HAL
open science

Rare dileptonic B0(s) meson decays at LHCb

Alessandro Mordà

To cite this version:

Alessandro Mordà. Rare dileptonic B0(s) meson decays at LHCb. High Energy Physics - Experiment [hep-ex]. Aix-Marseille Universite, 2015. English. NNT : . tel-01252492v2

HAL Id: tel-01252492

https://theses.hal.science/tel-01252492v2

Submitted on 14 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ D'AIX-MARSEILLE

ECOLE DOCTORALE PHYSIQUE ET SCIENCES DE LA MATIÉRE

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

Centre de Physique des Particules de Marseille (CPPM) UMR 7346
Centre de Physique Théorique (CPT) UMR 7332

Thèse présentée par

Alessandro Mordá

en vue d'obtenir le grade universitaire de docteur

Rare dileptonic $B^0_{(s)}$ meson decays at LHCb

Soutenue le 28/09/2015 devant le jury composé de:

Eli Ben-Haim	LPNHE, Paris	Examinateur
Marta Calvi	Universitá Milano-Bicocca	Examinateur
Jérôme Charles	CPT, Marseille	Co-directeur de thése
Eric Kajfasz	CPPM, Marseille	President du jury
Emi Kou	LAL, Orsay	Rapporteur
Giampiero Mancinelli	CPPM, Marseille	Directeur de thése
Karim Trabelsi	KEK, Tsukuba	Rapporteur

To my family

Abstract. The $B^0_{(s)} \to \ell\bar{\ell}$ decays are generated by Flavor Changing Neutral Currents, hence they can proceed only through loop processes. For this reason, and because of an additional helicity suppression, their branching ratios are predicted to be very small in the Standard Model (SM). Nevertheless their rates can be modified by the appearance of New Physics (NP) particles inside the loops. In SM their rates are predicted with a very small uncertainty, and from the comparison between the measured values and their theoretical prediction hints on the NP realization patterns can be inferred. A part of the original work presented in this thesis has been devoted to the optimization of the Multi Variate Analysis (MVA) classifier for the search of the $B^0_{(s)} \to \mu^+\mu^-$ with the full dataset collected at LHCb during the first run of the LHC (corresponding to an integrated luminosity of $\sim 3fb^{-1}$). This dataset has also been combined with the one collected by the CMS experiment to obtain the first observation of $B^0_s \to \mu^+\mu^-$. In view of the update of the analysis aiming to improve the sensitivity for the $B^0 \to \mu^+\mu^-$ mode, a new isolation variable, exploiting a topological vertexing algorithm, has been developed and additional studies for a further optimization of the MVA classifier performances have been done. Another original part of the work presented in this thesis concerns the definition of an analysis chain for the search of the $B^0_{(s)} \to \tau^+\tau^-$ modes, which are still largely unexplored. In this thesis the final state where both τ go into three charged π and a ν_τ has been studied. The presence of two ν in the final state of the decay makes the reconstruction of the τ momenta of the two τ. Nevertheless the possibility of measuring the two decay vertexes of the τ, as well as the B candidate production vertex, allows to impose geometrical constraints that can be used in the reconstruction of the τ momenta. In particular, a new algorithm for the full reconstruction of each event of these momenta and of related variables has been presented and discussed.

Resumé. Les désintégrations rares $B^0_{(s)} \to \ell\bar{\ell}$ sont générées par des courants neutres avec changement de la saveur. Pour cette raison, ainsi qu'à cause de la suppression d'hélicité, leurs taux de désintégration sont très petits dans le Modèle Standard (MS), mais la présence de particules virtuelles de Nouvelle Physique peut radicalement modifier cette prédiction. Dans le MS la prédiction théorique des taux de désintégration est très précise, et la comparaison des valeurs mesurées avec ces prédictions théoriques peuvent donner des indications sur la structure de la Nouvelle Physique. Une partie du travail original présenté dans cette thèse est dédié à l'optimisation de l'algorithme d'Analyse Multi Varié (MVA) pour la recherche de la désintégration $B^0_{(s)} \to \mu^+\mu^-$ avec l'échantillon collecté par l'expérience LHCb pendant la première période de fonctionnement du LHC. Cet échantillon a été combiné avec celui collecté par l'expérience CMS et pour la première fois la désintégration $B^0_{(s)} \to \mu^+\mu^-$ a été observée. En vue d'améliorer la sensibilité au mode $B^0 \to \mu^+\mu^-$, une nouvelle variable d'isolation, qui utilise un algorithme de reconstruction topologique inclusif, a été développée. De nouvelles études ont également été menées pour augmenter la performance des analyses multivariées. Une autre partie du travail original présenté dans cette thèse concerne la définition d'une chaine de sélection pour la recherche des désintégrations $B^0_{(s)} \to \tau^+\tau^-$, qui restent encore inexplorées. Dans cette thèse l'état final où les deux τ vont en trois π chargées et un τ est étudié. La présence des deux ν dans l'état final de la désintégration rend difficile une reconstruction des impulsions des deux τ. Cependant, la possibilité de mesurer les deux vertex de désintégration des τ ainsi que le vertex d'origine du candidat B, permet d'imposer des contraintes géométriques qui peuvent être utilisées dans la reconstruction des impulsions des deux τ. En particulier, un nouvel algorithme pour la reconstruction complété, événement par événement, de ces impulsions et de leurs variables associées est présenté et discuté.

Acknowledgments

By experience, this is the only part of the thesis that almost every person will read carefully and fully understand. I want then to thank all the persons which accompanied me during this three-years path, whose days took the colors of enthusiasm and frustration, tiredness and excitement, angry and happiness.

First of all I want to thank all the members of my jury: Ms. Kou and Mr. Trabelsi for the careful read of my manuscript (during summer); Mr. Ben-Haim and Ms. Calvi for their availability in coming to Marseille and their interest in the work I've done. I want to thank Mr. Kajfasz who gave me the possibility to work at the CPPM and at the CPT, in the framework of an interdisciplinary project (the OCEVU LabEx) between theoretical and experimental physics. And finally the latest two members of my jury, my two advisors. Giampiero, who has been an example of how to move in the busy environment of the LHCb collaboration and of the High Energy Physics family. From the very first moment he believed in me, sometimes even more than myself. I know that he had really to "fight" to give me the possibility to do my thesis at CPPM. All of you can judge from the results presented in this work if his efforts have been worth to be done. And Jerome for his patience, his expertise and his availability to make the bridge between theoreticians and experimentalists. There hasn't been a discussion with him from which I couldn't learn something useful for my work. Thanks to both of you, *mes chefs*.

I would like to thank all the members of the LHCb group at CPPM: Renaud for having welcomed me in the LHCb group at CPPM; Justine, Julien, Olivier, for their availability and their support during all these years; Andrei for its help and its expertise on GRID and informatics issues; the postdocs, Simon, Andrey, and Kristof with whom it has been a pleasure to share the work and jokes. The young students Morgan and Joan as well the "seniors", Mathieu and Cosme, whose works represent the basis of some of the results presented in this thesis. I want to thank Elie for his humor, his jokes, for his availability to always discuss of physics as well as any other topic, and for his constant interest in the prospects of us, young researchers.

Finally, among the members of the LHCb group in CPPM, a particular thanks goes to Walaa. We started and finished at the same time, sharing many of the stages of this long path. I don't even remember how many travels for conferences, schools and workshops we have done together. I wish her all the happiness and success in work and life.

I have to thank all the persons of the LHCb collaboration with whom I collaborated all during these three years: Francesco, Marco-Olivier, Flavio, Matteo P., Matteo R., Gaia, Barbara, and Tom.

I want to thank Kazuya for his humor and all the time spent together, and, together with him, all the PhD students of the CPPM with whom I shared the free time and the coffee breaks in the laboratory: Asma, Aurore, Benoit, Margot, Rima, and Sebastien. During this last year I had many physics discussion (rigorously only during breaks!) with Venu: it has always been a pleasure to discuss with him of our work (and even more).

I spent most of my free time with my italian company, thanks to whom it has been possible to recreate a piece of *belpaese* here in Marseille. I want to thank Matilde, Clara, Marcello, Sara, Matteo, Rossana, and Alessandro for all what they gave me all during these three years. My week-ends would not have been so funny without the company of Eunice, Girish, Anais, Jack, Anna, Ankita, whose sunny smile has always lightened this happy company, and Akankshi, with

her optimism and enthusiasm.

A lot of the social life of these three years of PhD couldn't have been possible without Daniele. Thanks to his availability in welcoming everybody at his place, it has been possible to set up the "*Week-ends Entertainment*" working group, for which we are proud of having been the very first conveners of the "*Saturday & Sunday Lunches*" sub-working group. I want then to thank him for his support during all of these three years, both in work and in the daily life.

A particular thought goes to that person which at the very beginning of this adventure has been a bright ray of light which cleaved the cloudy sky of my daily life. Human affairs are (unfortunately) not as "simple" as the physical laws that we, scientists, try to understand each day. I would preferred a different epilogue of the days spent together, but I prefer retain the lights of the shared joys instead of the shadows of the regrets. "*Io mi dico è stato meglio lasciarci che non esserci mai incontrati*"*.

I lived in a city, Marseille, unique because of its beauty. A person can like it or not, but cannot remain indifferent in front of it. I've been lucky and liked it from the very first sight. I thank then, in a symbolic way, this city who welcomed me and where I've never had the feeling of living in a foreign place.

This work has been carried out thanks to the support of the OCEVU Labex (ANR-11-LABX-0060) and the A*MIDEX project (ANR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French government program managed by the ANR.

* "It has been better to say you 'goodbye', than have never meet you". F. De André

Contents

1 Introduction and motivations **1**
1.1 Successes and limitations of the Standard Model . . . 1
1.1.1 The Standard Model of elementary particles physics . . . 1
1.1.2 Peculiar aspects of the Standard Model and indirect probes of New Physics 10
1.1.3 Limits of the Standard Model . . . 11
1.1.4 Flavor observables and the NP scale: the "Flavor Puzzle" . . . 13
1.2 $B^0_{(s)} \to \ell\bar{\ell}$ decays . . . 14
1.2.1 The effective hamiltonian for $B^0_{(s)} \to \ell\bar{\ell}$ . . . 14
1.2.2 $B^0_{(s)} \to \ell\bar{\ell}$ observables . . . 16
1.2.3 SM predictions . . . 18
1.2.4 $B^0_{(s)} \to \ell\bar{\ell}$: current bounds . . . 19
1.3 Current anomalies in B-meson decay observables and hints of NP in third generation 20
1.4 Conclusions . . . 21

2 The LHCb detector **23**
2.1 The LHC accelerator complex . . . 23
2.2 The $b\bar{b}$ production cross section at LHC . . . 26
2.3 The LHCb detector . . . 26
2.3.1 The tracking system . . . 30
2.3.2 The Particle Identification . . . 34
2.3.3 The Trigger system . . . 35
2.4 Conclusions . . . 37

3 $B^0_{(s)} \to \mu^+\mu^-$ **41**
3.1 Overview of the analysis . . . 41
3.1.1 Signal features . . . 42
3.1.2 Trigger selection . . . 43
3.1.3 Loose selection . . . 44
3.1.4 Tight selection . . . 48
3.2 BDT optimization . . . 48
3.2.1 Introduction to Boosted Decision Trees . . . 48
3.2.2 Optimization of the input variables set and tuning parameters . . . 51
3.2.3 Correlation with the invariant mass . . . 53
3.2.4 The final BDT classifier . . . 66
3.3 Signal yield and limit extraction . . . 68
3.3.1 Signal PDF calibration . . . 69
3.3.2 Backgrounds PDF calibration . . . 70
3.3.3 Normalization . . . 74
3.3.4 Results . . . 77
3.4 Combination with the CMS result . . . 80
3.5 Interpretation of the results. . . . 84

3.6 LHCb prospects for the next Runs of the LHC . . . 87
3.7 Preparing current analysis improvements . . . 88
3.7.1 The ZVtop algorithm . . . 88
3.7.2 Definition of Isolation Variables using ZVtop . . . 89
3.7.3 Combination of isolation variables . . . 96
3.7.4 BDT classifier with new isolation variables . . . 100
3.7.5 Study of the background composition of events falling in last BDT bin . 103
3.7.6 Pointing related variables . . . 105
3.7.7 Conclusions . . . 108
3.8 Conclusions . . . 109

4 $B^0_{(s)} \to \tau^+\tau^-$ 111
4.1 $B^0_{(s)} \to \tau^+\tau^-$ at LHCb . . . 111
4.1.1 The $\tau \to 3\pi\nu_\tau$ decay . . . 113
4.1.2 The $(3\pi, 3\pi)$ final state analysis . . . 114
4.2 Candidate reconstruction . . . 115
4.2.1 Preselection . . . 115
4.2.2 Background composition after preselection . . . 117
4.2.3 Background characterization . . . 118
4.2.4 Analysis strategy . . . 120
4.2.5 Signal MC generated samples . . . 120
4.3 Tight selection . . . 121
4.4 Normalization . . . 127
4.5 Full reconstruction of $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ events . . . 128
4.5.1 The one-dimensional case . . . 128
4.5.2 The real case . . . 129
4.5.3 A new approach . . . 130
4.5.4 Calculations . . . 134
4.5.5 Choice of the "right" solution . . . 135
4.5.6 θ approximation . . . 138
4.5.7 Discussion about discriminating variables . . . 142
4.5.8 Conclusion and prospects . . . 146
4.6 BDT classification . . . 146
4.6.1 BDT distributions . . . 149
4.6.2 Discussion about the observables . . . 151
4.7 BDT PDF calibration . . . 151
4.7.1 Calibration with the SS sample . . . 151
4.7.2 Control regions in OS . . . 156
4.8 Conclusions and prospects . . . 156

5 Conclusions and Prospects 159

Appendices 161

A Muons isolation variables 163

B Data-MC comparison for isolation inputs 165

C Impact of H related variables on signal-background discrimination. 167

D Solution of the system for $B^0_{(s)} \to \tau^+\tau^-$ signal reconstruction 171

E Latest developments of the full reconstruction of the $B^0_{(s)} \to \tau^+\tau^-$ events **175**

Chapter 1

Introduction and motivations

This chapter introduces the theoretical framework and the main reasons motivating the searches for $B^0_{(s)} \to \ell\bar{\ell}$ decays.

After a summary of the main concepts needed to build the Standard Model (SM) of elementary particles (Sec.1.1.1), its specific aspects will be presented (Sec.1.1.2). The Operator Product Expansion (OPE) approach will be introduced in Sec.1.2.1, and this formalism will be applied to the study of the observables of the $B^0_{(s)} \to \ell\bar{\ell}$ processes (Sec.1.2.2). The current experimental status will also be reviewed in Sec.1.2.4.

1.1 Successes and limitations of the Standard Model

The Standard Model of elementary particles physics is the theory currently used to describe interactions between the elementary particles observed. Far away from being a complete satisfactory model, its predictions have been nonetheless tested at an impressive level of accuracy.

In this section a pedagogical overview of the SM of elementary particle physics will be presented, together with the main motivations to look for its extensions. For a more complete and systematic presentation of the SM the reader is referred to Refs.[1, 3].

1.1.1 The Standard Model of elementary particles physics

The formal framework the SM is based on, is the Quantum Field Theory (QFT): particles are associated to the fluctuations of a related *field*, *i.e.* a set of functions defined on the four-dimensional space-time:

$$particle \;\leftrightarrow\; \Phi(x^\mu) \equiv \{\phi_i(x^\mu)\} \; .$$

The number of components of the field $\Phi(x^\mu)$ is determined by the spin of the related particle; the components $\{\phi_i(x^\mu)\}$ of each field transform linearly into each other according to a representation of the Lorentz group $(SU(2) \otimes SU(2))$ corresponding to the spin of the particle.

The Lagrangian. According to the principles of classical mechanics, the dynamic and the evolution of the fields are determined by the stationary points of the *action* $\mathcal{S}$ defined as the integral over the space time of a *Lagrangian density* $\mathcal{L}$,

$$\mathcal{S} \equiv \int d^4x \mathcal{L} \; . \tag{1.1}$$

$\mathcal{L}$ is given by the difference between a kinetic term $\mathcal{L}_{kin}$ and a *potential* $\mathcal{V}$, *i.e.*

$$\mathcal{L} \equiv \mathcal{L}_{kin} - \mathcal{V} \; , \tag{1.2}$$

both of them expressed as a function of the dynamical degrees of freedom of the theory, namely the set $\{\psi_i\}$ of fields related with each observed elementary particle. In particular the kinetic term is quadratic in the fields ψ and contains at least one derivative of the fields, while the potential is usually a polynomial of degree greater than 3 in the fields $\{\psi_i\}$ and, possibly, of their derivatives.

A particle is described by the fluctuation of the field around a configuration, called *vacuum* state, which minimizes the potential $\mathcal{V}(\{\psi_i\})$.

By means of the QFT and symmetry principles, the SM describes the dynamic of the so far observed elementary particles. These can be divided into three categories: the matter fields, the vector bosons, and the Higgs boson. They will be presented in the following.

Matter fields. The matter fields are associated with the spin $\frac{1}{2}$ elementary particles. All massive elementary spin $\frac{1}{2}$ particles can exist in two states of *chirality*: *left* or *right*. This translates into the fact that the field ψ (also called *spinor*) associated to the fermion can be expressed as the sum of two more fundamental objects, called *chiral* (or *Weyl*) spinors $\psi_{L,R}$, defined by the chirality operators $P_{L,R}$:

$$P_{L,R} \equiv \frac{1 \mp \gamma^5}{2}, \tag{1.3}$$

as follows

$$\psi_{L,R} \equiv \frac{1 \mp \gamma^5}{2} \psi \ . \tag{1.4}$$

These states transform each other under a parity transformation $\vec{x} \to -\vec{x}$.

Only massless fermions are eigenstates of chirality, while massive particles always exist in a superposition of the two chirality states.

The kinetic term of the lagrangian describing a generic fermion field is given by the Dirac term ($\sigma^\mu \equiv (1, \sigma^i)$, $\bar{\sigma}^\mu \equiv (1, -\sigma^i)$ where σ^i are the common Pauli matrices)

$$\mathcal{L}_{kin}^{fermion} = i\bar{\psi}_L \partial_\mu \bar{\sigma}^\mu \psi_L + i\bar{\psi}_R \partial_\mu \sigma_\mu \bar{\sigma}^\mu \psi_R \ . \tag{1.5}$$

The effects of the particle mass in the lagrangian are described by a term quadratic in the field in the form

$$\mathcal{L}_{mass} = -m_\psi \bar{\psi}_L \psi_R \ + \ h.c. \ . \tag{1.6}$$

The most simple way to take into account the mass of the particle, is to insert it "as it is", *i.e.* by adding the "hard" mass term of eq.(1.6). Nevertheless, this terms couples the left and right handed components of the fermionic field, and as it will be shown in the following, is inconsistent with the requirement of the invariance of the SM Lagrangian under the action of a given symmetry group. This is one of the hints suggesting that the masses of the elementary matter fields (as well as of the gauge bosons, that will be introduced later) are generated in a dynamical way, through the so-called Higgs mechanism, described later on.

The matter fields are classified into two families: *leptons* and *quarks*. Both feature the same structure, consisting of three replicas (or *flavors*) of the same pattern of fields:

- the basic pattern in the *quark sector* consists of an *up* type (with electric charge $q_u = +\frac{2}{3}$) and a *down* type (with electric charge $q_d = -\frac{1}{3}$) quarks. Both being massive, they exist in the two chirality states, and, for reasons that will be clear soon, they can be arranged into a left doublet $Q_L = (u_L, d_L)$ and two right singlets u_R, and d_R. Each quark field exists in turn in three states of an additional *color* charge.

This is the Dirac mass term for a fermion. There is also a second way to include the mass of a fermion which needs only one chiral spinor. This is the so-called Majorana mass term, and could be relevant to include the effects of the neutrino masses, for whose existence there are experimental evidences.

- the *leptonic sector* contains the charged lepton ℓ (which has an electric charge $q_\ell = -1$ and is described by its two chirality states ℓ_R and ℓ_L) and its neutrino ν_ℓ. In the minimal SM neutrinos are massless and thus exist in only one chirality state, which has been observed as the left handed one $\nu_{\ell L}$. As for the quark sector, these states are usually grouped into a left doublet $L_L = (\nu_{\ell L}, \ell_L)$ and the right singlet ℓ_R. The three replicas are called electron (e), muon (μ), and tau (τ) families. Leptons are neutral with respect to the color charge.

Each elementary particle introduced above comes with its own antiparticle

Multiplets and symmetries. As already anticipated, the elementary fields introduced before are arranged into different multiplets, whose elements transform linearly into each other under the action of a given transformation which has the mathematical structure of a Lie group (in more formal terms, this is expressed by saying that fields belongs to irreducible representations of some transformation groups). To each elementary field a given *charge* is associated, which tells if the field, or the multiplet it belongs to, is affected by a given transformation.

The transformations defined by a Lie group are parametrized by a set of parameters, one for each generator (*i.e.* independent actions) of the transformation group.

In general, transformations can be divided in two categories:

- *global*: where the parameters describing the transformation are constants, *i.e.* they do not depend on the particular space-time point.

- *gauge* (or *local*): where the parameters are functions of the space-time. A gauge transformation is also a global transformation, while the *viceversa* is not true.

The Noether theorem [4] establishes a connection between continuous symmetries among the degrees of freedom of a system and the existence of conserved charges. In particular it states that for each invariance of the Lagrangian under a *global* transformation acting on its degrees of freedom Ψ there exist:

- a current $J^\mu(\Psi)$ whose four-divergence is zero, *i.e.* $\partial_\mu J^\mu(\Psi) = 0$

- a conserved charge $Q(\Psi) \equiv \int dx^0 J^0(\Psi)$, *i.e.* $\partial_t Q(\Psi) = 0$.

In other words, the impossibility of distinguishing between two configurations, say Ψ and Ψ', of the fields, related each other by a transformation under a given group (*i.e.* $\Psi \leftrightarrow \Psi'$) leads to the existence of quantities which are conserved. The viceversa is also true, *i.e.* the observation of quantities which are conserved in the evolution of a given system is an indication of a *global* symmetry between its degrees of freedom.

So for instance, the fact that the total number of leptons of each flavor is conserved in their interactions is a clear indication of a symmetry of the lagrangian involving these fields (or , at least, some of their components). For that reason, the charged leptons ℓ and their neutrinos ν_ℓ enter the lagrangian as a (left) doublet in a form which is invariant under global transformations mixing these fields.

The lagrangian of the SM is required to be invariant under transformations of the fields generated by the actions of the group

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \ : \tag{1.7}$$

- the color group $SU(3)_c$ acts on the color multiplet of the quarks

An important remark concerns the fact that the quark and leptonic sectors are completely decoupled, in the sense that there are no transformations which mix quarks and leptons together.

Field	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
$Q^i_L = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix}$	3	2	1/6
u^i_R	3	1	2/3
d^i_R	3	1	-1/3
$L^i_L = \begin{pmatrix} \nu^i_\ell \\ \ell^i_L \end{pmatrix}$	1	2	$-1/2$
ℓ^i_R	1	1	-1

Table 1.1: Properties of the elementary fermions. The index i refers to the flavor, and goes from 1 to 3.

- the electroweak group $SU(2)_L \otimes U(1)_Y$ acts on all the elementary fields and is factorized into the $SU(2)_L$ sub-group, which acts only on the left doublets (and for that reason is responsible of the parity violation of the weak interactions), and the hypercharge rotation $U(1)_Y$ which acts on both left and right handed fields.

In Tab.1.1 the quantum numbers of the elementary fermions are summarized. As far as their quantum numbers are concerned, there are no distinctions between the different families both in the leptonic and in the quark sectors.

The fact that the mass term in eq.(1.6) couples the left and right handed components of the fermion fields makes this term not invariant under the action of the chiral group $SU(2)_L$, which affects the left component of the field while leaving the right one unchanged. In addition, the fact that the $SU(2)_L$ transformations mix particles with different masses (*i.e.* $\ell_L \leftrightarrow \nu_{\ell,L}$ and $u_L \leftrightarrow d_L$), is an additional indication that masses cannot be described through the lagrangian term in eq.(1.6), and their introduction in the SM lagrangian requires a more subtle mechanism. Thus, in the SM all elementary matter fields enter the lagrangian as massless fields, *i.e.* only through the term (1.5).

Symmetries and interactions. The main feature of the SM is that interactions between particles arise as a natural consequence of symmetries among them. In particular, in the SM the transformations generated by the group (1.7) are required to be a gauge symmetry of the world, namely of the SM lagrangian. In other terms this means that at *each point* of the space and at *each time* the matter fields can be transformed (according to the symmetry group and the representation they belong to) into each other in different ways, still leaving the world invariant.

This requirement has two fundamental consequences:

- new spin-1 particles must exist. These particles are associated to the fluctuations of the so-called *gauge connections* of the symmetry group (1.7) of the SM.
- interactions between the fermions and these particles arise.

From a formal point of view, these gauge-connections are introduced to make the kinetic derivative term (1.5) invariant when fields are transformed according to a local transformation, *i.e.*

$$\Psi_i(x^\mu) \to \mathcal{M}_{ij}(x^\mu)\Psi_j(x^\mu) \tag{1.8}$$

where $\Psi_i(x^\mu)$ denotes the set of all matter fields, $\mathcal{M}$ is the transformation matrix acting on them, and a sum over repeated indexes is implied.

The kinetic term of the lagrangian is invariant under the action of a constant transformation matrix $\mathcal{M}$, because of the fact that $\bar{\Psi}$ and Ψ bring respectively "factors" $\mathcal{M}^{-1}$ and $\mathcal{M}$ which give the identity matrix when multiplied by each other. In a symbolic way:

$$\bar{\psi}\partial\psi \to \mathcal{M}^{-1}\bar{\psi}\partial\mathcal{M}\psi = (\mathcal{M}^{-1}\mathcal{M})\bar{\psi}\partial\psi = \bar{\psi}\partial\psi \tag{1.9}$$

On the other hand, this is no more true for a space-time dependent matrix, because then the derivative will also act on the elements of the matrix $\mathcal{M}$ which can not be factorized anymore. The invariance under local transformations can still be recovered if the derivative ∂_μ is replaced by the so-called *covariant derivative*, *i.e.*

$$\partial_\mu \to D_\mu \equiv \partial_\mu - ig_s G^a_\mu t^a - ig_2 W^b_\mu T^t - ig_1 B_\mu Y \tag{1.10}$$

where

- t^a $(a = 1, 2, ..., 8)$ are the $SU(3)_c$ generators, T^t $(t = 1, 2, 3)$ are the generator of $SU(2)_L$ and Y is the generator of $U(1)_Y$
- g_s, g_1, and g_2 are the gauge-couplings
- G^a_μ, W^t_μ, B_μ are the gauge connections. These fields describe the eight gluons mediating the strong force, and four electroweak gauge bosons: $W^{1,2,3}$ and B, respectively for the chiral $SU(2)_L$ and hypercharge $U(1)_Y$ sub-groups. These gauge connection are in turn representation of the respective transformation group, which transform in the so-called *adjoint representation* of the group. Being four-vectors these field transform as a spin-1 representation of the Lorentz group.

With the replacement $\partial_\mu \to D_\mu$ the transformation law (1.9) is true also for local transformations, *i.e.*:

$$\bar{\psi}D\psi \to \mathcal{M}(x)^{-1}\bar{\psi}D\mathcal{M}(x)\psi = (\mathcal{M}(x)^{-1}\mathcal{M}(x))\bar{\psi}D\psi = \bar{\psi}D\psi \tag{1.11}$$

When replacing the derivative ∂_μ with the covariant derivative D_μ in the Dirac lagrangian, new *interaction* terms appear. These are given by the product of a Lorentz four vector fermion-antifermion bilinear (also called *current*) contracted with the gauge connections. Depending on the electric charge of the gauge boson which couples with the fermion current, there can be *neutral* or *charged* currents. An important remark concerns the fact that these terms are diagonal in the flavor space, both for quarks and for leptons.

In order to allow for gauge fields to propagate, a kinetic (gauge invariant!) term for these fields $\mathcal{L}_{gauge}$ is added to the fermionic lagrangian:

$$\mathcal{L}_{gauge} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} \tag{1.12}$$

where $F^a_{\mu\nu}$ is the so-called *field strength* and is defined for the generic gauge connection F_μ of a group symmetry with structure constants f^a_{bc} as

$$F^a_{\mu\nu} \equiv \partial_\mu F^a_\nu - \partial_\nu F^a_\mu - f^a_{bc}F^b_\mu F^c_\nu \,. \tag{1.13}$$

The observation that weak forces have a short range requires the introduction of a mass term for the electroweak gauge connections. Nevertheless the requirement of the invariance under the gauge group (1.7) does not allow for a hard mass term in the form

$$\mathcal{L}^{mass}_{gauge\ bosons} = \frac{m_F^2}{2}F_\mu F^\mu \,. \tag{1.14}$$

As for fermions, the masses of the gauge bosons must be generated dynamically, as it is explained in the following paragraphs.

This is the representation with the same dimension of the group.

Field	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$	1	2	-1/2

Table 1.2: Quantum numbers of the Higgs field with respect to the gauge group of the SM.

Higgs field and spontaneous breaking of the electroweak symmetry The lagrangian obtained from (1.5) with the replacement (1.10) and by adding the kinetic term (1.12) does not take yet into account the fact that particles have masses. It consists only of a kinetic term and of the interaction terms between matter and the gauge bosons, but it does not yet contain the quadratic term in the fields needed to take into account the mass in the evolution of the fields. In the SM the masses of the particles are generated by the Higgs-Brout-Englert mechanism [5, 6].

A $SU(2)_L$ doublet of scalar complex fields Φ (neutral under $SU(3)_c$ transformations) is introduced:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} . \tag{1.15}$$

The quantum numbers with respect to the gauge symmetry group are reported in Tab.1.2.

The dynamic of this field is described by the following lagrangian term

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \tag{1.16}$$

containing a kinetic term for Φ plus a potential $V(\Phi^\dagger\Phi)$, quadratic in the field Φ:

$$V(\Phi^\dagger \Phi) = \mu^2 (\Phi^\dagger \Phi) + \lambda (\Phi^\dagger \Phi)^2 \tag{1.17}$$

where μ^2 and λ are two real parameters.

If $\mu^2 > 0$, the potential (1.17) has a unique minimum $\bar{\Phi} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. If instead $\mu^2 < 0$, $V(\Phi^\dagger\Phi)$ has several minima and the Higgs field gets a non vanishing vacuum expectation value (vev) which can be cast in the standard form:

$$\bar{\Phi} = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \tag{1.18}$$

where $v \equiv \frac{\sqrt{-\mu^2}}{\lambda}$.

Among the known particles, the Higgs field, being a scalar, is the only one that can have a non vanishing vev.

No matter the sign of the μ^2 parameter, the Higgs potential is symmetric under the electroweak group $SU(2)_L \otimes U(1)_Y$. Nevertheless, depending on the sign of μ^2 the ground state can or not respect the symmetry of the potential. In particular, the vacuum state for $\mu^2 > 0$ exhibits the same symmetry of the potential, while this does not happen if $\mu^2 < 0$. In this case, indeed, the vacuum state breaks the symmetry of the lagrangian. This situation where the symmetry of the lagrangian is not preserved by its vacuum state is called *spontaneous symmetry breaking* (SSB).

Indeed, any non scalar field acquiring a non vanishing vev would violate the assumed Lorentz invariance of the vacuum state.

This case is also referred to as a symmetry realized *à la* Wigner-Weyl.

SSB and Higgs mechanism: bosons and fermion masses. The SSB of the electroweak group allows to introduce in the theory the mass terms both for fermions and for gauge bosons.

By requiring in the Higgs potential (1.17) $\mu^2 < 0$, the Higgs field gets a non-zero vev and can then be cast in the form:

$$\Phi = \bar{\Phi} + \delta\Phi \tag{1.19}$$

with $\delta\Phi$ the fields describing the fluctuation of the Higgs field around its vacuum state.

The covariant derivative in eq.(1.16) gives rise to quadratic terms in the electroweak gauge bosons:

$$\mathcal{L}^{(2)}_{gaugebosons} = \frac{g_2^2 v^2}{8}(W^1W^1 + W^2W^2) + \frac{v^2}{8}(g_2W^3 - g_1B)^2 \tag{1.20}$$

plus interaction terms between them and the Higgs field fluctuation $\delta\Phi$.

The fields $W^{1,2}$ can be arranged as

$$W^{\pm}_{\mu} = \frac{W^1_{\mu} \pm iW^2_{\mu}}{\sqrt{2}} \tag{1.21}$$

giving the fields associated to the $W^{\pm}$ vector bosons. Their mass is given by

$$m_W = \frac{g_2 v}{2} \ . \tag{1.22}$$

The second term in eq.(1.20) can be diagonalized in order to obtain the mass eigenstates. The two eigenvectors are given by

$$Z_{\mu} = W^3_{\mu}\cos\theta_W - B_{\mu}\sin\theta_W \tag{1.23}$$

$$A_{\mu} = W^3_{\mu}\sin\theta_W + B_{\mu}\cos\theta_W \ . \tag{1.24}$$

where the sine and cosine of the mixing angle θ_W (also called "Weinberg angle") are defined as:

$$\sin\theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \ , \ \cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \ . \tag{1.25}$$

The combination Z_{μ} is associated with the Z^0 boson, while the A_{μ} with the photon.

The masses of these two eigenstates are:

$$m_Z = \frac{M_W}{\cos\theta_W} \tag{1.26}$$

$$m_{\gamma} = 0 \ . \tag{1.27}$$

Only the $SU(3)_c$ gauge connections remain massless because the Higgs field does not couple with them.

Mass terms for the fermion fields can be introduced in the lagrangian of the SM by adding an interaction term between the fermions and the Higgs fields. This interaction takes the form of a Yukawa term

$$\mathcal{L}_{Yukawa} = Y_u^{ij}\bar{u}^i_R\Phi^T\epsilon Q^j_L - Y_d^{ij}\bar{d}^i_R\Phi^{\dagger}Q^j_L - Y_{\ell}^{ij}\bar{\ell}^i_R\Phi^{\dagger}L^j_L \ + \ h.c \ , \tag{1.28}$$

where

- i, j are indexes on the flavor space, while color indexes are implied
- $\epsilon \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

- Y_u^{ij}, Y_d^{ij}, Y_ℓ^{ij} are three generic 3×3 matrices in the flavor space, parameterizing the strength of the coupling of the Higgs with the fermion fields. These matrices generate the differences between the different flavors in the quark and in the leptonic sector.

The interaction term in eq.(1.28) is invariant under the SM gauge symmetry group. By expanding the Yukawa lagrangian around the non vanishing vev of the Higgs field $\bar{\Phi}$, the interaction between Φ and the fermion fields provides mass terms for the fermions trough the following three matrices:

$$M_u^{ij} = vY_u^{ij}/\sqrt{2} \tag{1.29}$$
$$M_d^{ij} = vY_d^{ij}/\sqrt{2} \tag{1.30}$$
$$M_\ell^{ij} = vY_\ell^{ij}/\sqrt{2} \tag{1.31}$$

plus interaction terms between fermions and $\delta\Phi$ in the form

$$\mathcal{L}_{Yukawa} = Y_u^{ij}\bar{u}_R^i\delta\Phi^T\epsilon Q_L^j - Y_d^{ij}\bar{d}_R^i\delta\Phi^\dagger Q_L^j - Y_\ell^{ij}\bar{\ell}_R^i\delta\Phi^\dagger L_L^j \; + \; h.c \tag{1.32}$$

Because the matrices Y^{ij} are not diagonal the flavor and mass eigenstates are not aligned. The matrices $M_{u,d,\ell}^{ij}$ can be expressed as a function of a diagonal matrix $\mathcal{M}$ with real non negative elements through two independent unitary *global* transformations, *i.e.* $M = \mathcal{U}_L\mathcal{M}\mathcal{U}_R^\dagger$. Inserting this expression into the Yuwawa term obtained after the SSB, the flavor eigenstates u, d, ℓ can be expressed as a function of the mass eigenstates u', d', ℓ' through the following relations:

$$u_L = \mathcal{U}_L^u u'_L \; , \; u_R = \mathcal{U}_R^u u'_R \tag{1.33}$$
$$d_L = \mathcal{U}_L^d d'_L \; , \; d_R = \mathcal{U}_R^d d'_R \tag{1.34}$$
$$\ell_L = \mathcal{U}_L^\ell \ell'_L \; , \; \ell_R = \mathcal{U}_R^\ell \ell'_R \; . \tag{1.35}$$

The mass matrices $\mathcal{M}_{u,d,\ell}$ are given by

$$\mathcal{M}_u = Diag(m_u, m_c, m_t), \tag{1.36}$$
$$\mathcal{M}_d = Diag(m_d, m_s, m_b), \tag{1.37}$$
$$\mathcal{M}_\ell = Diag(m_e, m_\mu, m_\tau) \; . \tag{1.38}$$

It's important to notice that the two elements of the quark left doublet transform in two different ways, so that

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix} = \begin{pmatrix} \mathcal{U}_L^u u'_L \\ \mathcal{U}_L^d d'_L \end{pmatrix} = \mathcal{U}_L^u \begin{pmatrix} u'_L \\ V_{CKM} d'_L \end{pmatrix} \tag{1.39}$$

where the Cabibbo-Kobayashi-Maskawa matrix V_{CKM} is defined as

$$V_{CKM} \equiv \mathcal{U}_L^{\dagger u}\mathcal{U}_L^d \tag{1.40}$$

The transformations (1.33)-(1.35) do not affect the kinetic term for the fermion fields nor the interaction terms with the gluons and the neutral electroweak gauge bosons. Indeed these terms, being diagonal in each fermion field, both for their left and right handed components, remain diagonal also in the basis given by the mass eigenstates. On the other hand, the charged electroweak interaction terms mix up and down type quarks and thus they introduce a mixing between the different quark flavors, being

$$\frac{g_2}{\sqrt{2}}\bar{u}_L^i\gamma^\mu d_L^i W_\mu^+ = \frac{g_2}{\sqrt{2}}\bar{u}_L'^i\gamma^\mu V_{CKM}^{ij} d_L'^j W_\mu^+ \tag{1.41}$$

The CKM matrix is a 3×3 unitary matrix and is parametrized by nine real parameters. This number can be reduced through a redefinition of the quark field phases, and V_{CKM} can be expressed as a function of only four parameters, three mixing angles and a unitary complex phase.

