7.2}_{-5.8}$
3 exp. (no 1 st bin)	$11.2^{+7.0}_{-5.6}$	$10.8^{+7.2}_{-5.7}$
Rookeyspdf	$9.5^{+7.0}_{-5.6}$	$10.5^{+7.4}_{-6.0}$
Binned histogram	$10.37^{+7.1}_{-5.6}$	$11.2^{+7.4}_{-5.9}$

Systematic uncertainties

Each *pdf* model has an intrinsic source of systematic uncertainty related to the knowledge of the nuisance parameters that define it. In the following we describe the method employed to estimate the systematic uncertainties associated to the parameterization of each *pdf*.

The parameters characterizing a given *pdf* are generated 500 times according to a Gaussian distribution centered at its mean value and width equal to its uncertainty. Then, using each of the new 500 generated *pdf*, we fit the simulated data sample described above and extract the yields.

In a first step, only the parameters defining the *pdf* of one type of event and component (mass or BDT) are generated, fixing the rest of parameters. This procedure allows to estimate the systematic uncertainty arising from each source. Fig. 6.11 shows the yield for $B_s^0 \rightarrow \mu^+ \mu^-$ events obtained by varying the parameters that define the BDT background and signal *pdfs*. The distribution of events is fit with two Gaussian distributions to account for asymmetric uncertainties. The positive and negative systematic errors are the widths of the Gaussians that fit the lower and the higher region with respect to the peak, respectively.

Accounting for a systematic due to the nuisance parameters that define the BDT background *pdf* and a further systematic due to the difference of yields observed when using different models might translate into double counting the uncertainties. In order to avoid this, we estimate which contribution is larger, and, if the negative uncertainty associated to the background BDT nuisance parameters is larger than the difference between models, only the former is accounted for in the final systematic uncertainty. In an opposite scenario, we estimate a systematic uncertainty due to all the parameters but those defining the background BDT *pdf*. Then, the obtained value is added in quadrature with the contribution arising from the difference observed among BDT background models.

Tab. 6.4 reports the yields obtained with the four different approaches already shown in Tabs. 6.2 and 6.3. For each of this yields we report the statistic and the systematic uncertainties. The latter is separated in four contributions. The total uncertainty is the sum in quadrature of all the contributions listed in Tab. 6.4.

Finally, the relative uncertainties corresponding to the total negative errors are reported on the third column in Tab. 6.4. These values measure the sensitivity of each approach in distinguishing an SM-like yield over the background only hypothesis. The smallest of these

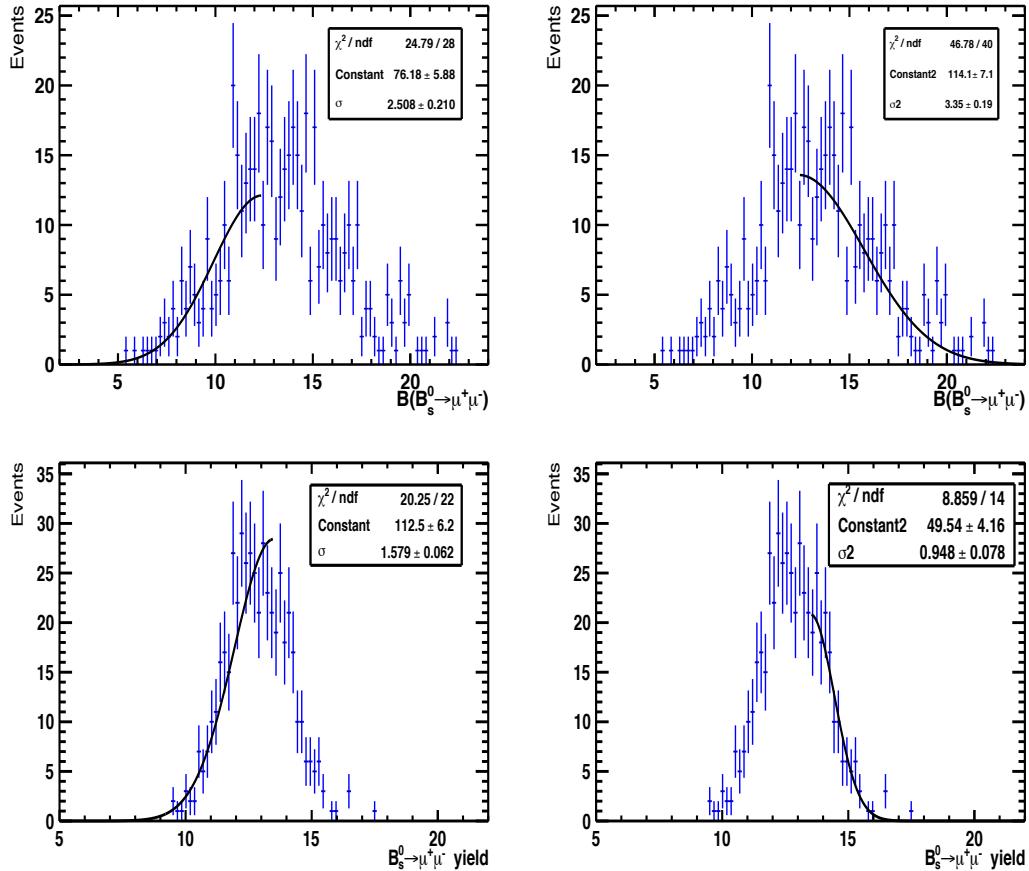


Figure 6.11: Distribution of $B_s^0 \rightarrow \mu^+\mu^-$ yield obtained generating the nuisance parameters of different components of the fit according to Gaussian distributions. The plots on the top show the systematic uncertainty arising from the nuisance parameters defining the BDT background component, while the plots on the bottom are obtained by varying the parameters defining the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BDT component. Both contributions are fit with two Gaussian distributions (one for the lower and one for the upper edge of the peak) to estimate asymmetric uncertainties.

Table 6.4: Yields and related uncertainties obtained with four different fit strategies: (a) fixing all the nuisance parameters and requiring a positive prior on the yields of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$; (b) fixing the parameters and not requiring a prior; (c) letting free the parameters and requiring a prior; and (d) letting free the parameters and not requiring any prior constraint. The systematic uncertainty is divided in different sources: nuisance parameters variation excluding the BDT background (syst1); BDT background nuisance parameters only (syst2); different BDT background model (syst3); and BDT range [0.25-1] (syst4). The third column gives the total systematic uncertainties, and the forth column reports the relative errors on the total negative uncertainties obtained with each fitting strategy.

	Results	Total syst	$\frac{\sigma^-(N_{B_s^0})}{N_{B_s^0}}$
(a)	$12.4 \begin{pmatrix} +6.9 \\ -5.4 \end{pmatrix}_{\text{stat}} \begin{pmatrix} +3.0 \\ -2.4 \end{pmatrix}_{\text{syst1}} \begin{pmatrix} +1.2 \\ -1.7 \end{pmatrix}_{\text{syst2}} \begin{pmatrix} +0.0 \\ -2.9 \end{pmatrix}_{\text{syst3}} \begin{pmatrix} +0.0 \\ -1.1 \end{pmatrix}_{\text{syst4}}$	$\begin{pmatrix} +3.0 \\ -3.9 \end{pmatrix}$	0.537
(b)	$13.3 \begin{pmatrix} +7.1 \\ -5.7 \end{pmatrix}_{\text{stat}} \begin{pmatrix} +3.2 \\ -2.5 \end{pmatrix}_{\text{syst1}} \begin{pmatrix} +2.8 \\ -1.4 \end{pmatrix}_{\text{syst2}} \begin{pmatrix} +0.0 \\ -2.7 \end{pmatrix}_{\text{syst3}} \begin{pmatrix} +0.0 \\ -1.3 \end{pmatrix}_{\text{syst4}}$	$\begin{pmatrix} +3.2 \\ -3.9 \end{pmatrix}$	0.519
(c)	$13.1 \begin{pmatrix} +7.2 \\ -5.6 \end{pmatrix}_{\text{stat}} \begin{pmatrix} +2.4 \\ -1.8 \end{pmatrix}_{\text{syst1}} \begin{pmatrix} +0.0 \\ -0.0 \end{pmatrix}_{\text{syst2}} \begin{pmatrix} +0.0 \\ -2.9 \end{pmatrix}_{\text{syst3}} \begin{pmatrix} +0.0 \\ -1.9 \end{pmatrix}_{\text{syst4}}$	$\begin{pmatrix} +2.4 \\ -3.9 \end{pmatrix}$	0.524
(d)	$13.3 \begin{pmatrix} +7.2 \\ -5.8 \end{pmatrix}_{\text{stat}} \begin{pmatrix} +2.2 \\ -1.2 \end{pmatrix}_{\text{syst1}} \begin{pmatrix} +0.0 \\ -0.0 \end{pmatrix}_{\text{syst2}} \begin{pmatrix} +0.0 \\ -2.7 \end{pmatrix}_{\text{syst3}} \begin{pmatrix} +0.0 \\ -2.5 \end{pmatrix}_{\text{syst4}}$	$\begin{pmatrix} +2.2 \\ -3.9 \end{pmatrix}$	0.523

numbers corresponds to strategy (b), not requiring any prior and fixing the parameters, and consequently this is the strategy that we follow to fit our data sample. This decision was taken before un-blinding the events in the region $\pm 60 \text{ MeV}/c^2$ around the B_s^0 and B^0 masses.

Moreover, we estimate a possible systematic uncertainty due to the small correlation of the BDT and the invariant mass in signal. For the performed study, based on fitting a sample of purely $B_s^0 \rightarrow \mu^+ \mu^-$ signal events where the parameters defining the signal invariant mass are extracted directly from simulated events, we obtain a shift in the yield smaller than 1%. This shift is negligible and hence not accounted for as a systematic uncertainty.

In order to visualize the results of the fit for the generated sample, Fig. 6.12 shows the projections of the fit in the BDT and invariant mass axes, excluding the first BDT bin for a better visualization of the fit yields.

6.3 Yields from 1 fb^{-1} of data

The procedure described in the previous section allows to extract the signal yields in our data sample and account for all systematic uncertainties. Tab. 6.5 shows the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields obtained with the three aforementioned background BDT models and fitting the range $\text{BDT} \in (0.25-1)$. The largest discrepancy between the results obtained with the baseline procedure and the other procedures arises when modeling the background BDT with a kernel estimator (syst3).

In Fig. 6.13 we display the projections of the fit in the BDT and invariant mass axes, excluding the first BDT bin for a better visualization of the fit yields.