A widely used parametrization for V_{CKM} is the one proposed by Wolfenstein [7]. In this case V_{CKM} is expressed as a function of four real parameters A, ρ, λ, η as follows:

$$\begin{aligned} V_{CKM} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\ &= \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \ . \end{aligned} \tag{1.42}$$

The current measured values for the parameters A, λ, ρ, and η are [8]:

$$\begin{aligned} \lambda = 0.22537 \pm 0.00014 \quad &, \quad A = 0.814^{+0.023}_{-0.024} \ , \\ \rho = 0.117 \pm 0.021 \quad &, \quad \eta = 0.353 \pm 0.013 \ . \end{aligned} \tag{1.43}$$

The mixing mechanism described by the *CKM* matrix is the *only* source of mixing between flavors in the SM and it can take place at tree level only in *charged* currents, *i.e.* through the interaction vertex in eq.(1.41).

Discrete symmetries of the SM and CP violation The Lagrangian of the SM is made up of quadratic terms in the fermionic and bosonic fields, and interactions terms in the form of the Lorentz product between fermionic bilinears (usually referred as *fermionic currents*) and the vector boson or Higgs fields. Each of these terms is invariant under the combined action of the following three *discrete symmetries*:

- *Time reversal* (T) which consists in the inversion of the time coordinate $t \to -t$;
- *Spatial inversion* (P) which inverts the spatial coordinates $\vec{x} \to -\vec{x}$. As a consequences, a left (right) handed fermion transforms into a right (left) handed one $\psi_{L,(R)} \to \psi_{R,(L)}$;
- Charge conjugation (C) which changes the sign of the electric charge of the field. The action of the charge conjugation on a left (right) handed fermion is given by $\psi_{L,R} \to \psi^c_{L,R} \equiv \pm i\sigma_2 \psi^\star_{L,R}$.

The invariance of the SM Lagrangian under the action of the CPT transformations is known as *CPT theorem* (for a more detailed explanation the reader is referred to Refs.[1]-[3]). All the Lagrangian terms except those describing the interaction of the massive gauge bosons with the fermionic currents are invariant under the action of the P,C,T transformations separately, and thus, trivially, under their combined action. On the other hand, the Lagrangian terms entering the electroweak sector are not invariant under the action of the P, C, and their combined action (referred as CP and which consist in the exchange of a particle with its own antiparticle) transformations, as it has been observed experimentally (see *e.g.* Refs.[9]-[12]). Under the formal point of view, the violation of the CP symmetry is due to the presence in the SM Lagrangian of complex parameters. The action of the CP transformation on any product, generically denoted by ϕ_i, of the SM fields is equivalent to the operation of complex conjugation, *i.e.* $\phi_i \to \phi_i^\star$. By writing in a very general form the Lagrangian $\mathcal{L}$ as follows (each a_i is a generic parameter of the Lagrangian):

$$\mathcal{L} = \sum_i a_i \phi_i + a_i^\star \phi_i^\star \tag{1.44}$$

the action of the CP transformation on $\mathcal{L}$ is given by

$$\mathcal{L} \to \mathcal{L}^{CP} = \sum_i a_i \phi_i^\star + a_i^\star \phi_i \ . \tag{1.45}$$

From this transformation law it can be seen that if *each* a_i parameter is real, then $\mathcal{L}$ is invariant; otherwise, if at least one of the a_i coefficients is complex, the invariance under CP is broken. The complex parameters entering the SM Lagrangian are the V_{CKM} matrix elements. Some complex phases of the V_{CMK} matrix elements can be reabsorbed through a redefinition of the SM field phases. In particular, it is possible to completely reabsorb the complex phases in V_{CMK} in presence of only two quark flavors; nevertheless, in the physical case where three quark flavors exist this is not possible anymore and the SM Lagrangian is left with one complex parameter, which is responsible for the violation of the CP physics. This mechanism has been pointed out in Ref.[30] by Kobayashi and Maskawa, and the CP violation induced by the presence of this complex phase in the V_{CMK} matrix is nowadays referred as "*KM mechanism*".

1.1.2 Peculiar aspects of the Standard Model and indirect probes of New Physics

The SM is characterized by some features which are quite specific. The most important of these aspects are:

- **Lepton Family Number conservation**: being the electro-weak interaction diagonal in the leptonic flavor, even after the redefinition of the lepton fields in the mass eigenstates, only leptons of the same flavor can interact with an electro-weak gauge boson. In particular only the following interaction terms are allowed: $Z^0\ell\bar{\ell}$, $\gamma\ell\bar{\ell}$, $W^-\ell\bar{\nu}_\ell$. This property is true not only at the tree level, but at each order in the perturbative expansion. Very small departures (well beyond the current experimental sensitivity) from this property are due to the neutrino oscillations (see next section), which allow for very small rates. For example, the predicted rate for the Lepton Flavor Violating decay $\mu^+ \to e^+\gamma$ is of the order $\sim 10^{-54}$ [13] and the current experimental limit (at 90% C.L.) on its branching fraction is [14]

$$\mathcal{BR}(\mu^+ \to e^+\gamma) < 5.7 \times 10^{-13} \ . \tag{1.46}$$

- **Lepton Flavor Universality (LFU)**: the coupling of the leptons to the electroweak gauge bosons are flavor independent. This property has been tested at colliders by looking at the W^+ decays. For instance the $\mathcal{BR}$ of the decays $W^+ \to \ell^+\nu_\ell$ with $\ell = e, \mu$ have been measured to be [8]:

$$\mathcal{BR}(W^+ \to e^+\nu_e) = 10.63 \pm 0.15 \tag{1.47}$$
$$\mathcal{BR}(W^+ \to \mu^+\nu_\mu) = 10.71 \pm 0.16 \tag{1.48}$$

  whose relative difference is smaller than 1%.

- **Flavor Changing Neutral Currents (FCNC) in the quark sector**: differently from the leptonic sector, flavor changing interactions can take place in the quark sector. The misalignment between the flavor and mass eigenstates of the quarks fields makes flavor changing currents arising, through the interaction term in eq.(1.41). Though, at tree level, only charged currents can change flavor while FCNC's are (accidentally) forbidden and can take place only at the loop level, through the emission of at least two off-shell heavy virtual states. For that reason all the processes involving transitions or interactions between quarks of the same *down* or *up* type are highly suppressed in the SM.

The first two of these aspects are related to the leptonic sector, and the discussion if one implies the other is still under investigation (see *e.g.* Refs.[15, 16]).

Even though all these properties have been tested at a very high precision level, they still represent one of the golden ways to probe possible extensions of the SM, required by the arguments presented in the next section. Indeed, these properties, being quite peculiar, are expected to be violated by any more general theory beyond the SM. One of the most promising ways to look for sources of NP beyond the SM is then represented by the precision measurements of observables related to processes sensitives to one or more of these features.

1.1.3 Limits of the Standard Model

Despite its non "fashionable" name, the SM is the most satisfactory description of the elementary particle dynamics available so far. Under the formal point view it is a renormalizable (and thus predictive) theory, and the quantum numbers of the fermions with respect to the gauge group in eq.(1.7) make the chiral anomalies vanishing at every order in the perturbation-theory, thus saving the consistency of the model. Under the experimental point of view its predictions have been verified to a very high precision level. Thanks to the mixing mechanism described by the CKM matrix, it can account for the CP violation observed in processes involving the K, D and B systems.

However, both on the theoretical and the experimental sides, there are several hints suggesting that this model should be interpreted as a low-energy manifestation of a more fundamental underlying theory. In particular, two kinds of motivations pushes for the search for NP sources. For sake of clarity, these motivations are here divided into two categories: the *completeness* and *naturalness* problems.

The completeness arguments have to do with the fact that the SM is actually an incomplete description of the world, and it does not take into account the outcome of some experimental observations. These motivations strongly indicate that the SM *must* be modified in order to include these effects. The major experimental hints requiring an extension of the SM are:

- observation of the mixing of neutrinos [18]. The oscillation between neutrinos of different flavors require the introduction of a mass term for these particles. The nature of this mass term, *i.e.* Majorana or Dirac, is still matter of study and no evidence in favor of one of the two hypothesis has still been obtained. In addition, the oscillations of neutrinos make possible Flavor Changing Currents also in the leptonic sector, as it happens between the quarks. Nevertheless, due to magnitude of these oscillations, this mixing, described through the so-called Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, has a smaller strength than in the quark sector, thus giving an even stronger suppression of Lepton Flavor Violating processes;

- cosmological observations [20, 21, 22] suggesting that new kinds of matter must exists. This *dark* matter (DM) interacts with the ordinary barionic particles only through gravity and its dynamic must be inserted in the framework of the SM as well. In addition, the existence of a dark energy density driving the observed expansion of the Universe is a further motivation for looking for extension of the SM;

- the amount of the observed matter-antimatter asymmetry in the Universe suggesting that extra sources of CP violation must exist. Indeed, even though the SM can account, through the KM mechanism, for this asymmetry, the predicted amount of CP violation in the SM is not enough to explain the observed amount of matter in the Universe;

Especially if compared with the name of its extensions.

To be precise, chiral anomalies vanish at one loop level, and then, thanks to Adler-Bardeen theorem [17], at every order in the loop expansion.

One of the "golden" way to probe the nature of the neutrino mass term is the observation of the so-called neutrinoless double β decay (see *e.g.* [19]) which can probe the Majorana nature of neutrinos.

- need of a quantum theory of gravity. Indeed gravitational interaction is well understood at the classical level and successfully described through the theory of general relativity, but a *quantum* theory of gravity has not yet been formulated. Every attempt to describe gravity as the result of an invariance under local transformations of the fundamental degrees of freedom (as it is done for the electroweak and strong interactions) turned out to fail so far.

The naturalness arguments, on the other hand, are not related to the consistency or completeness of the SM, hence they do not actually require a modification of the theory. These arguments just point out the fact that the SM, as it is, features some very particular patterns for which there are no reasons *a-priori* to exist. Nevertheless, these "accidental" features, that in the SM are just interpreted as the result of a *fine tuning* of the parameters of the theory, can be a natural consequence of the particular structure of the underlying fundamental theory. The most relevant of these naturalness problems are:

- smallness of the Higgs mass. The Higgs boson is the only spin-0 elementary particle. Differently from the elementary fermions and spin-1 bosons (whose masses are protected by some *custodial* symmetries [23] against huge radiative contributions coming from loop processes), the mass of the Higgs boson receives radiative contributions which diverges as $\sim \Lambda^2$, being Λ the cut-off for the momentum flowing in the loops. If the SM is valid up to the Planck scale ($\Lambda_{Planck} \simeq 10^{19}\ GeV/c^2$) this will make the Higgs boson mass M_H being much higher of the observed value [24] $M_H = 125.09 \pm 0.24\ GeV/c^2$, being

$$\Delta m_H^2 \sim \frac{1}{16\pi^2}\Lambda_{Planck}^2 \simeq 10^{36}\frac{GeV^2}{c^4} \tag{1.49}$$

  This suggests that new degrees of freedom can be excited inside the loop at an energy scale $\Lambda_{NP} \sim$TeV, in order to cancel the SM divergent contributions. If this is not the case and SM is the true fundamental theory up to Λ_{Planck}, a cancellation between the bare Higgs mass and the SM contributions arising from the loop diagrams must take place.

- High number of free parameters in the flavor sector (*i.e.* the Yukawa sector). The flavor sector is the most puzzling in SM, if compared with the gauge and Higgs ones. For instance these last two sectors feature a small number of free parameters which are the three gauge couplings in the gauge sector, and the two parameters μ^2 and λ in the Higgs potential. On the other hand, there are sixteen free parameters entering the flavor sector of the SM lagrangian; these parameters describe the masses of the elementary matter fields (twelve Yukawa couplings) and the mixing between quarks (four parameters which parametrize the CKM matrix: three mixing angles and a CP-violating complex phase). In addition, these parameters feature a highly *hierarchical* pattern, both with respect to the fermion masses and to the magnitude of the CKM matrix elements, as it is shown in Figs.1.1 and 1.2.

- CP violation in QCD: the smallness of the θ-term. In principle a CP-violating term built up with the gluon field strength $G^a_{\mu\nu}$ can be added to the lagrangian of the SM through the so called θ-term:

$$\mathcal{L}_\theta = \theta\varepsilon_{\mu\nu\rho\sigma}G^a_{\mu\nu}G^a_{\rho\sigma}\ . \tag{1.50}$$

  This term would violate the CP invariance of the lagrangian (because of the antisymmetric tensor $\varepsilon_{\mu\nu\rho\sigma}$), being $\mathcal{L}_\theta \xrightarrow{CP} -\mathcal{L}_\theta$. Nevertheless, since no evidence for CP violation in the strong interactions is observed (such, for instance, the existence of an electric dipole moment for the neutron), the θ coefficient turns out to be very small. The reason of the

In particular the gauge symmetry forbids hard mass terms for the gauge fields, while the chiral symmetry forbids the Dirac mass term for fermions.

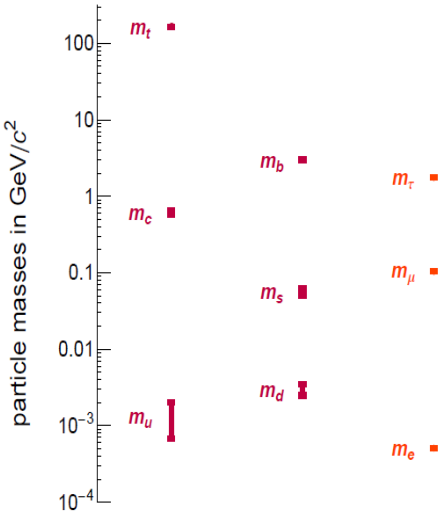


Figure 1.1: Schematic pattern of the fermion masses. Neutrinos are not included.

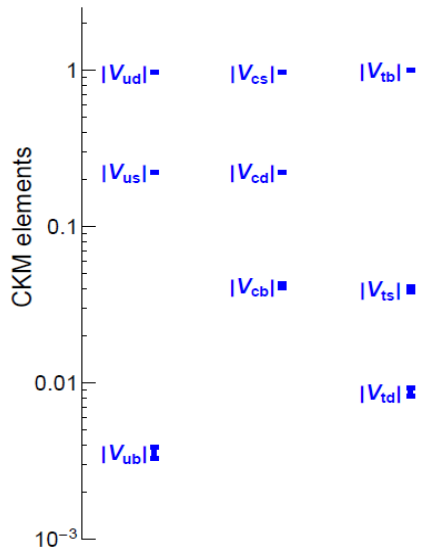


Figure 1.2: Schematic pattern of the magnitude of the CKM matrix elements.

smallness of this parameters is usually referred to as the *strong CP problem*. One possible solution for this problem has been provided by the Peccei-Queen model [25]. In this model the θ parameter is "promoted" to a dynamical (spin-0) field, called *axion*, which relaxes at the vanishing minimum of its potential, thus suppressing the interaction in eq.(1.50). Nevertheless, such a particle has never been observed so far and the strong CP problem still remains an open issue.

1.1.4 Flavor observables and the NP scale: the "Flavor Puzzle"

The observables from the flavor sector have played a major role in the construction of the SM. Since the 70's they have allowed to predict the existence and the properties of particles that were not yet observed:

- from the small branching fraction of the $K_L^0 \to \mu^+\mu^-$ mode Glashow, Iliopulos, and Maiani predicted, in 1970, the existence of the charm quark to suppress, through the GIM mechanism [26], this FCNC. In addition, from the measurement of the $K - \bar{K}$ frequency oscillation, Gaillard and Lee [27] predicted, in 1974, the value of the charm quark mass. The charm quark was discovered by two different teams at SLAC [28] and BNL [29] the same year;

- the observation of the CP violation in the $K - \bar{K}$ oscillations leads to the prediction of the existence of a third generation of quarks [30] in 1973. The down-type quark of this third generation, the bottom (or beauty) quark, has been experimentally observed at Fermilab in 1977 [31];

- the frequency of $B - \bar{B}$ oscillations leads in the late 80's to the prediction of the large quark top mass, experimentally observed at Fermilab in 1995 [32].

As in the past, flavor observables can still play a major role in the search for NP. As it is shown in Fig.1.3, the current lower bounds for Λ_{NP} obtained from flavor observables push the NP scale far beyond the TeV scale. At the same time, it has been shown that a natural solution of the fine tuning problem for the Higgs mass requires NP at a scale ∼TeV. The relevant question is then why the effects of this NP have not yet been observed in flavor observables? This is the

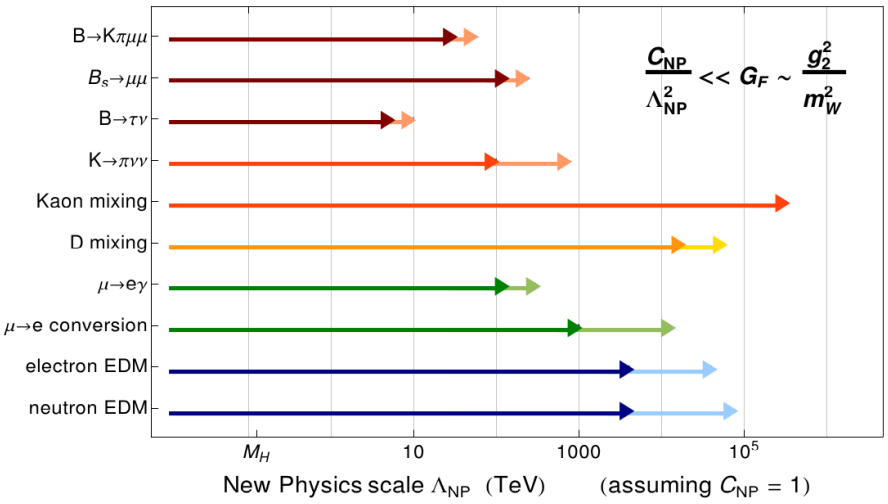


Figure 1.3: NP scale allowed from several flavor observables [33]. The dark and the light arrows correspond to the current and the future expected sensitivity (in the coming $\sim$ ten years) respectively.

so-called "flavor problem". This pending issue motivates the study of flavor observables related to processes not yet studied but that can be sensitive to NP effects. In particular, an increasing interest is given to processes involving third generation leptons.

1.2 $B^0_{(s)} \to \ell\bar{\ell}$ decays

Among the processes that can potentially reveal NP effects, one of the golden channels is represented by the rare dileptonic $B^0_{(s)} \to \ell\bar{\ell}$ decays. They are generated by the annihilation between the b and s (or d) quarks which constitute the B^0_s (B^0) meson. The Feynman diagrams contributing to this FCNC in the SM are shown in Fig.1.4.

The contributions of additional diagrams with new heavy intermediate particles in the loop, can modify the expected values in the SM of the observables related to these processes, such as their $\mathcal{BR}$. In order to provide a general description of such possible effects, the most common approach is represented by the so called "Operator Product Expansion" approach, that will be introduced in the following section.

1.2.1 The effective hamiltonian for $B^0_{(s)} \to \ell\bar{\ell}$

The Operator Product Expansion (OPE) approach (see *e.g.* [34]) allows a model-independent analysis of the process under study, without the need to specify the underlying theory.

The dynamics of the observable degrees of freedom are described using an effective hamiltonian $\mathcal{H}_{eff}$. This is given by the sum of several contributions, each of them being the product of an *effective operator* $\mathcal{O}$, describing the long distance physics, namely the particles in the initial and final states, times an *effective coupling* $\mathcal{C}$ (also called Wilson coefficient) which is obtained by integrating out the heavy degree of freedom of the fundamental underlying theory. These

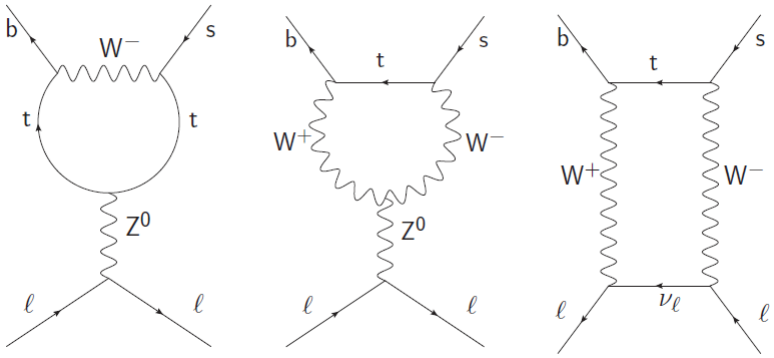


Figure 1.4: Feynman diagrams describing the di-leptonic decay of a B_s^0 meson. The first two diagrams from the right are the so-called Z penguins, while the last one is the box diagram.

coefficients can be computed pertubatively at a given energy scale μ (usually taken equal to the top quark mass) and then extrapolated, through the Renormalization Group Equations (RGE) to the energy scale relevant for the process under study (in the case of B mesons decays, it is the b-quark mass).

The relevant feature of the OPE approach is that it allows to separate different scales effects entering a given process; the effects of NP modify only the values of the Wilson coefficients.

A typical example of such description is the theory of the β decay proposed by Fermi [35]. In this case the $n \to pe^-\bar{\nu}_e$ transition is described by an effective four-fermion operator $(\bar{p}n)(\bar{e}\nu_e)$ with an effective coupling given by the "Fermi constant" G_F. In the fundamental theory (*i.e.* the SM) this effective interaction arises after integrating the heavy degree of freedom represented by the W boson mediating the fundamental transition $d \to ue^-\bar{\nu}_e$.

Depending on the particular process under study, different effective operators will contribute to the observables related to the process. It is important to stress the fact that the observables related to those transitions, such as annihilation or transitions rates, or angular observables, are functions of the Wilson coefficients, and their measurement can help to shed light into the NP realization pattern.

The effective operators $\mathcal{O}$ have dimension greater than four. For this reason they appear in $\mathcal{H}_{eff}$ multiplied by a negative power of an energy scale Λ_{NP} which is the typical energy scale of the NP.

For the dileptonic decays of a B_s^0 meson, the effective hamiltonian made up of the most general lowest dimension (six) effective operators, describing the annihilation between the $\bar{b}$ and s (d) quarks forming the B_s^0 (B^0) meson, is given by

$$\mathcal{H}_{eff}|_{bs\ell\ell} = \mathcal{G}_F(\mathcal{C}_{10}^{\ell}\mathcal{O}_{10}^{\ell} + \mathcal{C}_{10}^{\prime\ell}\mathcal{O}_{10}^{\prime\ell} + \mathcal{C}_S^{\ell}\mathcal{O}_S^{\ell} + \mathcal{C}_S^{\prime\ell}\mathcal{O}_S^{\prime\ell} + \mathcal{C}_P^{\ell}\mathcal{O}_P^{\ell} + \mathcal{C}_P^{\prime\ell}\mathcal{O}_P^{\prime\ell}) \,. \tag{1.51}$$

where

$$\mathcal{G}_F \equiv -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^{\star}\frac{e^2}{16\pi^2} \tag{1.52}$$

and the effective operators $\mathcal{O}_{10,P,S}^{(\prime)}$ are defined as follows (q_d is the light down-type quark, s or d):

$$\mathcal{O}_{10}^{\ell(\prime)} = (\bar{b}\gamma^{\mu}P_{L(R)}q_d)(\bar{\ell}\gamma_{\mu}\gamma^5\ell) \tag{1.53}$$

$$\mathcal{O}_S^{\ell(\prime)} = m_b(\bar{b}\ P_{L(R)}q_d)(\bar{\ell}\ell) \tag{1.54}$$

$$\mathcal{O}_P^{\ell(\prime)} = m_b(\bar{b}P_{L(R)}q_d)(\bar{\ell}\gamma^5\ell) \tag{1.55}$$

The effective operators $\mathcal{O}^{(\prime)}_{10,P,S}$ are in the form of a current-current product

$$\mathcal{O}^{\ell(\prime)}_{10,P,S} = J^{10,P,S}_q \times J^{10,P,S}_\ell \tag{1.56}$$

where $J^{10,P,S}_q$ and $J^{10,P,S}_\ell$ are respectively the quark and leptonic currents with different Lorentz structures. The Wilson coefficients $\mathcal{C}^{\ell(\prime)}_{10,P,S}$ are in the most general scenario independent. Nevertheless, as it has been pointed out in Ref.[36], some constraints among them can be imposed by requiring the invariance of the effective operators $\mathcal{O}^{(\prime)}_{10,P,S}$ under the electroweak symmetry group $SU(2)_L \times U(1)_Y$. In particular, the following constraints are found:

$$\begin{aligned} \mathcal{C}^{\ell}_S &= -\mathcal{C}^{\ell}_P \\ \mathcal{C}^{\ell\prime}_S &= \mathcal{C}^{\ell\prime}_P \ . \end{aligned} \tag{1.57}$$

A final remark concerns the fact that only the axial current operator $\mathcal{O}^{\ell(\prime)}_{10}$ contributes to the decay. This happens because the amplitudes from the vectorial current operator

$$\mathcal{O}^{\ell(\prime)}_9 = (\bar{b}\gamma^\mu P_{L(R)} q_d)(\bar{\ell}\gamma_\mu \ell) \tag{1.58}$$

vanish because of the conservation of the electromagnetic current $\bar{\ell}\gamma_\mu\ell$, being:

$$\begin{aligned} &\langle \ell\bar{\ell}|\mathcal{O}^{\ell(\prime)}_9|B^0_{(s)}\rangle = f_{B^0_{(s)}} p^\mu_{B^0_{(s)}} \cdot \bar{\ell}\gamma_\mu\ell = \\ &f_{B^0_{(s)}} p^\mu_{\ell+\bar{\ell}} \cdot \bar{\ell}\gamma_\mu\ell = f_{B^0_{(s)}} \partial^\mu(\bar{\ell}\gamma_\mu\ell) = 0 \ . \end{aligned} \tag{1.59}$$

1.2.2 $B^0_{(s)} \to \ell\bar{\ell}$ observables

Using the effective hamiltonian in eq.(1.51), the expression for the time-integrated untagged rate, summed over the helicities of the leptons, of the $B^0_{(s)} \to \ell\bar{\ell}$ decay is given by the following expression [37]:

$$\begin{aligned} \mathcal{BR}(B^0_q \to \ell^+\ell^-) =& \frac{\tau_{B_q} G_F^4 M_W^2 \sin^4\theta_W}{8\pi^5} \times |V^\star_{tb}V_{tq}|^2 f^2_{B_q} M_{B_q} m^2_\ell \beta_{m_\ell} \\ &\times (|\mathcal{P}|^2 + |\mathcal{S}|^2) \times \frac{1 + y_q \mathcal{A}^{\ell\ell}_{\Delta\Gamma}}{1 - y_q^2} \\ &= \mathcal{BR}(B^0_q \to \ell^+\ell^-)|^{CP}_{t=0} \times \frac{1 + y_q \mathcal{A}^{\ell\ell}_{\Delta\Gamma}}{1 - y_q^2} \end{aligned} \tag{1.60}$$

where:

- $f^2_{B_q}$ is the B^0_q meson decay constant, defined as $\langle 0|B^0_q\rangle$;
- τ_{B_q} and M_{B_q} are the B^0_q lifetime and mass, G_F is the Fermi constant, M_W is the $W^\pm$ boson mass, θ_W is the Weinberg angle;
- m_ℓ is the mass of the lepton in the final state;
- β_{m_ℓ} is the phase space volume of the final state, *i.e.*

$$\beta_{m_\ell} \equiv \sqrt{1 - \frac{4m^2_\ell}{M^2_{B^0_q}}} \ ; \tag{1.61}$$

- y_q and $\mathcal{A}^{\ell\ell}_{\Delta\Gamma}$ are defined as follows

$$y_q \equiv \frac{\Gamma^q_L - \Gamma^q_H}{\Gamma^q_L + \Gamma^q_H} \tag{1.62}$$

$$\mathcal{A}^{\ell\ell}_{\Delta\Gamma} \equiv \frac{\Gamma^{q,\ell\ell}_L - \Gamma^{q,\ell\ell}_H}{\Gamma^{q,\ell\ell}_L + \Gamma^{q,\ell\ell}_H} \tag{1.63}$$

being $\Gamma^q_{L,H}$ and $\Gamma^{q,\ell\ell}_{L,H}$ the B^0_q total width and the $B^0_q \to \ell\bar{\ell}$ partial width for the light and heavy eigenstate respectively; these quantities are needed to correctly compare the untagged $B^0_q \to \ell^+\ell^-$ decay rate summed over the helicities of the leptons at $t = 0$ $\mathcal{BR}(B^0_q \to \ell^+\ell^-)|^{CP}_{t=0}$ (which is the quantity computed theoretically) with the measured time-integrated branching fraction which is instead the observable which is experimentally measured. The difference between these two expressions for the $\mathcal{BR}$ arises because of the interplay between the $B^0_q - \bar{B}^0_q$ mixing and the $B^0_q \to \ell\bar{\ell}$ decay. The product $y_q\mathcal{A}^{\ell\ell}_{\Delta\Gamma}$ is given by

$$y_q\mathcal{A}^{\ell\ell}_{\Delta\Gamma} = \frac{(1-y_q^2)\tau_{\ell\ell} - (1+y_q^2)\tau_{B_q}}{2\tau_{B_q} - (1-y_q^2)\tau_{\ell\ell}} \ , \tag{1.64}$$

where the effective $B^0_q(t) \to \ell\bar{\ell}$ lifetime $\tau_{\ell\ell}$ is expressed as a function of the CP-averaged decay width $\langle\Gamma(B^0_q(t) \to \ell\bar{\ell})\rangle \equiv \Gamma(B^0_q(t) \to \ell\bar{\ell}) + \Gamma(\bar{B}^0_q(t) \to \ell\bar{\ell})$ as

$$\tau_{\ell\ell} \equiv \frac{\int_0^\infty t\langle\Gamma(B^0_q(t) \to \ell\bar{\ell})\rangle dt}{\int_0^\infty \langle\Gamma(B^0_q(t) \to \ell\bar{\ell})\rangle dt} \ . \tag{1.65}$$

The quantity $\mathcal{A}^{\ell\ell}_{\Delta\Gamma}$ can assume, in principle, any value in the range $[-1, 1]$. The SM prediction is $\mathcal{A}^{\ell\ell}_{\Delta\Gamma} = 1$ [37], but it has never been constrained so far. Experimentally, it can be measured using the effective lifetime $\tau_{\ell\ell}$, through the relation in eq.(1.64).

- P and S are Wilson coefficient dependent combinations defined as (m_b is the b-quark mass)

$$\mathcal{P} = (\mathcal{C}^\ell_{10} - \mathcal{C}'^\ell_{10}) + \frac{M^2_{B_q}}{2m_\ell}\frac{m_b}{m_b + m_q}(\mathcal{C}^\ell_P - \mathcal{C}'^\ell_P) \ , \tag{1.66}$$

$$\mathcal{S} = \beta_{m_\ell}\frac{M^2_{B_q}}{2m_\ell}\frac{m_b}{m_b + m_q}(\mathcal{C}^\ell_S - \mathcal{C}'^\ell_S) \ . \tag{1.67}$$

Discussion on the $\mathcal{BR}$. In the SM only the C^ℓ_{10} Wilson coefficients in eq.(1.60) is non negligible. This is due to the fact that in the SM only left-handed currents interact weakly. The computed value for the C^ℓ_{10} in the SM case is

$$\mathcal{C}^{\ell SM}_{10} \equiv \mathcal{C}^{SM}_{10} = -4.31 \ . \tag{1.68}$$

A contribution to $\mathcal{C}_S$ coming from the mediation of a Higgs boson is suppressed by a factor $\sim 10^3$ with respect to the contribution due to $\mathcal{C}_{10}$.

The SM contribution receives an additional suppression due to the factor $\frac{m_\ell}{M_{B_q}}$ in front of it. This is the so-called *helicity suppression*, which makes the $\mathcal{BR}$ vanishing for massless leptons in the final state. This is a consequence of the momentum and spin conservation in the decay of a spin-0 particles into two spin-$\frac{1}{2}$ daughters. To better explain why this happens it is helpful to analyze the decay in the spin-0 mother's rest frame. In this reference, the two daughter particles are emitted back-to-back along a given direction, say z, that can be used as quantization axis

for their spin; in order to conserve the spin projection along this z-direction, the spin projections of the two daughters must be opposite. This means that the two daughters have the same helicity, defined as the spin projection along the momentum directions (*i.e.* $h \equiv \vec{s} \cdot \frac{\vec{p}}{|p|}$). Now, a massless fermion and its antiparticle have opposite chirality. Since in the massless limit chirality eigenstates are also helicity eigenstates, it follows that the decay of a spin zero particle into a pair of massless spin-$\frac{1}{2}$ particle-antiparticle is forbidden.

In NP scenarios, the other Wilson coefficients as well can be non negligible and the $\mathcal{C}_{10}$ coefficient can receive a (lepton flavor dependent) shift $\delta C_{10}^{\ell NP}$.

An important remark concerns the fact that only one hadronic non-perturbative parameter enters the expression of the $\mathcal{BR}$. This is the $B^0_{(s)}$ meson decay constant $f^2_{B_{(s)}}$, and it is currently the most important source of uncertainty in the theoretical prediction. The important fact to stress here is that this parameter appears as an overall factor, hence it cancels out in ratios of observables. This is also true for the elements of the V_{CKM} matrix.

From eq.(1.60), it can be seen that the $\mathcal{BR}$ can only constrain the differences $(\mathcal{C}^{\ell}_{10,P,S} - \mathcal{C}^{\ell\prime}_{10,P,S})$. For this reason the $\mathcal{BR}(B^0_{(s)} \to \ell\bar{\ell})$ observable is not sensitive to NP effects shifting in the same "direction" the $\mathcal{C}^{\ell}_{10,P,S}$ and the $\mathcal{C}^{\ell\prime}_{10,P,S}$ coefficients. In order to constrain their sum, other processes must be exploited, for instance the decays $B \to K\ell\bar{\ell}$.

Another remark concerns the fact that depending on the NP scenario, the value of the $\mathcal{BR}$ can be higher but also lower than the SM prediction. In particular, if NP enters only the $\mathcal{C}_S^{(\prime)}$ coefficient, the value of the $\mathcal{BR}$ can only be higher than the SM prediction; if NP appears in the pseudoscalar operator $\mathcal{O}_P$ then the $\mathcal{BR}$ can also have a lower value with respect to the SM prediction, depending on the relative phase between the two terms "$2\frac{m_\ell}{M_{B_q}}(\mathcal{C}^{\ell}_{10} - \mathcal{C}^{\ell\prime}_{10})$" and "$(\mathcal{C}^{\ell}_{P} - \mathcal{C}^{\ell\prime}_{P})$".

The ratio $\frac{\mathcal{BR}(B^0\to\mu^+\mu^-)}{\mathcal{BR}(B^0_s\to\mu^+\mu^-)}$. An additional observable to be measured is given by the ratio of the branching fractions for the $B^0 \to \ell\bar{\ell}$ and $B^0_s \to \ell\bar{\ell}$ modes [41] (see also [42]):

$$\mathcal{R} \equiv \frac{\mathcal{BR}(B^0 \to \mu^+\mu^-)}{\mathcal{BR}(B^0_s \to \mu^+\mu^-)} = \frac{\tau_B}{\tau_{B_s}} \frac{M_B}{M_{B_s}} \frac{f_B^2}{f_{B_s}^2} \left|\frac{V_{td}}{V_{ts}}\right|^2 . \tag{1.69}$$

This ratio is sensitive to departures from the Minimal Flavor Violating (MFV) hypothesis for NP scenarios.

1.2.3 SM predictions

The last updated SM predictions for the branching fraction of the $B^0_{(s)} \to \ell\bar{\ell}$ modes have been computed in Ref.[39]. These predictions include both electromagnetic and QCD corrections, at $\mathcal{O}(\alpha_{em})$ and $\mathcal{O}(\alpha_s^2)$ respectively.

For the $B^0 \to \ell\bar{\ell}$ modes the predicted $\mathcal{BR}$ are:

$$\mathcal{BR}(B^0 \to e^+e^-) = (2.48 \pm 0.21) \times 10^{-15} \tag{1.70}$$
$$\mathcal{BR}(B^0 \to \mu^+\mu^-) = (1.06 \pm 0.09) \times 10^{-10} \tag{1.71}$$
$$\mathcal{BR}(B^0 \to \tau^+\tau^-) = (2.22 \pm 0.19) \times 10^{-8} \tag{1.72}$$

while for the $B^0_s \to \ell\bar{\ell}$ modes are found

$$\mathcal{BR}(B^0_s \to e^+e^-) = (8.54 \pm 0.55) \times 10^{-14} \tag{1.73}$$
$$\mathcal{BR}(B^0_s \to \mu^+\mu^-) = (3.65 \pm 0.23) \times 10^{-9} \tag{1.74}$$
$$\mathcal{BR}(B^0_s \to \tau^+\tau^-) = (7.73 \pm 0.49) \times 10^{-7} \tag{1.75}$$

As has been pointed out in Ref.[38], this particular dependence is a direct consequence of the conservation of parity in strong interactions.

For each $B^0_{(s)}$ mode, the ratio of the $\mathcal{BR}$ for two final different final states is given by

$$\frac{\mathcal{BR}(B^0_{(s)} \to \ell\bar{\ell})}{\mathcal{BR}(B^0_{(s)} \to \ell'\bar{\ell'})} = \frac{m_\ell^2}{m_{\ell'}^2} \cdot \frac{\beta_{m_\ell}}{\beta_{m'_\ell}} \tag{1.76}$$

where β_{m_ℓ} is defined in eq.(1.61).

The ratio $\frac{m_\ell^2}{m_{\ell'}^2}$ is the helicity suppression factor while the term $\frac{\beta_{m_\ell}}{\beta_{m_{\ell'}}}$ comes from the available phase space in the decay. This ratio is an additional observable theoretically clean, since the uncertainties from the non-perturbative hadronic form factors and from the knowledge of the CKM matrix elements cancel in the ratio. A departure from the theoretical prediction for these quantities would be a clear indication of LFU violating extensions of the SM.

Concerning the ratio $\mathcal{R}$ defined in eq.(1.69), the SM prediction is (see [49] and references therein)

$$\mathcal{R} = 0.0295^{+0.0028}_{-0.0025} \ . \tag{1.77}$$

1.2.4 $B^0_{(s)} \to \ell\bar{\ell}$: current bounds

The modes with two electrons in the final state have the smallest branching fractions (because of the higher helicity suppression). The measurement of their $\mathcal{BR}$ is out of the experimental reach within the near future, and only the following Upper Limits ($\mathcal{UL}$) (at 90% C.L.) are available so far [43]:

$$\mathcal{BR}(B^0 \to e^+e^-) \ < \ 8.3 \times 10^{-8} \tag{1.78}$$
$$\mathcal{BR}(B^0_s \to e^+e^-) \ < \ 2.8 \times 10^{-7} \tag{1.79}$$

The $B^0_{(s)} \to \mu^+\mu^-$ modes are the "cleanest" ones under the experimental point of view. Their searches have started more than 30 years ago, and the first evidence for the $B^0_s \to \mu^+\mu^-$ decay has been obtained by the LHCb Collaboration [44] analyzing the first $2fb^{-1}$ collected during the first run of the LHC. The measured $\mathcal{BR}$ was:

$$\mathcal{BR}(B^0_s \to \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9} \ . \tag{1.80}$$

The same analysis set an Upper Limit ($\mathcal{UL}$) at 95% Confidence Level on the branching fraction of the $B^0 \to \mu^+\mu^-$ mode:

$$\mathcal{BR}(B^0 \to \mu^+\mu^-) < 9.4 \times 10^{-10} \ . \tag{1.81}$$

These values are consistent with the SM predictions and rule out huge NP effects. The latest results on this decay from the LHCb collaboration (together with the combination with the results from CMS) will be presented in Chapter 3.

The modes $B^0_{(s)} \to \tau^+\tau^-$ have the highest branching fraction but are still largely unexplored, being much more challenging from an experimental point of view (as it will be shown in Chapter 4). The BABAR collaboration measured an $\mathcal{UL}$ for the branching fraction of the $B^0 \to \tau^+\tau^-$ decay using a sample of $(232 \pm 3) \times 10^6$ $B\bar{B}$ events produced in the decay $\Upsilon(4S) \to B\bar{B}$ [45]. The reported limit at 90% Confidence Level is

$$\mathcal{BR}(B^0 \to \tau^+\tau^-) < 4.3 \times 10^{-3} \ . \tag{1.82}$$

A direct search for the $B^0_s \to \tau^+\tau^-$ mode has never been performed and the only available limits on its $\mathcal{BR}$ come from indirect constraints. These have been obtained in two ways:

- from an analysis of B decays with large missing energy at LEP [46] which gives an $\mathcal{UL}$ of 5%;
- from the measurement of the ratio of the $B^0_{(s)}$ lifetimes from LHCb [48] which gives an $\mathcal{UL}$ of 3%

Even though several orders of magnitude above the SM prediction, these bounds are the only ones available for these modes. Despite the practical difficulty of reconstructing modes with τ's in the final state (due to missing energy carried away by the ν), these modes are acquiring an increasing importance and are triggering a lot of interest in connection to hints of LFU violating NP scenarios. In particular, as it will be shown in the next section, decays involving τ's can still show sizable NP effects even if the corresponding modes with μ or electrons were in agreement with SM predictions.

1.3 Current anomalies in B-meson decay observables and hints of NP in third generation

The possibility that NP could give sizable contributions to observables related to processes involving third generation leptons has been considered for several years, well before the start of the LHC.

In this section the main motivations for the search of the $B^0_{(s)} \to \tau^+\tau^-$ will be briefly reviewed.

The anomalous like-sign dimuon charge asymmetry. A reason of interest in the rates of processes generated by the effective four-fermions operator $(\bar{b}s)(\bar{\tau}\tau)$ raised in relation with the observed anomalous like-sign di-muon asymmetry A^b_{SL} observed by the DØ and CDF Collaborations [52, 55]. The weighted average for A^b_{SL} is

$$A^b_{SL} = (-74.1 \pm 19.3) \times 10^{-4} \tag{1.83}$$

and must be compared with the SM prediction [56]:

$$A^{b(SM)}_{SL} = (-2.3 \pm 0.4) \times 10^{-4} \ . \tag{1.84}$$

The discrepancy between the measured value and the theoretical prediction is $\sim 3.8\sigma$. In Ref.[57] has been pointed out the fact that the enhancement in A^b_{SL} can be explained only by four-fermions operators of the type $(\bar{b}\Gamma s)(\bar{\tau}\Gamma\tau)$ and $(\bar{b}\Gamma s)(\bar{c}\Gamma c)$ (where Γ is a generic matrix with spinorial indexes). In Ref.[51] the possible enhancement of the $B^0_{(s)} \to \tau^+\tau^-$ rate has been studied in two specific models:

- in a leptoquark model an enhancement no larger than 0.3% is allowed for the $\mathcal{BR}(B^0_s \to \tau^+\tau^-)$;
- in a model with a light Z' which does not couple with light quarks an enhancement up to $\sim 5\%$ is still possible for $\mathcal{BR}(B^0_s \to \tau^+\tau^-)$.

Nevertheless none of the previous scenarios account for the observed discrepancy in A^b_{SL}, and only the model with the light Z' can help to reduce the tension still without solving it.

LFU violation. The interest in modes with τ leptons in the final state is further strengthened by the recent observation of anomalies in B decays involving $b \to s\ell\bar{\ell}$ transitions. In particular, these anomalies have triggered a lot of interest in relation to LFU violation in B decays. In this framework, the study of the $B^0_{(s)} \to \tau^+\tau^-$ mode acquires an increasing importance.

Among the observed anomalies, the most relevant motivating the interest in final states with third generation leptons are:

- R_K *anomaly.* The LHCb collaboration has measured the ratio of the decay rate of the $B^+ \to K^+\ell\bar{\ell}$ modes for $q^2 \in [1,6]\ GeV^2/c^4$ (where q^2 is the invariant mass square of the dimuon system). The result is [58]:

$$R_K \equiv \frac{\mathcal{BR}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{BR}(B^+ \to K^+e^+e^-)} = 0.745^{+0.090}_{-0.074}(stat) \pm 0.036(syst) \tag{1.85}$$

  and must be compared with the SM prediction which is $R_K = 1$. The result (1.85) is 2.6σ away from the SM.

- $B \to D^{(\star)}\tau\nu$ *anomalies.* These anomalies appear in the tauonic current interactions. The interesting observables are the ratios ($\ell = \mu, e$)

$$R_{D^{(\star)}} = \frac{\mathcal{BR}(\bar{B} \to D^{(\star)}\tau\bar{\nu}_\tau)}{\mathcal{BR}(\bar{B} \to D^{(\star)}\ell\bar{\nu}_\ell)} \tag{1.86}$$

  which in the SM are predicted to be [59, 60]:

$$R_D^{SM} = 0.296 \pm 0.016\ , \tag{1.87}$$
$$R_{D^\star}^{SM} = 0.252 \pm 0.003\ . \tag{1.88}$$

  The BABAR [61], Belle [62, 63, 64] and LHCb [65] Collaborations have measured such ratios, and their average gives the following results [66]:

$$R_D = 0.391 \pm 0.041(stat) \pm 0.028(syst)\ , \tag{1.89}$$
$$R_{D^\star} = 0.322 \pm 0.018(stat) \pm 0.012(syst)\ , \tag{1.90}$$

  which result in a combined tension of 3.9σ with respect to the SM predictions.

Taken singularly, none of these measurements has a significance high enough to claim for deviations from the SM. Nevertheless, if the central values are confirmed after the updated analysis with the new data from the RunII of the LHC, they points towards a NP scenario with LFU violation. The consequences of such scenarios have been analyzed in the framework of a MLFV extension of the SM in Ref.[16]. In particular for the modes with two τ in the final state a boost of $\sim 10^3$ is predicted, thus bringing the $\mathcal{BR}(B_s^0 \to \tau^+\tau^-)$ at a level of $\sim 1 \times 10^{-3}$, which is reachable in the near future.

Lifetime differences in the $B_q^0 - \bar{B}_q^0$ system. In Ref.[50], NP scenarios giving an enhancement of the lifetime differences in the $B_q^0 - \bar{B}_q^0$ system have been studied; in this model dependent analysis, the enhancement of $\Delta\Gamma_s/\Gamma_s$ is correlated with an enhancement of the $\mathcal{BR}(B_{(s)}^0 \to \tau^+\tau^-)$. In particular, in a NP scenario with scalar Leptoquarks an enhancement up to $\sim 18\%$ and $\sim 0.4\%$ respectively for the B_s^0 and the B^0 modes is predicted.

1.4 Conclusions

In this Chapter a brief introduction to the Standard Model of elementary particle physics has been presented. The crucial point of the SM is that the interactions among the elementary particles is a natural consequence of a symmetry of the fundamental lagrangian under local transformations mixing them. The requirement of the invariance of the Lagrangian under the

The predicted value for $\mathcal{BR}(B^0 \to \tau^+\tau^-)$ is just below the current experimental limit from BABAR [45].

gauge group of the SM forbids hard mass terms (both for fermions and for gauge connections) and the masses of elementary particles must be generated in a dynamical way through the Higgs mechanism.

Despite the high predictive power of the SM, there are several hints indicating that it must be considered as an effective low-energy realization of a more fundamental theory.

Among the processes that could reveal NP effects there are the rare dileptonic $B^0_{(s)} \to \ell\bar{\ell}$ decays. To describe in a model independent way these processes, the Operator Product Expansion formalism has been introduced and a discussion about the relevant observables has been presented. Even though the golden channel under the experimental point of view is the mode with two μ in the final state, several recent anomalies observed in $b \to s\ell\bar{\ell}$, motivates the search for the more challenging decay to two τ in the final state. Indeed, in LFU violating scenarios, NP effects can show up in these modes even if the corresponding decays with lighter leptons were in agreement with the SM predictions.

Chapter 2

The LHCb detector

LHCb is a dedicated B-physics experiment installed at CERN at the Large Hadron Collider (LHC). It is mainly devoted to the study of CP-violation and the search for rare decays of b-hadrons. These are created in the hadronization process of the $b\bar{b}$ pairs produced in the collisions between the proton beams accelerated with the LHC.

The LHC has been designed to accelerate two proton beams circulating in opposite directions at an energy of 7 TeV each (corresponding to an energy of 14 TeV in their center of mass system); nevertheless during the first run of the LHC, in the years 2011-2012, protons have been accelerated up to an energy of 3.5 and 4 TeV respectively in 2011 and 2012 (corresponding to an energy in the center of mass of 7 and 8 TeV respectively). Only from June 2015, after the Long Shutdown I, the LHC has started working at an energy in the center of mass of 13 TeV.

In this chapter the LHC environment will be presented together with a brief overview of the LHCb detector.

2.1 The LHC accelerator complex

The LHC is a two rings superconducting proton accelerator. It is installed in the 27 km tunnel used previously for the LEP program. The beams accelerated in the LHC are structured in proton bunches and are guided by different varieties of magnets:

- 1232 dipole magnets of 15 meters in length, which bend the beams
- 392 quadrupole magnets of 5-7 meters long, which focus the beams.

Before being injected in the LHC rings, the proton beams are accelerated through the accelerating machines already present in the CERN site. Protons are obtained from hydrogen atoms from which the electrons are stripped through an intense electric field. Protons are accelerated up to 50 MeV by the LINAC before being boosted up to 1.8 GeV by the Proton Synchroton Booster. Through the Proton Synchroton the proton beams are accelerated up to an energy of 25 GeV before reaching the last accelerator, the Super Proton Synchroton (SPS) which boosts them up to 450 GeV. Then protons are injected in the LHC accelerator rings. All this chain is schematically shown in Fig.2.1.

Along the LHC are installed four main experiments. Two of them are general-purpose experiments: "A Toroidal LHC ApparatuS" (ATLAS) and the "Compact Muon Solenoid" (CMS). The other two experiments are dedicated to the study of the quark-gluon plasma and QCD in extreme conditions of pressure and temperature, "A Large Ion Collider Experiment" (ALICE), and to precision measurements in b and c-hadron decays, LHCb.

Instantaneous and integrated Luminosity. The most important parameter characterizing an accelerator is the *instantaneous luminosity* $\mathcal{L}$ defined as

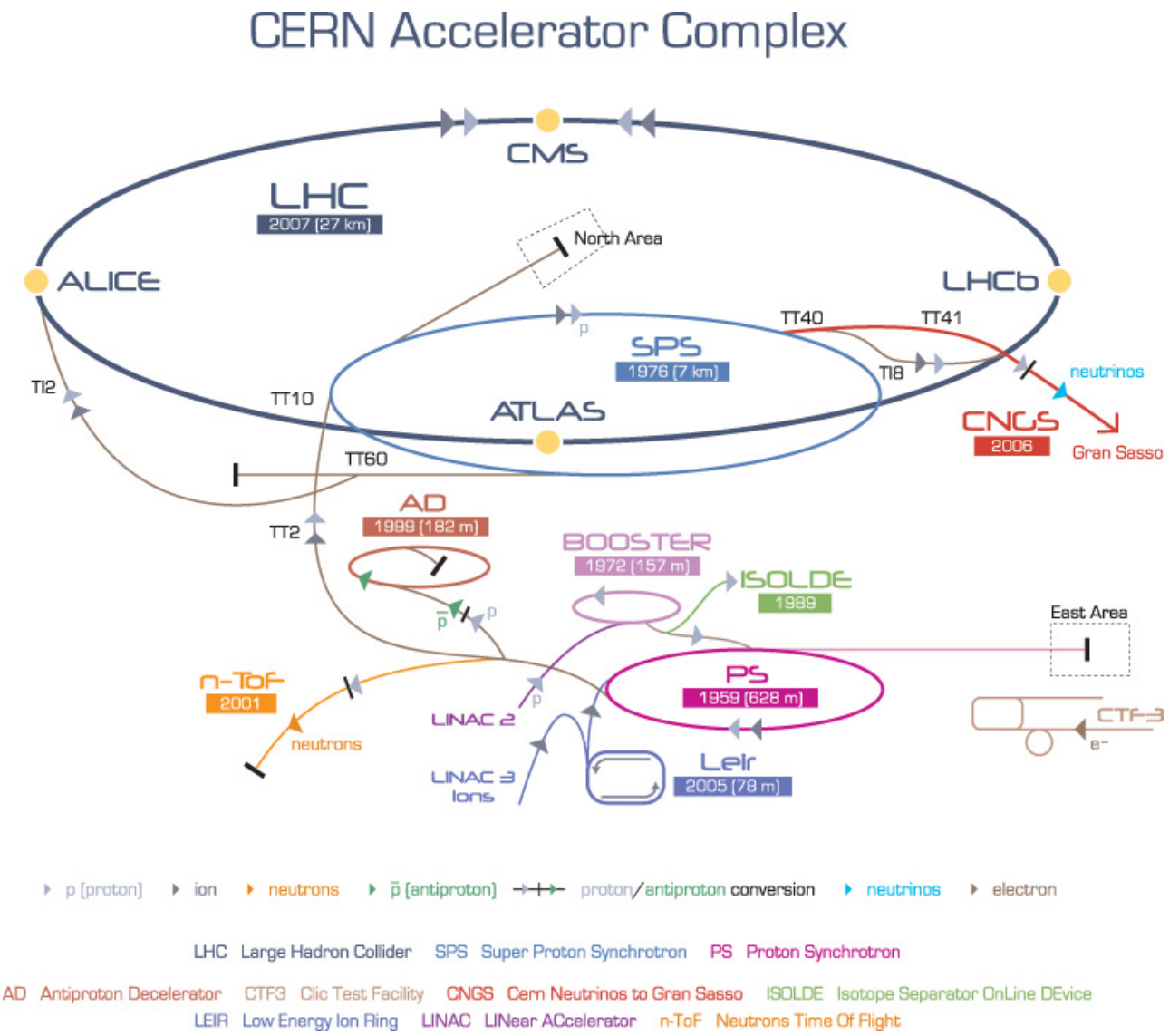


Figure 2.1: The CERN accelerator complex.

Parameter	2011	2012	Nominal
Center of mass energy (TeV)	7	8	14
Particles per bunch (N)	1.2×10^{11}	1.2×10^{11}	1.2×10^{11}
Number of Bunches (n)	1800	1800	2808
Bunch revolution frequency [kHz] f_{rev}	11	11	11
$\sigma^{\star}_{x,y}$ [μm]	~ 60	$\sim$60	$\sim$15
Bunch spacing [ns]	50	50	25
Instantaneous luminosity [$cm^{-1}s^{-1}$] ($\mathcal{L}$)	$3.65 \cdot 10^{33}$	$3.65 \cdot 10^{33}$	10^{34}

Table 2.1: LHC parameters during the 2011 and 2012 data taking periods, and their nominal values.

$$\mathcal{L} = \frac{N^2 n f_{rev}}{4\pi\sigma_x\sigma_y} F \tag{2.1}$$

where

- N and n are the number of protons per bunch and the number of bunches respectively;
- f_{rev} the bunch revolution frequency;
- $\sigma_{x,y}$ the sizes of the beam in the transverse plane to the beam axis;
- F a geometrical factor taking into account the crossing angle between the two colliding beams and separations of the beams at the interaction point.

The designed instantaneous luminosity of the LHC is $\mathcal{L} = 10^{34} cm^{-1}s^{-1}$ and is reached using $\sim$ 2835 proton bunches per beam crossing each 25ns, which corresponds to a spatial separation between two bunches of $\sim$ 7.5 m. Each bunch consists of an average number of 10^{11} protons. In Tab.2.1 the main parameters of the LHC are reported, for different running configurations.

The instantaneous luminosity can also be expressed as a function of the average number of interactions per bunch crossing μ as follows [67]:

$$\mathcal{L} = \frac{\mu n f_{rev}}{\sigma_{ine}} \tag{2.2}$$

being σ_{ine} the inelastic pp cross section (which is equal to 73.5 ± 3.1 mb at $\sqrt{s} = 7$ TeV).

Each of the four experiments at the LHC works with different running conditions by tuning the geometrical factor F and the beam section sizes $\sigma_{x,y}$. At the LHCb detector the constraints from the detector readout and the trigger thresholds require to keep the average number of collisions μ of 1.7.

The collisions between the protons beams, together with other processes (like interaction of the protons with gas molecules present in the beam pipe), reduce progressively the beam population. On average, the beam lifetime is about ten hours from the first collisions. For that reason also the instantaneous luminosity will decrease with time. In order to have stable data taking conditions, in LHCb (and ALICE), the decrease of the beam population is compensated by modifying the geometrical factor F bringing the two beams closer to each other. This technique is usually referred as luminosity leveling [69]. In ATLAS and CMS the geometric factor is maintained at its maximum value, Fig.2.2 shows the instantaneous luminosity as a function of time for the four LHC experiments.

The instantaneous luminosity can be integrated over time, giving the *integrated luminosity* L:

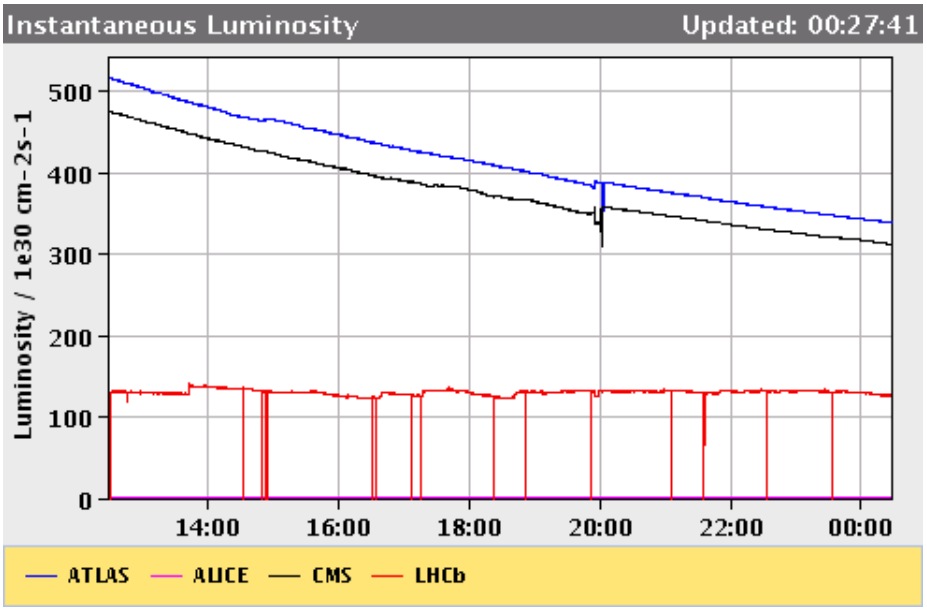


Figure 2.2: Instantaneous luminosity as a function of time for the four LHC experiments.

$$L \equiv \int dt \mathcal{L} \, . \tag{2.3}$$

Fig.2.3 shows the integrated luminosity delivered by LHCb experiment during the first two years of data taking.

2.2 The $b\bar{b}$ production cross section at LHC

The production of $b\bar{b}$ pairs at the LHC is the result of the interactions among the partons during the collisions between the protons. In particular the three processes responsible for the $b\bar{b}$ pair production are [70, 71, 72]:

- pair production, accounting for 16% of the total production, which is due to the elementary processes: $q\bar{q} \to b\bar{b}$ ($q\bar{q}$ annihilation) and $gg \to b\bar{b}$ (gluon fusion);
- flavor excitation, representing 54% of the total production. It takes place when one of the two $b\bar{b}$ virtual quarks in the proton quark pair undergoes an interaction with a parton of the other proton;
- gluon splitting (or separation), contributing for 27% of the total production;

The Feynman diagrams describing the above mentioned processes are shown in Fig.2.4.

The angular distribution of the $b\bar{b}$ pairs produced in the collisions has a non trivial distribution. Indeed, the $b\bar{b}$ pair is very closed to the direction of the colliding protons. The angular correlation between the polar angle of the two b quarks with respect to the direction of the incoming proton is shown in Fig.2.5.

2.3 The LHCb detector

The LHCb detector is a one arm spectrometer in the forward direction covering an angle comprise between 10 and 250 mrad around the beam pipe. The angular coverage is usually expressed as a function of the pseudo-rapidity η defined as follows:

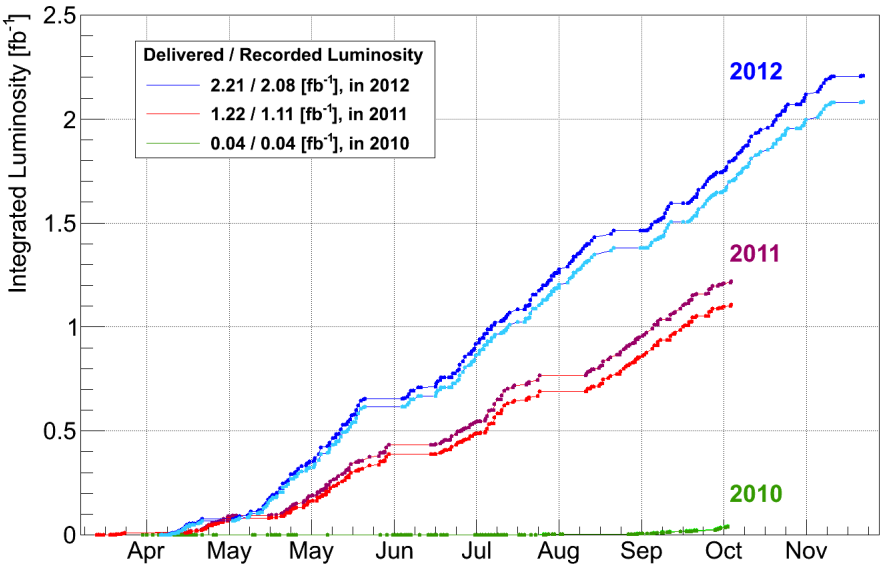


Figure 2.3: Integrated luminosity delivered and recorded by the LHCb experiment during the first two years of data taking.

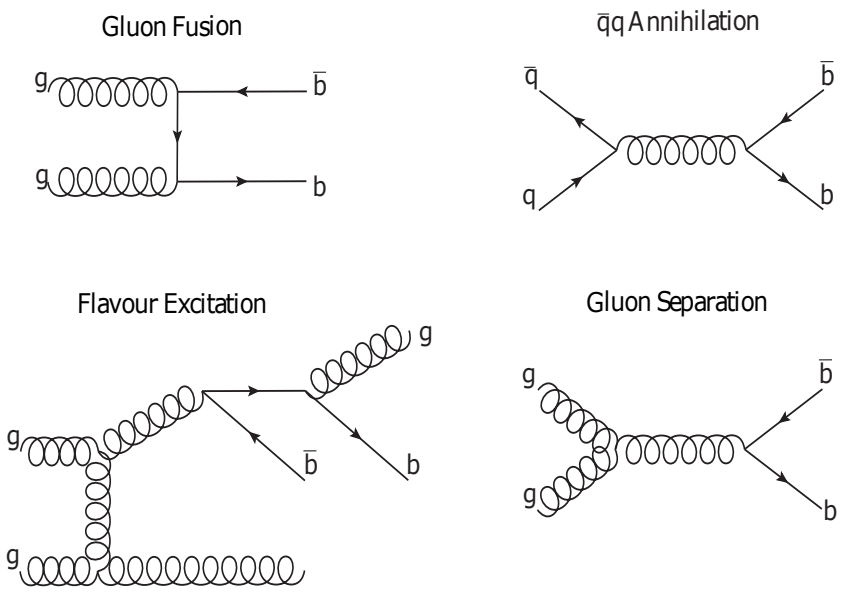


Figure 2.4: $b\bar{b}$ production mechanism in proton-proton collisions at the LHC.

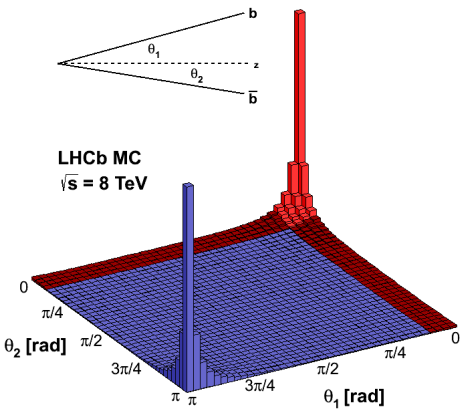


Figure 2.5: Angular correlation between the two b quarks produced in proton-proton collision.

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \tag{2.4}$$

In terms of η the LHCb acceptance is between 2 and 5.

A schematic representation of the LHCb detector is shown in Fig.2.6.

The particular forward geometry is dictated by the angular distribution of the $b\bar{b}$ pairs produced in the proton-proton collisions. Indeed, despite its reduced angular acceptance, the detector covers the region in the polar angle with the highest rate of $b\bar{b}$ events, as it is shown in Fig.2.7

The LHCb detector uses information collected from different sub-detectors to:

- reconstruct charged tracks through the tracking system devices (described in Sec.2.3.1)
- identify the tracks through Particle IDentification (PID) dedicated sub-detectors (presented in Sec.2.3.2).

In order to reduce the data flow coming from the detector and select only the potentially interesting events containing a b-hadron decay, the LHCb detector is also equipped of a multi-level trigger system briefly presented in Sec.2.3.3.

In a very schematic way, the functioning of the detector can be summarized in the following steps:

- the b-hadrons are produced in the collision point, where the two proton beams cross each other. The Vertex Locator surrounds the interaction region in order to reconstruct where exactly the two protons interacted;
- the produced b-hadron travels inside the Vertex Locator for a distance of $\sim$1 cm; then it decays into its daughter particles that travel all along the different sub-detectors which collect the relevant information for their track reconstruction and particle identification;
- a hardware trigger system, based on information coming from some sub-detectors, selects only potentially interesting events, which will be then fully reconstructed and analyzed off-line.

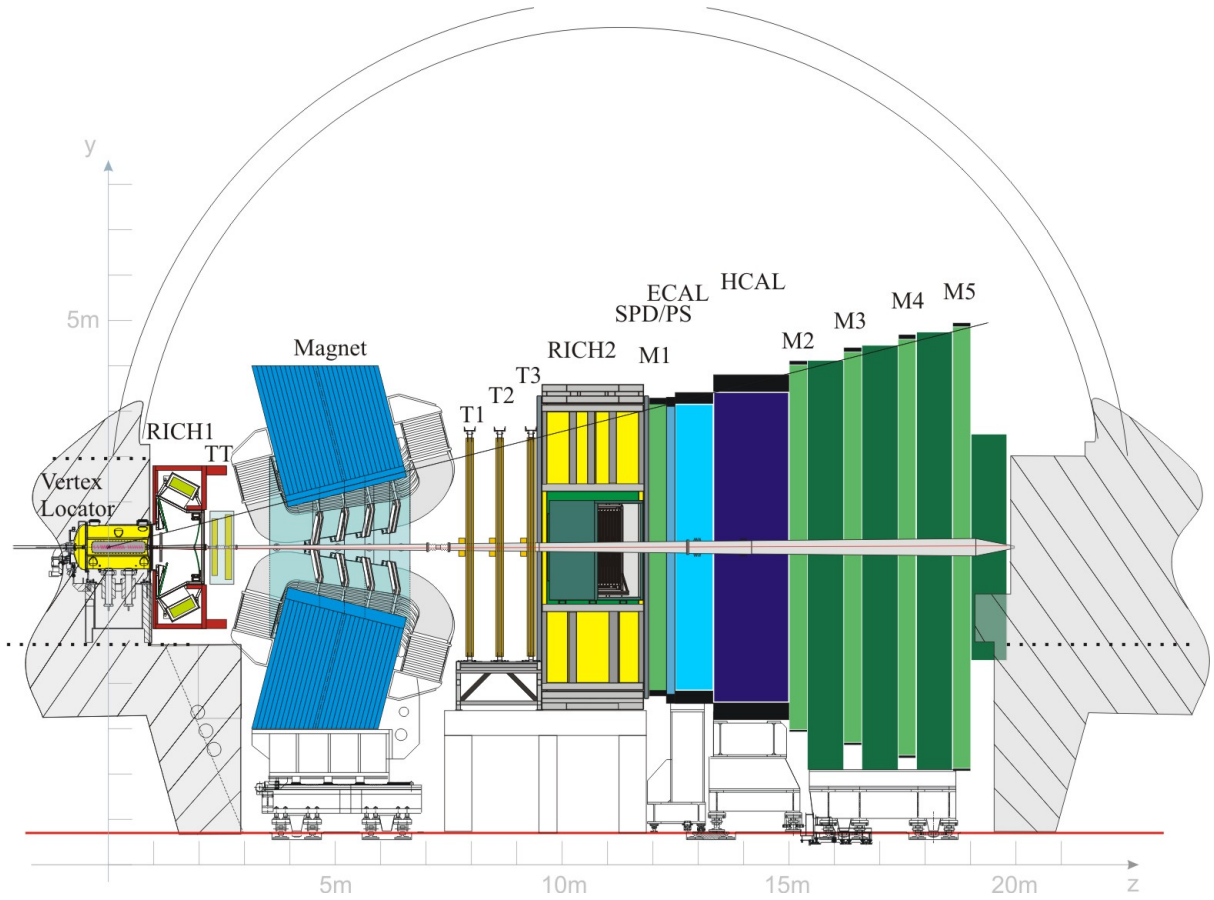


Figure 2.6: Lateral section of the LHCb detector. Particles are created at the interaction points between the two proton beams, at the origin of the z axis, which is inside the Vertex Locator (VELO).

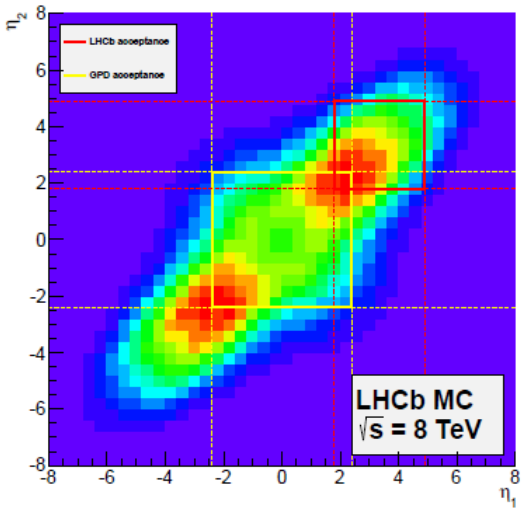


Figure 2.7: Correlation between the pseudo-rapidities of the two b quarks. The red square is the region covered by the LHCb detector, while the yellow one is the region covered by CMS and ATLAS.

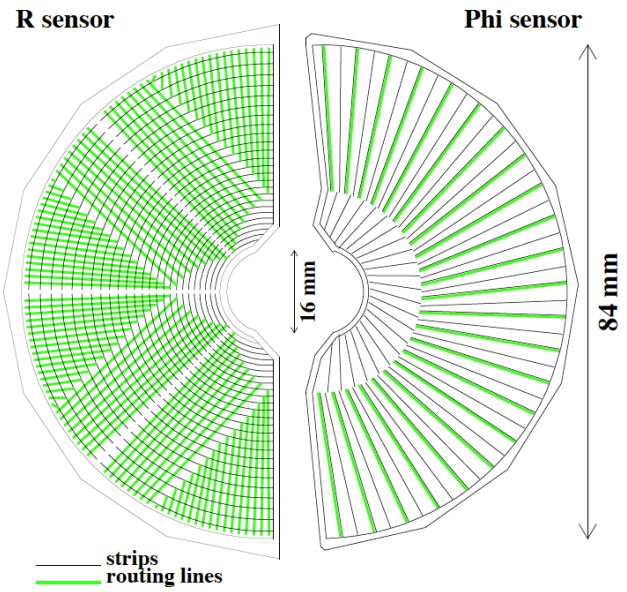


Figure 2.8: VELO sensors geometries.

2.3.1 The tracking system

One of the main features of b-hadrons is their long lifetime. For instance, the lifetime of a B_s^0 meson is about 1ps. This results, thanks also to the high boost of the $b\bar{b}$ pairs produced, in a large separation ($\sim$1cm) between the primary interaction vertex of the protons and the b-hadron decay vertex. In order to exploit as much as possible this property in the search for b-hadron decays and reject background events, an efficient vertexing system is required. In addition, an efficient tracking is required not only to resolve the different vertices in the event, but also to achieve a good momentum resolution of the tracks, which will reflect in a good mass resolution. In the following the various sub-detectors forming the LHCb tracking system will be presented.

The Vertex Locator. The Vertex Locator (VELO) is located around the crossing points of the two proton beams and its purpose is to provide information on the trajectories of charged particles in the region close to the interaction point. Through this information it allows to reconstruct the position of the primary and secondary vertices. The VELO sensors are shown in Fig.2.8: they are silicon modules with two different segmentations: one in ϕ and the other in r. The complete VELO detector is shown in Fig.2.9

The VELO contains also a Pile-Up veto stations, which provides to the hardware trigger system informations about the number of interactions in the event.

In a typical event at the LHCb around 30-35 tracks per interaction vertex are reconstructed. In these conditions, the resolution on the primary vertex is $\sim$12μm in the plane transverse to the beam axis, and $\sim$65μm along the direction of the beam axis.

The magnet. The magnet is used to measure the tracks' momenta. It generates a magnetic field of 4 Tesla and, by bending the trajectory of charged tracks, it allows to measure their momenta with a relative error $\delta p/p \sim 0.4\%$. The polarity of the magnet is periodically inverted in order to study left-right asymmetries detections effects and to evaluate related systematic errors for CP-violation related analysis.

Tracker stations. The tracking stations are divided in the following two categories:

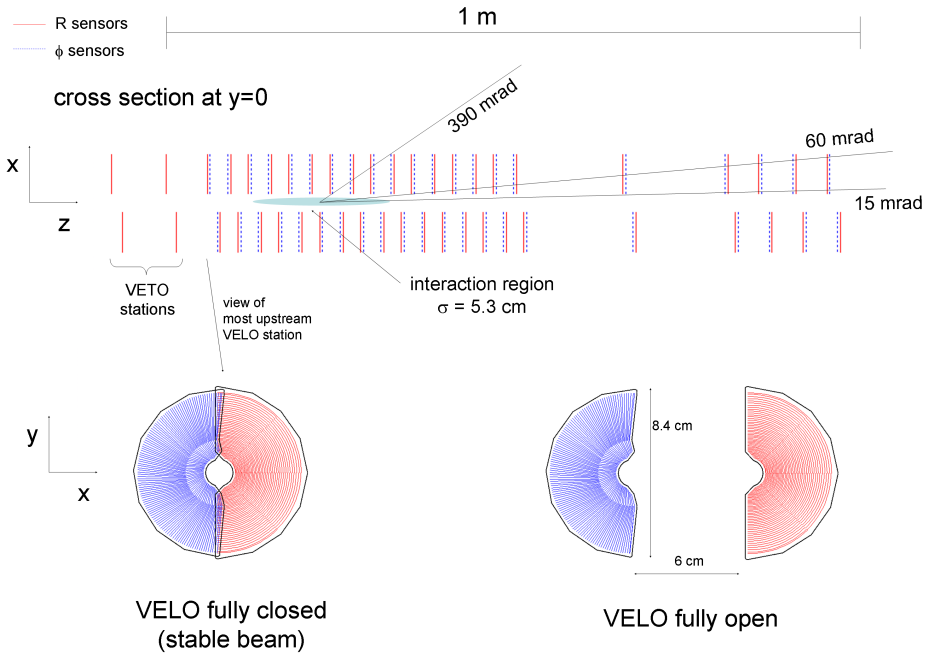


Figure 2.9: Schematic illustration of the VELO detector of LHCb.

T-stations are placed upstream and downstream the magnet. Depending on the detection principle, the trackers are divided into silicon trackers (TT and IT1-3) and outer trackers (OT1-3).

- *Tracker Turicensis* (TT) are placed upstream the magnet and downstream the RICH1 detector (see next section). The main goal of this system is to determine the momentum of tracks having a high impact parameter with respect to the reconstructed primary interaction vertex of the two proton beams (PV). They are made of silicon strips with a high granularity in order to achieve a good track resolution in the high tracks density region near the beam axis.