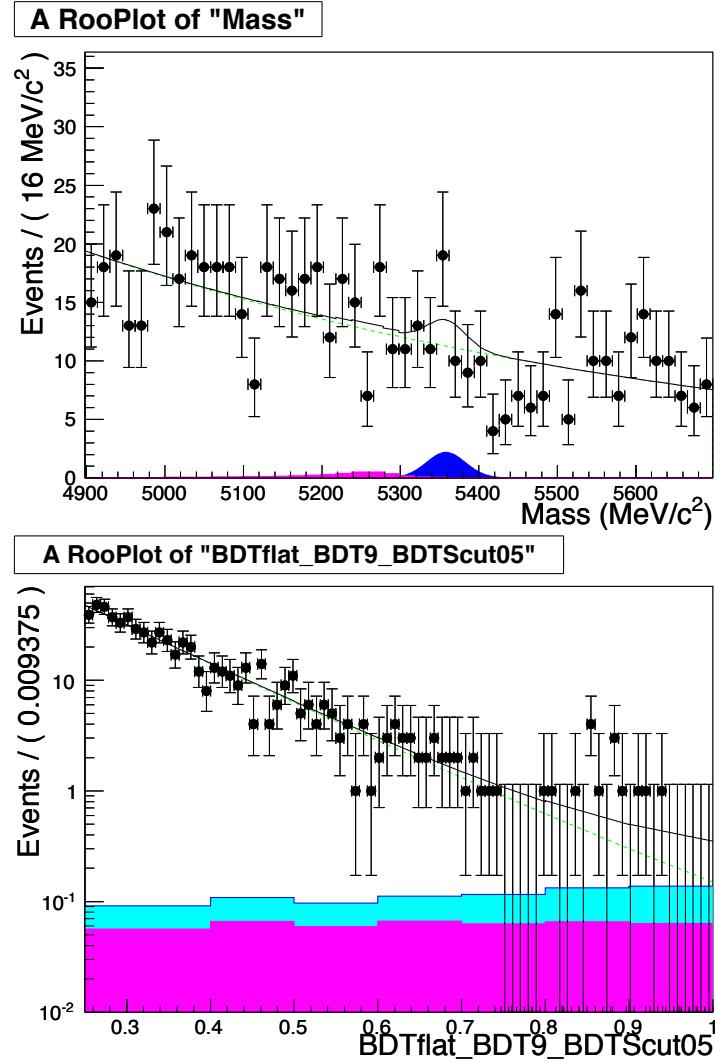


Figure 6.12: Projections of the fit on the BDT (top) and invariant mass (bottom) axes, excluding $\text{BDT} < 0.25$, obtained with the generated data sample. The blue areas corresponds to the $B_s^0 \rightarrow \mu^+ \mu^-$ yield, the yellow areas to the $B^0 \rightarrow \mu^+ \mu^-$ yield and the pink areas to the peaking background yield.

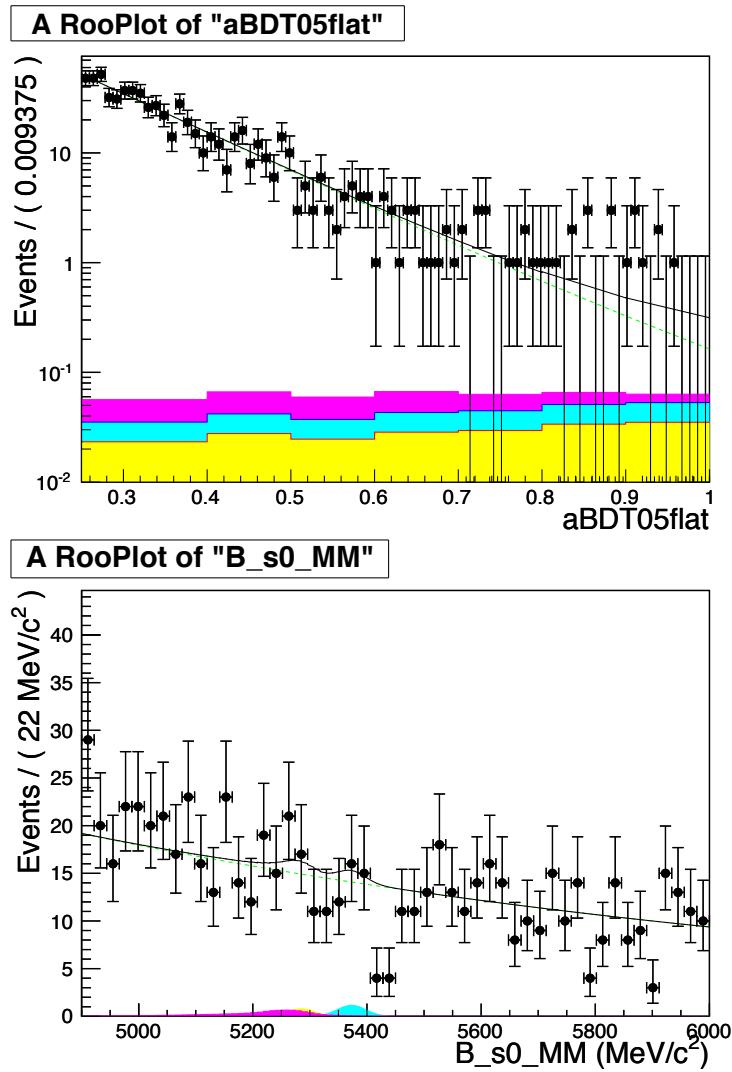


Figure 6.13: Projections of the fit on the BDT (top) and invariant mass (bottom) axes, excluding $\text{BDT} < 0.25$. The blue areas corresponds to the $B_s^0 \rightarrow \mu^+ \mu^-$ yield, the yellow areas to the $B^0 \rightarrow \mu^+ \mu^-$ yield and the pink areas to the peaking background yield.

Table 6.5: Fit yields obtained for 1 fb^{-1} of data obtained with different BDT background models: in the first row, we report the results obtained with the baseline model defined with three exponentials; in the second row, three exponentials excluding the first BDT bin; in the third row, kernel estimator; and in the fourth row, binned pdf with the same 8 bins as signal BDT.

	Yield $B_s^0 \rightarrow \mu^+ \mu^-$	Yield $B^0 \rightarrow \mu^+ \mu^-$
3 exponentials	$4.5^{+5.1}_{-3.5}$	$3.6^{+6.3}_{-4.5}$
3 exp. (no 1 st BDT bin)	$4.6^{+5.4}_{-3.7}$	$3.1^{+6.6}_{-4.6}$
RooKeysPdf	$2.2^{+5.1}_{-3.5}$	$0.1^{+5.9}_{-4.0}$
Binned	$4.1^{+5.2}_{-3.6}$	$0.9^{+6.3}_{-4.6}$

Tab. 6.6 reports, in the first five rows, the systematic uncertainties due to each *pdf* model obtained fitting 500 generated event samples with Gaussian-distributed nuisance parameters. For the numbers given in these first five rows, only the parameters related to each component vary, the others are fixed to their mean values. Then, the same table gives the systematic uncertainty associated to all the nuisance parameters (seventh row) and to all the parameters, fixing those defining the BDT background (sixth row).

Although, the uncertainty associated to the background BDT parameters (syst2) is again the dominant among them, it is smaller than the contribution due to the difference between models (syst3) (see Tab 6.5). For this reason we estimate the uncertainty due to all parameters except those defining the background BDT *pdf* (syst1), and then add these results in quadrature with the systematic uncertainty arising from the difference between models:

$$N_{B_s^0} = 4.5 \left({}^{+5.1}_{-3.5} \right)_{\text{(stat)}} \left({}^{+1.7}_{-2.3} \right)_{\text{(syst)}} \quad (6.3)$$

$$N_{B^0} = 3.6 \left({}^{+6.3}_{-4.5} \right)_{\text{(stat)}} \left({}^{+3.4}_{-3.6} \right)_{\text{(syst)}} \quad (6.4)$$

Figs. 6.14 and 6.15 display the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields obtained by varying the parameters related to different sources of systematic uncertainty. These figures contain the same information as Tab. 6.6.

A number of 1000 simulated samples were generated according to the same yields extracted from data. Fig. 6.16 shows the number of fits giving a particular yield, uncertainty and pull distribution for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$. The plots in the figure show that the errors and the pulls are almost Gaussian. From the pulls distributions we estimate an small overestimate of the statistical uncertainties of 5% and 7%, and no correction is applied to the final results.

6.4 Conclusions

This chapter describes the extended maximum likelihood fit used to obtain the signal yields. The validation of the fit method using simulated data proved the lack of significant biases

Table 6.6: Systematic uncertainties associated to each component of the fit. The third row shows the contribution to the systematic uncertainty coming from the number of $B_{(s)}^0 \rightarrow h^+ h^-$ TIS events for all BDT bins. This contribution enters the uncertainty related to the signal and peaking background BDT. The sixth row gives the systematic uncertainties related to all components, and the seventh the systematic arising from all components, excluding those defining the BDT background component.

Model	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow \mu^+ \mu^-$
Signal mass	+0.10 -0.07	+0.23 -0.17
Background mass	+0.02 -0.01	+0.15 -0.06
Number of TIS	+0.38 -0.40	+0.50 -0.62
Background BDT	+1.6 -1.4	+3.4 -2.2
Peaking background mass	+0.09 -0.00	+0.00 -0.09
All	+1.7 -1.4	+3.4 -2.3
All-(Backg BDT)	+0.36 -0.41	+0.57 -0.65

in the statistical uncertainty estimation, and the BDT background modeling as the main contributor to the systematic uncertainty.

The final result for the $B_s^0 \rightarrow \mu^+ \mu^-$ yield is less than half the predicted SM \mathcal{B} value, although compatible with the SM as reported in the next section.