- 3 *T-stations* placed downstream the magnet and before the RICH2 detector. Depending on the distance with respect to the beam axis, these stations are divided into:

  – *Inner Tracker* (IT) covering the region closest to the beam axis. Similarly to the TT, it consists of silicon detectors with a high granularity in order to better resolve tracks in the high charged tracks density region close to the beam pipe

  – *Outer Tracker* (OT) covering the rest of the area of the T tracker. It is used to measure the momenta over a wider acceptance. It is made of straw tubes because of the smaller track density in the outer region on the T-stations

Fig.2.10 shows the tracking stations, while the whole tracking system is shown in Fig.2.11.

Track reconstruction. Depending on the tracking station where they are detected, the following kinds of tracks are defined (see also Fig.2.12):

- Velo tracks leave hits only in the VELO stations and are used to reconstruct the primary vertices;

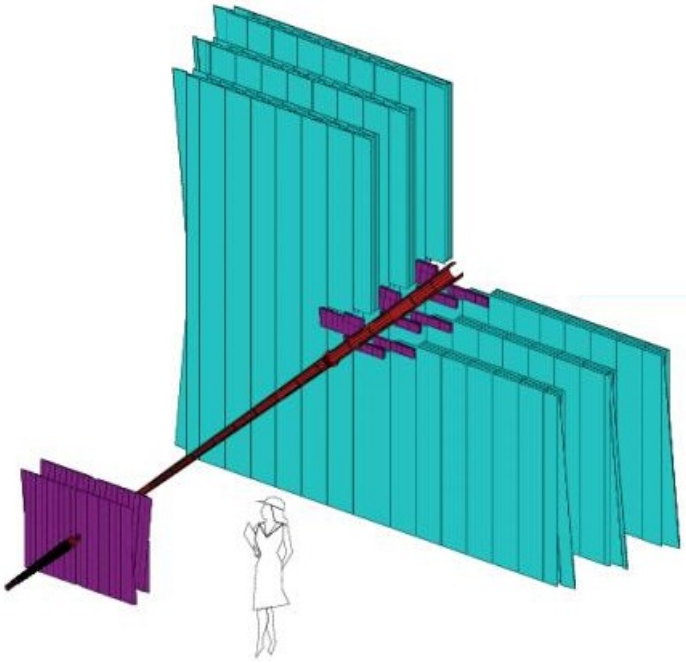


Figure 2.10: The T-stations. The different colors indicates different technologies: in violet the silicon tracker in blue the straw tubes.

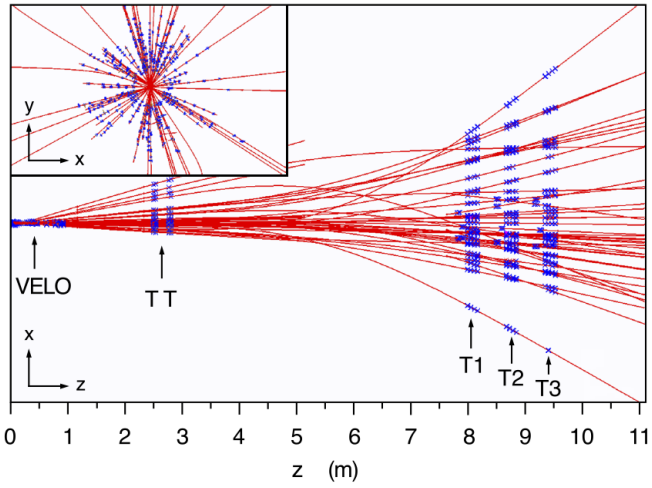


Figure 2.11: Schematic representation of the tracking system functioning.

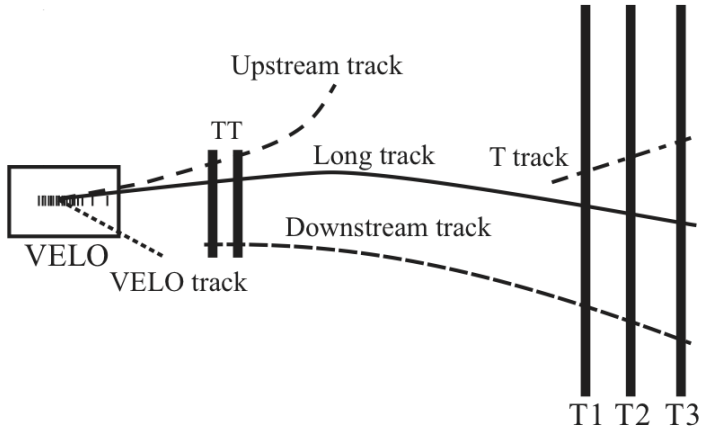


Figure 2.12: Track types in LHCb.

- T tracks traverse only the T stations and are usually generated by secondary interactions of the produced particles in the collision with the material of the detector;
- Upstream tracks leave hits in the VELO and the TT stations;
- Downstream tracks traverse the TT and T stations;
- Long tracks leave hits across the whole tracking system, allowing the most precise measurement of their momentum and impact parameter. Most physics analysis use these tracks.

Track reconstruction algorithm. The LHCb track reconstruction algorithm is divided into two steps:

- a track candidate is built by associating hits in the T stations to the track candidate. The association of the track candidate with the hits is done with different strategies [73];
- for each track candidate, the corresponding track is fitted using a Kalman fitter technique [74]. The quality of a track is quantified by the χ^2 of the fit, the pulls of the track parameters, and a covariance matrix .

Because of the several strategies adopted to build the track candidates the same particle can be associated to different track candidates, called *clone tracks*. This multiplicity is solved by running a "*clone killer*" algorithm [78]. In addition, some track candidates can be built up of random hits in the detector. These fake tracks are usually referred to as "*ghost*" tracks, and a likelihood method to quantify the probability that a given track candidate be a ghost has been defined in Ref.[79].

The efficiency of the tracking system and reconstruction algorithm is estimated to be around 96% for Long Tracks.

Vertex and tracks related geometrical variables Once the tracks and vertexes have been reconstructed, variables quantifying their quality and their geometrical properties are computed in order to be used in the physics analyses.

Vertex positions are reconstructed by fitting together the reconstructed tracks that are compatible with being originating from the same point. The quality of the reconstructed vertexes is quantified by the fit χ^2. To quantify the distance between two reconstructed vertexes in the event (for instance between a PV and a candidate B meson decay vertex) the following two variables are used:

- "*flight dstance*" (FD) defined as the distance between the two vertexes,
- "*flight distance* χ^2" (FDχ^2), defined as the χ^2 of the vertex obtained by fitting together the two vertexes.

The most relevant variables related to the reconstructed tracks are:

- "*impact parameter*" with respect to the reconstructed interaction vertex (IP). This is defined as the distance of closest approach of the reconstructed PV and the track. Decay vertexes of b or c-hadron decays are well displaced with respect to the PV and tracks originating from such decays are thus expected to have a large value of the IP;
- "*impact parameter* χ^2" ($IP\chi^2$), defined for each reconstructed track as the change in the χ^2 of the reconstructed PV after including the considered track to the set of those (here referred as "$\{tracks_{PV}\}$") used to fit the PV:

$$IP\chi^2 \equiv \chi^2(\{tracks_{PV}\} + \text{track}) - \chi^2(\{tracks_{PV}\}). \tag{2.5}$$

  If the considered track has been originated from a b or c-hadron decay the change in the χ^2 of the PV is expected to be large.

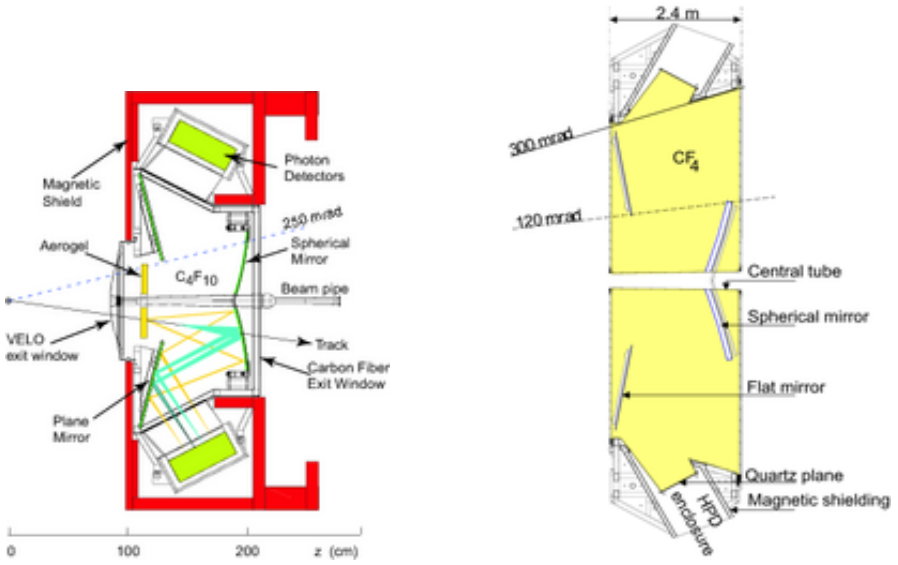
2.3.2 The Particle Identification

The particle identification of tracks is a key ingredient to study B decays. For instance, the study of the $B^0_{(s)} \to \mu^+\mu^-$ requires a good muon identification but also an efficient separation between π and K for the invariant mass PDF calibration for the signal from the control channel $B \to hh^{(\prime)}$ (where $h^{(\prime)} = \pi, K$). This particle identification is achieved at LHCb using the information from the sub-detectors presented in the following.

The RICH detectors. The Ring Imaging Cherenkov (RICH) detectors are used to identify pions, kaons, and protons. A RICH detector allows to measure the velocities of particles through the ring of Cherenkov light produced by charged particles when passing inside a *radiator* medium. LHCb uses two RICH detectors: the RICH1 and RICH2 situated respectively upstream and downstream the magnet. While the RICH1 covers all the angular acceptance of LHCb, the RICH2 covers a limited acceptance. Also the momentum range of the particles they can discriminate is different: from 1 to 60 GeV for the RICH1 and from 15 to 100 GeV for RICH2. Figs.2.13(a)-2.13(b) shows schematically the two RICH detectors.

The calorimeter system. The calorimeter system is designed to estimate the energy of charged and neutral particles and at the same time to provide information for their identification. In particular, the calorimeter system is made up of the following sub-detectors:

- the Scintillator Pad Detector (SPD) is used to distinguish charged particles from neutral ones;
- the Pre-Shower detector (PS) allows to distinguish photons and electrons from hadrons;
- the Electromagnetic Calorimeter (ECAL) detects electron and photons through their electromagnetic shower. Using the information from the PS, it allows an energy measurement with a resolution $\delta E/E = 9\%/\sqrt{E} \oplus 0.8\%$ (E expressed in GeV);
- the Hadronic Calorimeter (HCAL) is situated upstream the ECAL and allows a measurement of the energy of the hadrons of $\delta E/E = 69\%/\sqrt{E} \oplus 9\%$ (E expressed in GeV);



(a) (b)

Figure 2.13: RICH detectors: RICH1 (2.13(a)) and RICH2 (2.13(b))

The muon system. This sub-detector consists of five stations (M1-M5). In particular the M1 station is placed upstream while the remaining (M2-M5) downstream with respect to the calorimeter system, as shown in Fig.2.14. Given a reconstructed track, a candidate muon is found when hits in the muon stations match the direction of a track reconstructed in the tracking system. In particular a test statistics for the muon or non-muons hypothesis is built looking at the average distance of the closest hits in the muon station to the extrapolation of the track direction and the boolean variable IsMuon is defined.

2.3.3 The Trigger system

In the LHC running conditions presented in Sec.2.1, the total inelastic pp cross section is ∼70 mb. This value is well above the cross section for the $b\bar{b}$ production, as shown in Fig.2.15. For this reason, a selective trigger system is required before performing physics analysis. The aim of the LHCb trigger system is to select events containing B and D decays with hadrons and leptons in the final state while rejecting as much as possible the background events created in the proton-proton collisions. In particular, the trigger system has to reduce the event rate down to 5kHz (from the starting frequency of 40 MHz of the proton-proton collision). This rate is the one allowed from the stockage resources.

The trigger system of LHCb is splitted into two levels: a level zero (L0) hardware trigger which performs an on-line selection reducing the rate down to 1MHz, and a High Level Trigger (HLT) which is a software trigger which refines the L0 selection and further reduces the rate down to the required 5kHz. The set of trigger decisions is encoded in a Trigger Configuration Key (TCK) which uniquely identifies the requirements used for the selection of the events.

The Level-0 trigger. The L0 trigger exploits simple features of B decays. Using the information from the calorimeters and the muon system, it returns a positive decision depending on:

- transverse energy of the candidate in the calorimeters (L0hadron, L0electron, L0photon),
- transverse momentum of the candidates in the muon system (L0muon, L0dimuon).

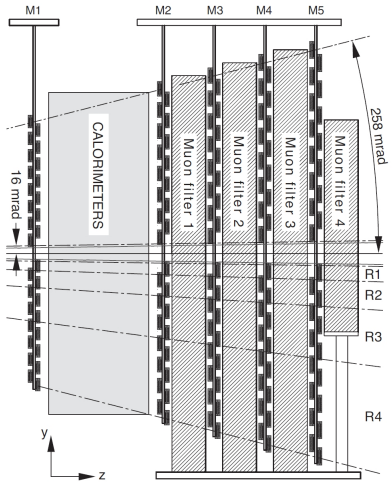


Figure 2.14: Representation of the muon stations.

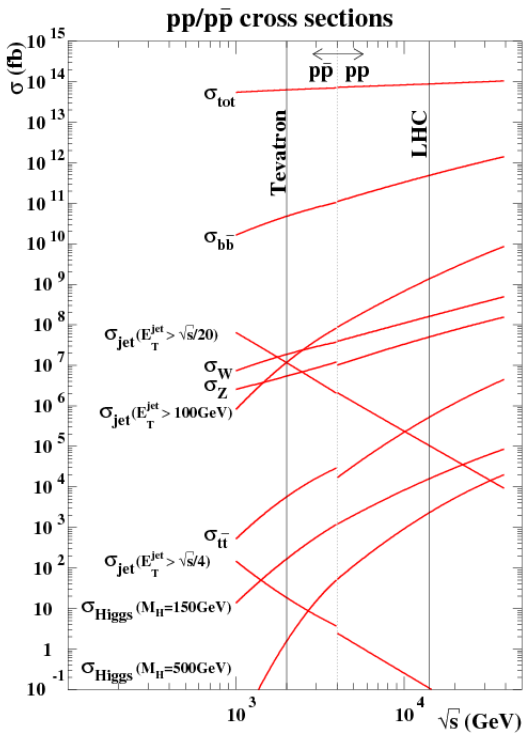


Figure 2.15: *pp* cross section for different processes as a function of the energy in the center of mass of the two proton beams ($\sqrt{s}$).

Line		Requirements	GEC
L0SingleMuon	TCK1 TCK2-3	$p_T > 1.48$ GeV/c $p_T > 1.76$ GeV/c	n.SPD hits < 600
L0DiMuon	TCK1 TCK2-3	$\sqrt{p_{T1} \times p_{T2}} > 1.3$ GeV/c $\sqrt{p_{T1} \times p_{T2}} > 1.6$ GeV/c	n.SPD hits < 900
L0Hadron		$E_T > 3.5$ GeV	n.SPD hits < 600

Table 2.2: L0 trigger requirements for the relevant lines of the analysis presented in this work.

Line	TCK	Requirements		
Hlt1DiMuonLowMass	TCK1 TCK2-3	$IP\chi^2 > 3$ $IP\chi^2 > 6$	$p_{T\mu} > 0.5$ GeV/c $p_{T\mu} > 0.5$ GeV/c	$m_{\mu\mu} > 1.0 GeV/c^2$ $m_{\mu\mu} > 1.0 GeV/c^2$
Hlt1DiMuonHighMass		$p_{T\mu} > 0.5$ GeV/c	$m_{\mu\mu} > 2.5$ GeV/c^2	
Hlt1TrackMuon		$IP\chi^2 > 16$	$IP > 100\mu m$	$p_{T\mu} > 1$ GeV/c
Hlt1TrackAllL0		$IP\chi^2 > 16$	$IP > 100\mu m$	$p_{T\mu} > 1.85$ GeV/c
GEC		VELO Hits $< 10^4$	IT Hits < 3000	OT Hits < 15000

Table 2.3: HLT1 trigger requirements for the lines relevant for the analysis presented in this thesis.

In addition, in order to reject very busy events in term of the number of tracks (which will require long time to be processed by the High Level Trigger), a Global Event Cut (GEC) is applied. This cut exploits the following information:

- number of PV's estimated through the Pile-Up system in the VELO,
- multiplicity of charged tracks measured by the SPD.

In Tab.2.2 the L0 trigger requirements for the lines relevant in the analysis presented in this thesis are reported.

The High Level Trigger. The HLT refines the selection done from the L0 trigger by exploiting the information from the all sub-detectors. The HLT is in turn splitted into two levels:

- the HLT1 uses only the information from the VELO and the tracking stations. In Tab.2.3 the HLT1 trigger requirements for the lines relevant in the analysis presented in this thesis are reported.
- the HLT2 performs a full reconstruction of the event, thus making the final choice if retain or reject a given event. For the analysis presented here the main HLT2 trigger line is the "Hlt2Topo" which uses the output of a Multivariate discriminant (see Sec.3.2.1) to select the interesting events. More details about this line can be found in Ref.[75]. The other lines used are reported in Tab.2.4.

A schematic representation of the trigger system is shown in Fig.2.16.

2.4 Conclusions

In this Chapter the experimental setup for the studies presented in this thesis has been presented. After a brief introduction about the LHC complex accelerator, the main $b\bar{b}$ production mechanism have been presented. Then the LHCb detector has been briefly described in its main components:

Line	Requirements		
Hlt2DiMuonJPsi	$\text{IsMuon}_{\mu_1,\mu_2}$= true	$\lvert m_{\mu\mu} - m_{J/\psi}\rvert < 120$ MeV/c^2	
Hlt2DiMuonDetached	$\text{IsMuon}_{\mu_1,\mu_2}$= true	$\lvert m_{\mu\mu} - m_{J/\psi}\rvert < 120$ MeV/c^2	DLS>3
Hlt2DiMuonBmm	$\text{IsMuon}_{\mu_1,\mu_2}$= true	$m_{\mu\mu} > 4.5$ GeV/c^2	DLS>3
GEC	VELO tracks < 350		

Table 2.4: HLT2 trigger requirements for the lines relevant for the analysis presented in this thesis. "DLS" is the distance between the primary and secondary vertexes divided by its error.

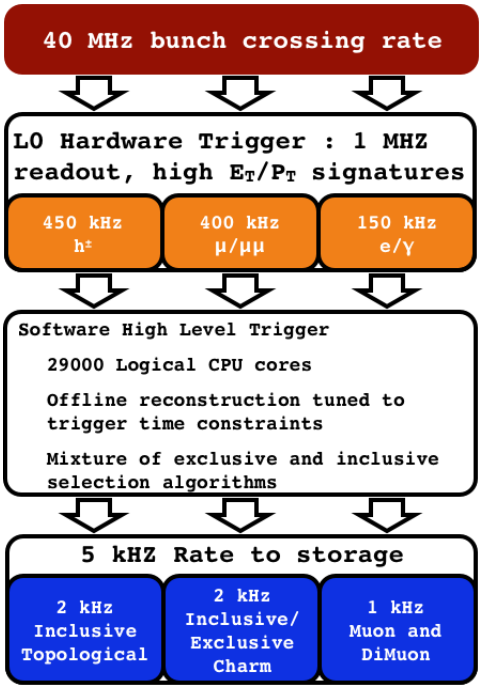


Figure 2.16: Scheme of the LHCb trigger architecture.

- the tracking system which allows to reconstruct with an excellent precision the interaction points as well as the secondary vertices in the event, and at the same times provides a very precise measurement of the charged tracks momentum;
- the RICH, the calorimeters, and the muon stations whose information are exploited in the particle identification algorithms and by the trigger system;
- the trigger system which allows to select only the potentially interesting events featuring a b-hadron decay, while rejecting the huge amount of background events produced in the proton-proton collision.

The good performances achieved by the LHCb detector are crucial to perform searches for rare decays, especially $B^0_{(s)} \to \mu^+\mu^-$. The good tracking algorithm performances together with the efficient particle identification allow to efficiently reconstruct very tiny signals and to achieve a good discrimination against the different sources of background.

Chapter 3

$B^0_{(s)} \to \mu^+\mu^-$

The $B^0_{(s)} \to \mu\mu$ analysis presented in this chapter is based on a dataset of 3.1 fb^{-1} collected during the first LHC Run in 2011-2012. 1 fb^{-1} has been collected at a center of mass energy of $\sqrt{s} = 7$ TeV during the last months of 2010 and in 2011, while the remaining 2.1 fb^{-1} have been collected in 2012 at $\sqrt{s} = 8$ TeV. The results presented in this chapter have been published in Refs.[77, 49].

The Standard Model (SM) expected rates for the $B^0 \to \mu^+\mu^-$ and the $B^0_s \to \mu^+\mu^-$ modes imply that in one billion of B^0_s only three B^0_s mesons decay into the di-muon channel, while only one B^0 in ten billions goes into a dileptonic final state. The first evidence for the $B^0_s \to \mu^+\mu^-$ decay has been obtained by the LHCb collaboration using a dataset of $2.1 fb^{-1}$ [44].

To measure the branching fraction ($\mathcal{BR}$) of the $B^0_{(s)} \to \mu\mu$ decay, the number of observed signal events in the dataset is extracted and then converted into a value for the $\mathcal{BR}$ using some normalization channels.

In Section 3.1 a general overview of the analysis is presented. In Section 3.2 the optimization of the performances of the Boosted Decision Tree (BDT) classifier is presented. Section 3.3 covers the $\mathcal{BR}$ measurement and limit extraction and in Section 3.4 the combined analysis of LHCb and CMS is presented. Finally in Section 3.7 the definition and optimization of new isolation variables and of the final Boosted Decision Tree classifier in view of future rounds of the analysis are presented.

3.1 Overview of the analysis

The $\mathcal{BR}$ of a given decay is proportional to the total number of observed signal events $N_{B_{(s)}}$ and inversely proportional to the total number $N_{B_{(s)}}$ of $B^0_{(s)}$ produced. Taking into account an efficiency factor ϵ_{sig} describing the effects due to the detector acceptance as well as the whole selection process, the following relation holds

$$\mathcal{BR}_{sig} = \epsilon_{sig}^{-1} \frac{N_{sig}}{N_{B_{(s)}}}. \tag{3.1}$$

The goal of the analysis is thus to select in the collected dataset those events being potentially the $B^0_{(s)} \to \mu\mu$ decays, count them and convert through relation (3.1) in a value for the $\mathcal{BR}$. To do that the number of $B^0_{(s)}$ mesons in the dataset must be extracted and the efficiency factor ϵ_{sig} must be computed.

The total number of B mesons in the sample is proportional to the integrated luminosity L through the relation

$$N_{B_q} = L \cdot \sigma_{b\bar{b}} \cdot f_q \cdot \epsilon_q \ , \tag{3.2}$$

where

- $\sigma_{b\bar{b}}$ is the $b\bar{b}$ pair cross section
- f_q is the hadronization fraction, *i.e.* the fraction of b-quarks hadronizing into a B_q meson
- ϵ_q is an efficiency factor due to the detector acceptance.

In the analysis presented here, nevertheless, the total number of $B_{(s)}$ is obtained in a different way. In particular eq. (3.1) is used for some *normalization channels* of already well known $\mathcal{BR}$. By inverting this relation the value of $N_{B_{(s)}}$ is obtained as a function of the number N_{Norm} of normalization channel events present in the dataset. More explicitly, the relation between the $\mathcal{BR}$ and the number of observed signal events N_{sig} is:

$$\mathcal{BR}_{sig} = \frac{\epsilon_{Norm}}{\epsilon_{sig}} \frac{\mathcal{BR}_{Norm}}{N_{Norm}} \cdot \frac{f_{Norm}}{f_{sig}} \cdot N_{sig} = \alpha_{Norm} \cdot N_{sig} \tag{3.3}$$

where $\frac{f_{Norm}}{f_{sig}}$ is the ratio of the hadronization probabilities for a b-quark into the signal and the decaying B meson in the normalization channel. The *normalization factor* α_{Norm} has been defined as the conversion factor between the observed number of signal events and the signal $\mathcal{BR}$.

The particular form of eq.(3.3) motivates the choice of the normalization channels. In particular they are chosen in order to be as much as possible similar to the signal, both with respect to the topology and the particle content of the final state. In such a way, possible systematics and biases in the reconstruction, trigger, and selection processes cancel out in their ratio. In addition these channels must be precisely measured.

The analysis proceeds in two steps: the $B^0_{(s)} \to \mu\mu$ candidate selection is achieved through a preliminary loose selection followed by a refined selection based on the output of a Multi Variate Analysis (MVA) classifier (see *e.g.* Ref.[80]). This first selection aims to remove as many obvious background events as possible while keeping a very high signal efficiency, in order to increase the sensitivity S which is defined as

$$S = \frac{N}{\sqrt{N+B}} \tag{3.4}$$

where N and B are the number of signal and background events respectively. In order to avoid any further removal of signal events from the sample, the selected events are classified with respect to two independent variables: the invariant mass of the di-muon system $m_{\mu\mu}$ and a variable describing the geometry of the event. This variable is the output of a second MVA classifier using kinematical and topological variables related to the signal candidate. The $B^0 \to \mu^+\mu^-$ and the $B^0_s \to \mu^+\mu^-$ signal yields are extracted through a simultaneous unbinned extended maximum likelihood fit in the $m_{\mu\mu}$ variable in eight bins of the output of the MVA classifier. If an excess of signal candidate events is observed, its significance is evaluated. If this is greater than 3σ a value for the $\mathcal{BR}$ is measured, if not, the observed pattern of events is compared with the expected one for several $\mathcal{BR}$ hypotheses and an upper limit on the $\mathcal{BR}$ is computed.

3.1.1 Signal features

Fig.3.1 shows a typical $B^0_{(s)} \to \mu\mu$ decay. The $B^0_{(s)}$ meson is produced at the interaction point between the two proton beams, which will be referred to as the Primary Vertex (PV). Thanks to its large lifetime (see Tab.3.1) it flies inside the VELO for $\sim$ 1cm before decaying into the two μ of the final state.

The signature of a $B^0_{(s)} \to \mu\mu$ decay in an event is thus the presence of two tracks identified as muons forming a good Secondary Vertex (SV) well displaced with respect to any other PV in the event; the sum of the momenta of the two tracks must be collinear with the direction defined by the $B^0_{(s)}$ production and decay vertexes, and the invariant mass of the two tracks $m_{\mu\mu}$ must be compatible with the one of the B^0_s or B^0 mesons.

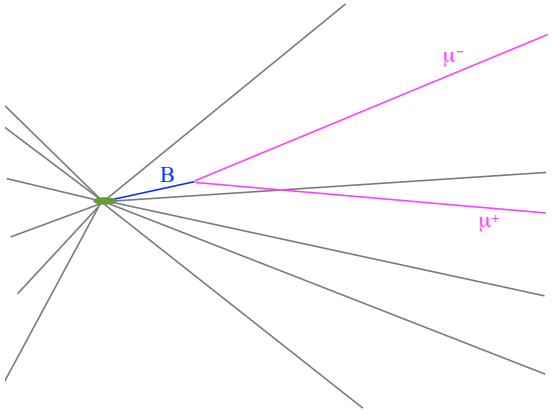


Figure 3.1: Schematic representation of a $B^0_{(s)} \to \mu\mu$ event.

Meson	quark content	Mass (MeV/c^2)	Mean life τ (s^{-12})
B^0_s	$s\bar{b}$	5366.3 ± 0.24	$1.472^{+0.024}_{-0.026}$
B^0	$d\bar{b}$	5279.5 ± 0.17	1.525 ± 0.009

Table 3.1: Properties of the B^0_s and B^0 mesons.

Since (as it will be shown in the next section) both the signal selection and the final classification rely on requirements on the quality of the reconstructed vertex formed by the two muons and the topology of the reconstructed $B^0_{(s)} \to \mu\mu$ candidates, the following two normalization channels are chosen

- $B^+ \to J/\psi(\to \mu\mu)K^+$: which, like the $B^0_{(s)} \to \mu\mu$ signal, contains two muons coming from the same vertex in the final state;
- $B^0 \to \pi^\pm K^\mp$ which, being a 2-body decay, has the same topology of the $B^0_{(s)} \to \mu\mu$ signal.

3.1.2 Trigger selection

At the trigger level, events featuring the presence of a $B^0_{(s)} \to \mu\mu$ decay candidate are selected by the following trigger requirements (see Sec.2.3.3):

- at the L0 level the high transverse momentum (p_T) of the muons coming from a B decay is exploited. In particular the following two triggers are used:
  - L0Muon,
  - L0Dimuons;
- at the HLT1 level the trigger selection relies on the single muon trigger Hlt1TrackMuon or the di-muon triggers Hlt1DiMuonLowMass and Hlt1DiMuonHighMass;
- the HLT2 trigger is based on Hlt2DiMuonBmm for the signal. For the $B^+ \to J/\psi K^+$ the Hlt2DiMuonJPsi is used (except for the last $470 fb^{-1}$ for which the Hlt2DiMuonDetached is used), while for the $B \to hh^{(\prime)}$ the Hlt2Topo2Body, Hlt2B2hhX and Hlt2B2hh decisions are used.

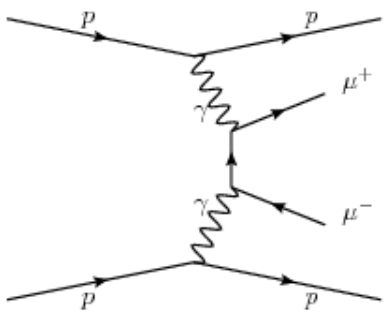


Figure 3.2: Feynman diagram for the production mechanism of the exclusive di-muons process.

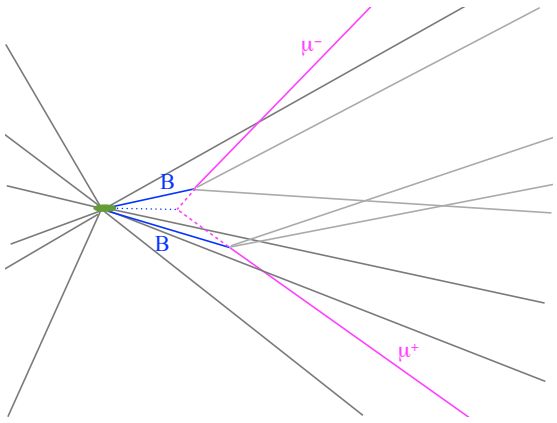


Figure 3.3: Schematic representation of a combinatoric background event for $B^0_{(s)} \to \mu\mu$.

3.1.3 Loose selection

The collected dataset contains a huge amount of background events, namely events appearing to be very similar to the signal but coming from other processes. The most important sources of backgrounds are:

- exclusive di-muons processes $pp \to p\mu\mu p$. The production mechanism is shown in Fig.3.2.
- Prompt combinatorial candidates where (at least) one of the two muons comes from the primary vertex.
- $B^0_{(s)} \to \mu\mu\gamma$ decays which can fake the $B^0_{(s)} \to \mu\mu$ signal because of the two muons coming from the same vertex. In addition, being a 3-body decay, this process is not helicity suppressed as in the case of the $B^0_{(s)} \to \mu\mu$ signal and can represent a potentially dangerous source of background.
- B cascade decays $B \to D(\to \mu X)\mu X$ which are due to the decay of the B into a muon and charmed meson, which subsequently decays into another muon and other (neutral) tracks
- Combinatorial background due to semileptonic decays of the two B mesons produced in the collision, as depicted in Fig.3.3, and which reconstructed as coming from the same fictitious vertex.
- $B \to hh'$ with $h, h' = \pi, K$ decays, where the two hadrons are misidentified because of decays in flight.

Mode	Requirements
$B^0 \to \pi^- \mu^+ \nu_\mu$	$m_{\pi\mu} > 4.5\ GeV/c^2$
$B_s^0 \to K^- \mu^+ \nu_\mu$	$m_{K\mu} > 4.5\ GeV/c^2$
$B^+ \to \pi^+ \mu^+ \mu^-$	both μ in the acceptance
$B^0 \to \pi^0 \mu^+ \mu^-$	both μ in the acceptance
$\Lambda_b^0 \to p \mu^- \bar{\nu}_\mu$	$m_{p\mu} > 4.5\ GeV/c^2$
$B_c^+ \to J/\psi(\to \mu^+\mu^-)\mu^+\nu_\mu$	$m_{\mu\mu} > 4.5\ GeV/c^2$
$\Lambda_b^0 \to p\pi$	
$\Lambda_b^0 \to pK$	

Table 3.2: List of the main exclusive modes studied in the present analysis, together with the generator cuts applied during the production of the related MC samples.

- Other hadronic or semileptonic decays of a B meson, where some of the daughter particles escape the detection or are misidentified as a muon, can contribute to the background. In Tab.3.2 the exclusive modes taken into account in the present analysis are listed.

In order to remove these background events from the dataset and increase the signal purity, a selection achieved in two steps has been designed.

A first loose selection aims to reduce the dataset to a manageable level by removing all the obvious background events. The main requirements of this selection are the following:

- **tracks quality requirements**: good track χ^2, small ghost probability , small Kullback-Leibler (KL) distance (by which tracks that have been duplicated at the reconstruction level are removed);
- **vertex quality and geometry requirements**: small distance of closest approach (DOCA) between the two tracks, good χ^2 of the vertex formed by the two muons. High flight distance significance of the candidate B meson, small impact parameter significance (IPχ^2) of the B and high IPχ^2 of the two daughters.

These cuts allow to remove the background events for which the two muons making the B candidates originate from two different vertexes.

- **kinematical requirements**: candidates are required to have an invariant mass $m_{\mu\mu}$ in a window of ± 500 MeV around the central value of the B_s mass, a minimum transverse momentum of the μ and the $B^0_{(s)}$ with respect to the direction defined by the beam axis (pT) greater than 0.25 GeV/c^2 and 0.5 GeV/c^2 respectively.

These kinematical cuts allow to remove the sources of (exclusive) backgrounds where the two muons in the final state are produced together with other undetected or unreconstructed particles. In particular the cut on the invariant mass $m_{\mu\mu}$ is very efficient to remove the $B_{(s)} \to \mu\mu\gamma$ background. Indeed, even if it is a three body decay not helicity suppressed, the di-muon system hardly passes the cuts on $m_{\mu\mu}$, being [81]

$$\mathcal{BR}(B_s^0 \to \mu\mu\gamma)_{E(\gamma)<60\ MeV/c^2} = 1.6 \times 10^{-12}\ . \tag{3.5}$$

The cut on the transverse momentum pT of the $B^0_{(s)}$ candidate is very efficient to remove the $pp \to p\mu\mu p$ processes. Indeed, even though each of the two muons has a high transverse momentum, their sum will result in a very soft p_T, being produced in opposite direction [82].

- **PID requirements** on the muon tracks. The variables used are (see Ref.[83] for a more detailed description):

i.e. the probability that a track has been reconstructed from hits due to different particles

- $DLL(\mu(K), \pi)$ defined as the logarithm of the ratio between the probability $P_{\mu(K)}$ for the track to be compatible with the ''$\mu(K)$-hypothesis" and the one P_π to be compatible with the π hypothesis:

$$DLL(\mu(K), \pi) = \ln\left(\frac{P_{\mu(K)}}{P_\pi}\right)$$

- the IsMuon variable which looks for hits in the muon stations around the extrapolation of the reconstructed track.

These requirements are meant to reduce backgrounds coming from misidentified hadrons.

In Tab.3.3 all the requirements of this loose selection are reported in detail.

Table 3.3: Selection for $B^0_{(s)} \to \mu^+\mu^-$, $B^0_{(s)} \to h^+h^-$, and $B^+ \to J/\psi K^+$ channels; DOCA is the distance of closest approach between the two tracks, VDS is the secondary vertex flight distance significance, and DLL is the combined PID likelihood to discriminate different particle hypotheses.

Cut	applied on	value	applied on	value
		$B^0_s \to \mu^+\mu^-$ and $B^0_{(s)} \to h^+h^-$		$B^+ \to J/\psi K^+$
track χ^2/ndf	μ / h	<3	μ / h	< 3
ghost prob		< 0.3		
DOCA		<0.3 mm		<0.3 mm
IPχ^2		>25		>25
p_T		> 0.25 and < 40 GeV/c		> 0.25 and < 40 GeV/c
p		<500 GeV/c		<500 GeV/c
IsMuon	μ only	true	μ only	true
vertex χ^2	$B_{(s)}$	<9	J/ψ	<9
VDS		> 15		> 15
ΔM		$\lvert M(hh, \mu\mu) - m_B \rvert <$ 60 MeV/c^2		$\lvert M(\mu\mu) - m_{J/\psi} \rvert <$ 60 MeV/c^2
IPχ^2	$B_{(s)}$	< 25	B^+	< 25
t		$< 9 \cdot \tau(B^0_s)$		$< 9 \cdot \tau(B^0_s)$
BDTS		> 0.05		> 0.05
DLL$(K - \pi)$		< 10		
DLL$(\mu - \pi)$		> -5		
ΔM				$\lvert M(J/\psi K) - m_B \rvert <$ 100 MeV/c^2
p_T (B^0_s)	$B^0_s \to \mu^+\mu^-$ $B^0_{(s)} \to h^+h^-$	> 0.5 GeV/c		

The loose selection presented is very efficient to remove all prompt background (thanks to the p_T requirement), the $B_{(s)} \to \mu\mu\gamma$ (thanks to the requirement on the di-muon invariant mass) and a large fraction of $B \to hh'$.

3.1.4 Tight selection

The sample selected with the criteria presented in the previous section is still polluted by a huge amount of combinatorial candidates, as well as of physical exclusives modes. A refined selection, based on a Boosted Decision Tree (BDT) algorithm (see Sec.3.2.1), has thus been designed to further increase the signal purity of the sample. The selection criteria based on the output of this BDT, hereafter called "BDTS", allows a background rejection of 70% in the whole di-muon invariant mass fit region while keeping an efficiency on the signal of 95%.

The BDT algorithm has been trained using the following set of six input variables:

- the impact parameter and the impact parameter χ^2 of the B candidate,
- the χ^2 of the B (or the J/ψ vertex, for the normalization channel $B^+ \to J/\psi(\to \mu\mu)K^+$),
- the angle between the direction of the momentum of the B candidate and the direction given by the B candidate production and decay vertexes (DIRA),
- DOCA between the two μ tracks,
- the smallest impact parameter of the two μ tracks with respect to *any* reconstructed primary vertex in the event.

To train the BDT classifier a sample of simulated signal events has been used, while a sample of events falling in the invariant mass sidebands $[4800 - 5000] \cup [5500 - 6000] MeV/c^2$ of the di-muon pair is used as background training sample.

The selected events are classified with respect to two variables, the invariant mass of the di-muon system $m_{\mu\mu}$ and the output of another MVA classifier which uses topological and kinematical input variables as well. The more these two variables are able to discriminate signal against the background, the higher the sensitivity of the analysis will be.

3.2 BDT optimization

One of the important improvements in the analysis presented here is the optimization of the MVA classifier used to extract the signal yields and the limit. In this section a brief introduction to the MVA techniques is given and the optimization of the performances of the BDT is then described.

3.2.1 Introduction to Boosted Decision Trees

A Multi Variate Classifier combines the information from different variables $\{x_i\}$ (called *input variables* of the classifier) differently distributed for signal and background events in a unique output variable $\mathcal{X} = \mathcal{F}(\{x_i\})$. The distributions of the output variable $\mathcal{X}$ for signal and background events will be much more separated than the ones of the single input variables, resulting in an improvement of the signal-background discrimination. In particular, depending on the output value $\mathcal{X}$ of the MVA classifier, it is possible to classify an event of unknown category (*i.e.* signal or background) as being signal or background-like. The function $\mathcal{F}(\{x_i\})$ does not have an analytic parametrization and its particular mapping from the set $\{x_i\}$ to the output value $\mathcal{X}$ is the result of a *learning algorithm* (or, equivalently, of a *learning machine*). Indeed in order to make the classifier learning how to classify events of unknown category using the information from the single input variables, a training must be done on the so-called *Training*

samples, one representative of signal and another for background. On these samples, the algorithm "learns" which are the distinctive features of signal and background events with respect to the input variables $\{x_i\}$.

In more details, the training process establishes the set of decisions (based on the distributions of the input variables $\{x_i\}$) by which the signal-background classification is achieved.

The most elementary example is a simple cut: given a variable $\bar{x}$ with known Probability Distribution Functions (PDFs) $f_{sig}(\bar{x})$ and $f_{bkg}(\bar{x})$ for signal and background events respectively, an optimal value $\bar{x}_{cut}$ can be fixed by requiring the maximization of a given criteria, *e.g.* the sensitivity defined below:

$$S(\bar{x}_{cut}) = \frac{\int_{\bar{x}_{min}}^{\bar{x}_{cut}} f_{sig}(y)dy}{\sqrt{\int_{\bar{x}_{min}}^{\bar{x}_{cut}} f_{sig}(y)dy + \int_{\bar{x}_{min}}^{\bar{x}_{cut}} f_{bkg}(y)dy}}. \tag{3.6}$$

Practically, while looking for the value $\bar{x}_{cut}$ maximizing $S(\bar{x}_{cut})$, a scan on m values of $\bar{x}$ is done, and the one giving the highest value for $S(\bar{x})$ is taken as $\bar{x}_{cut}$. Of course, the highest the value of m, the better the efficiency of the selected cut will be.

In a Boosted Decision Tree the previous process is iterated several times as follows (see Fig.3.4):

- the initial sample is splitted in two subsamples (defined by the cuts on the most discriminating variable x_1 among the input ones) with different signal purities S;
- then the two subsamples are further splitted using cuts on the most discriminating variable among the remaining ones (which can be, in principle, differents for *each* subsample);
- this process is iterated a certain number of times (called *Depth* of the tree) till a given criterium is fulfilled (*e.g.* no more input variables to use, too few events left in the subsample, or maximum number of splittings achieved);
- each of these intermediate sub-samples is called *node* of the tree, while each of the subsamples at the end of the splitting process is called *leaf*. Depending on its signal purity, a given node or leaf can be signal or background-like. The whole decision/splitting process is called *decision tree*.

Since the composition of the training events is known, it is possible to identify those events, which, being signal (background), are classified as background (signal) at the end of the decision process. The tree of decision can be repeated n-times assigning to misidentified events at the iteration i a higher weight with respect to the correctly classified ones. In this way at the $(i+1)^{th}$ iteration the decision processes will take more into account such events and thus all cuts will be tuned in a more refined way to correctly classify the previously misidentified events. This process is called *boosting* of the decision tree. The set of decision trees obtained after n iterations is called *forest*.

The most important advantages of the MVA classification are:

- a higher efficiency with respect to a simple chain of rectangular cuts, since the optimal set of cut values $\{\bar{x}^i_{cut}\}$ with respect to the input variables are fixed in a "global" way, looking at the interplay between the different input variables,
- the exploitation of the *correlations* between the input variables, which are not fully used in a chain of rectangular cuts.

The final output variable $\mathcal{X}$ is, by construction, more discriminating than each of the single input variables. The performances are usually evaluated on two independent signal and background samples (called *Test* samples) with respect to the ones used for the training of the operator, using the so-called Receiving Operating Characteristic (ROC) curve. This is built from

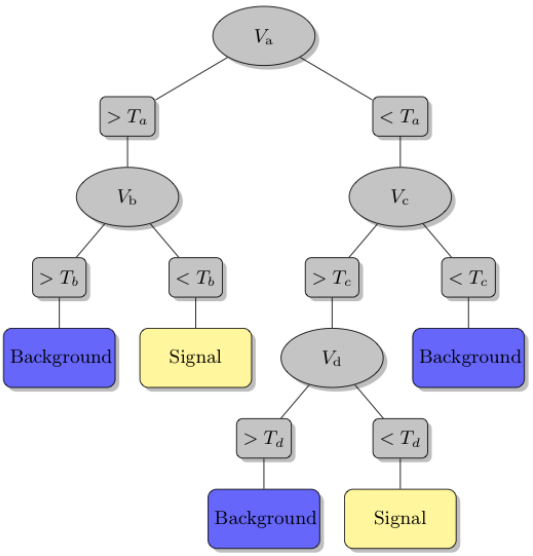


Figure 3.4: Schematic representation of a Decision Tree of maximal Depth equal to 3. V_i are the input variables, while T_i are the values of the cuts.

the (known) PDFs for the output variable $\mathcal{X}$ for signal ($F_{sig}(\mathcal{X})$) and background ($F_{bkg}(\mathcal{X})$) of the Test sample events as the pair of values

$$(\epsilon_{sig}(\mathcal{X}), 1 - \epsilon_{bkg}(\mathcal{X})) \tag{3.7}$$

where $\epsilon_{sig(bkg)}(\mathcal{X})$ is the efficiency for a given cut on $\mathcal{X}$ for signal and background respectively, *i.e.*

$$\epsilon_{sig(bkg)}(\mathcal{X}) = \int_{\mathcal{X}_{min}}^{\mathcal{X}} F_{sig(bkg)}(\mathcal{X}) d\mathcal{X} \ . \tag{3.8}$$

The performances of a classifier depend on three factors:

- the set of input variables $\{x_i\}$: the more the input variables are differently distributed and correlated between signal and background, the better will be the signal-background separation of the BDT output $\mathcal{X}$;
- the training parameters of the classifier: as explained, during the training of the classifier several parameters control the learning process, summarized in the following:
  - the number of decision trees in the forest (**nTrees**)
  - the minimum number of events required in a leaf (**EvtsMin**)
  - the maximum allowed depth of the decision tree (**MaxDepth**)
  - the number of steps during the optimization of the node cut (**nCuts**)
  - the parameter of the boosting algorithm (**AdaBoostBeta** or β);
- the size of the training samples for signal and background.

By optimizing the values of these training parameters, the performances of a BDT can be improved considerably, for example by increasing the number of trees in the forest, the depth of each decision tree and the number of cuts. The counterpart of this improvement is the risk of *overtraining* the classifier. This means that the learning process is very sensible to statistical fluctuations in the training sample, so that the classifier becomes very efficient in the

signal-background discrimination for the samples with which it has been trained, but looses in performances when classifying events in a statistically independent sample. Even if it is not a problem in itself it's better to reduce it as much as possible. In particular the overtraining can be cured :

- using large training samples for signal and background, in order to reduce the dependence on statistical fluctuations;
- tuning the BDT parameters to make the decision process as less as possible driven by the statistical fluctuations in the training sample. The most relevant parameters to this extent are the $nCuts$ and the $MaxDepth$, and, even if it has a smaller impact, the $nTrees$.

While the increase in the size of the training sample both reduces the overtraining and improves the signal background discrimination, the tuning of the BDT parameters in order to reduce the overtraining of the classifier can lead to a degradation of the performances of the classifier. For this reason, when optimizing the performances of a classifier, the best possible compromise between signal-background discrimination and overtraining of the classifier must be found.

The TMVA package. The tools used for the analysis presented here is the Toolkit for Multi Variate Analysis (TMVA) package [84]. This Toolkit implements several MVA techniques, in addition to the BDT algorithm, such as Neural Network (NN) classification algorithms.

No matter the particular MVA algorithm, the output given by the TMVA classifier is a number in the range $[-1, 1]$: it's close to -1 for background-like events or to 1 for signal-like events.

Flattening of the BDT . The distributions $F_{sig(bkg)}$ of the output variable $\mathcal{X}$ of the Boosted Decision Tree are transformed through the following change of variable, called *flattening*:

$$\mathcal{X} \to \mathcal{X}_{flat} \equiv \frac{\int_{\mathcal{X}_{min}}^{\mathcal{X}} F_{sig}(y)dy}{\int_{\mathcal{X}_{min}}^{\mathcal{X}_{max}} F_{sig}(y)dy} \tag{3.9}$$

By definition, the values of the variable $\mathcal{X}_{flat}$ are then constantly distributed in the interval $[0, 1]$ for signal events, while they are peaking at zero for background events.

Another way to interpret the flattening transformation is a rebinning (with differents bin sizes) of the histogram describing the distribution of the $\mathcal{X}$ variable. The rebinning is done in such a way that the fractions of signal events be equal in each bin.

3.2.2 Optimization of the input variables set and tuning parameters

One of the main improvements of the analysis presented here with respect to the previous ones consists in the use of an improved MVA classifier. To this extent, a larger sample of background combinatorial MC generated events has been generated. This sample corresponds to an integrated luminosity $\sim 50\ fb^{-1}$ (to be compared with the sample used for the previous rounds of the analysis, equivalent only to $\sim 0.5 fb^{-1}$). In term of number of events, 90164 generic MC $b\bar{b} \to \mu\mu X$ events have been used as a proxy for background, while 683671 $B_s^0 \to \mu\mu$ MC generated events for signal.

BDT input variables. The classifier used for the previous $B_{(s)}^0 \to \mu\mu$ analysis was trained using the following set of nine input variables:

- the B candidate meson proper time (τ),
- the impact parameter of the B (IP(B)),

- the transverse momentum of the B candidate ($p_T(B)$),
- the B candidate isolation based on the CDF definition ($I_{CDF}(B)$) [88], *i.e.* ,

$$I_{CDF} = \frac{p_T(B)}{p_T(B) + \sum_{track \in cone} p_T(track)} \tag{3.10}$$

  where the ''*cone*" is defined by the relation

$$\sqrt{\delta\eta^2 + \delta\phi^2} < 1.0 \tag{3.11}$$

  being $\delta\eta$ and $\delta\phi$ the difference in pseudorapidity and ϕ coordinate of the given track with respect to the B candidate,
- the minimum impact parameter significance of the muons with respect to any primary vertex (IP$\chi^2(\mu)$),
- the distance of closest approach of the two muons ($DOCA$),
- the isolation of the two muons (defined in Appendix A) with respect to any other track in the event ($I(\mu)$),
- the minimum transverse momentum of the two muons ($min_{pT}(\mu)$),
- the cosine of the angle between the muon momentum in the di-muon rest frame and the vector perpendicular to the plane defined by the B direction and the beam axis ("polar angle").

Only the $DOCA$ and the IP(B) are in common with the BDT classifier used for the tight selection.

Thanks to the larger available statistics of generic $b\bar{b} \to \mu\mu X$ MC events the effects of the following new input variables for the BDT classifier (in addition to the previous ones) have been tested:

- the absolute value of the difference between the pseudorapidity of the two muons ($\Delta\eta$),
- the absolute value of the difference between the spherical ϕ coordinate of the two muons ($\Delta\phi$),
- the angle between the B candidate momentum and the "thrust momentum" of the B (''other B angle"). The "thrust momentum" is defined as the sum of the momenta of all long tracks coming from a B and excluding those coming from long lived particles. These tracks are selected through requirements on the IPχ^2 with respect to the B PV (0.4 <IPχ^2 < 40), on their transverse momentum p_T (200 < $p_T(track)$ < 30000 MeV), and on the ratio of their transverse momentum and the one of the two muons ($\frac{p_T(track)}{p_T(\mu^\pm)} >$ 0.0001). If there is no track in the event satisfying these requirements the thrust momentum is set to 0. Said in a different way, the thrust momentum is the sum of the momenta of all tracks consistent with originating from the decay of the other b-hadron in the event,
- the angle between the direction of the μ^+ and the thrust momentum above defined in B rest frame (''other B boost angle").

These variables had already been studied for the previous round of the analysis [86] but due to the lack of MC statistics their performances couldn't be evaluated with enough precision and had been discarded at the moment.

The Figure of Merit of classifiers. The optimization of the BDT classifier has been performed with respect to the BDT tuning parameters and the set of input variables. To compare the performances of different classifiers, a so-called Figure Of Merit (FOM) must be defined. One possible choice (that had been adopted for previous rounds of the analysis) is to take the FOM as the integral of the ROC curve; nevertheless this method does not take fully into account the importance of the most sensitive bins.

For this round of the analysis a new FOM is used, and its computation is done as follows:

- for each bin of the BDT under optimization the expected number of background n_b and signal+background events n_{s+b} in the B_s^0 mass signal window are computed. The total number of signal and background events are taken to be the ones expected in the first half of 2011 $1.1 fb^{-1}$ dataset, while the efficiencies are taken from the signal and background test samples.

- under the assumption that in each BDT bin these numbers fluctuate as a poissonian distribution, the following quantity is computed

$$\Delta LQ_i = 2 \cdot \ln \left(\frac{\frac{Poiss(n_b, n_b)}{Poiss(n_{s+b}, n_b)}}{\frac{Poiss(n_b, n_{s+b})}{Poiss(n_{s+b}, n_{s+b})}} \right) \; ; \tag{3.12}$$

- the FOM for a given BDT operator is given by the sum over all the BDT bins of the ΔLQ_i.

The higher the value of the FOM, the higher the signal-background separation of the BDT classifier will be.

3.2.3 Correlation with the invariant mass

The BDT classifier is required to be as much as possible uncorrelated with the other variable used to classify the events in the dataset, *i.e.* the di-muon invariant mass $m_{\mu\mu}$, at least for the combinatorial background events. The reason has to do with the extrapolation of the combinatorial background invariant mass shape inside the signal mass windows, which, as it will be shown in the next section, is obtained by interpolating the distribution in the invariant mass sidebands. For this reason, a BDT classifier more performant for events falling in one or both the invariant mass sidebands and less performant for events falling in the signal region with respect to $m_{\mu\mu}$ will cause a bias in the estimation of the combinatorial background inside the signal mass window. Two kinds of correlation between the BDT output and the di-muon invariant mass can arise:

- linear correlation due to a BDT classifier more discriminating for events falling, for example, in the left $m_{\mu\mu}$ sideband and less performant for events falling in the right $m_{\mu\mu}$ sideband. This kind of correlation can still be dealt with during the fit and it has been shown (see Ref.[76]) that effects of correlations up to $\sim 10\%$ are taken into account by the systematic errors of the fit. This kind of correlation is quantified by the linear coefficient $\frac{\partial \mathcal{X}}{\partial m_{\mu\mu}}$;

- peaking correlation arising when the BDT classifier is more performant for events falling in both the left and right invariant mass sidebands, and less performant for those events with an invariant mass in the signal mass windows. This particular kind of correlation must be avoided at all, because it creates a false peak in the signal mass region. This peaking correlation can be estimated looking at the coefficient between the BDT output and the absolute value of the difference between the di-muon invariant mass and the B_s^0 mass, *i.e.* $\frac{\partial \mathcal{X}}{\partial |m_{\mu\mu} - M_{B_s^0}|}$.

	nTrees	β	EvtsMin	MaxDepth	nCuts
BDT9 $2.1fb^{-1}$	250	1	400	3	20
BDT9 optimized	1000	0.75	4200	8	20

Table 3.4: Tuning parameters of two BDT classifiers using the set of nine input variables used in the previous round of the analysis. In first line are reported the ones of the classifier used for the $2.1fb^{-1}$ paper (which were optimized on a background sample of equivalent integrated luminosity of 0.50 fb^{-1}, in second line the optimized values using the $50fb^{-1}$).

	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
BDT9 $2.1fb^{-1}$	3.6	1.12
BDT9 retrained	2.58	0.56

Table 3.5: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events for classifiers using nine input variables and trained on training samples of different sizes ($0.5fb^{-1}$ and $50fb^{-1}$).

Correlations of the BDT output with the invariant mass arise when one or more of the input variables are correlated with the mass. For example, variables related to the pointing are correlated with the reconstructed invariant mass of the two muons in the final state, both for signal and physical background events. Indeed, the emission of neutral or unreconstructed particles together with the two muons causes a shift of $m_{\mu\mu}$ towards lower values and the total momentum of the visible system is no more well aligned with the one of the reconstructed B meson candidate.

Another reason for the correlation of the output of the BDT classifier with the invariant mass lies in the values of the BDT tuning parameters. As it will be shown soon, if these are too optimized, the classifier can spot out the correlations between some of the input variables (even starting from variables which are not so strongly correlated with the invariant mass) thus giving an output correlated with $m_{\mu\mu}$.

Effects of BDT training parameters. In the first line of Tab.3.4 are reported the values of the BDT classifier's tuning parameters used for the previous round of the analysis (which uses as input variables the first nine ones introduced earlier).

The optimization of the tuning parameters, using the larger training sample available for this round of the analysis, gives higher optimal values for the nTrees and MaxDepth parameters, as it can be seen from the second line of Tab.3.4.

This translates in an improvement of the performances of the BDT, but, at the same time, a peaking correlation between the output of the classifier and the dimuon invariant mass appears, as can be seen from Fig.3.5.

To understand the reasons of such a correlation the following tests have been done:

- to test the effects of the training sample dimension, a BDT classifier with the tuning parameters used for the previous round has been trained on a larger training sample. The results are shown in Fig.3.6 and the values for the correlation of the BDT output with the invariant mass are reported in Tab.3.5. The use of a bigger training sample only improves the performances of the classifier and is not responsible for the peaking correlation.

- to test the effect of the increased number of trees a scan on nTrees has been performed, while keeping fixed all others parameters. The result of such scan is reported in Fig.3.7 and in the first three lines of Tab.3.6. As in the previous case, the increasing number of

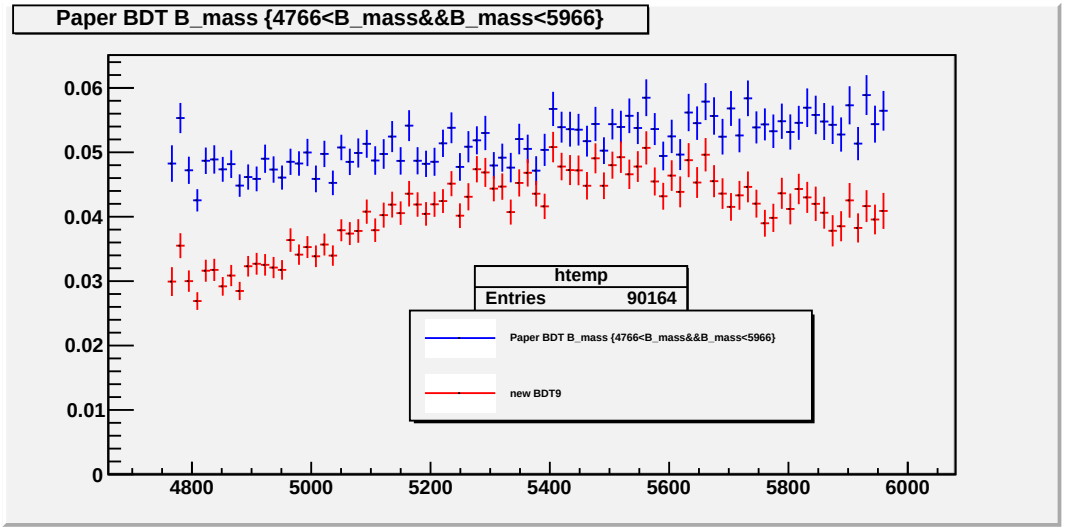


Figure 3.5: Average values of the output of different BDT classifiers as a function of the $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events: in blue of a BDT trained with the tuning parameters of the classifier used for the previous round of the analysis that have been optimized using a background sample of integrated luminosity equivalent to 0.5 fb^{-1}; in red of a classifier using tuning parameters optimized using a background sample of integrated luminosity equivalent to 50 fb^{-1}.

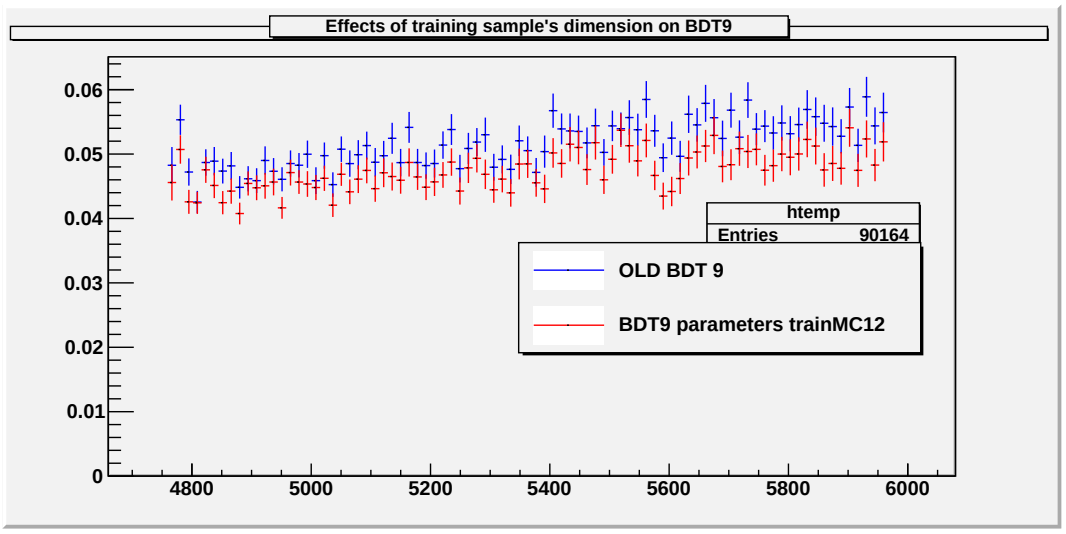


Figure 3.6: Average values of the output of two different BDT classifiers, using nine input variables, the tuning parameters in the first row of Tab.3.4, and trained on training samples of different sizes (blue $\sim 0.5 fb^{-1}$, red $\sim 50 fb^{-1}$), as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

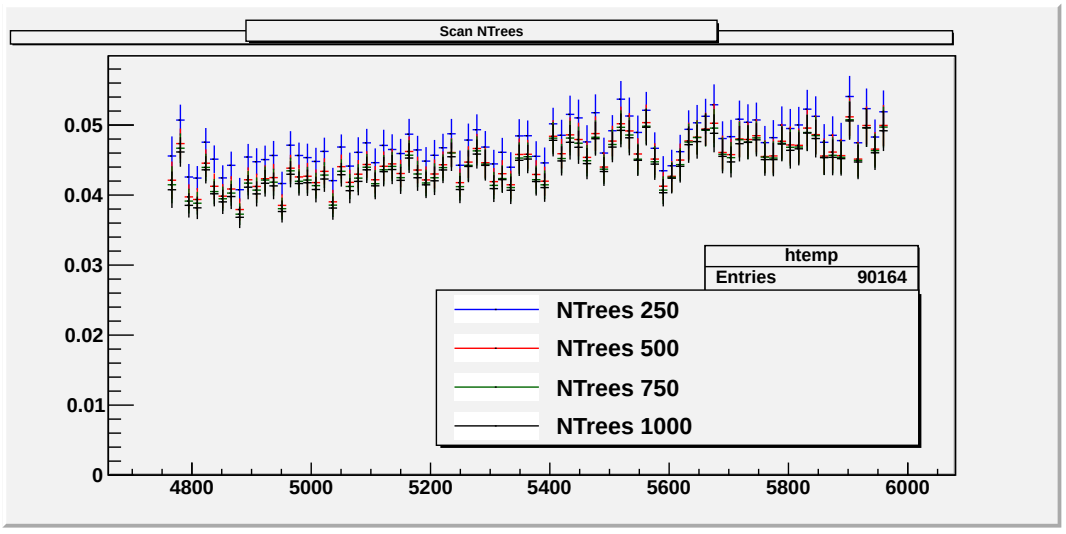


Figure 3.7: Average values of the output of different BDT classifiers using nine input variables and the tuning parameters in the first row of Tab.3.4 except for the "nTree" tuning parameter as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

nTrees	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
500	3.16	0.78
750	3.30	0.78
1000	3.43	0.82
MaxDepth	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
3	2.58	0.56
8	11.38	12.41

Table 3.6: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events for classifiers using nine input variables and different values of the tuning parameters.

trees in the forest is not what causes the peaking correlation, and in fact it only improves the performances.

- finally, the effect of the increased depth of each tree (MaxDepth) has been studied, by scanning over two possible values while keeping fixed all other parameters. The result of such scan is reported in Fig.3.8 and in the last two lines of Tab.3.6. In this case it's evident that the peaking correlation is due to the high value of the MaxDepth parameter.

The same behavior is seen when using the enlarged set of thirteen input variables, as it can be seen in Fig.3.9 and Tab.3.7

Effects of BDT input variables. Since this correlation already appears when using the reduced set of nine input variables (and is even more pronounced with thirteen input variables), it's among those nine variables that the one(s) causing this correlation must be found. The ones which can give rise to the aforementioned correlation are the transverse momentum of the B candidate, the minimum transverse momentum of the two muons, and the "polar angle". These variables are strongly correlated with $m_{\mu\mu}$ as it is shown in Figs.3.10-3.11.

To test the effects of these variables on the output of the BDT classifier, three classifiers using a set of eight input variables (obtained from the nine input variables set used for the previous publication by removing, one at time, the minimum transverse momentum of the muons, the

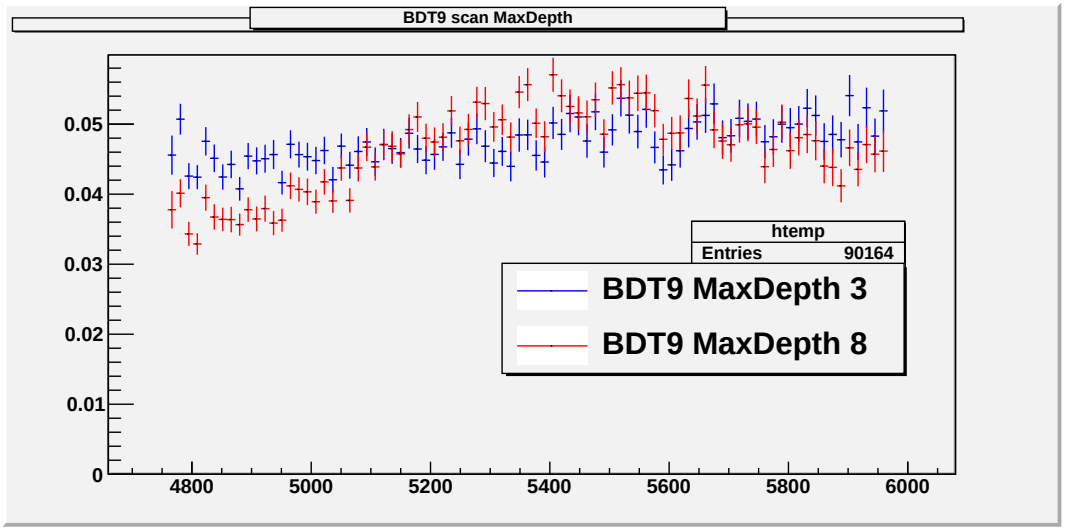


Figure 3.8: Average values of the output of different BDT classifiers using nine input variables and the tuning parameters in the first row of Tab.3.4 except for the "MaxDepth" tuning parameter as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

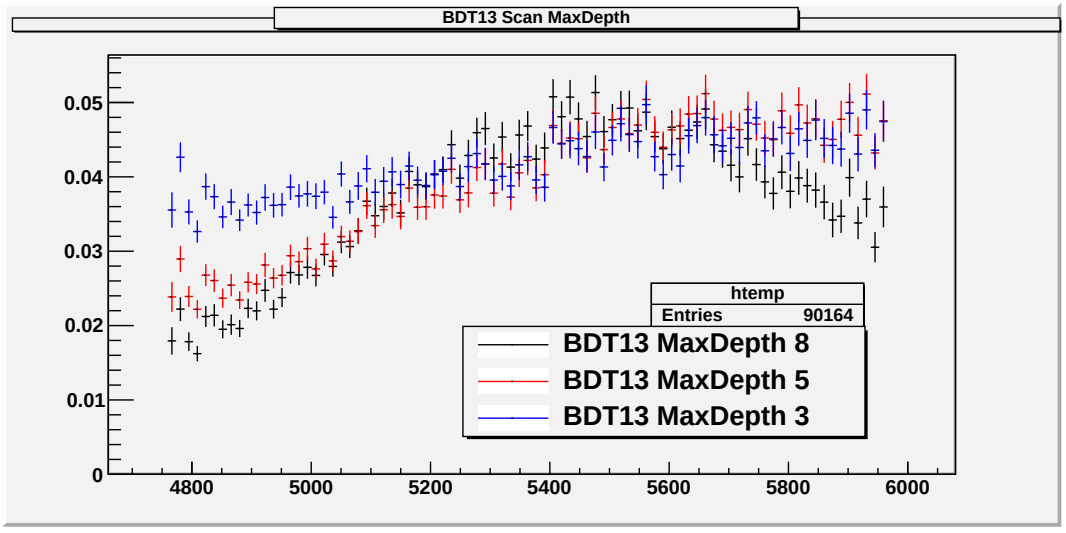


Figure 3.9: Average values of the output of different BDT classifiers using thirteen input variablesand the tuning parameters in the first row of Tab.3.4 except for the "MaxDepth" tuning parameter as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

MaxDepth	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
3	5.21	1.46
5	12.86	5.67
8	11.20	15.40

Table 3.7: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events for classifiers using thirteen input variables and differing only for the MaxDepth tuning parameter values.

Variable	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
$min_{p_T}\mu$	21.19	3.86
$p_T(B)$	7.16	1.55
polar angle	0.4	0.29
$\lvert\Delta\eta\rvert$	2.13	0.54
$\lvert\Delta\phi\rvert$	4.03	0.71

Table 3.8: Values (in %) of the correlations with $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events of some of the BDT input variables.

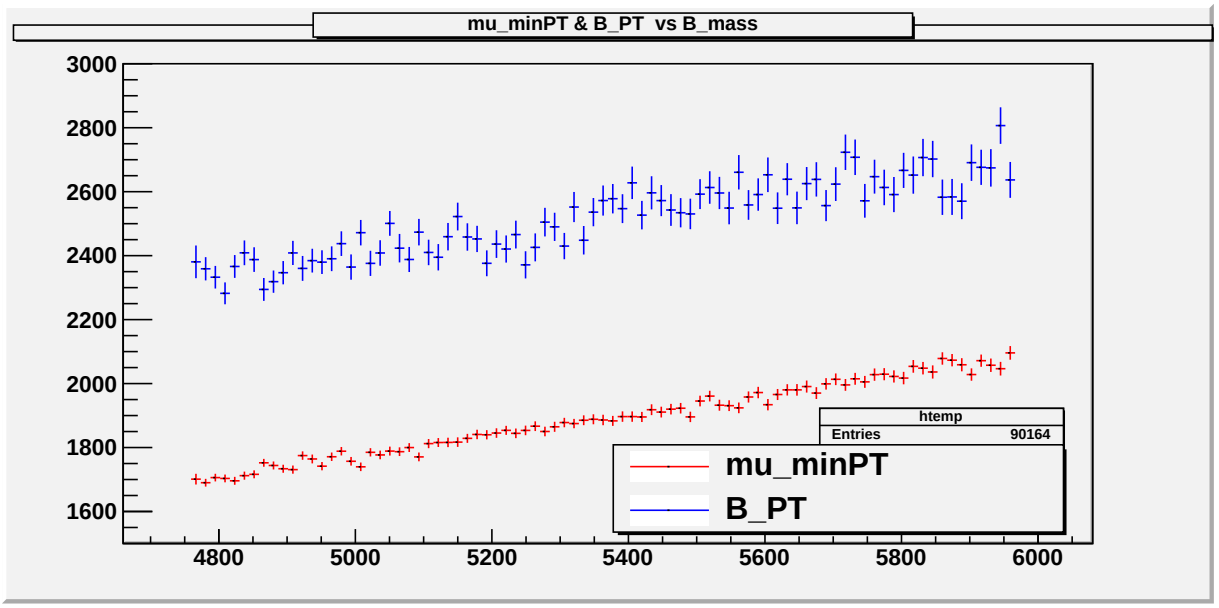


Figure 3.10: Average values of $min_{p_T}\mu$ (mu_minPT in the plot) and $p_T(B)$ (''B_PT" in the plot) as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

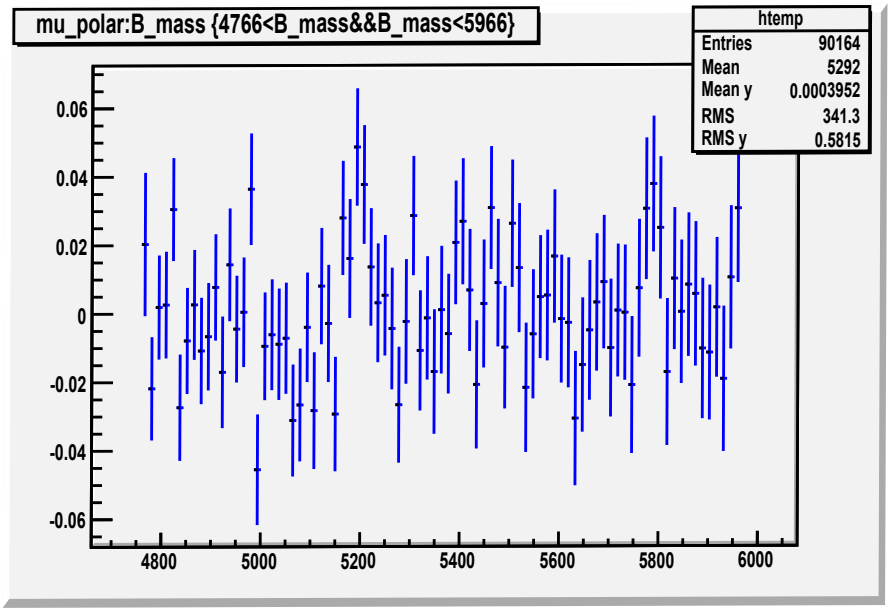


Figure 3.11: Average value of the ''polar_angle" variable as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

BDT	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0} \rvert$
BDT9	6.35	6.62
BDT8 (BDT9 without $min_{p_T}(\mu)$)	0.41	0.34
BDT8 (BDT9 without $p_T(B)$)	6.01	3.64
BDT8 (BDT9 without "polar angle")	6.02	1.17

Table 3.9: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events for classifiers using nine and eight input variables.

transverse momentum of the B candidate and the polar angle) and the optimized BDT tuning parameters listed in the second row of Tab.3.4 have been tested. The result of the test is shown in Fig.3.12 and the values of the correlations with the invariant mass are reported in Tab.3.9.