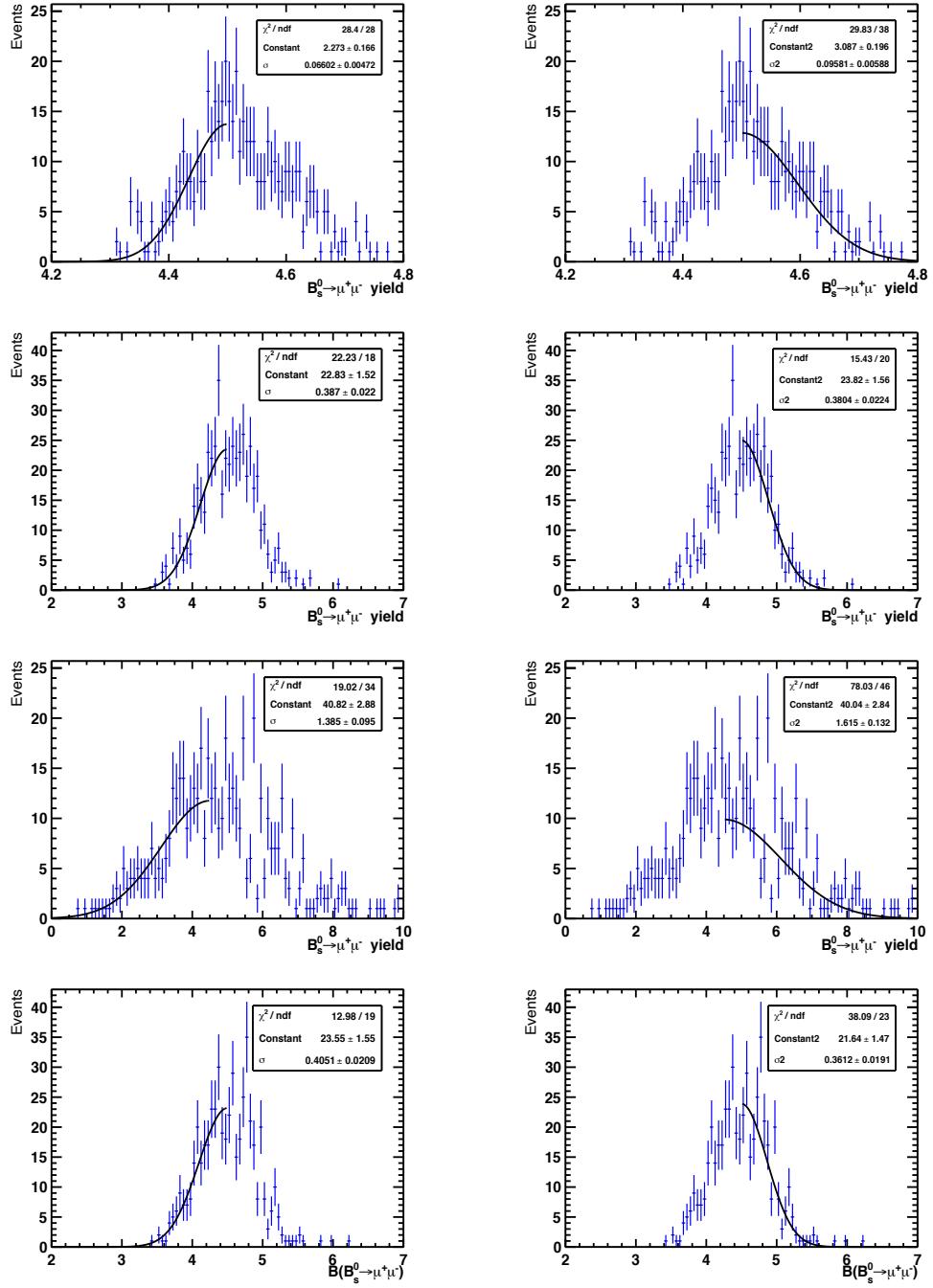


Figure 6.14: Distribution of $B_s^0 \rightarrow \mu^+\mu^-$ yields, obtained generating the nuisance parameters of different components of the fit according to Gaussian distributions. From top to bottom: signal mass component; signal BDT; background BDT; and all the parameters but those defining the background BDT. All contributions are fit with two Gaussian distributions (one for the lower and one for the upper edge of the peak) to estimate asymmetric uncertainties.

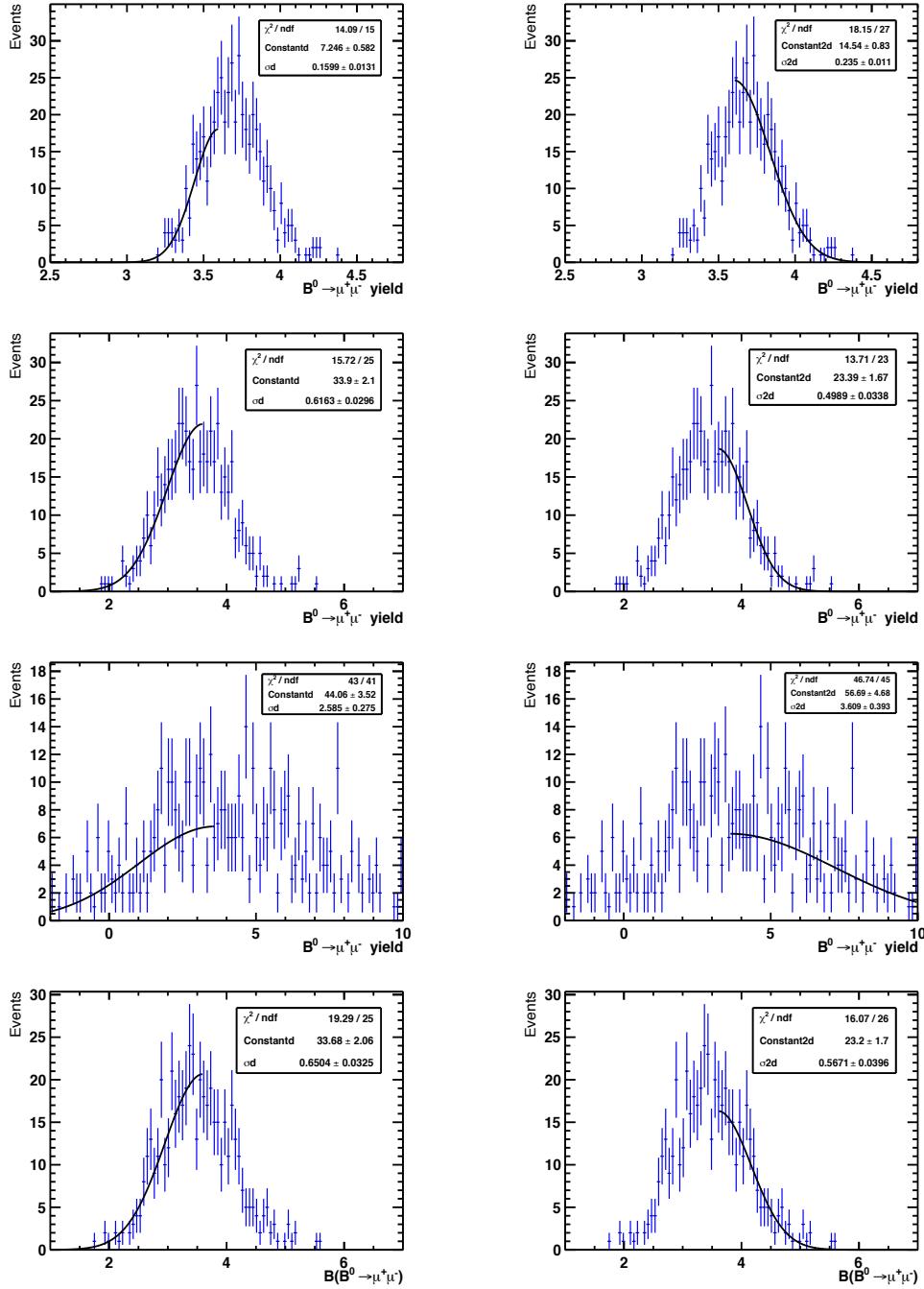


Figure 6.15: Distribution of $B^0 \rightarrow \mu^+ \mu^-$ yields, obtained generating the nuisance parameters of different components of the fit according to Gaussian distributions. From top to bottom: signal mass component; signal BDT; background BDT; and all the parameters but those defining the background BDT. All contributions are fit with two Gaussian distributions (one for the lower and one for the upper edge of the peak) to estimate asymmetric uncertainties.

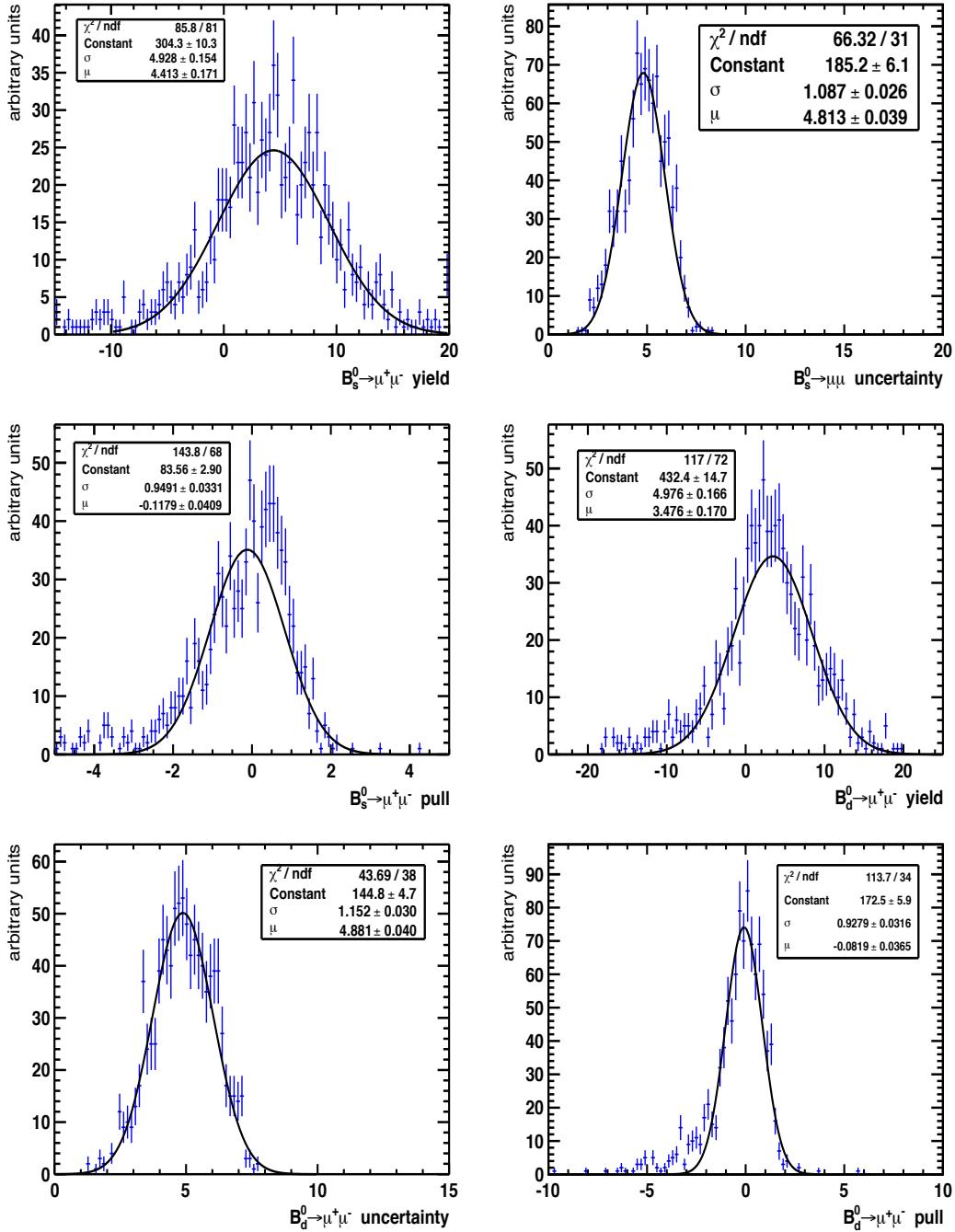


Figure 6.16: Distributions of signal yields, uncertainties and pulls from 1000 toy experiments.

Chapter 7

The measurement of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$

Sec. 3.1 describes the approach used to normalize the number of signal $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ events. The decays used for this normalization are: $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$, $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)\phi(\rightarrow K^+ K^-)$, and $B^0 \rightarrow K\pi$ (including charged conjugated processes). In order to translate the number of observed events into a \mathcal{B} we need to compute the factor multiplying the signal yield. In Sec. 7.1 we analyze the different contributions entering such factor, while Sec. 7.2 reports the values of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ we obtain with our 2011 data sample.

7.1 Normalization of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$

The signal (sig) branching fraction can be extracted using the following formula:

$$\mathcal{B} = \mathcal{B}_{nc} \times \frac{\epsilon_{nc}^{rec} \epsilon_{nc}^{sel|rec} \epsilon_{nc}^{trig|sel}}{\epsilon_{sig}^{rec} \epsilon_{sig}^{sel|rec} \epsilon_{sig}^{trig|sel}} \times \frac{f_{nc}}{f_{sig}} \times \frac{N_{sig}}{N_{nc}} = \alpha \times N_{sig}, \quad (7.1)$$

where α denotes the normalization factor and nc normalization channel. f_{sig} and f_{nc} refer to the probabilities that a b -quark hadronizes into a B_s^0 (B^0) or into the b -meson of the normalization channel. The overall efficiencies are separated in different factors: ϵ^{rec} is the reconstruction efficiency of all the final state particles of the decay in the acceptance of the detector; $\epsilon^{sel|rec}$ the efficiency of selecting reconstructed candidates; and $\epsilon^{trig|sel}$ the efficiency of triggering reconstructed and selected candidates.