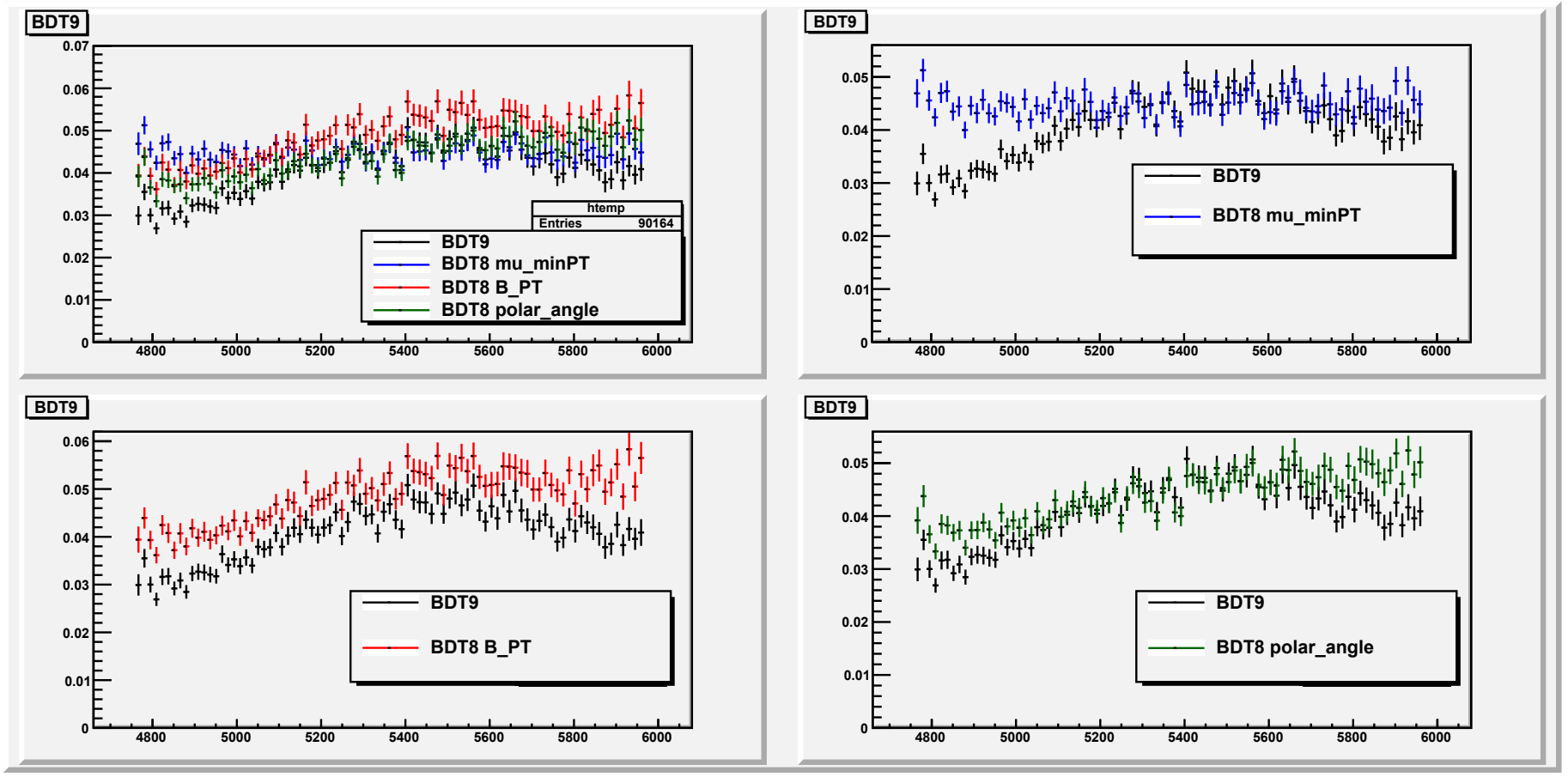


Figure 3.12: Average values of the output of BDT as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events for different classifiers using nine (in black) and eight input variables. These are obtained by the set of nine by removing one at time the minimum transverse momentum of the two μ (blue), the transverse momentum of the B candidate (red), and the polar angle (green). The peaking correlation obtained with nine input variables disappears if the minimum transverse momentum of the two μ is removed, while it is still present as far this variable is used as a input of the BDT.

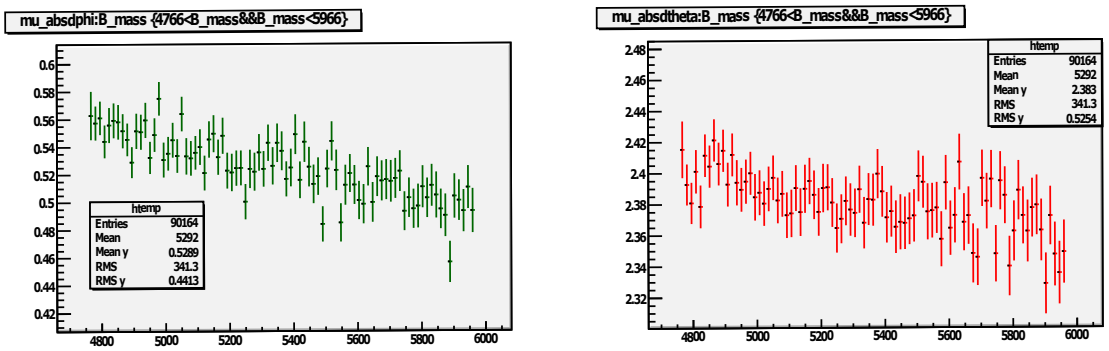


Figure 3.13: Average values of the angular variables $|\Delta\phi|$ (green) and $|\Delta\eta|$ (red) (on the y-axis) as a function of $m_{\mu\mu}$ (on the x-axis) for $b\bar{b} \to \mu\mu X$ MC generated events.

As far as the set of input variable is concerned, the peaking correlation is due to the minimum transverse momentum of the two muons, since, when it is removed, the peak disappears. The same happens when the enlarged set of input variables is used. In this case also the two new variables $|\Delta\eta|$ and $|\Delta\phi|$ as well are correlated with $m_{\mu\mu}$ as Fig.3.13 shows. The effects of the removal of such variables, as well as of the $min_{pT}(\mu)$, from the set of thirteen input variables are shown in Fig.3.14 and Tab.3.10.

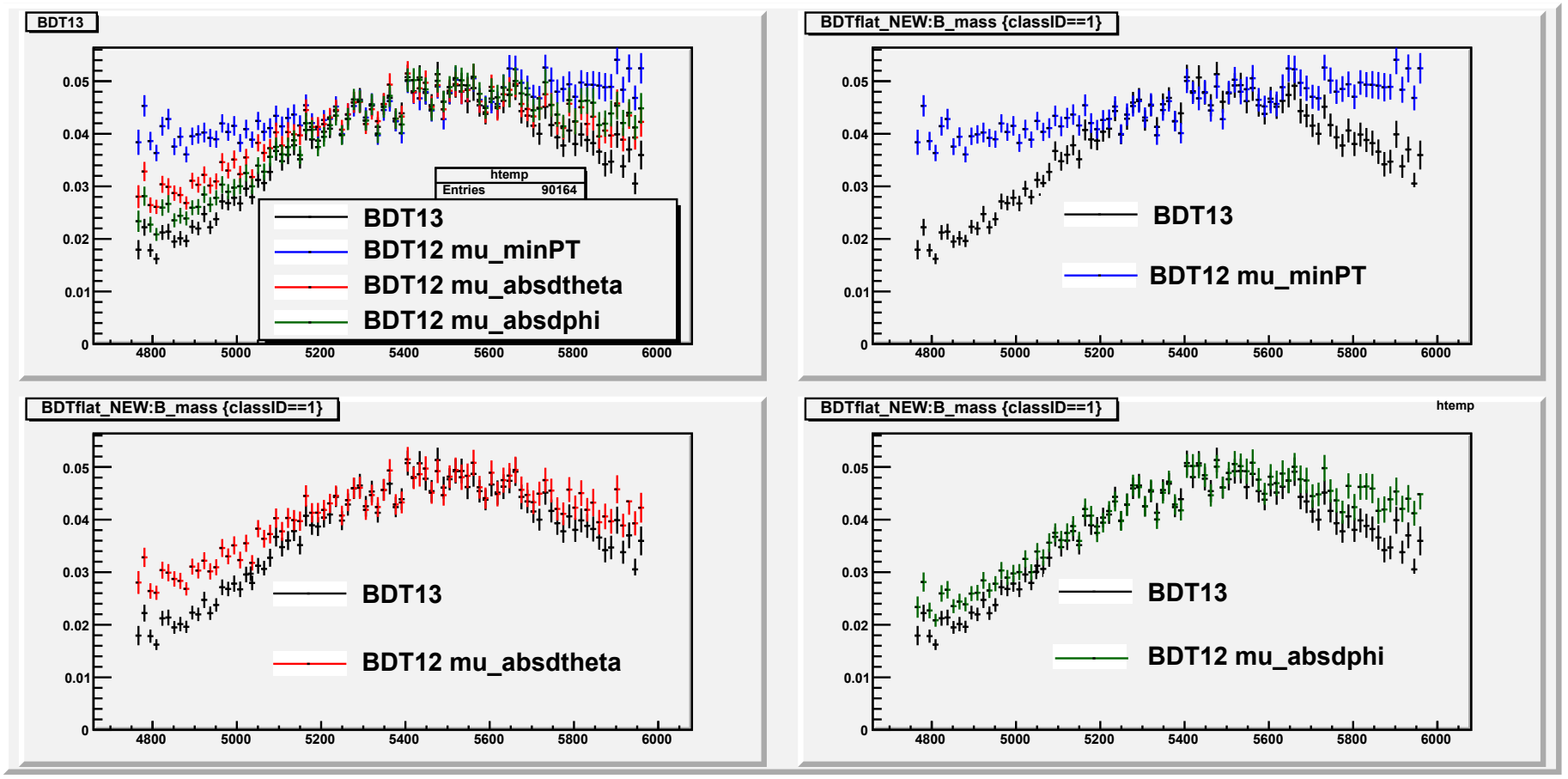


Figure 3.14: Average values of the output of BDT as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events for different classifiers using thirteen (in black) and twelve input variables. These are obtained by the set of thirteen variables by removing one at time the minimum transverse momentum of the two μ (blue), the $|\Delta\eta|$ (red), and the $|\Delta\phi|$ (green). Also with the enlarged set of thirteen input variable, the peaking correlation disappears if the minimum transverse momentum of the two μ is removed, while it is still present if this variable is used as a input of the BDT.

BDT	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s}\rvert$
BDT13	11.4	12.9
BDT12 (BDT13 without $min_{p_T}(\mu)$)	5.43	0.98
BDT12 (BDT13 without $\lvert\Delta\theta\rvert$)	8.15	7.44
BDT12 (BDT13 without $\lvert\Delta\phi\rvert$)	11.45	8.52

Table 3.10: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for classifiers using thirteen and twelve input variables for $b\bar{b} \to \mu\mu X$ MC generated events.

	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
BDT9	6.35	6.62
MLP9	5.49	0.58
BDT13	11.4	12.9
MLP13	5.38	0.53

Table 3.11: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for classifiers using nine and thirteen input variables for $b\bar{b} \to \mu\mu X$ MC generated events.

It's important to stress that even though the peaking correlation disappears, when removing the minimum transverse momentum of the two muons from the input variables, the output of the classifier is still slightly linearly correlated with the invariant mass $m_{\mu\mu}$.

A final remark concerns the fact that the peaking correlation between the invariant mass $m_{\mu\mu}$ and the BDT classifier's output appearing when the $min_{p_T}(\mu)$ is used as input variable is peculiar of the BDT method. Indeed this correlation disappears when, instead of a BDT operator, a Neural Network classifier is used. In this case, using nine or thirteen input variables, the output of the MVA classifier is only linearly correlated with $m_{\mu\mu}$. The results are shown in Figs.3.15 and 3.16 and the values of the correlation with the invariant mass are reported in Tab.3.11. The Neural Network classifier does not feature the peaking correlation with $m_{\mu\mu}$, even if the $min_{p_T}(\mu)$ variable is used in the training. Nevertheless, as it is shown in Figs.3.17-3.18 the signal background discrimination is not as good as for the BDT algorithm.

Conclusions. To summarize, the previous studies show that:

- the peaking correlation between the BDT output and the invariant mass of the di-muon system is due both to variables correlated with the invariant mass $m_{\mu\mu}$ (especially $min_{p_T}(\mu)$) and to the BDT parameter MaxDepth. In particular, it has been shown that the presence of the $min_{p_T}(\mu)$ variable together with a high value (equal or greater than 5) of the MaxDepth parameter make this correlation appearing. An important remark concerns the fact that, even though the $min_{p_T}(\mu)$ was used in the BDT with nine input variables used for the previous analysis, such classifier didn't show such a correlation (as it is shown in Fig.3.5). The reasons for that are mostly due to to the low value (equal to 3, see 3.4) of the MaxDepth parameter, and to the smaller available statistics used to optimize the tuning parameter of that classifier. A BDT with thirteen input variables, with the inclusion of the $min_{p_T}(\mu)$, and a low value of the MaxDepth parameter, turns out to have the same performances of the classifier using twelve input variables and a higher value of the MaxDepth parameter, but features a larger linear correlation of its output with the invariant mass $m_{\mu\mu}$;

- the linear correlation is "intrinsically" due to the other variables and it's not an "artifact" of the classifier or its optimization process, being still present when using a Neural Network

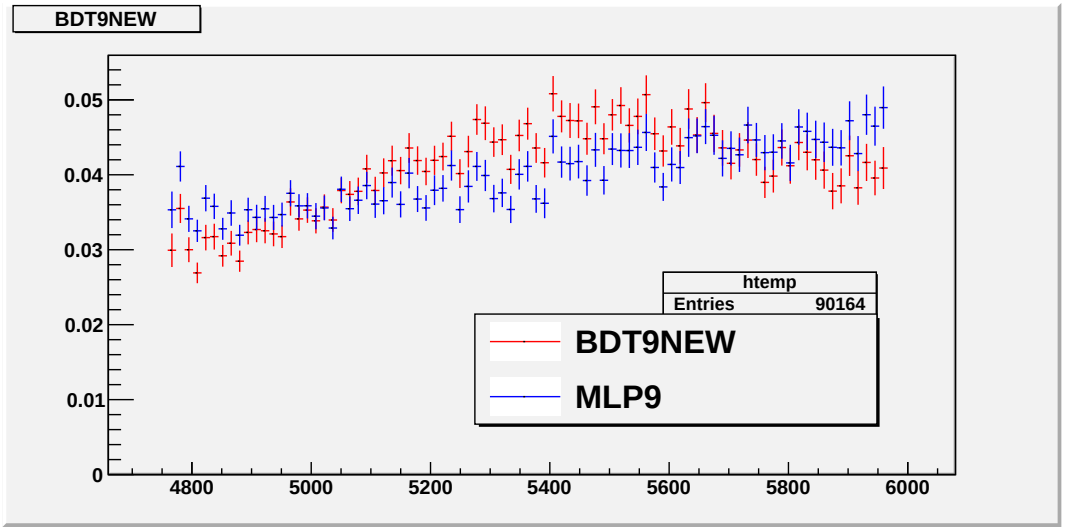


Figure 3.15: Average values of the output of a BDT and a Neural Network classifier (MLP) using the same nine input variables as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

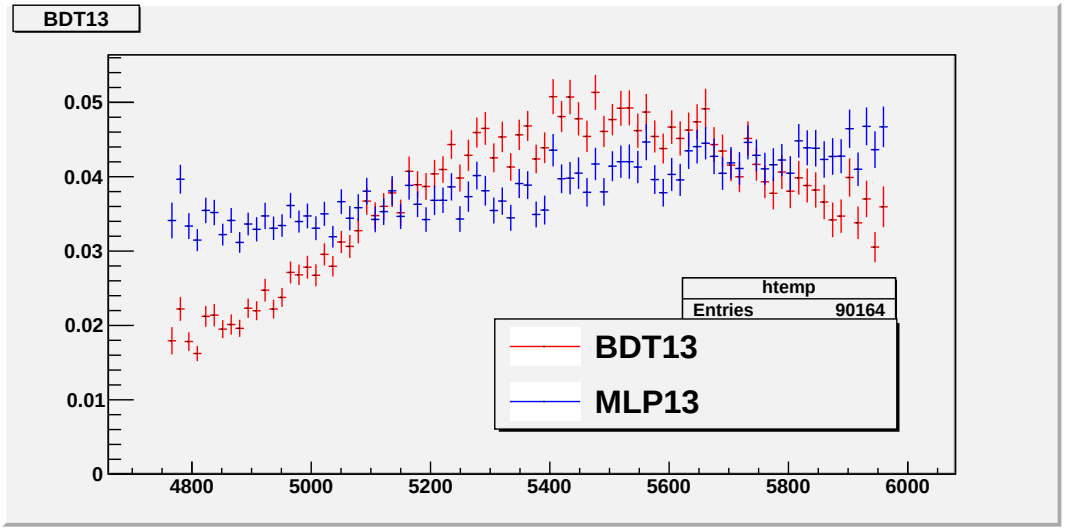


Figure 3.16: Average values of the output of a BDT and a Neural Network classifier (MLP) using the same thirteen input variables as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

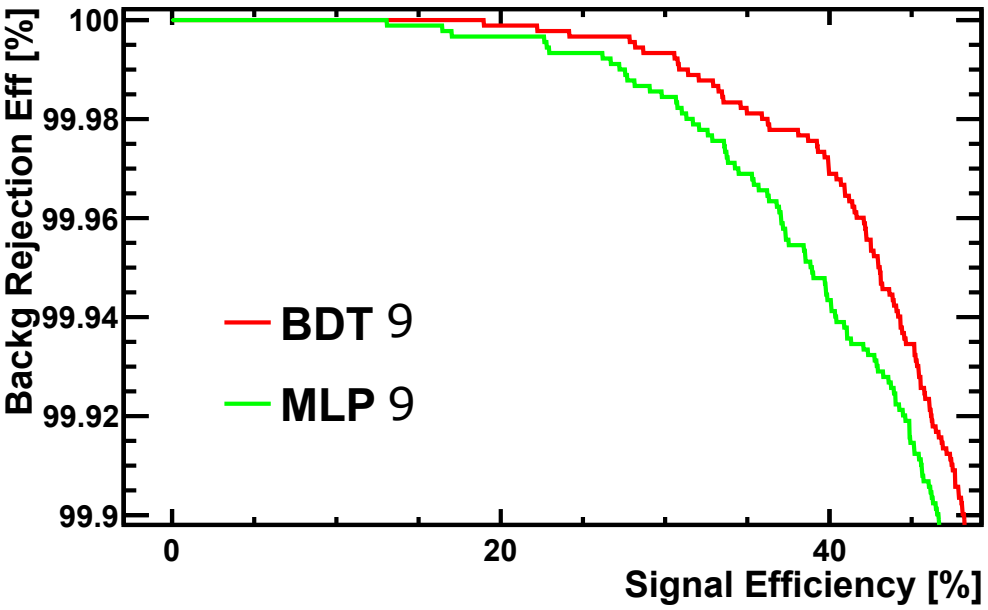


Figure 3.17: ROC curves of a BDT and a Neural Network (MLP) classifiers trained with the set of nine input variables used for the previous round of the analysis.

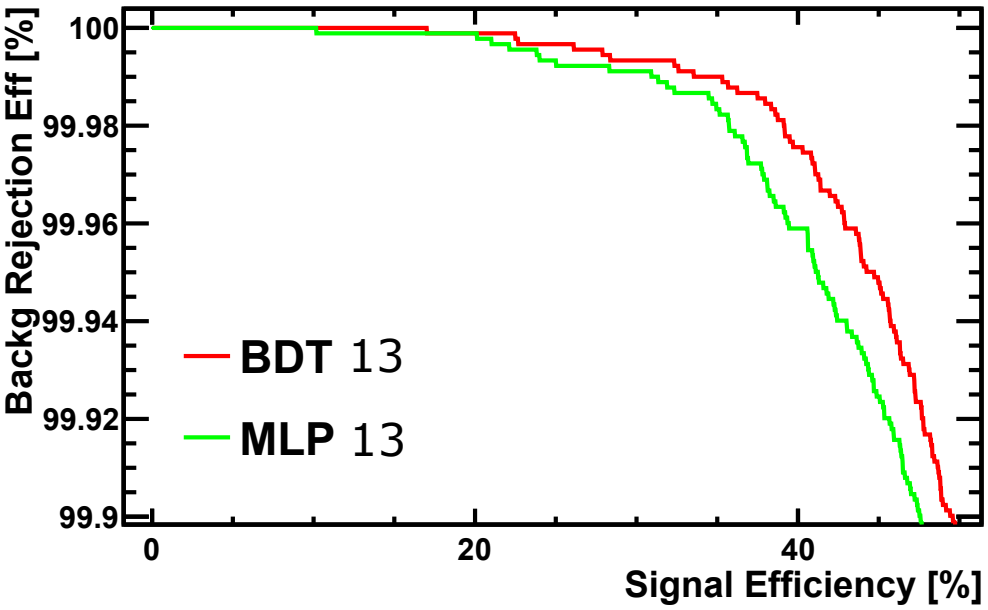


Figure 3.18: ROC curves of a BDT and a Neural Network (MLP) classifiers trained with the enlarged set of thirteen input variables.

nTrees	EvtsMin	MaxDepth	nCuts	β
1000	4500	6	30	0.5

Table 3.12: BDT parameters used to train the MVA classifier with twelve input variables.

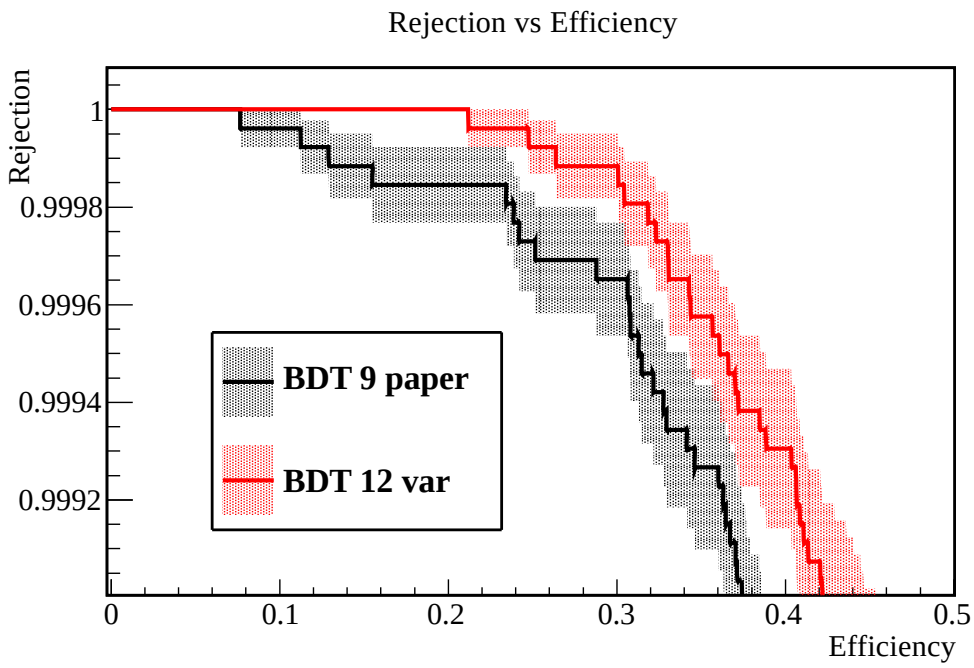


Figure 3.19: ROC curves of the BDT operators used for the previous $B^0_{(s)} \to \mu\mu$ analysis and the new optimized operator with twelve input variables.

classifier.

3.2.4 The final BDT classifier

The final BDT classifier is then built up of twelve input variables:

- eight of the nine input variables used for the previous round of the analysis ($min_{p_T}(\mu)$ being removed due to peaking correlation with the invariant mass),
- four additional "angular" variables introduced in the previous section ($|\Delta\eta|$, $|\Delta\phi|$, ''other B angle", ''other B boost angle").

The BDT tuning parameters have been optimized according to the performances found on generic $b\bar{b} \to \mu\mu X$ Monte Carlo sample for the background and $B^0_s \to \mu\mu$ Monte Carlo sample for the signal. A scan over the values of the tuning parameters listed in the previous section has been done, and the set maximizing the value of the above mentioned FOM has been adopted to train the final classifier. The best values of the tuning parameters are reported in Tab.3.12

The value of the FOM for this classifier is 32.6, to be compared with that of the BDT classifier used for the previous publication which is equal to 18.6. The improvement in the performances of the new classifier is also shown in the ROC curves of the two operators shown in Fig.3.19.

The values of the linear and peaking correlations of the classifier's output with respect to $m_{\mu\mu}$ are reported in Tab.3.13. The correlation of $B^0_s \to \mu\mu$ MC signal in the signal mass windows (values in parenthesis) is negligible, while the correlation at low $m_{\mu\mu}$ is due to radiative loss of energy from the muons in the final state. The correlation for background events is negligible, and the BDT value as a function of the invariant mass of the two muons is shown in Fig.3.20.

Sample	$m_{\mu\mu}$	$\lvert m_{\mu\mu} - M_{B_s^0}\rvert$
$B_s^0 \to \mu\mu$	5.3% (0.32%)	5.0% (0.26%)
$b\bar{b} \to \mu\mu X$	1.2%	0.02%

Table 3.13: Values (in %) of the BDT output correlations with $m_{\mu\mu}$ for the final BDT classifier with twelve input variables. The correlation of $B_s^0 \to \mu\mu$ MC signal in the signal mass windows (values in parenthesis) is negligible.

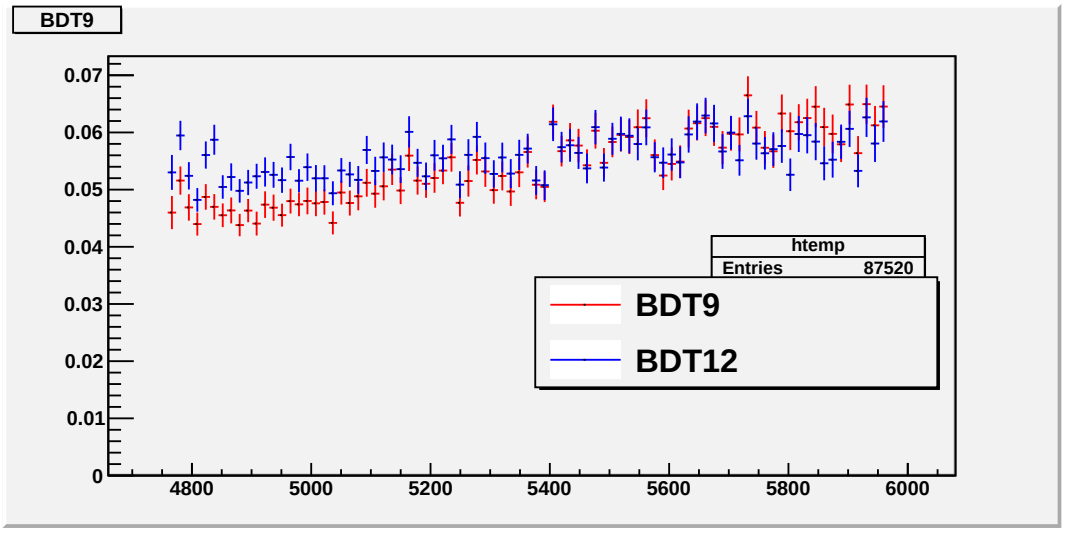


Figure 3.20: Average values of the output of the BDT classifiers trained with the new set of twelve input variables (blue) and the set of the previous analysis (red) as a function of $m_{\mu\mu}$ for $b\bar{b} \to \mu\mu X$ MC generated events.

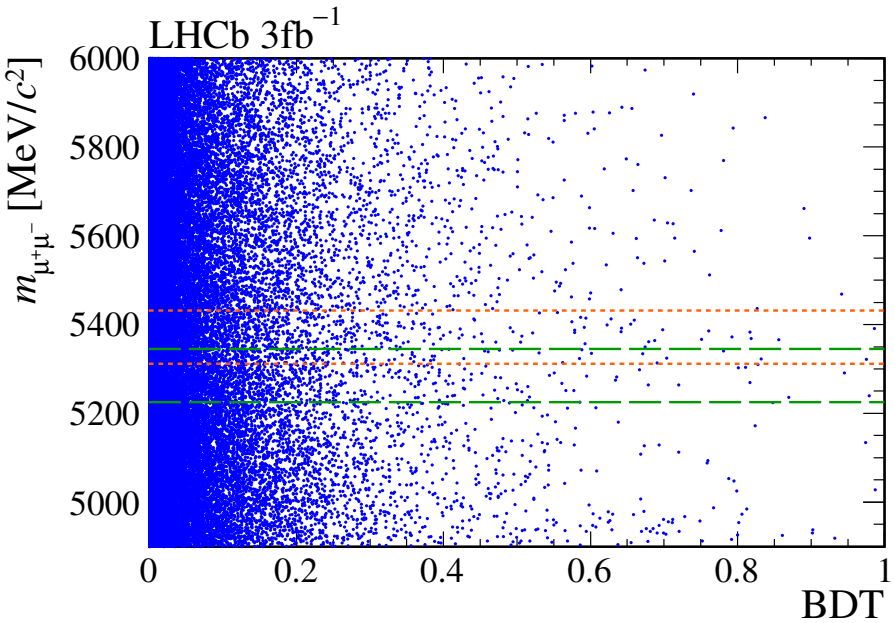


Figure 3.21: 2 dimensional plane in the BDT output (x-axis) and the invariant mass $m_{\mu\mu}$ (y-axis). The orange and green lines indicate the B_s^0 and the B^0 signal windows respectively.

3.3 Signal yield and limit extraction

Once the BDT classifier has been defined, the selected events are classified in a 2 dimensional plane according to the output of the BDT classifier and the invariant mass $m_{\mu\mu}$, as shown in Fig.3.21. Only candidates with an invariant mass higher than 4900 MeV/c^2 have been analyzed. This lower boundary is chosen to exclude all backgrounds coming from the b cascade decays $b \to c(\to \mu X)\mu X'$.

The number of observed signal $B^0 \to \mu\mu$ and $B_s^0 \to \mu\mu$ events in the dataset is extracted through a simultaneous unbinned likelihood fit in the invariant mass projection in eight BDT bins. In particular, for each BDT bin, the observed pattern of events is fitted with a function

$$f_{tot}(m_{\mu\mu}) = N_{bkg}F_{bkg}(m_{\mu\mu}) + N_{B^0}F_{B^0}(m_{\mu\mu}) + N_{B_s^0}F_{B_s^0}(m_{\mu\mu}) \ , \tag{3.13}$$

where N_{bkg,B^0,B_s^0} and $F_{bkg,B^0,B_s^0}(m_{\mu\mu})$ are the yields and the invariant mass PDFs of the background, the $B^0 \to \mu\mu$, and $B_s^0 \to \mu\mu$ events respectively. The functions $F_{bkg,B^0,B_s^0}(m_{\mu\mu})$ are normalized to one, *i.e.*

$$\int F_{bkg,B^0,B_s^0}(m_{\mu\mu})dm_{\mu\mu} = 1 \ . \tag{3.14}$$

To perform the fit, the knowledge of the functions $F_{bkg,B^0,B_s^0}(m_{\mu\mu})$, as well of the event fraction in each BDT bin, is required. In the following the definition and calibration of the BDT and invariant mass PDFs will be shown, both for signal and the different sources of background.

The BDT binning has been optimized [89] in order to reach the highest sensitivity for the SM signal. The eight BDT bins have the following boundaries:

$$0.0, 0.25, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$$

and are characterized by different signal purities, and the last bins will be the most sensitive ones.

In each BDT bin, the invariant mass range $[4900, 6000]$ MeV/c^2 can be divided into four regions

	B^0	B_s^0
μ	$5284.90 \pm 0.10_{stat} \pm 0.20_{syst}$ MeV/c^2	$5371.85 \pm 0.17_{stat} \pm 0.19_{syst} MeV/c^2$
σ	$22.83 \pm 0.07_{stat} \pm 0.42_{syst}$ MeV/c^2	$23.24 \pm 0.08_{stat} \pm 0.44_{syst} MeV/c^2$
α	$2.065 \pm 0.005_{stat} \pm 0.010_{syst}$	
n	$1.118 \pm 0.013_{stat} \pm 0.038_{syst}$	

Table 3.14: Crystall Ball parameters for the B^0 and the B_s^0 modes.

- the **left sideband** defined by the range $[4900\ \text{MeV}/c^2, m_{B^0} - 60\text{MeV}/c^2]$,
- the $\boldsymbol{B^0}$ **signal mass window**, defined as the region in the range $[m_{B^0} - \text{MeV}/c^2, m_{B^0} + \text{MeV}/c^2]$,
- the $\boldsymbol{B_s^0}$ **signal mass window**, defined as the region in the range $[m_{B_s^0} - 60\ \text{MeV}/c^2, m_{B_s^0} + \text{MeV}/c^2]$,
- the **right sideband** defined in the range $[m_{B_s^0} + 60\ \text{MeV}/c^2, 6000\ \text{MeV}/c^2]$.

3.3.1 Signal PDF calibration

The signals mass PDF is assumed to be a Crystal Ball function [90] ($\mathcal{CB}$ in the following), *i.e.* a Gaussian function with an exponential left tail to take into account radiative loss of energy in the final state:

$$\mathcal{CB}(m) = \begin{cases} \left[1 + \alpha\frac{(m-\mu)}{n\sigma} - \frac{\alpha^2}{n}\right]^{-n} \cdot e^{\left[-\frac{1}{2}\alpha^2\right]} & for \quad \frac{m-\mu}{\sigma} < -\alpha \\ e^{-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2} & for \quad \frac{m-\mu}{\sigma} \geq -\alpha \end{cases} . \tag{3.15}$$

The Crystal Ball function is parametrized by 4 quantities:

- the central value μ of the gaussian component,
- the mass resolution σ of the gaussian component,
- transition point α between the gaussian and the exponential component in units of the resolution σ,
- exponent n of the exponential component.

The central values μ_{B^0} and $\mu_{B_s^0}$ as well as the mass resolutions σ_{B^0} and $\sigma_{B_s^0}$ for the B^0 and the B_s^0 modes respectively are obtained through a fit in the invariant mass distribution of the di-hadronic decays $B_{(s)}^0 \to K^+\pi^-$, $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$. The values for the resolutions have also been cross-checked interpolating the invariant mass resolution of charmonium and bottomium resonances decaying in two muons, as shown in Fig.3.22.

The parameters α and n are obtained through simulated $B_{(s)}^0 \to \mu\mu$ events and are assumed to be the same for the B^0 and the B_s^0 mode. The final results for the two Crystal Ball functions' parameters are summarized in Tab.3.14.

Comparing the μ and σ values in Tab.3.14, it's possible to see that the two B^0 and B_s^0 masses are well resolved.

The BDT PDF of the signal is calibrated on data using $B_{(s)}^0 \to h^+h^-$ candidates in data. Indeed, as the BDT is build up of geometrical variables, the di-hadronic decays of $B_{(s)}^0$ well reproduce the $B_{(s)}^0 \to \mu\mu$ decay topological and kinematical features. To further reduce differences between the di-hadronic decays and the $B_{(s)}^0 \to \mu\mu$ events due to the reconstruction and selection process, only $B \to hh^{(\prime)}$ events for which the two hadrons in the final state are also

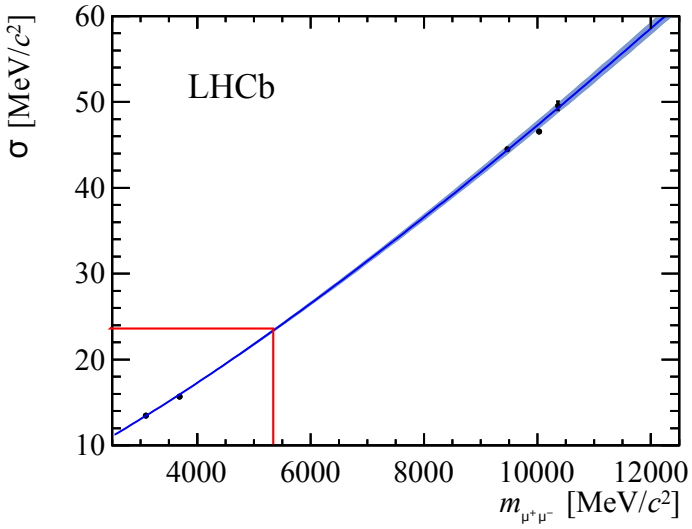


Figure 3.22: Invariant mass resolution as a function of the charmonium or bottomium resonance mass. The red lines indicates the di-muons mass resolution at the value of the B_s^0 mass.

inside the Muon detector acceptance are used, and the BDT shape is corrected to take into account differences due to μ-identification and trigger requirements between the two channels; to further reduce discrepancies between the signal and the calibration sample, an additional correction to the $B \to hh^{(\prime)}$ BDT shape is added in order to take into account differences in the lifetime acceptance of the BDT bins [91].

The number of $B^0_{(s)} \to h^+h^-$ events in each BDT bin is extracted from data through a fit in the di-hadron invariant mass m_{hh}. To compute the invariant mass m_{hh} a mass hypothesis on the final state particles (kaons or pions) must be done. This has been done according to the values of the PID variables $\Delta LL_{K-\pi}$. The invariant mass $B^0_{(s)} \to h^+h^-$ signal shape is described by a Crystal Ball function. The combinatorial background is parametrized through an exponential, while background from partially reconstructed B decays is described by the *RooPhysBkg* function defined as follows [92]:

$$\begin{aligned} f(m, m_0, c_p, \sigma_p) =& N \cdot \int_{-\infty}^{+\infty} m' \left(1 - \frac{m'^2}{m_0^2}\right) \Theta(m_0 - m') e^{-c_p m'} \\ & \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{m-m'}{\sigma_p}\right)^2} dm' \end{aligned} \tag{3.16}$$

where N is a normalization factor and Θ is the step function.

3.3.2 Backgrounds PDF calibration

The most important source of background events in the invariant mass range is due to combinatorial reconstructed candidates. The invariant mass distribution of these candidates spans all over the invariant mass range.

In addition to this source of background, other physical modes pollute the left invariant mass sideband as well as the B^0 signal windows. These sources of background can be divided into the following three categories:

- $B^0_{(s)} \to hh^{(\prime)}$ with the two hadrons misidentified as muons. The main process responsible for such a misidentification is the decay in flight of the two hadrons which take place

outside the VELO. This background is not located in the B_s^0 mass windows, because of the shift in $m_{\mu\mu}$ due to the missing energy carried out by the two neutrinos. Nevertheless, and just because of that shift, this source of background pollutes the B^0 mass window, representing the main obstacle for an observation of the $B^0 \to \mu^+\mu^-$ mode.

- semileptonic $B^0_{(s)}$ decays with one misidentified hadron:
  - $B^0 \to \pi^-\mu^+\nu_\mu$. The $\mathcal{BR}$ of this mode is $(1.44 \pm 0.05) \cdot 10^{-4}$ [93]. The presence of the unreconstructed neutrino shifts both the invariant mass as well as the BDT output towards lower values, because of the correlation of the pointing variable with the invariant mass.
  - $B_s^0 \to K^-\mu^+\nu_\mu$. The predicted $\mathcal{BR}$ of this mode is $(1.27 \pm 0.49) \cdot 10^{-4}$ [95]. The expected contribution of this mode is smaller than the previous one, because of the larger mass shift and a lower hadronization probability f_s
  - $\Lambda_b \to p\mu^-\bar{\nu}_\mu$. The $\mathcal{BR}$ of this mode was unknown at the time of the analysis presented here and a theoretical estimation by [96] has been used:

$$\mathcal{BR}(\Lambda_b^0 \to p\mu^-\bar{\nu}) = 3.3^{+1.5}_{-1.2} \cdot 10^{-4} \times \left(\frac{V_{ub}}{3.5 \cdot 10^{-3}}\right)^2 . \tag{3.17}$$

    By inserting the latest average of V_{ub} the following value is found to be:

$$\mathcal{BR}(\Lambda_b^0 \to p\mu^-\bar{\nu}) = (4.75 \pm 2.11) \cdot 10^{-4} . \tag{3.18}$$

- semileptonic $B^0_{(s)}$ decays with two muons coming from the same vertex
  - $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$. These decays are generated by FCNC. While the $B^+ \to \pi^+\mu^+\mu^-$ decay has been observed with a $\mathcal{BR}(B^+ \to \pi^+\mu^+\mu^-) = (2.3 \pm 0.6_{stat} \pm 0.1_{syst}) \cdot 10^{-8}$ [97], the $B^0 \to \pi^0\mu^+\mu^-$ has not yet been observed. It's expected to be of the same order of the charged mode, and an extimation for its value is obtained using theoretically predicted values [95] for $\mathcal{BR}(B^+ \to \pi^+\mu^+\mu^-)|_{theo} = (1.95^{+0.61}_{-0.48}) \cdot 10^{-8}$ and $\mathcal{BR}(B^0 \to \pi^0\mu^+\mu^-)|_{theo} = (0.91^{+0.33}_{-0.28}) \cdot 10^{-8}$ as follows:

$$\mathcal{BR}(B^0 \to \pi^0\mu^+\mu^-) = \frac{\mathcal{BR}(B^0 \to \pi^0\mu^+\mu^-)|_{theo}}{\mathcal{BR}(B^+ \to \pi^+\mu^+\mu^-)|_{theo}} \cdot \mathcal{BR}(B^+ \to \pi^+\mu^+\mu^-) \tag{3.19}$$

  - partially reconstructed $B_c^+ \to J/\psi(\to \mu\mu)\mu^+\nu_\mu$. Due to the larger mass of the B_c meson this decay can in principle pollute the B_s^0 mass window. The hadronization fraction of a b quark into a B_c meson is, even with a large uncertainty, two orders of magnitude smaller than the one into a B meson. In addition, thanks both to the unreconstructed neutrino and the extra charged muon track, the pointing and isolation variables of such decay will be background like.

The invariant mass and BDT PDF of these exclusives modes are obtained using simulated data. The expected yields are estimated by normalizing to the $B^0 \to J/\psi K^+$ decay. Even though a lot of these backgrounds do not pollute the signal mass windows, being shifted in the left invariant mass sideband, their contribution in this region can still spoil the evaluation of the combinatorial background when extrapolating its shape interpolating the left and right invariant mass sidebands.

The yields from the $B^0_{(s)} \to hh^{(\prime)}$ backgrounds doubly misidentified hadrons are obtained directly by the explicit selection.The invariant mass and BDT shape are obtained with data driven methods.

The first measurement of the $\mathcal{BR}$ of this mode has been obtained by the LHCb Collaboration [94] in 2015.

The invariant mass shape of the combinatorial background is obtained by interpolating in each BDT bin the left and right invariant mass sidebands with a simple exponential function. In order to deal with the above mentioned sources of physical backgrounds in the extrapolation of the combinatorial one, the following choices have been made:

- the $B^0_{(s)} \to hh^{(\prime)}$, $B^0 \to \pi^-\mu^+\nu_\mu$ and $B^{0(+)} \to \pi^{0(+)}\mu^+\mu^-$ modes have been added as separate PDFs in the fit. Indeed, they represent the most important source of physical background ($\sim 81\%$) in the invariant mass range and high BDT region;
- the $B^0_s \to K^-\mu^+\nu_\mu$ mode mass shape is very similar to the one of the $B^0 \to \pi^-\mu^+\nu_\mu$ mode, thus its contribution has just been added to that of this mode;
- the $\Lambda_b \to p\mu^-\bar{\nu}_\mu$ mode has just been taken into account as a systematic error in the fit;
- the $B^+_c \to J/\psi(\to \mu\mu)\mu^+\nu_\mu$ mode has a negligible yield with respect to the other channels and its mass shape is quite well approximated by the exponential shape of the combinatorial background.

The result of the fit to the invariant mass sidebands is shown in Fig.3.23.

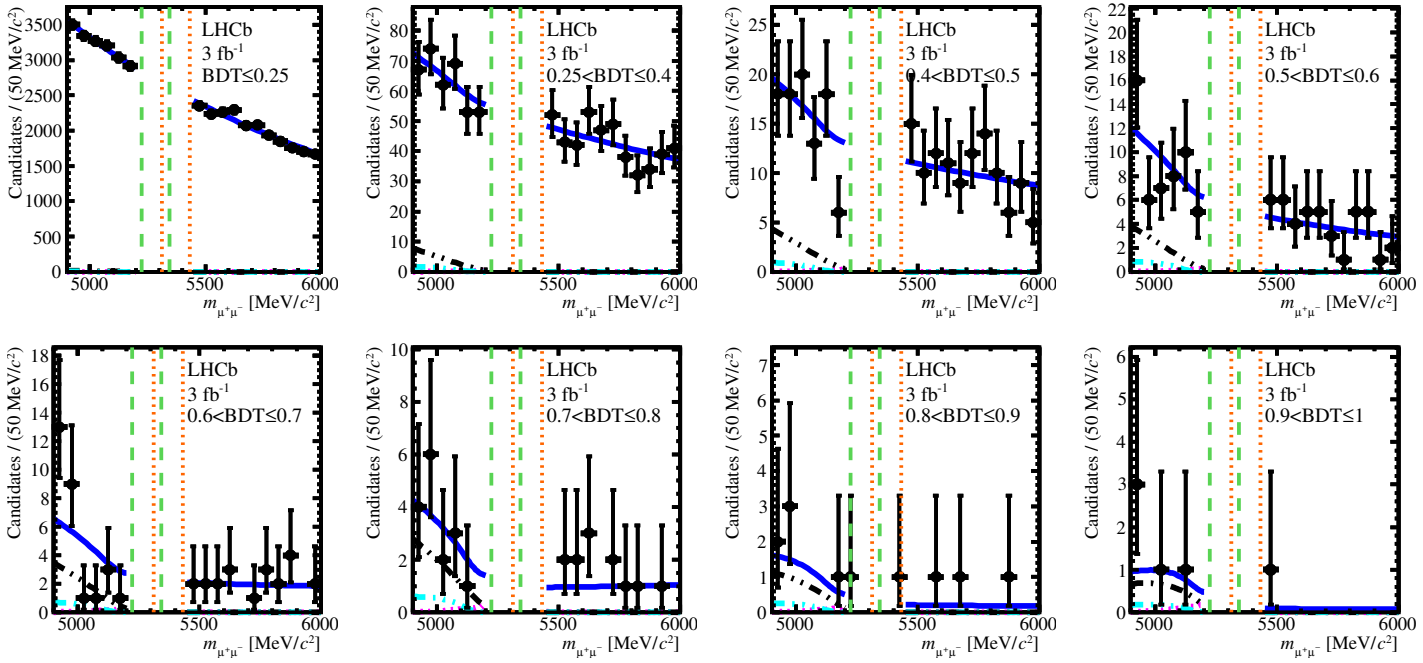


Figure 3.23: Invariant mass shape fit of combinatorial background in the eight BDT bins. The contribution from the peaking $B^0_{(s)} \to hh^{(\prime)}$ is the dashed pink line, the one from $B^0_{(s)} \to \pi^+(K^-)\mu^+\nu_\mu$ is the dashed black line, the one from $B^{0(+)} \to \pi^{0,(+)}\mu^+\mu^-$ is the dashed cyan line, and the fit result is indicated by the continuous blue line.

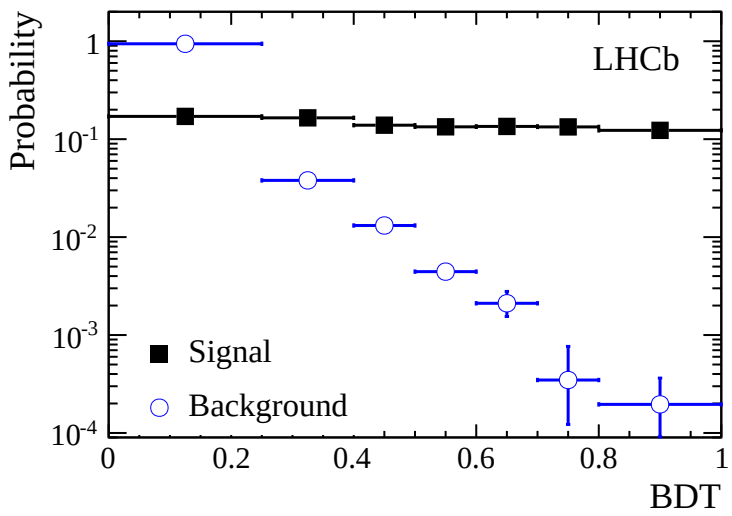


Figure 3.24: BDT PDF shapes for combinatorial background (blue circles) and signal (taken from the $B \to hh'$ control sample) (black squares).

The BDT shape is taken from events falling in the invariant mass sidebands and its distribution, together with the one for signal taken from the $B \to hh'$ control sample, is shown in Fig.3.24.

3.3.3 Normalization

To translate the observed number of signal events into a value for the $\mathcal{BR}$ the total number of B^0 and B_s^0 mesons produced must be known. As already mentioned in Sec.3.1 these numbers are obtained through normalization channels of well known $\mathcal{BR}$, for which the same relation (3.3) holds. In particular, the normalization coefficient (otherwise called *single event sensitivity*) $\alpha_{(s)}$ is given by the following relation:

$$\alpha_{(s)} \equiv \frac{\mathcal{BR}_{norm}}{N_{norm}} \cdot \frac{f}{f_{(s)}} \cdot \frac{\epsilon_{norm}^{GEN}\epsilon_{norm}^{REC\&SEL|GEN}\epsilon_{norm}^{TRIG|SEL}}{\epsilon_{sig}^{GEN}\epsilon_{sig}^{REC\&SEL|GEN}\epsilon_{sig}^{TRIG|SEL}} \,. \tag{3.20}$$

In the previous equation the efficiencies for signal and control channels have been factorized as the products of three partial efficiencies related to different steps of the selection process:

- ϵ^{GEN} is the efficiency due to the detector acceptance (defined by a polar angle θ in the range $[10, 400]$ mrad). This efficiency is computed on samples of MC generated events and its values are reported for different modes in Tab.3.15
- $\epsilon^{REC\&SEL|GEN}$ is the efficiency to reconstruct and select events falling inside the detector acceptance. This quantity depends on the detector acceptance, on the phase space of the particles in the final state and on the efficiencies of the algorithm used to reconstruct the tracks, and of the selection procedure. The ratio $\frac{\epsilon_{norm}^{REC\&SEL|GEN}}{\epsilon_{sig}^{REC\&SEL|GEN}}$ is computed using

$B_s^0 \to \mu^+\mu^-$	$B^+ \to J/\psi K^+$	$B^0 \to K^+\pi^-$
(17.75±0.09)%	(15.78±0.08)%	(17.73±0.09)%

Table 3.15: Detector acceptance efficiencies ϵ^{GEN} for signal and normalization channels

Sample	$\frac{\epsilon^{REC\&SEL\|GEN}_{B^+\to J/\psi K^+}}{\epsilon^{REC\&SEL\|GEN}_{B^0_{(s)}\to\mu\mu}}$	$\frac{\epsilon^{REC\&SEL\|GEN}_{B^0\to K^+\pi^-}}{\epsilon^{REC\&SEL\|GEN}_{B^0_{(s)}\to\mu\mu}}$
2011	0.473±0.011	0.840±0.026
2012	0.478±0.009	0.847±0.024

Table 3.16: Values of the ratio of the reconstruction and selection efficiencies $\epsilon^{REC\&SEL|GEN}$ for normalization and signal events.

Sample	$\frac{\epsilon^{TRIG\|SEL}_{B^+\to J/\psi K^+}}{\epsilon^{TRIG\|SEL}_{B^0_{(s)}\to\mu\mu}}$	$\frac{\epsilon^{TRIG\|SEL}_{B^0\to K^+\pi^-}}{\epsilon^{TRIG\|SEL}_{B^0_{(s)}\to\mu\mu}}$
2011	0.937 ± 0.030	0.0587 ± 0.0024
2012	0.955 ± 0.020	0.0501 ± 0.0021

Table 3.17: Values of the ratio of the trigger efficiencies $\epsilon^{TRIG|SEL}$ for normalization and signal events.

MC simulated samples and differences between data and MC are taken as systematic uncertainties. The reconstruction process is affected by systematic effects and biases. Even though a lot of them are strongly correlated for signal and normalization channels (thus canceling in their ratio), there are still some residual discrepancies which must be carefully estimated:

- the $B^+ \to J\psi K^+$ channel is affected by the presence of an extra track in the final state. Being a 3-body decay the kinematic of the event differs considerably from the one of the signal;
- the $B^0 \to \pi^- K^+$ channel is affected by the different acceptances of the muon and tracking system, and the Particle Identification (PID) requirements. To reduce the effects of the different acceptance, only events where both hadrons in the final state are inside the muon system acceptance are selected to compute the normalization factor.

In Tab.3.16 the final values of the ratio of such efficiencies for signal and normalization channels are reported.

- $\epsilon^{TRIG|SEL}$ is the trigger efficiency for events which have been reconstructed and selected. This quantity is computed using a data-driven method [98]. In Tab.3.17 the values for such ratios are reported.

The branching fractions of the normalization modes are reported in Tab.3.18

The remaining quantities appearing in the previous expression for $\alpha_{(s)}$ are computed as follows:

- the N_{Norm} for the two normalization channels are extracted through a fit to the invariant mass distribution after the selection procedure has been applied to the normalization

$\mathcal{BR}(B^+ \to J/\psi(\to \mu^+\mu^-)K^+)$	$\mathcal{BR}(B^0 \to K^+\pi^-)$
$(6.025\pm0.205)\cdot10^{-5}$	$(1.94\pm0.06)\cdot10^{-5}$

Table 3.18: Branching fraction of the normalization channels.

Sample	$N_{B^+\to J/\psi K^+}$	$N_{B^0\to K^+\pi^-}$
2011	355232 ± 608	10809 ± 439
2012	761122 ± 891	26749 ± 447

Table 3.19: Normalization channels' yields.

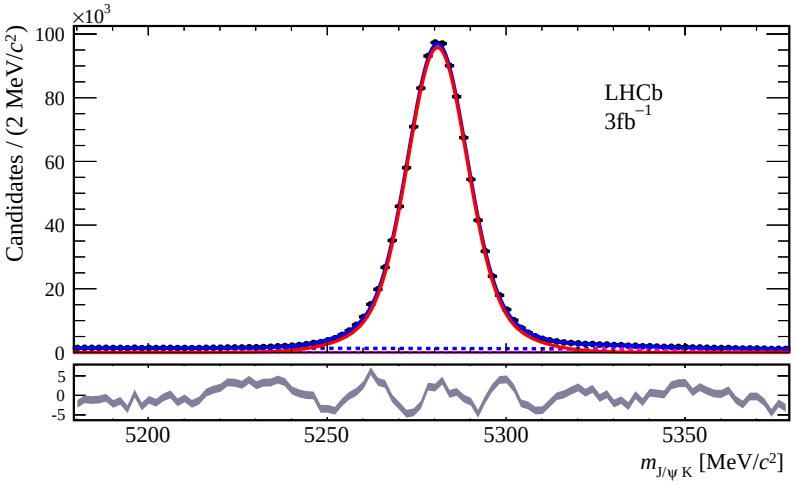


Figure 3.25: Invariant mass distribution for the $B^+ \to J/\psi K^+$ normalization channel events in data, used to extract the yield $N_{B^+\to J/\psi K^+}$. The red line is the mass shape of the $B^+ \to J/\psi K^+$ events, the blue dashed line is the combinatorial background component.

channel samples. Fig.3.25 shows the result of the fit to the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ events, while the total numbers of observed normalization channels' events are reported in Tab.3.19

- the value for the ratio $\frac{f_s}{f_d}$ has been measured by LHCb in two ways [99, 100]:
  - hadronic measurement: by comparing the relative abundances of $B_s^0 \to D_s^-\pi^+$, $B^0 \to D^-K^+$, and $B^0 \to D^-\pi^+$
  - semileptonic measurement: using the $B^0_{(s)} \to D_{(s)}X$ decays

  The value of the combination is

$$\frac{f_s}{f_d} = 0.259 \pm 0.015 \,. \tag{3.21}$$

Using all previous inputs the normalization factors $\alpha_{(s)}$ are computed for each normalization mode in the 2011 and 2012 dataset. Results are reported in Tab.3.20

For each normalization channel the global normalization factor for the combined 2011 and 2012 dataset is obtained by the relation

$$N_{B^0_{(s)}\to\mu\mu} = N^{2011}_{B^0_{(s)}\to\mu\mu} + N^{2012}_{B^0_{(s)}\to\mu\mu} = \left(\frac{1}{\alpha^{2011}} + \frac{1}{\alpha^{2012}}\right) \cdot \mathcal{BR}(B^0_{(s)} \to \mu^+\mu^-) \tag{3.22}$$

from which it follows that

$$\frac{1}{\alpha_x} \equiv \left(\frac{1}{\alpha_x^{2011}} + \frac{1}{\alpha_x^{2012}}\right) \,. \tag{3.23}$$

The final values for the normalization coefficients α and α_s are taken as the weighted average of the normalization factors for each normalization mode. The values for the B^0 and the B_s^0

Sample	2011	2012
$\alpha^{B^+\to J/\psi K^+}_{B^0\to\mu\mu}$ $(\times 10^{-11})$	7.74±0.36	3.51±0.18
$\alpha^{B^0\to K^+\pi^-}_{B^0\to\mu\mu}$ $(\times 10^{-11})$	7.62±0.77	3.57±0.30
$\alpha^{B^+\to J/\psi K^+}_{B^0_s\to\mu\mu}$ $(\times 10^{-10})$	2.99±0.23	1.35±0.11
$\alpha^{B^0\to K^+\pi^-}_{B^0_s\to\mu\mu}$ $(\times 10^{-10})$	2.94±0.35	1.38±0.15

Table 3.20: Values for $\alpha_{(s)}$ in the 2011 and 2012 datasets.

modes are the following:

$$\alpha = (2.83 \pm 0.09) \cdot 10^{-11} \tag{3.24}$$

$$\alpha_s = (8.93 \pm 0.62) \cdot 10^{-11} \ . \tag{3.25}$$

Using the previous values for normalization coefficients $\alpha_{(s)}$, the expected numbers of $B^0_{(s)} \to \mu^+\mu^-$ events in the Standard Model $\mathcal{BR}$ hypothesis and in the mass range $[4900, 6000]\ MeV/c^2$ are

$$N^{exp}_{B^0\to\mu\mu} = 4.5 \pm 0.4 \tag{3.26}$$

$$N^{exp}_{B^0_s\to\mu\mu} = 39.5 \pm 4.2 \ . \tag{3.27}$$

3.3.4 Results

The compatibility of the observed pattern of events with a given hypothesis for the $\mathcal{BR}$ is quantified through the CL_s method [101, 102]. This method relies on the measurement of the CL_{s+b} and CL_b quantities which quantify the compatibility of the observed pattern of events with the signal plus background and background only expectation, for a given $\mathcal{BR}$ hypothesis. The BDT range and the invariant mass signal window are divided in 72 2D bins, the already known 8 for the BDT and 9 for the invariant mass. The inputs for the computation of the CL_{s+b} and CL_b values are the expected number of combinatorial and physical background events and the signal fractions according to the BDT PDF.

The comparison of the distributions of the observed events and the expected background events allows to calculate the p-value (*i.e.* $1-\mathrm{CL}_b$) which is the probability that the observed pattern of event is generated by a fluctuation of the background. The results for the B^0 and B^0_s modes are:

- $B^0 \to \mu^+\mu^-$: p-value of 5.6% corresponding to a significance less than 3σ; the number of observed events is then compatible with the background only expectation and an upper limit on its $\mathcal{BR}$ is set with the CL_s method. The distribution of the CL_s as a function of different $\mathcal{BR}$ hypothesis is shown in Fig.3.26

  The final result for the Upper Limit at 95% confidence level is

  $$\mathcal{BR}(B^0 \to \mu^+\mu^-) < 7.4 \times 10^{-10} \ . \tag{3.28}$$

- $B^0_s \to \mu^+\mu^-$: p-value of $1.4{\cdot}10^{-4}$ corresponding to a significance of 4σ; in this case the number of observed events is no more compatible with the background only hypothesis and a value of the $\mathcal{BR}$ is then measured through an extended maximum likelihood fit in the invariant mass projection in eight bins of the BDT variable. The background is described by the PDF of the most relevant physical modes and an exponential, describing the combinatorial components. The free parameters of the fit are the $B^0 \to \mu^+\mu^-$ and $B^0_s \to \mu^+\mu^-$ branching fractions. The PDFs parameters as well as the event yields for each exclusive mode are fluctuated around their central value with gaussian constraints. The result of the fit in the eight BDT bins is shown in Figs.3.27 and 3.28.

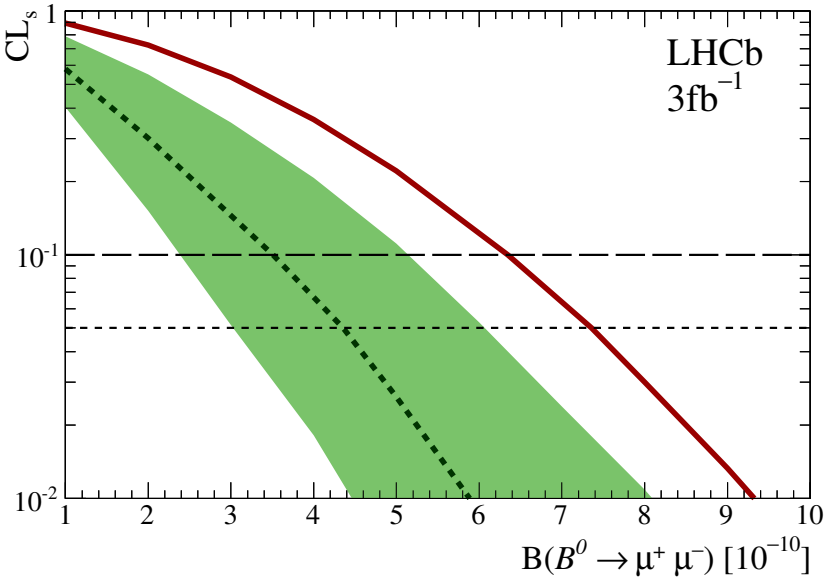


Figure 3.26: CL_s as a function of different hypotheses on $\mathcal{BR}(B^0 \to \mu^+\mu^-)$. The dashed black line is the median of the expected CL_s distribution for the background only hypothesis, the green band represents its 1σ range, and the solid red curve is the observed CL_s.

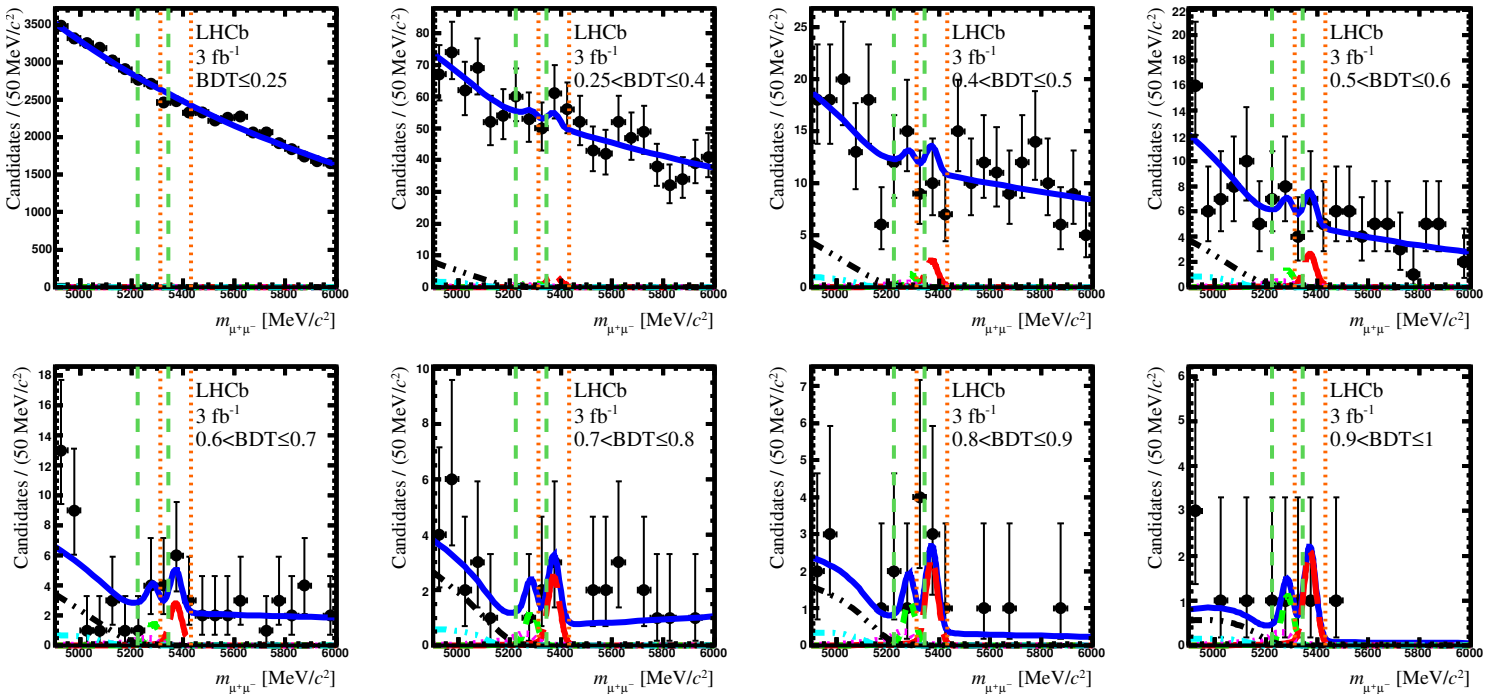


Figure 3.27: Fit results in the eight BDT bins. The contribution from the peaking $B^0_{(s)} \to hh^{(\prime)}$ is the dashed pink line, the one from $B^0_{(s)} \to \pi^+(K^-)\mu^+\nu_\mu$ is the dashed black line, the one from $B^{0(+)} \to \pi^{0,(+)}\mu^+\mu^-$ is the dashed cyan line; the $B^0_s \to \mu^+\mu^-$ signal is indicated in red, the $B^0 \to \mu^+\mu^-$ signal component in green, and the fit result is the continuous blue line.

The measured values of the branching fractions are

$$\mathcal{BR}(B_s^0 \to \mu^+\mu^-) = \left(2.9^{+1.1}_{-1.0}(stat)^{+0.3}_{-0.1}(syst)\right) \times 10^{-9} \tag{3.29}$$

$$\mathcal{BR}(B^0 \to \mu^+\mu^-) = \left(3.7^{+2.4}_{-2.1}(stat)^{+0.6}_{-0.4}(syst)\right) \times 10^{-10} \ . \tag{3.30}$$

3.4 Combination with the CMS result

The $B^0_{(s)} \to \mu\mu$ analysis presented here has been combined with the one from the CMS collaboration and has appeared in [49]. The common result has been obtained by combining the dataset collected by the two experiments during the first Run of the LHC in order to fully exploit the statistical power of the data. The LHCb analysis has already been presented in the previous sections. The CMS analysis follows the same strategy [103]: the $B^0_{(s)} \to \mu\mu$ candidates are reconstructed in events where two opposite charge tracks, identified as muons, come from a good vertex, well displaced with respect to any other primary vertex in the event. Then a BDT algorithm trained using topological and kinematic variables is used to fight the combinatorial background and to classify events as signal or background-like. The rejection of peaking backgrounds, coming from partially reconstructed b-hadron decays and requiring also the misidentification of one or more particles as μ, is achieved applying PID algorithms.

The common dataset is divided into 20 categories according to the experiment (LHCb or CMS) and the output of the BDT value. The dataset of CMS has in turn being splitted with respect to the data taking period (2011 or 2012) as well and, because of the large variation of the mass resolution with the angle between the μ-track and the beam axis, the μ-track detection region (barrel or endcap). In total 8 categories come from the LHCb dataset and 12 from the CMS one.

The yields of signal events are converted into a value of the $\mathcal{BR}$ using the $B^+ \to J/\psi(\to \mu\mu)K^+$ normalization channel. The yields of signal events are extracted from an extended maximum likelihood fit to the invariant mass $m_{\mu\mu}$ of the two muons in the range $[4.9, 5.8]$ GeV/c^2 in all the 20 categories simultaneously. The common parameters in the fit are the branching fractions of the two modes.

In Fig.3.29 the $m_{\mu\mu}$ distributions of the selected candidates in each of the 20 categories are shown, with superimposed the results of the fits.

In Fig.3.30 a weighted $m_{\mu\mu}$ of the selected candidates in the whole dataset is shown: events in the category i are weighted by a factor $w_i = S_i/\sqrt{S_i + N_i}$, where S_i is the expected number of $B_s^0 \to \mu\mu$ signals and N_i is the number of background events under the B_s^0 mass peak.

The results of the combined fit show an excess of events both at the B_s^0 and at the B^0 mass value with respect to the background only expectation. The significance of the observed excess of events at the B_s^0 mass value is found to be 6.2σ, giving the first observation of the $B_s^0 \to \mu\mu$ decay; the corresponding measured value of the branching fraction is

$$\mathcal{BR}(B_s^0 \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9} \ . \tag{3.31}$$

Concerning the $B^0 \to \mu\mu$ mode, the significance of the observed excess of events is found to be 3.0σ (evaluated using the CL_s method), which constitute the first evidence of the $B^0 \to \mu\mu$ decay. The measured value of the branching fraction is found to be

$$\mathcal{BR}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10} \ . \tag{3.32}$$

In Fig.3.31 the likelihood contours of the $\mathcal{BR}(B^0 \to \mu^+\mu^-)$ versus the $\mathcal{BR}(B_s^0 \to \mu^+\mu^-)$ are shown. These contours are obtained by computing the $-2\Delta \ln L$ test statistic, as the difference in the log-likelihood function ($\ln L$) between fits with fixed values of the two branching fractions and the one from the nominal fit.

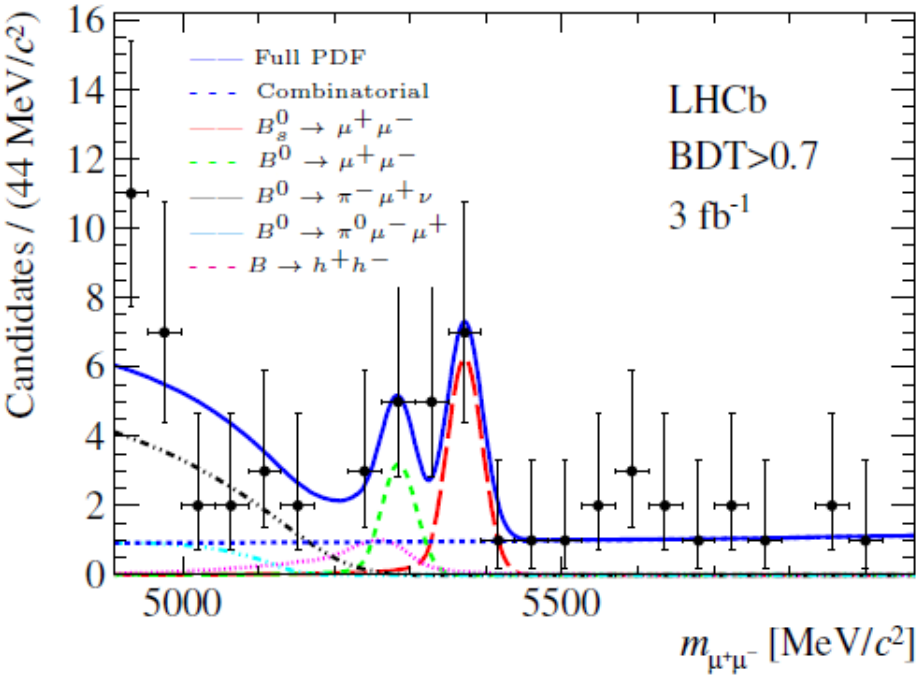


Figure 3.28: Fit results for events falling in the most sensitive region with BDT>0.7. The contribution from the peaking $B^0_{(s)} \to hh^{(\prime)}$ is the dashed pink line, the one from $B^0_{(s)} \to \pi^+(K^-)\mu^+\nu_\mu$ is the dashed black line, the one from $B^{0(+)} \to \pi^{0,(+)}\mu^+\mu^-$ is the dashed cyan line; the $B^0_s \to \mu^+\mu^-$ signal is indicated in red, the $B^0 \to \mu^+\mu^-$ signal component in green, and the fit result is the continuous blue line.

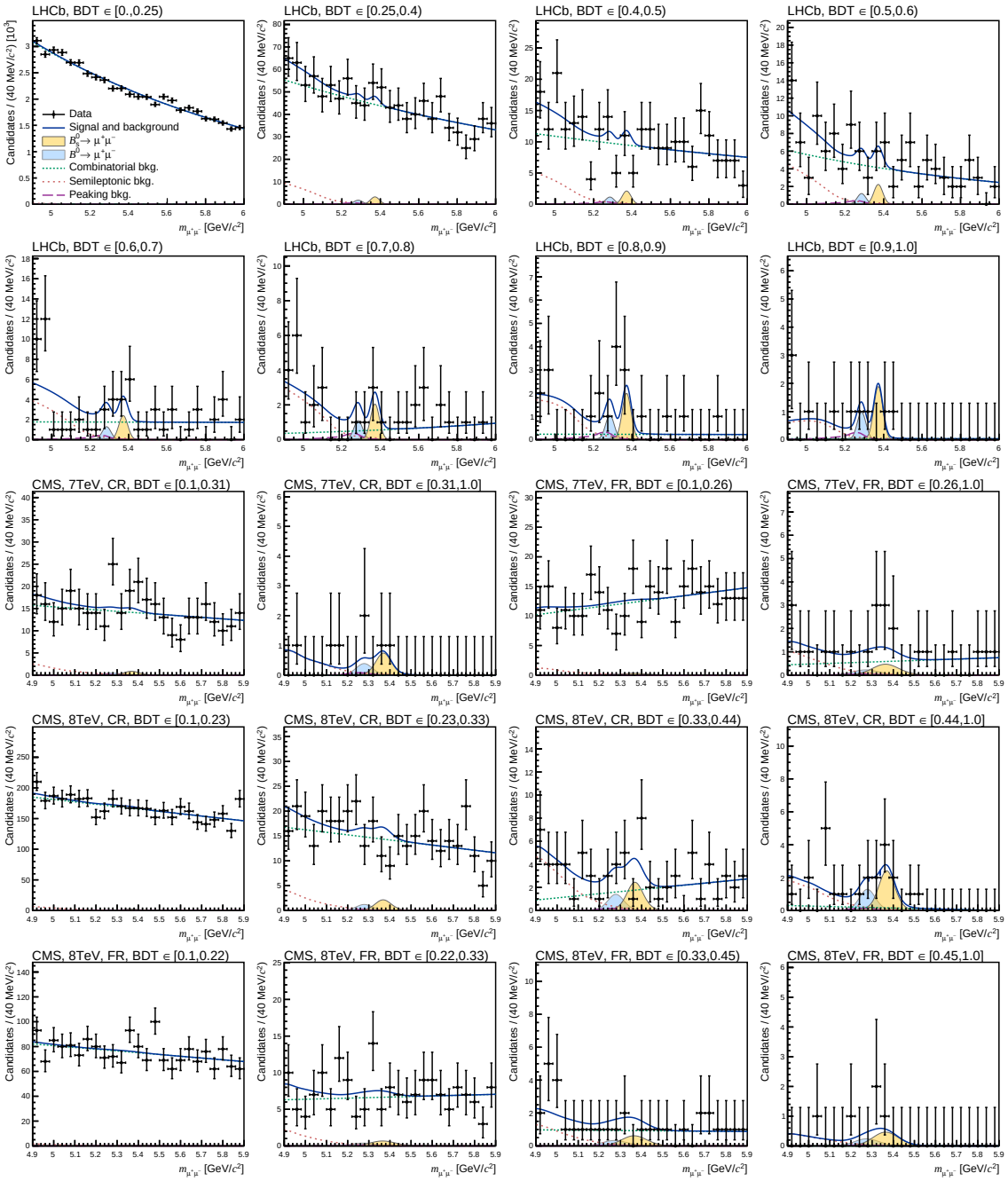


Figure 3.29: Invariant mass $m_{\mu\mu}$ distributions for events in all categories, with fit results superimposed.

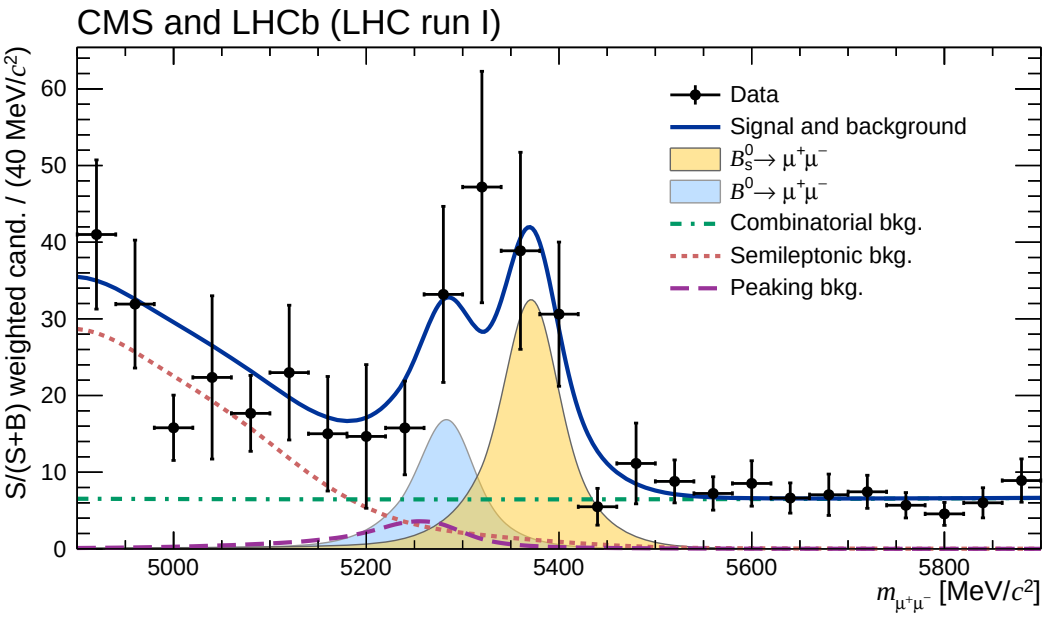


Figure 3.30: Weighted distribution of the invariant mass distribution $m_{\mu\mu}$ for events in all categories.