The criteria desired in a normalization channels are:

- low branching fraction uncertainty;
- similar trigger efficiency as signal;
- same number and type of final particles as signal;
- same fragmentation factor (f).

All possible normalization samples fail in fulfilling all these requirements, hence we use the three channels mentioned at the beginning of the chapter in order to decrease systematic uncertainties.

In the following paragraphs we give a brief overview of all the elements defining α (Eq. 7.1). For more details we refer the reader to Ref. [43].

Normalization channels' branching fractions The first column in Tab. 7.1 reports the branching fractions of the decays used in the normalization. The largest uncertainty corresponds to the decay $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \phi(\rightarrow K^+ K^-)$.

Ratio of production fractions LHCb determines the fragmentation fraction $\frac{f_s}{f_d}$ in two different ways: using the relative abundance of $B_s^0 \rightarrow D_s^- \pi^+$, $B^0 \rightarrow D^- K^+$, and $B^0 \rightarrow D^- \pi^+$ [50] and using semileptonic $B \rightarrow DX$ decays [51]. The average of the two LHCb results is:

$$\frac{f_s}{f_d} = \frac{f_s}{f_u} = 0.267^{+0.021}_{-0.020}. \quad (7.2)$$

Ratio of reconstruction efficiencies The reconstruction efficiency comprises not only the efficiency to reconstruct all the tracks in the final state, but also the detector acceptance efficiency, or fraction of tracks in the LHCb geometrical acceptance. The reconstruction efficiency is obtained using simulation and then validated in data.

The ratio of reconstruction efficiencies depends on the reconstruction efficiency of an extra track (kaon) when the normalization is performed using $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^+$ decays, of two extra tracks when using $B_s^0 \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) \phi(\rightarrow K^+ K^-)$, and of the different phase space of the muons from $J/\psi \rightarrow \mu^+ \mu^-$ and signal.

The acceptance and reconstruction efficiency for the extra track in $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^+$ can be probed using the ratio of events of $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^+$ with respect to $B^0 \rightarrow J/\psi K^{*0}$ which contains four tracks in the final state:

$$\frac{\epsilon_{REC}(B^+ \rightarrow J/\psi K^+)}{\epsilon_{REC}(B_s^0 \rightarrow \mu^+ \mu^-)} \simeq \frac{\epsilon_{REC}(B^0 \rightarrow J/\psi K^{*0})}{\epsilon_{REC}(B^+ \rightarrow J/\psi K^+)}. \quad (7.3)$$

The ratio data/MC reconstruction efficiencies has been estimated:

$$\frac{\epsilon_{data}^{REC}(B^+ \rightarrow J/\psi K^+)/\epsilon_{data}^{REC}(B^0 \rightarrow J/\psi K^{*0})}{\epsilon_{MC}^{REC}(B^+ \rightarrow J/\psi K^+)/\epsilon_{MC}^{REC}(B^0 \rightarrow J/\psi K^{*0})} = 1.03 \pm 0.04 \pm 0.06, \quad (7.4)$$

where the first error is due to statistics and it is dominated by the number of simulated events, the second is the uncertainty related to the branching fraction of $B^0 \rightarrow J/\psi K^{*0}$ and $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-) K^+$. This result validates the use of simulation for the reconstruction efficiency.

Ratio of selection efficiencies The ratio of selection efficiencies is extracted using simulation and then cross-checked in data.

The main differences between signal and normalization channels are related with the $IP\chi^2$ requirement on the extra tracks and the different kinematic properties for the muons from J/ψ and B_s^0 or B^0 .

The selection efficiencies have been computed using simulation in three different scenarios: unsmeared, smeared, and oversmeared by 50% samples to estimate possible residual effects.

The ratio of selection efficiencies for the three normalization channels is determined from the smeared MC:

$$\epsilon_{B^+ \rightarrow J/\psi K^+}^{SEL|REC} / \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{SEL|REC} = 0.84 \pm 0.01, \quad (7.5)$$

$$\epsilon_{B_s^0 \rightarrow J/\psi \phi}^{SEL|REC} / \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{SEL|REC} = 0.59 \pm 0.02, \quad (7.6)$$

$$\epsilon_{B^0 \rightarrow K\pi}^{SEL|REC} / \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{SEL|REC} = 1.11 \pm 0.01, \quad (7.7)$$

where the errors are from MC statistics.

Even though the absolute selection efficiencies between unsmeared and smeared simulated events vary between 5 and 7%, the ratio between signal and all three normalization channels stays constant within 1% because the efficiency change in the signal goes in the same direction as in the corresponding normalization channel, and therefore cancels in the ratio.

The second column of Tab. 7.1 gives the ratio of the product of reconstruction and selection efficiencies between signal and normalization channels.

Ratio of trigger efficiencies Data driven techniques are used to determine the trigger efficiencies. Detached $J/\psi \rightarrow \mu^+ \mu^-$ decays in data, after re-weighting the momentum and the p_T of muons to match the expected distributions of $B_s^0 \rightarrow \mu^+ \mu^-$, allow to extract the signal efficiency. We use the trigger unbiased $B_{(s)}^0 \rightarrow h^+ h^-$ events to estimate $\epsilon_{B^0 \rightarrow K\pi}^{trig|sel}$, with the assumption that all $B_{(s)}^0 \rightarrow h^+ h^-$ modes have the same trigger efficiency.

To obtain the trigger efficiency from data we use:

$$\epsilon^{trig} = \frac{N^{trig}}{N^{sel}} = \frac{N^{TIS}}{N^{sel}} \frac{N^{trig}}{N^{TIS}} = \epsilon^{TIS} \frac{N^{trig}}{N^{TIS}}. \quad (7.8)$$

Both N^{trig} and N^{TIS} are observables and ϵ^{TIS} is the efficiency to trigger with no information from the signal. This last quantity can be obtained directly from data $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$:

$$\epsilon^{TIS} = (5.1 \pm 0.9 \pm 0.4)\%, \quad (7.9)$$

where the first error is statistical and it is dominated by the number of TIS events. The second error is the systematic uncertainty which is evaluated using simulation from the difference of the trigger efficiencies between the true value and the one obtained using Eq. 7.8.

To estimate the trigger efficiency for the signal, an efficiency map as a function of the largest p_T and largest IP of the muons from $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$ is computed. Applying this efficiency map to the muon spectrum of $B_s^0 \rightarrow \mu^+ \mu^-$ simulated sample leads to:

$$\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^{TRIG|SEL} = (91.4 \pm 0.4 \pm 3.9)\% \quad (7.10)$$

Table 7.1: Summary of the components and their uncertainties entering in the normalization factors for the three considered channels.

\mathcal{B} $(\times 10^{-5})$	$\frac{\epsilon_{cal}^{REC}}{\epsilon_{sig}^{REC}} \frac{\epsilon_{cal}^{SEL REC}}{\epsilon_{sig}^{SEL REC}}$	$\frac{\epsilon_{cal}^{TRIG SEL}}{\epsilon_{sig}^{TRIG SEL}}$	N_{cal}	$\alpha_{B_d \rightarrow \mu^+ \mu^-}^{cal}$ $(\times 10^{-11})$	$\alpha_{B_s \rightarrow \mu^+ \mu^-}^{cal}$ $(\times 10^{-10})$	
$B^+ \rightarrow J/\psi K^+$	6.0 ± 0.2	0.50 ± 0.01	0.95 ± 0.02	$340\,100 \pm 4500$	8.5 ± 0.4	3.2 ± 0.3
$B_s^0 \rightarrow J/\psi \phi$	3.4 ± 0.9	0.24 ± 0.01	0.95 ± 0.02	$19\,000 \pm 160$	11 ± 3	4.2 ± 1.1
$B^0 \rightarrow K\pi$	1.9 ± 0.1	0.86 ± 0.03	0.047 ± 0.003	$10\,100 \pm 920$	8.0 ± 1.0	2.9 ± 0.4

The systematic uncertainty arises from the combination of two errors: one associated to the method used in Eq. 7.8 applied to $B_s^0 \rightarrow \mu^+ \mu^-$ (3%) and the second due to 2.5% of $B_s^0 \rightarrow \mu^+ \mu^-$ events that are triggered not using the muon triggers.

The third column of Tab. 7.1 gives the values of the trigger efficiency ratios.

In the fourth column of Tab. 7.1 we also report the yields obtained for each normalization channel. The $B^+ \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) K^+$ invariant mass is modeled with a double Gaussian function for signal, while the background is empirically modeled with two functions: an exponential for the combinatorial background, and a Gaussian on the left-hand part of the invariant mass distribution for the physical background. The quoted number of signal events in Tab. 7.1 has a 1.3% systematic uncertainty associated to the difference of yields obtained when applying a background subtraction method.

The $B_s^0 \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) \phi (\rightarrow K^+ K^-)$ signal is modeled with a double Gaussian function and an exponential function for the background as well. In this case, the associated systematic uncertainty due to the background subtraction method is 0.4%.

Normalization factors Tab. 7.1 summarizes all different parameters defining the normalization factors.

The final normalization factors used to estimate the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ are a weighted average of those shown in Tab. 7.1:

$$\alpha_{B_s^0 \rightarrow \mu^+ \mu^-} = (3.19 \pm 0.28) \times 10^{-10}, \quad (7.11)$$

$$\alpha_{B^0 \rightarrow \mu^+ \mu^-} = (8.38 \pm 0.39) \times 10^{-11}. \quad (7.12)$$

7.2 $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ obtained for 1 fb^{-1}

Chap. 6 and Sec. 7.1 present all the necessary ingredients to extract $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ in the 1 fb^{-1} data sample:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (1.44^{+1.64}_{-1.14}) \times 10^{-9}, \quad (7.13)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (0.30^{+0.53}_{-0.38}) \times 10^{-9}, \quad (7.14)$$

where the errors are statistical only.

7.2.1 Estimate of the systematic uncertainties

In Sec. 6.2 we introduced a technique, based on the generation of Gaussian-distributed nuisance parameters, to estimate the systematic uncertainties due to the parameterization of the different *pdfs* and described all possible sources of systematic uncertainties.

We utilize the aforementioned technique to estimate the uncertainty due to all parameters; and to all the parameters but those related with the BDT background parameterization (see previous chapter). Now we include $\alpha_{B_s^0 \rightarrow \mu^+ \mu^-}$ and $\alpha_{B^0 \rightarrow \mu^+ \mu^-}$, and their uncertainties, in both computations.

Figs. 7.1 and 7.2 display the number of toy experiments as a function of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ compatible with Gaussian-distributed nuisance parameters. As described in the previous chapter, the systematic uncertainties arise from the widths of the Gaussians used to fit the lower and the upper edges of the distributions, also reported in Tab. 7.2. In this table, the quoted values on the first column corresponds to the systematic uncertainty when all the nuisance parameters are free; the second column gives the results when all the parameters but those that define the BDT for the background are left free; the third column specifies the systematic uncertainty associated with the implementation of a different model for the background BDT; the fourth column quotes the systematic uncertainty given the different peaking background mass shapes; and the last column reports the systematic uncertainty due to the BDT range fit. As explained through the last chapter, these are the main ingredients to estimate the final systematic uncertainty.

The positive systematic uncertainty arises from the first, forth and the fifth columns in Tabs. 7.2 added in quadrature. For the negative uncertainty estimation, we account that the difference between the fit result using 3 exponentials and the kernel estimator exceeds the systematic estimate from varying the background BDT nuisance parameters. For this reason, we use the values on the second, third and fifth columns added in quadrature.

The final results including the systematic uncertainty are:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (1.4^{+1.7}_{-1.4}) \times 10^{-9}, \quad (7.15)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (0.3^{+0.6}_{-0.5}) \times 10^{-9}. \quad (7.16)$$

7.3 Comparison of the measured $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ with the exclusion limits

The results in the previous section (and last chapter) conclude that there is no evidence of $B_s^0 \rightarrow \mu^+ \mu^-$ nor $B^0 \rightarrow \mu^+ \mu^-$ events. Therefore, the exclusion limits using a standard procedure such as the CL_s method [41] were computed and published in Ref. [4]. The classifier described in Chap. 5 is used to extract such limits.

In the aforementioned publication, the quoted most probable value for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, $(0.8^{+1.8}_{-1.3}) \times 10^{-9}$, is obtained via a different approach as described in this section. This

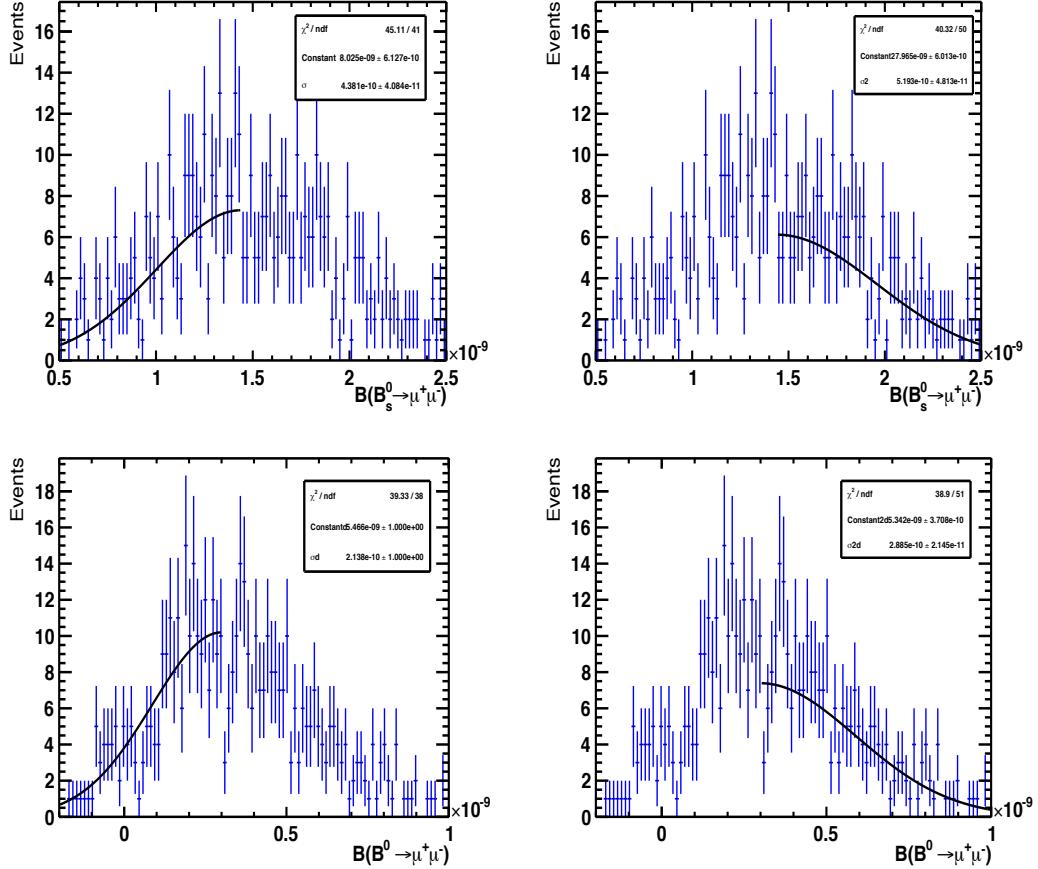


Figure 7.1: Distribution of the number of toy experiments in terms of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (top) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (bottom) obtained generating the nuisance parameters of all the components of the fit according to Gaussian distributions. The total number of toy experiments performed is 500.

method consists in a simultaneous unbinned fit in the usual eight bins of BDT in which the signal yields are constrained to the fractions of BDT reported in Fig. 6.3. One difference between this method and the presented extended maximum likelihood fit is that the background BDT is parameterized for the latter, allowing to take into consideration more information related to it. On the other hand, the simultaneous fit approach is more similar to the CL_s method (see next section) as the BDT is binned likewise. One advantage of the two-dimensional fit is that the systematic uncertainties related to each component can be studied separately. The results obtained with both methods agree within the uncertainties, as referred in Ref. [4].

Sec. 7.3.1 briefly recapitulates the method used to obtain the published limits, and Sec. 7.3.2 compares the results of the extended maximum likelihood fit with the exclusion limits.

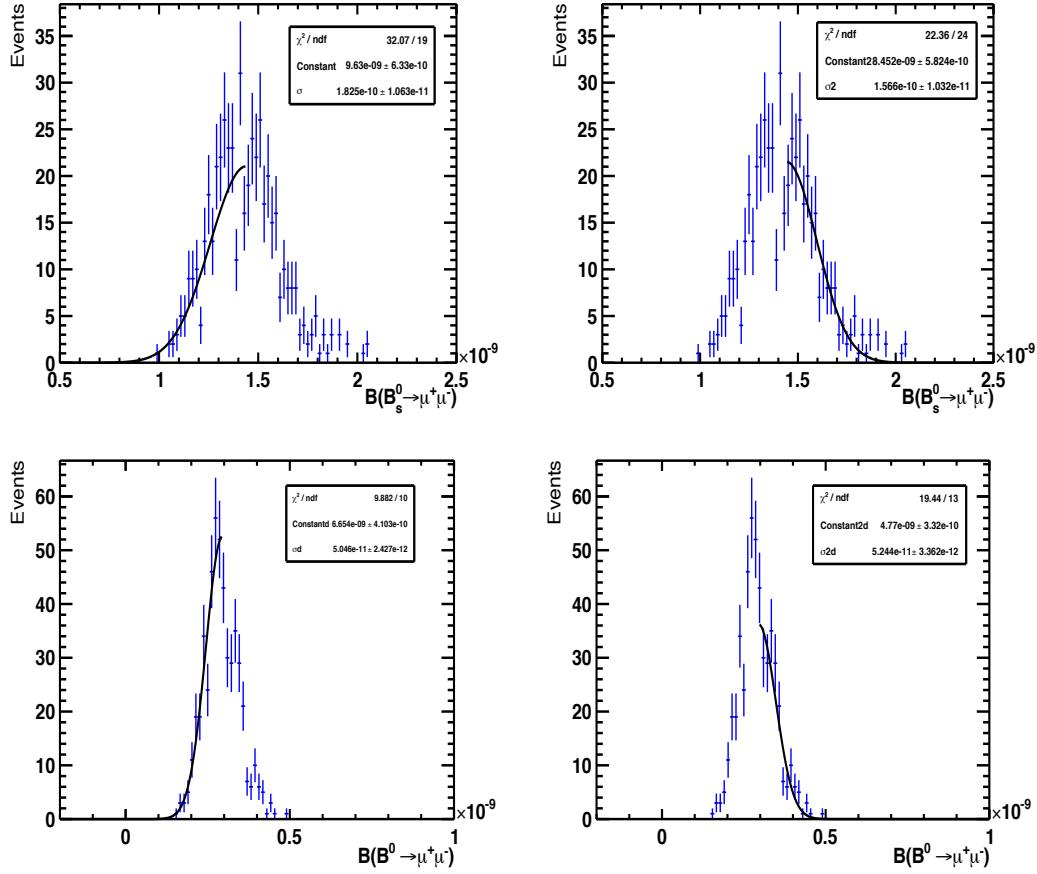


Figure 7.2: Distribution of the number of toy experiments in terms of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (top) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (bottom) obtained generating the nuisance parameters of all the components, excluding the background BDT component, according to Gaussian distributions. The total number of toy experiments performed is 500.

Table 7.2: Contributions to the final systematic uncertainties of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. In column (1) we report the systematic uncertainty due to all contributions, including α . In (2) the computation is as in (1), but fixing the parameters that define the background BDT. (3), (4) and (5) show the contributions due to the different BDT background model, the peaking background mass and the BDT range, respectively.

	(1)	(2)	(3)	(4)	(5)
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	+0.52 -0.45	+0.16 -0.18	+0.00 -0.73	+0.01 -0.03	+0.03 -0.00
$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	+0.29 -0.21	+0.05 -0.05	+0.00 -0.29	+0.03 -0.01	+0.00 -0.04

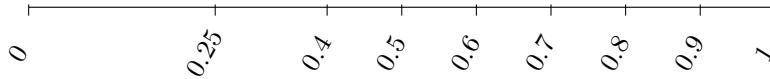


Figure 7.3: BDT binning.

Figure 7.4: Invariant mass binning in the B_s^0 mass region. The correspondent binning for B^0 is shifted by 88 MeV/ c^2 towards lower mass values.

7.3.1 Exclusion limits

The BDT classifier output described in Chap. 5 and the invariant mass for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ candidates are binned according to the ranges presented in Figs 7.3 and 7.4, in order to maximize the sensitivity of the analysis Refs. [43] and [52]. The binning for the BDT output is the same as in Chap. 5.

The compatibility of the observed events in each of the previous seventy-two bins, given a $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ hypothesis, is evaluated with the CL_s method [41]. This procedure, originally used in Higgs searches at LEP, accounts for the compatibility of a data sample with a signal plus background (CL_{s+b}) and background only hypothesis (CL_b).

Before detailing the results of the limits, we briefly describe the method followed to obtain them.

For a given branching fraction hypothesis, and for each bin, a large number of two types of pseudo-experiments is generated: one in which the data is background; and another containing background and signal. The number of signal events corresponds to the given branching fraction hypothesis ($\mathcal{B} = \alpha \times N_{signal}$). Each pseudo-experiment has its own test statistic $Q = \prod \mathcal{P}(d_i, s_i + b_i) / \mathcal{P}(d_i, s_i + b_i)$, where $\mathcal{P}(d_i, x)$ denotes the probability that the expected number of events of signal and background (x) fluctuates to give the observed d_i in a given bin i .

Fig. 7.5 [52] shows the distributions of $-2\ln Q$ for the signal plus background pseudo-experiments (blue) and background only (green). The red vertical line is the $-2\ln Q$ value for a particular data sample. This figure allows to visualize the values of CL_{s+b} (area under the blue curve at the right of the red line) and CL_b (area under the green distribution at the right of the red line). The CL_s for a particular data sample, and for a given branching fraction, is extracted with the CL_{s+b} and the CL_b : $CL_s = CL_{s+b} / CL_b$.

$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ limits

Fig. 7.6 displays the CL_s distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ obtained with the method described above for different branching fraction hypothesis for the unblinded 1 fb^{-1} data sample (solid blue curve). The yellow band constitutes the CL_s region compatible with 68% of performed toy experiments with a background plus SM signal hypothesis. The dashed black

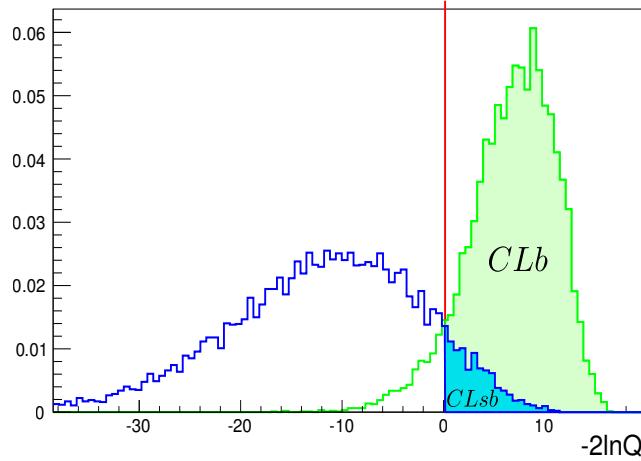


Figure 7.5: Distribution of the test statistics for a set of generated pseudo-experiments with a background only hypothesis (green) and signal plus background hypothesis (blue) for an assumed $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. The vertical line corresponds to the value of the test statistic for a certain data sample; the green area is the CL_b and the blue area CL_{s+b} [52].

line is the median of the CL_s extracted from such toy experiments. The measured limits at 95% CL ($CL_s=0.05$) and 90% CL ($CL_s=0.1$) correspond to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 4.5 \times 10^{-9}$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 3.8 \times 10^{-9}$, respectively.

Moreover, Fig. 7.7 shows the CL_s distribution for $B^0 \rightarrow \mu^+ \mu^-$. Here, the yellow band corresponds with the CL_s region compatible with 68% of toy experiments with a background only hypothesis. The measured limits at 95% CL and 90% are $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 10.3 \times 10^{-10}$ and $BR(B^0 \rightarrow \mu^+ \mu^-) < 8.1 \times 10^{-10}$.

7.3.2 Comparison of the measured $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ with the exclusion limit

We use a study based on simulated toy experiments to cross-check the compatibility of the fit results with the exclusion limits obtained with the CL_s method. The procedure is exactly the same as in Sec. 6.3 where we generated 1000 toys with the expected yields found in data. Here we generate 10000 toys instead. The reason to do this is clarified below.

Fig. 7.8 shows the distribution of toy experiments as a function of the number of $B_s^0 \rightarrow \mu^+ \mu^-$ compatible with the yields observed in data. We compute the fractional coverage, from yields above zero up to the expected SM value 10, to be of 83%. The yields corresponding to 90% and 95% confidence levels are also computed and reported in Tab. 7.3, after translating the yields into branching fractions. All above confidence levels have an associated uncertainty of 2%, whereas for the toy experiments in Chap. 6 the uncertainties were of about 6%.

The reported confidence levels in Tab. 7.3 have no systematic uncertainty associated. The comparison is therefore performed with the expected limits with no systematic uncertainties. Fig. 7.9 displays the CL_s as a function of branching fraction without considering systematic uncertainties for the unblinded data sample. The branching fractions corre-

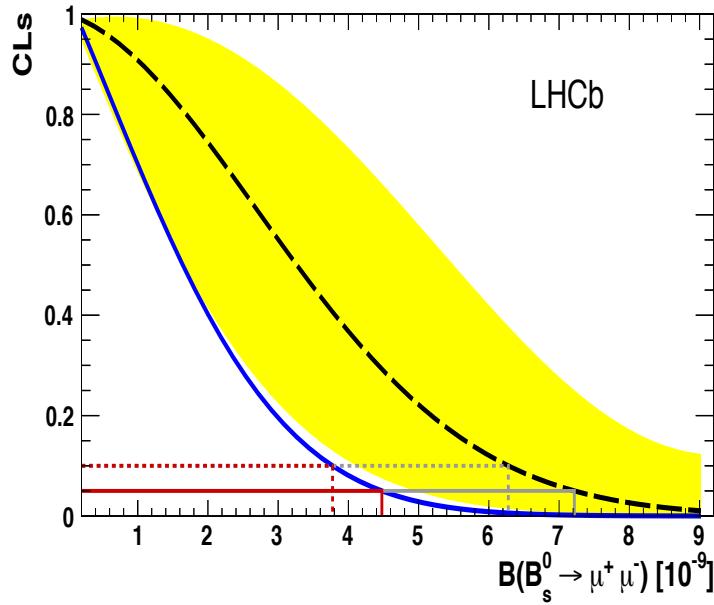


Figure 7.6: CL_s as a function of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. The dashed black line is the median of the expected CL_s value, while the yellow area covers the region of 68% compatibility with the mean value. The solid blue line corresponds to the observed CL_s . The 90% and 95% CL are given by the horizontal dashed and solid red lines, respectively.

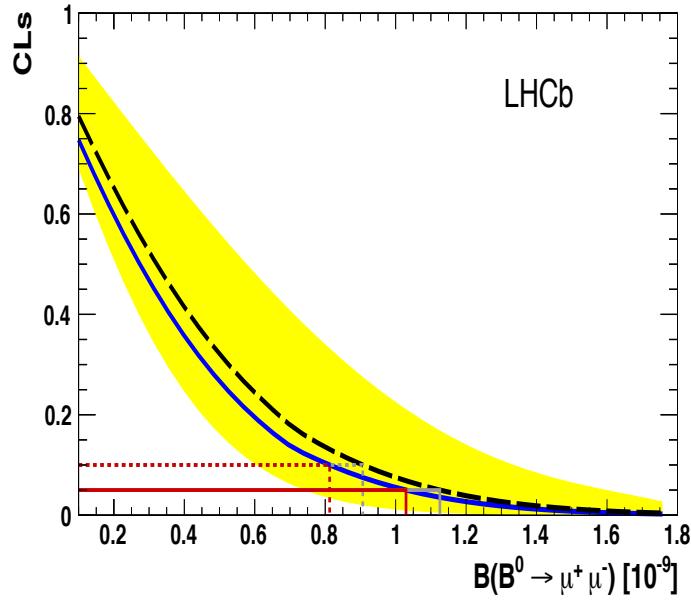


Figure 7.7: CL_s as a function of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. The dashed black line is the median of the expected CL_s value, while the yellow area covers the region of 68% compatibility with the mean value. The solid blue line corresponds to the observed CL_s . The 90% and 95% CL are given by the horizontal dashed and solid red lines, respectively.

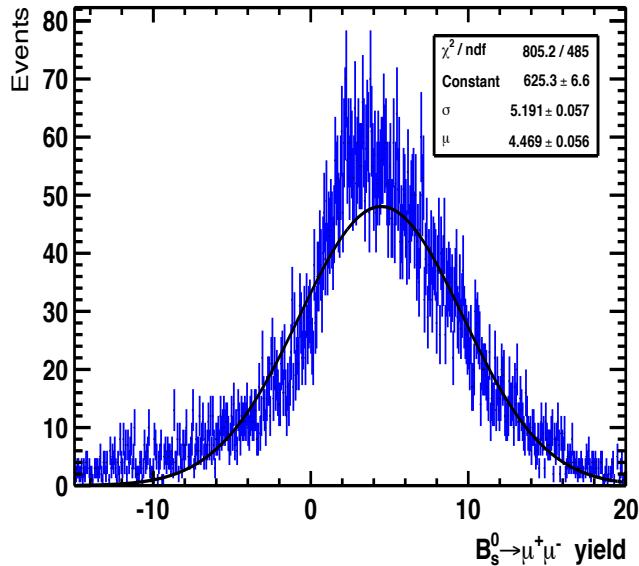


Figure 7.8: Distribution of number of toy experiments giving a certain $B_s^0 \rightarrow \mu^+ \mu^-$ yield. The total number of generated toys is 10000.

sponding to 83%, 90% and 95% CL are reported in Tab. 7.3, and are consistent with the confidence levels computed from the toy experiments.

7.4 Implications of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ on NP models

This section covers the impact of the measured $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ on few physics models beyond the SM.

Sec. 7.4.1 describes the computation of the allowed parameter space for the CMSSM (see Chap. 1) given the current exclusion limit, and under certain constraints. In Sec. 7.4.2 we present the impact on several BSM models.

7.4.1 Implications on CMSSM

The results presented in this section are obtained with SuperIso [53] which employs experimental flavor observables as input to extract predictions for several new physics scenarios. Here we focus on the constrained version of the MSSM, the CMSSM, which has five free parameters.

Fixing $\tan \beta$, the sign of μ and A_0 , we scan m_0 and $m_{1/2}$ in the range [0-2] TeV. The lightest supersymmetric particle LSP is forced to be neutral in the computation scan. Fig. 7.10 (a) shows in blue the allowed parameter space for $\tan \beta = 10$, while in grey we represent the direct constrain arising from $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. In the same figure, the white region represents the region excluded by current experimental data and by requiring a neutral LSP. We require $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 5 \times 10^{-9}$, corresponding to the observed limit (4.5×10^{-9})

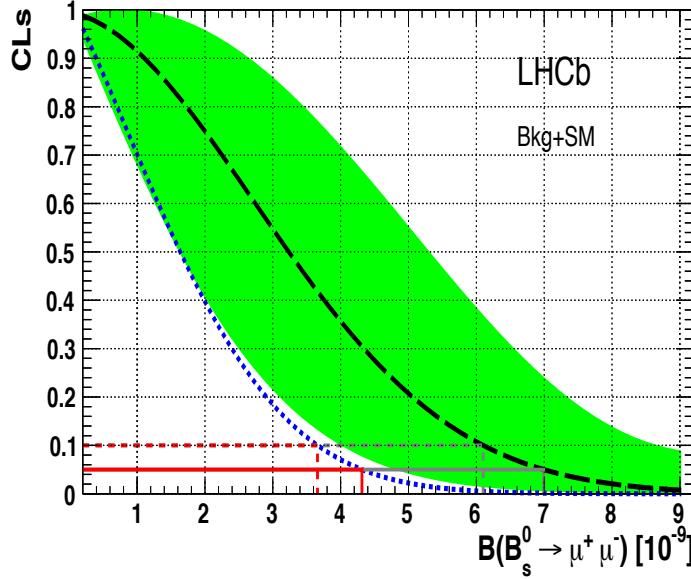


Figure 7.9: CL_s as a function of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. The dashed black line is the median of the expected CL_s value, while the yellow area covers the region of 68% compatibility with the mean value. The solid blue line corresponds to the observed CL_s . The 90% and 95% CL are given by the horizontal dashed and solid red lines, respectively.

Table 7.3: Exclusion limits extracted from the CL_s method and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ values arising from the fit at different confidence levels. All values do not take into account any systematic uncertainty.

C.L.	2D fit	CL_s
$(83 \pm 2)\%$	3.2×10^{-9}	3.2×10^{-9}
$(90 \pm 2)\%$	3.8×10^{-9}	3.7×10^{-9}
$(95 \pm 2)\%$	4.4×10^{-9}	4.3×10^{-9}

plus an 11% contribution to take into account systematic uncertainties due to theoretical calculations, mostly driven by the uncertainty in f_s .

The exclusion due to this observable, although negligible at low $\tan \beta$ (a), is significant for larger $\tan \beta$ values as shown in Fig 7.10 (b) for $\tan \beta = 50$.

7.4.2 Implications on general models

Fig. 7.11 shows $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ vs $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ predicted by the Minimal Flavor Violation hypothesis and by different NP scenarios: several MSSM constrained models and the minimal fourth generation framework with a single Higgs (so called SM4). This last model considers a sequential fourth generation of quarks and (t' , b' , τ' , and ν') which introduces new CP sources and may introduce a dynamical electroweak symmetry breaking.

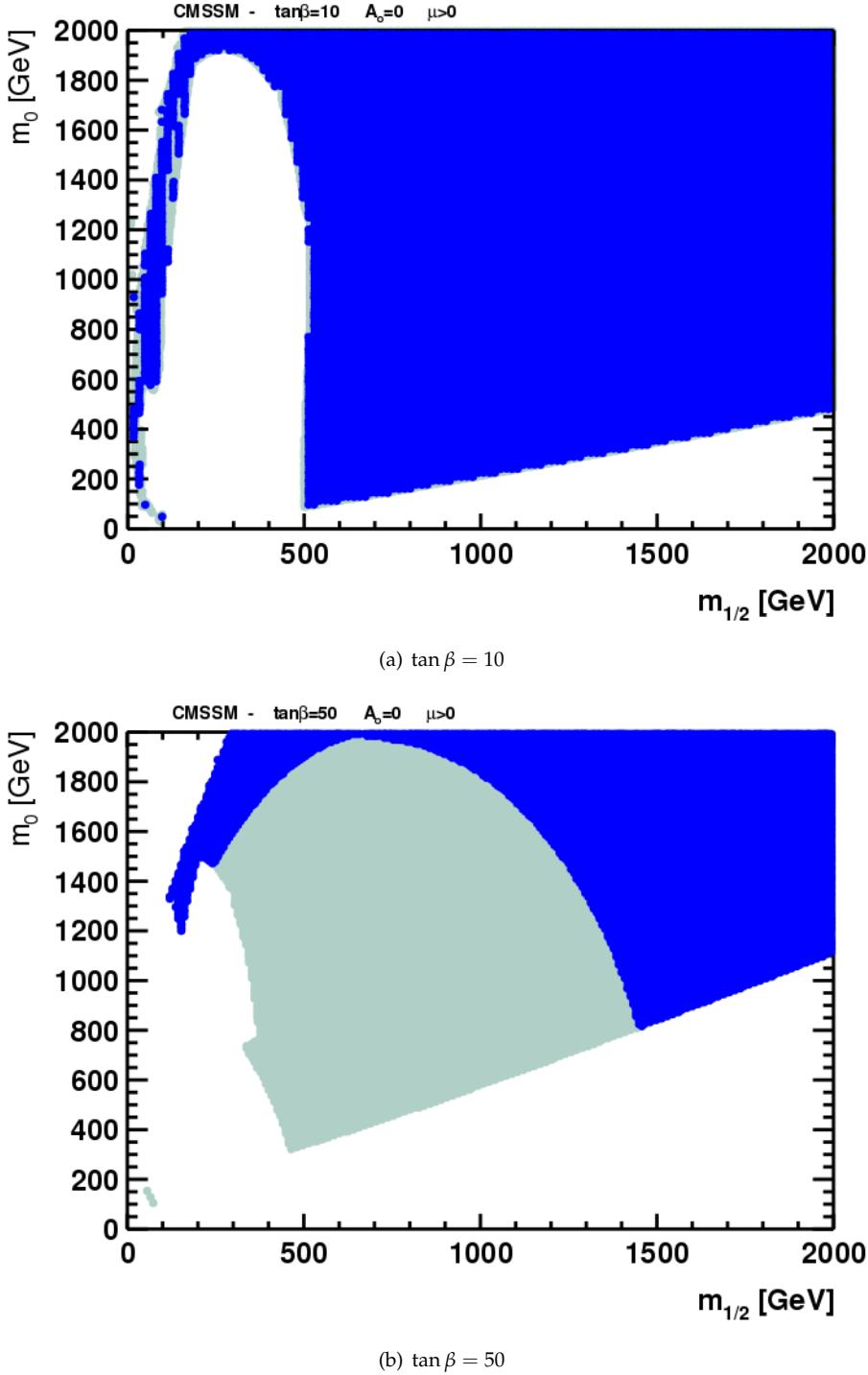


Figure 7.10: Parameter space in CMSSM for $\tan \beta = 10$ (a) and $\tan \beta = 50$ (b), for $A_0 = 0$ and sign of μ positive. The white areas are excluded by a charged lightest supersymmetric particle and by the Higgs boundaries. The grey areas correspond to the excluded areas giving the obtained $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ limit [4] and computed using Refs. [53], [54], and [55]. The blue areas represent the remaining allowed regions.

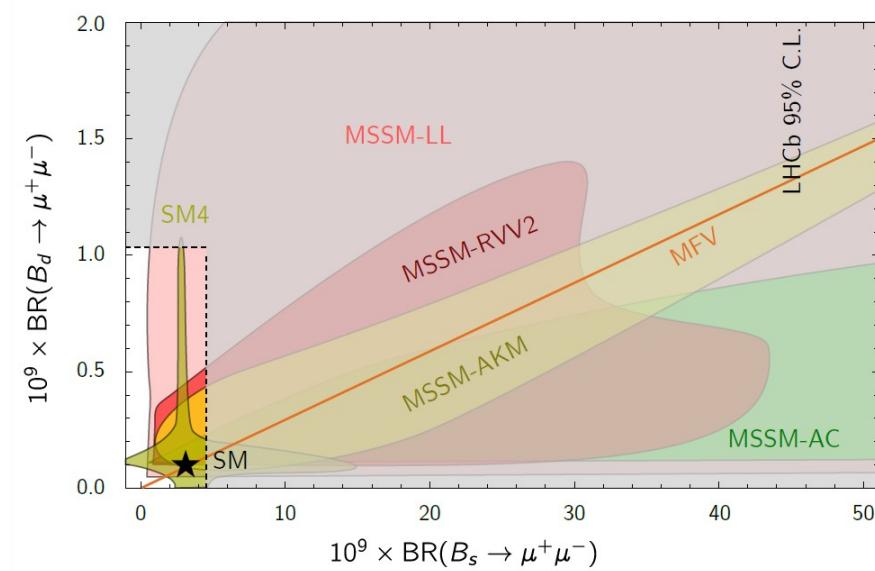


Figure 7.11: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ vs $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ predicted by the Minimal Flavor Violation (MFV) hypothesis and by different NP framework with a single Higgs (SM4). Superimposed to the theoretical predictions, the shade area corresponds to the LHCb exclusion limits [4]. The SM value is marked with a star.

In the same figure, we find superimposed the constraint directly arising from the LHCb limits. These limits drastically reduce the allowed region, and exclude most of the predicted values by the different MSSM models. Being close to the SM prediction, represented with a star in said figure, the allowed region predicted by SM4 is not as constrained as the supersymmetric scenarios.

7.5 Conclusions

This chapter represents the final step in the measurements of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$. The obtained results are:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (1.4^{+1.7}_{-1.4}) \times 10^{-9}, \quad (7.17)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (0.3^{+0.6}_{-0.5}) \times 10^{-9}. \quad (7.18)$$

Although compatible with the SM values at a 83% CL without systematic uncertainties, the most probable $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ value extracted with the extended maximum likelihood is slightly ($\sim 1\sigma$) lower than the SM.

Given the lack of signal evidence, the computed exclusion limits represent the most restrictive limit for these decays [4]. Such limits strongly constrain MSSM models.

Conclusion

At the end of my thesis work the LHCb collaboration has published the world best limits on $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ [4]. The improvement on the exclusion limit between the 2010 D0's public note [56] and the current limit is of almost one order of magnitude for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. Tab. 7.4 reports the exclusion limits obtained by different experiments since 2010.

Table 7.4: Compilation of the different exclusion limits at 95% CL for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, before and after the work presented in this document.

Publication	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$
D0 (2010) [56]	51×10^{-9}	-
CDF (2011) [57] ¹	43×10^{-9}	76×10^{-10}
LHCb (2011) [39]	56×10^{-9}	15×10^{-9}
ATLAS (2012) [58]	22×10^{-9}	-
LHCb (2012) [40]	15×10^{-9}	38×10^{-10}
CMS (2012) [59]	7.7×10^{-9}	18×10^{-10}
LHCb (2012) [40]	4.5×10^{-9}	10×10^{-10}

Not only the measurements of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ represent a complement to the searches performed at the LHC multi-purpose experiments, ATLAS and CMS , but they are also a benchmark of new physics models beyond the SM. For instance, the limit on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ strongly constraints the m_0 and $m_{1/2}$ CMSSM parameter space below 2 TeV for $\tan\beta = 50$.

The core of this thesis comprises two main topics: the background rejection with multivariate analysis techniques and the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ signal yields extraction.

This document describes the main sources of background expected for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, and presents the procedure for their reduction. An optimized multivariate classifier based on the boosted decision trees technique, and used in the selection phase of the analysis, allows to drastically reduce the $B_{(s)}^0 \rightarrow h^+ h^-$ background, and also rejects about 76% of the combinatorial $b\bar{b} \rightarrow \mu\mu X$ background keeping a signal efficiency of about 92%.

After the selection process, a further step in discriminating the signal from the background explores another BDT multivariate classifier, optimized to have a large background

¹In the same publication an excess of $B_s^0 \rightarrow \mu^+ \mu^-$ candidates is reported. For such excess, $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)=1.8^{+1.1}_{-0.9}$ is determined.

rejection in the low signal efficiency region, namely below 50%. This final classifier leads to a background rejection of 99.9%, after the aforementioned selection, for a signal efficiency of 50% according to simulation.

We estimate the signal yields present in our data sample with an extended maximum likelihood fit in invariant mass and BDT output. The validation of the fit using simulation reflects the proper estimation of the statistical uncertainties. Systematic uncertainties, mainly dominated by the parametrization of the background BDT classifier output, have been carefully studied and taken into account in the final results with 1 fb^{-1} of data:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (1.4 \left(\begin{array}{c} +1.6 \\ -1.1 \end{array} \right)_{\text{(stat)}} \left(\begin{array}{c} +0.5 \\ -0.8 \end{array} \right)_{\text{(syst)}}) \times 10^{-9}, \quad (7.19)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (0.3 \left(\begin{array}{c} +0.5 \\ -0.4 \end{array} \right)_{\text{(stat)}} \left(\begin{array}{c} +0.4 \\ -0.3 \end{array} \right)_{\text{(syst)}}) \times 10^{-9}. \quad (7.20)$$

The probability for the measured $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ to fall between zero and the expected SM value is 83%, according to simulation.

Perspectives concerning the analysis of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$

The acquired experience through the work presented in this document serves to summarize the main topics upon which we can improve in future analyses:

- the *multivariate selection (mvs)*, in principle developed with the intention of finding a better way to suppress the large $B_{(s)}^0 \rightarrow h^+ h^-$ background, helped in ameliorating the performances of the final discriminant classifier BDT. Therefore, an improvement of this *mvs*, together with a tighter selection requirement would potentially improve the trend of the analysis performances. Furthermore, a tighter selection requirement allows a better parameterization of the background BDT, reducing the systematic uncertainties associated with it, and, consequently, the uncertainty of the final branching fractions.
- The optimization of the discriminant BDT in the low signal efficiency region had an unavoidable drawback, which was the lack of simulated events in this region. During the preparation of this document, a large sample of simulated $b\bar{b} \rightarrow \mu\mu X$ events with high BDT output is under production. This sample will allow to study a new optimization of the classifier which could translate into a higher background rejection;
- The performances of the discriminant classifier are related with the number of events used in the training process. If limited by simulated statistics, a possible solution would be to train the BDT with data background events.

Assuming the same sensitivity as for the presented analysis, the probability of a 3σ observation of an SM-like $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ has been computed as a function of luminosity. Assuming an integrated luminosity of 2.5 fb^{-1} collected at the end of 2012, the probability of a 3σ observation is 50%. Moreover, the LHCb upgrade [60] will allow to reach uncertainties on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ of 0.15×10^{-9} with 50 fb^{-1} , smaller than the current SM prediction.

Bibliography

- [1] A. Buras. Minimal Flavor Violation and Beyond: Towards a Flavor Code for Short Distance Dynamics. 2011. [hep-ph/1012.1447].
- [2] CDF collaboration. Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays in 3.7 fb^{-1} of $p\bar{p}$ collisions with CDF II. *CDF Public Note* 9892.
- [3] D. M. Straub. New physics searches in flavor physics. 2011. [arXiv:1107.0266].
- [4] R. Aaij *et al.* [LHCb collaboration]. Strong Constraints on the Rare Decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. *Phys. Rev. Lett.*, **108**:231801, 2012.
- [5] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, **10**(531), 1963.
- [6] M. Kobayashi and T. Maskawa. CP violation in the Renormalizable Theory of Weak Interactions. *Prog. Theor. Phys.*, **49**(652), 1973.
- [7] S. L. Glashow, J. Iliopoulos and L. Maiani. Weak Interactions with Lepton Hadron Symmetry. *Phys. Rev.*, **D2**(1285), 1970.
- [8] R. Barate *et al.* [LEP Working Group for Higgs Boson Searches]. Search for the Standard Model Higgs Boson at LEP. *Phys. Lett.*, **B565**(61), 2003. [hep-ex/0306033].
- [9] ATLAS Collaboration. Combined search for the Standard Model Higgs boson using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s}=7 \text{ TeV}$ with the ATLAS detector at the LHC. *Phys. Lett.*, **B710**:49–66, 2012.
- [10] CMS Collaboration. Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s}=7 \text{ TeV}$. *Phys. Lett.*, **B710**:26–48, 2012.
- [11] F. Zwicky. Spectral Displacement of Extra Galactic Nebulae. *Helv. Phys. Acta*, **86**(110), 1937.
- [12] M. S. Roberts and A. H. Rots. Comparison of Rotation Curves of Different Galaxy Types. *Astronomy and Astrophysics*, **26**(483), 1973.
- [13] J. Dunkley *et al.* [WMAP collaboration]. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Likelihoods and Parameters from the WMAP Data. *Astrophysics J.*, **180**(306), 2009.
- [14] A. D. Sakharov. Violation of CP Symmetry, C Asymmetry, and Baryon Asymmetry of the Universe. *JETP Lett.*, **5**(24), 1967.

- [15] M. B. Gavela, P. Hernandez, J. Orloff and O. Pene. Standard Model CP Violation and Baryon Asymmetry. *Mod. Phys. Lett.*, **A9**(795), 1994.
- [16] P. Fayet, S. Ferrara. Supersymmetry. *Phys. Rept.*, **32**(249), 1977.
- [17] S. P. Martin. A Supersymmetry Primer. v6 - 2011. [hep-ph/9709356].
- [18] L. Susskind. Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory. *Phys. Rev.*, **D20**(2610), 1979.
- [19] N. Arkani-Hamed, S. Dimopoulos and G. Dvali. The Hierarchy Problem and New Dimensions at a Millimeter. *Phys. Lett.*, **B436**(263), 1998.
- [20] L. Randall and R. Sundrum. A Large Mass Hierarchy from a Small Extra Dimension. *Phys. Rev. Lett.*, **83**(3370), 1999.
- [21] N. Arkani-Hamed, A. G. Cohen and H. Georgi. Electroweak symmetry breaking from dimensional deconstruction. *Phys. Lett.*, **B513**(232), 2001. [hep-ph/0105239].
- [22] G. Isidori and D.M. Straub. Minimal Flavour Violation and Beyond. 2012. [hep-ph/1202.0464].
- [23] C. Bobeth et al. Analysis of neutral Higgs-boson contributions to the decays $B^0 \rightarrow l^+l^-$ and $B^0 \rightarrow Kl^+l^-$. *Phys. Rev.*, **D64**(074014), 2001.
- [24] J. Laiho, R. S. Van de Water, E. Lunghi. Lattice QCD inputs to the CKM unitarity triangle analysis. [hep-ph/0910.2928], 2009.
- [25] A. Schwartz and T. Gershon. <http://www.slac.stanford.edu/xorg/lfag/>.
- [26] A. Buras. BSM models facing the recent LHCb data: A first look. 2012. [hep-ph/1204.5064].
- [27] K. De Bruyn et al.. On Branching Ratio Measurements of B_s Decays. 2012. [hep-ph/1204.1735].
- [28] L. J. Hall and M. B. Wise. *Nucl. Phys.*, **B187**:397–408, 1981.
- [29] H. Logan and U. Nierste. $B_{s,d} \rightarrow l^+l^-$ in a Two Higgs Doublet Model. *Nucl. Phys.*, **B586**(39), 2000. [hep-ph/0004139].
- [30] M. Misiak et. al.. The First Estimate of $\mathcal{B}(\bar{B} \rightarrow X_s \gamma)$ at $\mathcal{O}(\alpha_s^2)$. *Phys. Rev. Lett.*, **98**(022002), 2007. [hep-ph/0609232].
- [31] A. Buras et al.. Higgs-mediated FCNC's: Natural Flavor Conservation vs. Minimal Flavour Violation. 2010. [hep-ph/1005.5310].
- [32] E. Witten. Dynamical Breaking of Supersymmetry. *Nucl. Phys.*, **B188**(513), 1982.
- [33] O. Buechmueller et. al.. Higgs and Supersymmetry. 2012. [hep-ph/1112.3564].
- [34] J. R. Ellis, K. A. Olive and Y. Santoso. Constraining Supersymmetry. *New J. Phys.*, **4**(32), 2002. [hep-ph/0202110].

- [35] J. R. Ellis, K. A. Olive and Y. Santoso. The MSSM Parameter Space with Non-Universal Higgs Masses. *Phys. Lett.*, **B539**(107), 2002. [hep-ph/0204192].
- [36] A. Alves *et al.* [LHCb collaboration]. The LHCb Detector at the LHC. *J. Inst.*, **3**(S08005), 2009.
- [37] L. Evans and P. Bryant. LHC machine. *Journal of Instrumentation*, **3**(08):S08001, 2008.
- [38] T. Sjostrand, L. Lonnblad, S. Mrenna and P. Skands. Pythia 6.3 physics and manual. 2003.
- [39] R. Aaij *et al.* [LHCb collaboration]. Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. *Phys. Lett.*, **B699**:330–340, 2011. [hep-ex/1103.2465].
- [40] R. Aaij *et al.* [LHCb collaboration]. Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. *Phys. Lett.*, **B708**:55–67, 2012. [hep-ex/1112.1600].
- [41] A. Read. Presentation of Search Results: The CLs Technique. *J. Phys.*, **G28**, 2002.
- [42] R. Aaij *et al.* [LHCb collaboration]. Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7 \text{ TeV}$. *Eur. Phys.*, **J(C71)**, 2011.
- [43] C. Adrover *et al.*. Search for the rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ with 1.02 fb^{-1} . *CERN-LHCb-ANA-2011-102*, 2012.
- [44] L. Breiman, J. H. Friedman, R. A. Olsen and C. J. Stone. Classification and Regression Trees. *Wadsworth, Stamford*, 1984.
- [45] TMVA. Toolkit for Multi Variable analysis with ROOT. <http://tmva.sourceforge.net>.
- [46] T. Aaltonen *et al.*. Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays with CDF-II. 2011. arXiv:0508036 [hep-ex].
- [47] C. Adrover, G. Mancinelli, M. Perrin-Terrin and J. Serrano. Measurement of the Branching Ratios of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ Decays using a Maximum Likelihood Unbinned Fit. *CERN-LHCb-INT-2011-023*, 2011.
- [48] Roger Barlow. Extended maximum likelihood. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **297**(3):496 – 506, 1990.
- [49] K. Cranmer. Kernel Estimation in High-Energy Physics. 2000. [hep-ex/0011057].
- [50] C. Carson *et al.*. Measurement of the relative yields of the decay modes $B_{(s)}^0 \rightarrow D_{(s)}^0 \pi$ and $B_{(s)}^0 \rightarrow D_{(s)}^0 K$ and determination of f_d/f_s for 7 Tev pp collisions. *CERN-LHCb-ANA-2010-010*, 2010.
- [51] M. Artuso *et al.*. Measurement of b -hadron production fractions in 7 TeV centre-of-mass energy pp collisions. *CERN-LHCb-ANA-2010-015*, 2010.
- [52] M. Perrin-Terrin, G. Mancinelli. Optimization of the binning of the discriminating variables used in the computation of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ upper limits with the modified frequentis approach. 2012. LHCb-INT-2012-003.

- [53] F. Mahmoudi. SuperIso v3.2: A program for calculating flavour physics observables in 2HDM and supersymmetry. 2011. [arXiv:0808.3144].
- [54] B.C. Allanach. SOFTSUSY: a program for calculating supersymmetric spectra. 2012. [arXiv:hep-ph/0104145].
- [55] P. Bechtle. HiggsBounds: Confronting Arbitrary Higgs Sectors with Exclusion Bounds from LEP and the Tevatron. 2011. [arXiv:0811.4169].
- [56] D0 Collaboration. Search for Rare Decay $B_s^0 \rightarrow \mu^+ \mu^-$. *Phys. Lett. B*, **693**(539), 2010.
- [57] CDF Collaboration. Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. *Phys. Rev. Lett.*, **107**(191801), 2011.
- [58] Atlas Collaboration. Search for $B_s^0 \rightarrow \mu^+ \mu^-$ decay. *CERN-PH-2012-067*, 2012.
- [59] CMS Collaboration. Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays. *BPH-11-2012-020*, 2012.
- [60] LHCb Collaboration. Framework TDR for the LHCb Upgrade. *CERN-LHCC-2012-007*, 2012.