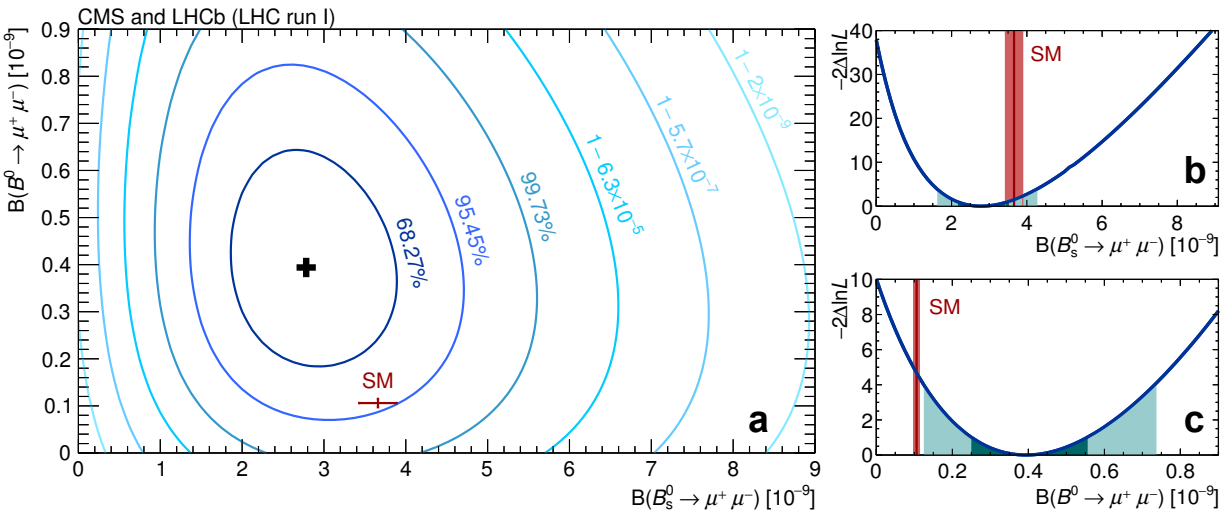


Figure 3.31: Likelihood contours of the $\mathcal{BR}(B^0 \to \mu^+\mu^-)$ versus the $\mathcal{BR}(B_s^0 \to \mu^+\mu^-)$ (a); one-dimensional profile likelihood scan for $\mathcal{BR}(B_s^0 \to \mu^+\mu^-)$ (b) and $\mathcal{BR}(B^0 \to \mu^+\mu^-)$ (c), obtained by fixing only one of the two parameters of interest and allowing all the others to vary in the fit. The dark and light cyan areas are the $\pm 1\sigma$ and $\pm 2\sigma$ regions. The SM prediction with its uncertainty is indicated by the red vertical band.

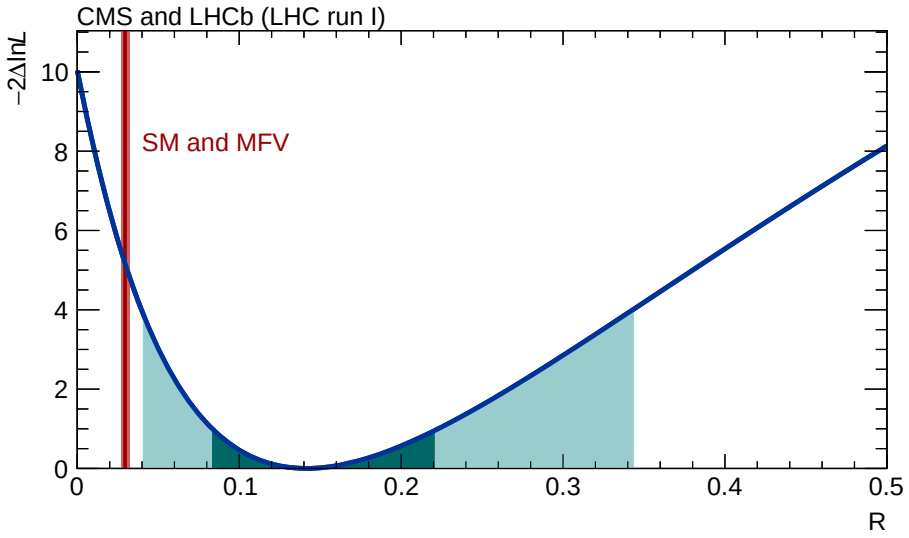


Figure 3.32: Test statistic $-2\Delta \ln L$ as a function of $\mathcal{R} = \mathcal{BR}(B^0 \to \mu\mu)/\mathcal{BR}(B_s^0 \to \mu\mu)$. The dark and light cyan areas are the $\pm 1\sigma$ and $\pm 2\sigma$ regions. The SM prediction with its uncertainty is indicated by the red line.

To better estimate the agreement with the SM prediction a fit where the parameters of interests are the ratio of the branching fraction with respect to their SM predicted value, *i.e.* the quantity

$$\mathcal{S}_{SM}^{B^0_{(s)}} \equiv \frac{\mathcal{BR}_{meas}(B^0_{(s)} \to \mu\mu)}{\mathcal{BR}_{SM}(B^0_{(s)} \to \mu\mu)} \ . \tag{3.33}$$

The obtained results for these quantities are:

$$\mathcal{S}_{SM}^{B^0} = 3.7^{+1.6}_{-1.4} \tag{3.34}$$

$$\mathcal{S}_{SM}^{B^0_s} = 0.76^{+0.20}_{-0.18} \tag{3.35}$$

The measured value of $\mathcal{S}_{SM}^{B^0}$ is 2.2σ away from the SM prediction, while $\mathcal{S}_{SM}^{B^0_s}$ is at 1.2σ from its SM prediction.

Finally, another relevant observable related to these modes is the ratio $\mathcal{R}$ of the branching fraction

$$\mathcal{R} = \frac{\mathcal{BR}(B^0 \to \mu\mu)}{\mathcal{BR}(B_s^0 \to \mu\mu)} \ . \tag{3.36}$$

This observable is relevant to constraint non Minimal Flavor Violating modes. In Fig.3.32 the one-dimensional likelihood scan is shown for this quantity.

The measured value for $\mathcal{R}$ is

$$\mathcal{R} = 0.14^{+0.08}_{-0.06} \tag{3.37}$$

which is compatible with its SM prediction at the 2.3σ level.

3.5 Interpretation of the results.

In this section the interpretation of the experimental results presented in the previous section will be discussed. Firstly the direct constraints on the Wilson coefficients obtained using the

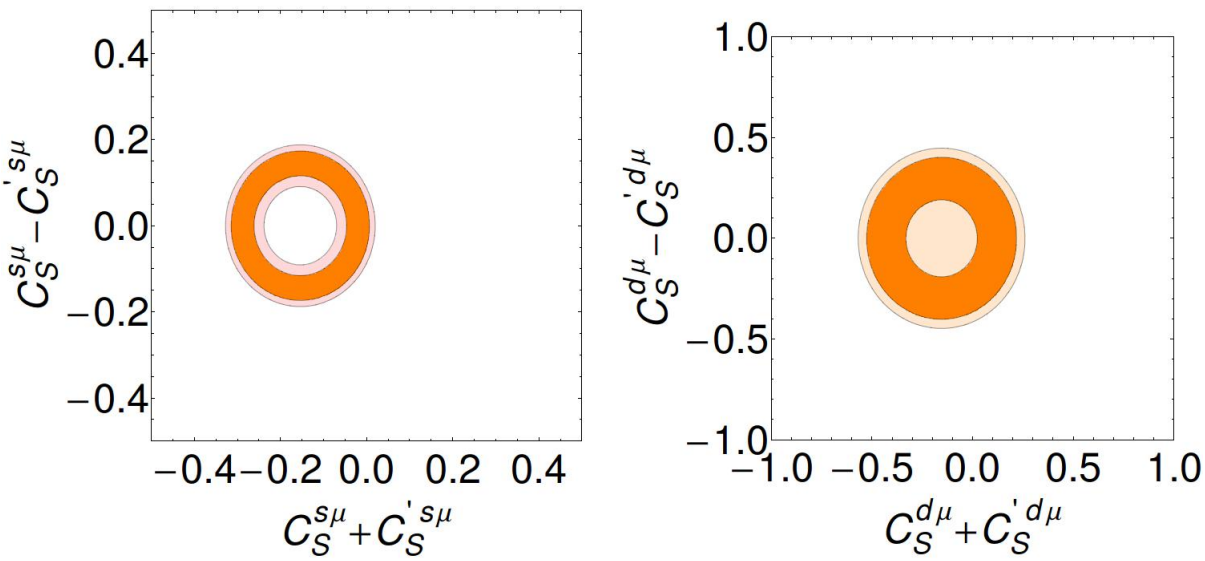


Figure 3.33: 1σ and 3σ contours on the $(\mathcal{C}_S^{\mu} - \mathcal{C}_S^{\mu\prime})$-$(\mathcal{C}_P^{\mu} - \mathcal{C}_P^{\mu\prime})$ plane with the constraints in eq.(1.57) obtained with the measured values in eqs.(3.31) and (3.31). $\mathcal{C}_{P,S}^{s\mu}$ and $\mathcal{C}_{P,S}^{d\mu}$ refer to the Wilson coefficients related to the $B_s^0 \to \mu\mu$ and $B^0 \to \mu\mu$ decays respectively. Picture from Ref.[36].

measured values of the branching fractions of the $B_{(s)}^0 \to \mu^+\mu^-$ modes, will be shown. Then the results of the global fits of the Wilson coefficients to the measured values of the $\mathcal{BR}(B_{(s)}^0 \to \mu^+\mu^-)$ and other $b \to s\ell\bar{\ell}$ transition related observables will be shortly reviewed.

Direct constraints on the Wilson Coefficients. The measured values for the $\mathcal{BR}(B_{(s)}^0 \to \mu\mu)$ decays are both in good agreement with previous measurements [44]. The measured central values still show slight deviations from the SM predictions presented in Sec.1.2.3. These deviations are not strong enough to claim a statistically significant discrepancy, but, if confirmed with more data, they can provide hints of possible NP scenarios to be explored. In particular, if the measured central values for the branching fractions do not change with more statistics, the following considerations are in order:

- the measured values hint a Quark Flavor Violating scenario, because the branching ratio of the B^0 and the B_s^0 modes are differently shifted with respect to the SM predictions; in particular the branching ratio of the B_s^0 mode lies below its SM prediction, while the branching ratio of the B^0 decay is higher than the SM prediction of a factor ~ 3.5;

- as a consequence of the previous point, the measured value of the ratio $\mathcal{R}$, points toward a non-MFV NP scenario, thus requiring new sources of flavor violation in addition to the Higgs-fermions Yukawa interaction.

The constraints in the $(\mathcal{C}_S^{\mu} - \mathcal{C}_S^{\mu\prime})$-$(\mathcal{C}_P^{\mu} - \mathcal{C}_P^{\mu\prime})$ plane in the case where eq.(1.57) holds are reported in Fig.3.33. In particular, these contributions are compatible with zero, and rule out NP effects in the scalars or pseudoscalars Wilson coefficients. Concerning the axial current operators, the measurements still leave rooms for a non vanishing contribution $\delta\mathcal{C}_{10}^{\mu} - \delta\mathcal{C}_{10}^{\mu\prime}$. In particular the following bounds can be obtained [16]:

- from $B_s^0 \to \mu\mu$: $\delta\mathcal{C}_{10}^{s\mu} - \delta\mathcal{C}_{10}^{s\mu\prime} \simeq 0.5$,

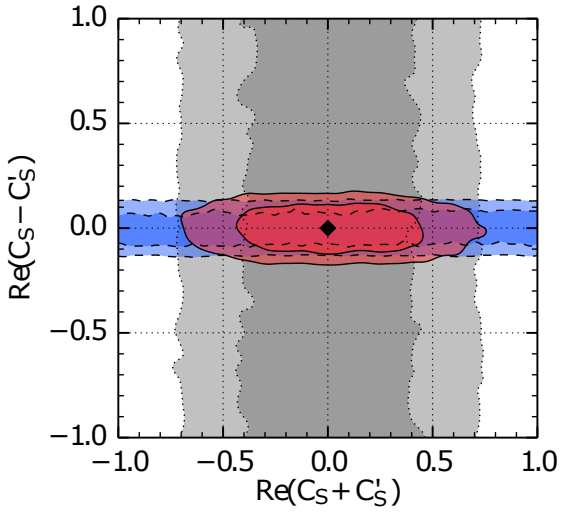


Figure 3.34: 1σ (dark) and 2σ (light) regions for the $Re(C_S - C'_S)$ and $Re(C_S + C'_S)$ allowed by the measured values of the $\mathcal{BR}(B_s^0 \to \mu\mu)$ and the "flat term" F_H^μ in the decay $B^0 \to K^{\star 0}\mu\mu$ (the reader is referred to Ref.[104] for its definition). The red regions are the 1σ (dark) and 2σ (light) regions allowed by the combination. The SM value is the black diamond. Picture from Ref.[104].

- from $B^0 \to \mu\mu$: $\delta\mathcal{C}_{10}^{s\mu} - \delta\mathcal{C}_{10}^{s\mu\prime} \simeq \mathcal{C}_{10}^{\mu SM}$.

What is evident is that the measured values allow to rule out huge NP enhancements, unless a fine tuning of the Wilson coefficients makes observable effects vanishing.

Global fit of Wilson coefficients from $b \to s\ell\bar{\ell}$ observables. Of particular interest is also the interplay between the constraints on the Wilson coefficients from the $B_{(s)}^0 \to \mu^+\mu^-$ decays and the ones obtainable from other processes generated by $b \to s\ell\ell$ transitions. As it has been pointed out in Sec.1.2.2, the $\mathcal{BR}$ observable is sensitive only to the differences of the left and right Wilson coefficients $C_{10,S,P}^{(\prime)}$. Further constraints can be obtained from the observables related to $b \to s\ell\ell$ transitions, which are instead sensitive to the sum of the $C_i^{(\prime)}$ $(i = 9, 10, P, S)$ Wilson coefficients. For example, Fig.3.34 shows the constraints on the real part of the sum and the difference of the scalar Wilson coefficients $\mathcal{C}_S^{(\prime)}$ introduced in eqs.(1.51), (1.54) coming from the $\mathcal{BR}(B_s^0 \to \mu\mu)$ and the co-called "flat-term" F_H^μ entering the angular distribution of the decay products in the $B^0 \to K^\star\mu^+\mu^-$ (for its definition the reader is referred to Ref.[104]). The measured values for the scalar Wilson coefficients are well compatible with the SM prediction.

The experimental results about the $\mathcal{BR}(B_{(s)}^0 \to \mu\mu)$ are also used as inputs of global fits to constraints the effective operators entering the $b \to s\ell\bar{\ell}$ transitions. In Ref.[105] a global fit of the NP contribution $\delta C_9/C_9^{SM}$ and $\delta C_{10}/C_{10}^{SM}$ to the Wilson coefficients has been done (for the observables used as inputs of the fit see caption of the Fig.3.35) and the allowed regions for these parameters are shown in Fig.3.35. From the fit results the SM value seems disfavored by 2.3σ. This discrepancy comes from the tension of the best-fit values of C_9 which in turn is due to the observed tensions in the $b \to s\ell\bar{\ell}$ transition related processes (notably the $B^{0,\pm} \to K^{(\star)\pm}\mu^+\mu^-$) . Nevertheless this tensions is not enough significant to be interpreted as a clear sign of NP and more data are needed in order to better interpret these results.

Fig.3.36 shows another global fit from Ref.[106] where the constraints from the $\mathcal{BR}$ and the angular observables are separately shown, together with the combination. This fit shows that

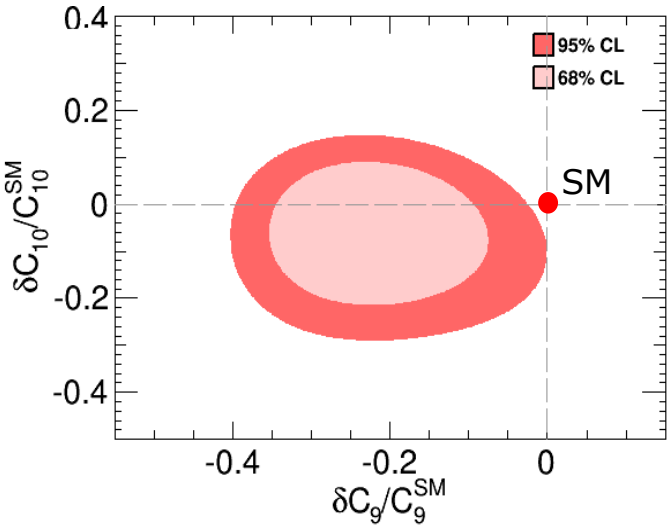


Figure 3.35: 1σ and 3σ regions for the δC_{10} and δC_9 shift in the value of the Wilson coefficients allowed by the measured values of the $\mathcal{BR}(B^0_{(s)} \to \mu\mu)$, R_K, the differential branching fraction of the exclusive modes $B^{0,(+)} \to K^{0,\star(+)}\mu\mu$, the branching ratio of the $B^0 \to K^\star ee$ decay, and the branching ratios of the inclusive modes $B^{0,(+)} \to X_s\mu\mu$ and $B^{0,(+)} \to X_s ee$. The red point is the SM value. Picture from Ref.[105].

the main tensions between the experimental measurements of the observables and their SM predictions come from the angular distributions.

3.6 LHCb prospects for the next Runs of the LHC

The data collected during the forthcoming Run II of the LHC will play a major role reducing the statistical uncertainties on the $B^0_{(s)} \to \ell\bar{\ell}$ related observables that have been measured so far. This will allow to enforce or alleviate the currently observed deviations from SM predictions. Indeed the comparison between the theoretical predictions and the measured values of the observables is still affected by the large experimental uncertainties. These are still dominated by the statistical errors, as it can be seen from eqs.(3.29)-(3.30), and the larger data sample collected during the forthcoming runs of the LHC will allow to reduce this error. The improvement of the LHCb analysis due to the larger available statistics has been studied using MC toys based on the current published analysis. In particular, at the end of the second run, where an integrated luminosity of 5 fb^{-1} will be collected by LHCb, the errors on $\mathcal{BR}(B^0_s \to \mu^+\mu^-)$ and $\mathcal{BR}(B^0 \to \mu^+\mu^-)$ will be decreased by a factor 2. These values are still a factor ~ 2 and ~ 10 for the B^0_s and the B^0 modes respectively, larger than the theoretical errors. The precision of the measurements can be further improved by optimizing the current analysis, and some ideas and tools to this extent will be presented in the next section.

At the same time, the increased statistics will allow to measure new observables that are complementary with the branching ratios, such as the effective lifetime defined in eq.(1.65). Indeed, as it has been pointed out in Ref.[37], this observable can still show up NP effects even though the measured branching ratio is in agreement with the SM prediction. The feasibility of the effective $B^0_s \to \mu^+\mu^-$ lifetime measurement with the dataset collected by LHCb during the second and third runs of the LHC has been carefully studied (for more details the reader is referred to Ref.[107]). In the dataset collected during the first run ~ 30 $B^0_s \to \mu\mu$ (~ 39 in the SM hypothesis $\mathcal{BR}$, see eq.(3.27)) are observed. By extrapolating from these values (*i.e.* assuming a scenario where no improvements in the analysis are done), in the dataset collected at the end

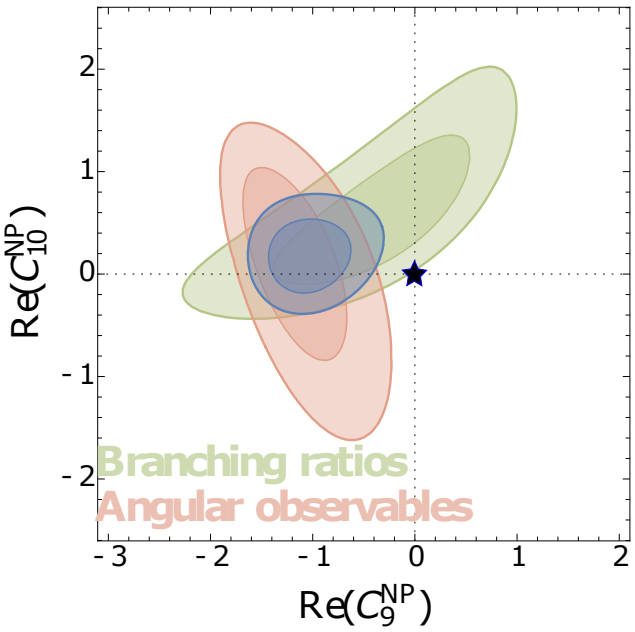


Figure 3.36: 1σ (dark) and 2σ (light) allowed regions for the NP contributions to C_{10}^{NP} and C_9^{NP} to the C_{10} and C_9 Wilson coefficients from the measured values of the $b \to s\ell\bar{\ell}$ transitions. The allowed values from the $\mathcal{BR}$ (green) and the angular observables (pink) are separated, and their combination is the blue region. The black star denotes the SM case. Picture from Ref.[106].

of the Run II ~ 50 $B_s^0 \to \mu\mu$ events (~ 65 in the SM hypothesis) are expected. This statistics can already allow to measure the effective lifetime with a precision, computed using MC toys, of $\sim 15\%$. After the Run III (assuming an integrated luminosity of $50 fb^{-1}$) ~ 500 signal events will be available (~ 650 in the SM $\mathcal{BR}$ hypothesis) and the lifetime measurement can be done with an estimated precision (computed using MC toys), of $\sim 8\%$.

3.7 Preparing current analysis improvements

In the present section two preliminary studies concerning the definition of a new isolation variable and a further optimization of the final BDT classifier will be presented. These studies aim to improve the previous analysis in view of a re-analysis of the $3.1 fb^{-1}$ collected during the first run of the LHC.

Among the input variables of the final BDT classifier, one of the most discriminating between signal and background is the isolation of the two muons (hereafter referred as "iso5"). Exploiting a topological vertex algorithm ("ZVtop" in the following) [109] another "independent" isolation variable has been defined and studied. Practically speaking, several variables related to the isolation of the muons, can be sorted out using the ZVtop algorithm. But in order to exploit all the discriminating power of these variables, they have been combined using a MVA operator. In the following the definition of these variables related to the muon isolation will be presented, followed by their combination into a unique and more discriminating variable through a BDT algorithm.

3.7.1 The ZVtop algorithm

The main idea of the topological vertex algorithm is to consider the reconstructed tracks in the event as a probability tube in three dimensions. The 3D probability density associated to the i-th track in the event is:

$$f_i(\vec{x}) \equiv e^{-\frac{1}{2}(\vec{x}-\vec{p})^T V^{-1}(\vec{x}-\vec{p})} \tag{3.38}$$

where $\vec{p}$ is the point of closest approach of a tracks with respect to a generic space-point $\vec{x}$ and V is the covariance matrix of the track at the point $\vec{p}$.

The main object of the ZVtop algorithm is the so-called Vertex Function ("Vf" hereafter) defined as the sum of all probability functions f_i associated to the N tracks in the event, *i.e.*

$$Vf(\vec{x}) \equiv \sum_{i=1}^{N} f_i(\vec{x}) - \frac{\sum_{i=1}^{N} f_i^2(\vec{x})}{\sum_{i=1}^{N} f_i(\vec{x})} \tag{3.39}$$

The Vf defined above is such that its value is close to n if $n+1$ tracks make a common vertex at a given space point $\vec{\vec{x}}$. The maxima of Vf represent candidate decay vertexes. In order to avoid (obvious) large maxima in correspondence of the primary interaction points, the tracks used to build up Vf are selected through a cut on their IPχ^2 with respect to any PV.

Usually, even if three or more tracks are very close each other at a given point, it's very unlikely that they perfectly overlap at that point. For this reason, the maxima of Vf are usually two-tracks vertexes, eventually quite close each other. The ZVtop algorithm performs thus a "clustering" of all the two-tracks reconstructed maxima into bigger vertexes. After this clustering, the algorithm attaches to each reconstructed vertex the tracks in the event, in order to build a candidate. One important feature of that algorithm is that a given track can be attached to only one reconstructed vertex.

For a more detailed description of the ZVtop algorithm, the reader is referred to Refs.[109, 110].

3.7.2 Definition of Isolation Variables using ZVtop

The variables related to the isolation of the muons that have been studied can be divided in two categories:

- Type-I isolations: variables making use only of the Vertex Function
- Type-II isolations: variables exploiting the vertex finder algorithm as well.

Some of these variables compare the ''results" found with the standard reconstruction with the ones obtained with the ZVtop algorithm.

The Vertex Function is built up with tracks satisfying a certain selection criteria on the IPχ^2, IP, Ghost Probability; analogously the set of tracks that will be attached to the Vf maxima are also selected through a cut on the same parameters. In addition other parameters enter the ZVtop algorithm, *e.g.* those related to the maximum finder algorithm.

The ones that have been optimized for the definition of the isolation variables are the following:

- "*Cut_Vf_IPS*" by which the tracks that are used to calculate the Vf are selected,
- "*Cut_IPS*" by which the tracks that must be attached to the vertexes found are selected,
- "*Cut_max_VF*" which is the threshold above which the Vf has a maximum. This parameter is meant to reject fake vertexes which can arise from wiggles in the Vf generated *e.g.* when two tracks get too close without crossing each other.

For the studies presented here the optimal values for these three parameters are reported in Tab.3.21.

These variables have been defined for the first time in [110].

Cut_Vf_IPS	Cut_IPS	Cut_max_VF
20	30	0.5

Table 3.21: List of parameters together with their final value that have been optimized for the definition of the ZVtop-based isolation variables.

Type-I isolations. The most important Type-I variable is the sum of the numbers of the Vf vertexes found along each muon track after the removal of the other muon.

Let $track(\mu^+)$ and $track(\mu^-)$ be the two muon tracks. Being "good" tracks they are often selected in the set of tracks used to build the Vf. By definition of the vertex function, if $track(\mu^+)$ and $track(\mu^-)$ cross each other then the Vf has a maximum at the crossing point. The idea is to look for extra maxima of the Vf along the muon tracks $track(\mu^+)$ and $track(\mu^-)$ in addition to the one they form at the B candidate position. The counting of these extra vertexes is done as follows:

- starting considering $track(\mu^+)$, the Vf is computed removing $track(\mu^-)$ from the set of tracks
- a scan of the Vf values along the $track(\mu^+)$ (from -3 cm up to $+5$ cm with respect to the position of the B candidate decay vertex) is perfomed and the number n_+ of maxima is found ;
- considering the second $track(\mu^-)$, the Vf is computed again removing the *other* $track(\mu^+)$ from the set of tracks,
- a scan of the Vf values along the $track(\mu^-)$ (from -3 cm up to $+5$ cm with respect to the position of the B candidate decay vertex) is performed and the number n_- of maxima is found ;
- the final isolation variable is the sum: $N_{tot} = n_+ + n_-$.

For signal this value must be peaked at zero, while for background this does not necessarily happen.

In addition to the previous one, other variables can be defined:

- value of the Vf at the B candidate position ("Vf_B");
- value of the absolute maximum of the Vf along all muons tracks ("Vfmax_Muons") and its position ("Position_Vfmax_Muons");
- value of the absolute maximum of the Vf along the muons tracks in the downstream region (with respect to B decay vertex) ("Vfmax_Muons_DownStream') and its position ("Position_Vfmax_Muons_DownStream");
- value of the absolute maximum of the Vf along the muons tracks in the upstream region (with respect to B decay vertex) ("Vfmax_Muons_UpStream') and its position ("Position_Vfmax_Muons_UpStream").

These variables and their discriminating power depend only on two of the three parameters listed above: the IPχ^2 of tracks used to build the Vertex Function (*Cut_IPS_VF*) and the value of the Vf threshold value above which the Vf has a maximum (*Cut_max_Vf*).

In Figs.3.37-3.38 the distributions of the previous variables for the $B_s^0 \to \mu\mu$ signal and for the generic $b\bar{b} \to \mu\mu X$ Monte Carlo background events are shown.

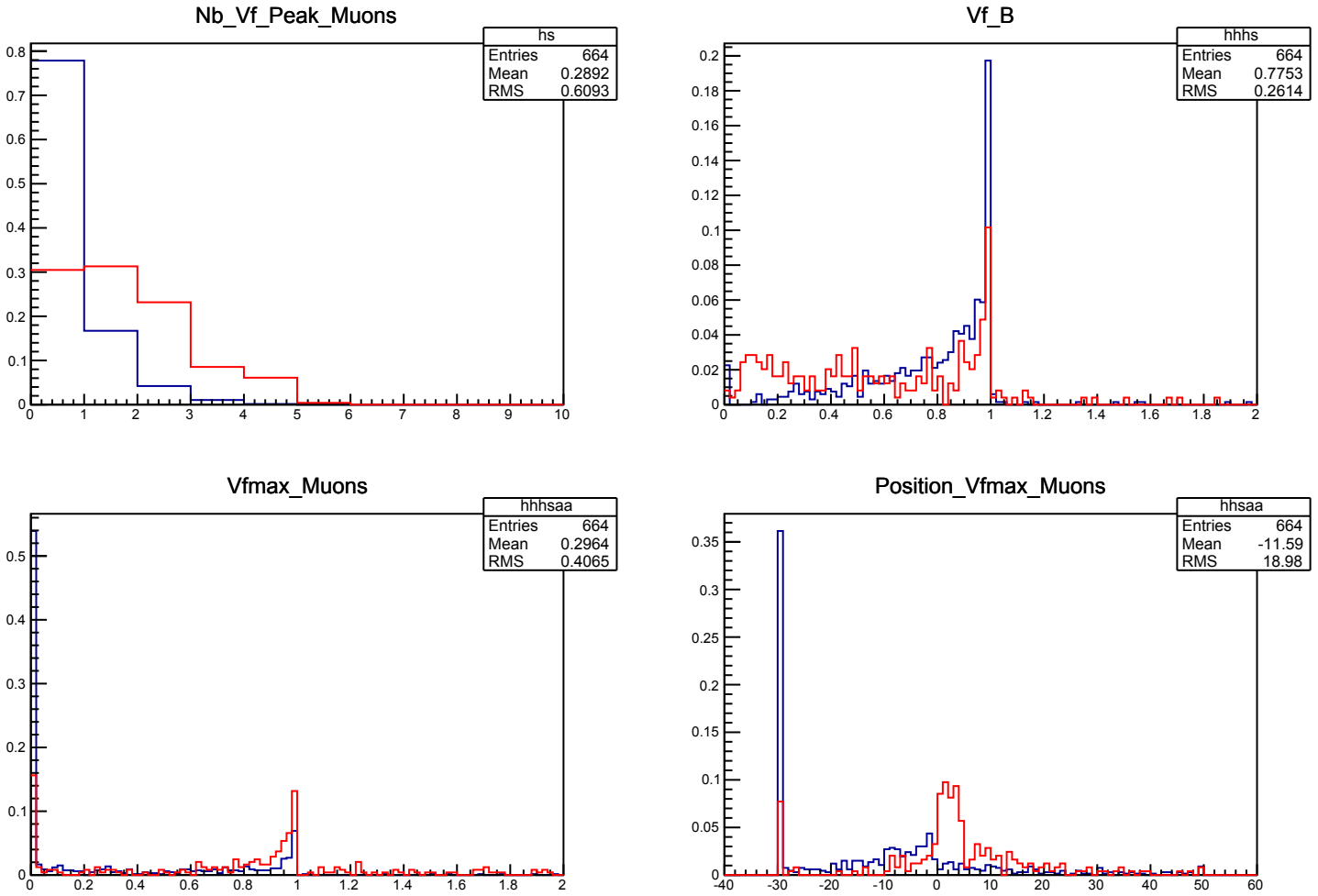


Figure 3.37: Isolation related variables computed with the topological vertexing algorithm. Top-left: the sum of the number of peaks of the Vf along the two muon tracks when the other track is removed; top-right: value of the Vf at the B candidate decay vertex position. In the bottom figures maximum value of the Vf along the two muon tracks (left) and its position in the z direction (right). In blue is the signal, in red the generic $b\bar{b} \to \mu\mu X$ background.

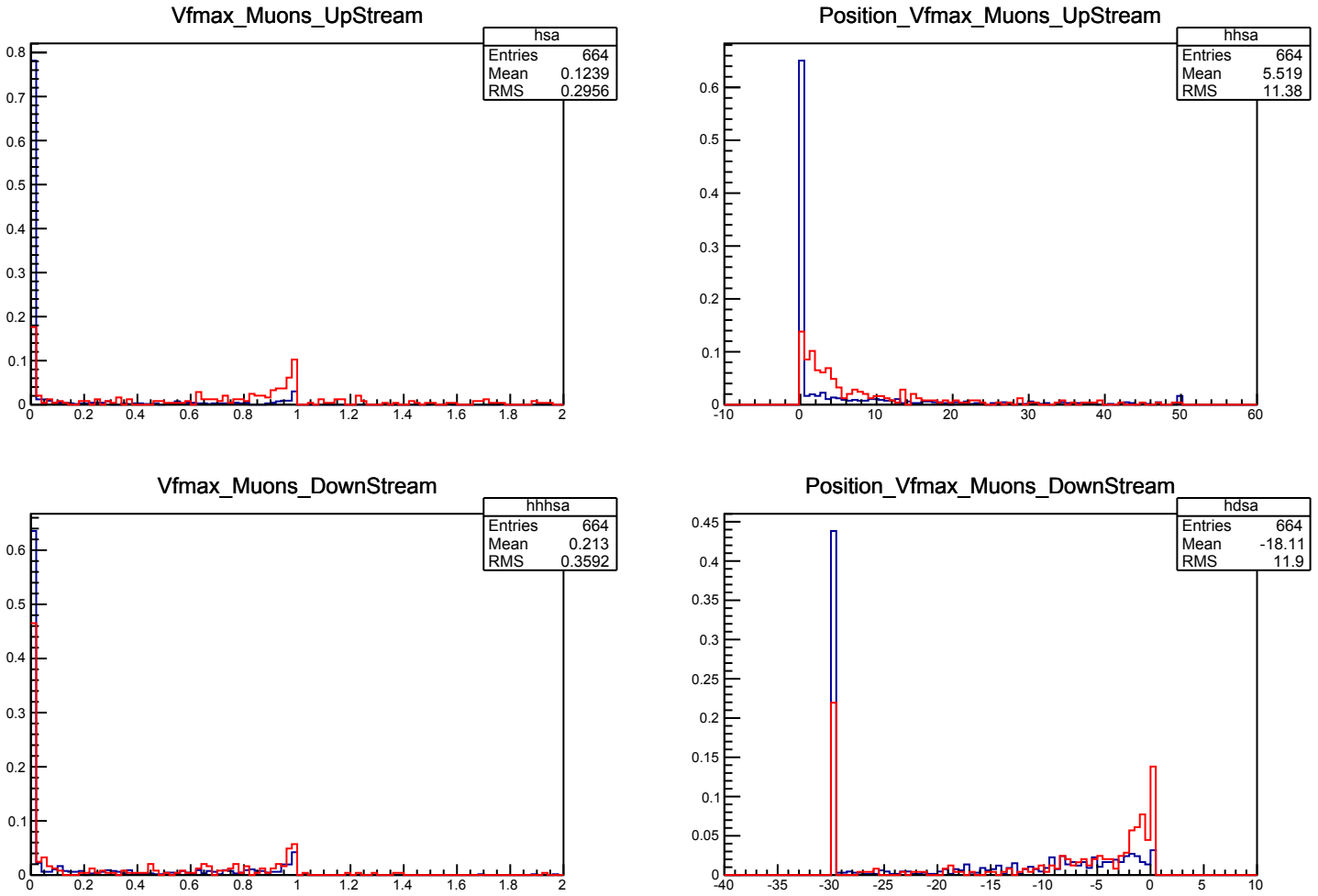


Figure 3.38: Isolation related variables computed with the topological vertexing algorithm: maximum value of the Vf along the two muon tracks in the upstream direction with respect to the B candidate decay vertex (top-left) and its position in the z direction (top-right); maximum value of the Vf along the two muon tracks in the downstream direction with respect to the B candidate decay vertex (bottom-left) and its position in the z direction (bottom-right). In blue is the signal, in red generic $b\bar{b} \to \mu\mu X$ background.

Type-II isolation variables. The Type-II variables make explicit use of the ZVtop algorithm, by which a selected set of tracks is attached to the vertexes of the Vf. For this reason the signal-background separation of these variables depends also on the third parameter introduced above, the *Cut_IPS*.

The type-II isolation variables are the following:

- "mu_zv_same" defined as follows:
  - 0 if the two μ tracks come from two different ZVtop-reconstructed vertexes (*i.e.* ZVtop finds two different origin vertexes for the 2 tracks)
  - 1 if the two μ tracks come from the same ZVtop-reconstructed vertex with only two tracks attached (*i.e.* ZVtop reconstructs the same standard-reconstructed vertex)
  - 2 if the two μ tracks come from the same ZVtop-reconstructed vertex with more than two tracks are attached

  For signal this variable must peak at 1, while for $b\bar{b} \to \mu\mu X$ events the ZVtop algorithm should attach the two tracks to different vertexes.

- "mu_zv_nvtx": total number of vertexes found by the algorithm

- "Mum_zv_ipall": min IP of the μ^- with respect to any other vertex found by ZVtop different from the one which is our candidate (and analogously for μ^+)

- "Mum_zv_ipsall": min IPχ^2 of the μ^- with respect to any other vertex found by ZVtop different from the one which is our candidate (and analogously for μ^+)

- "Mum_zv_chi2": the χ^2 of the ZVtop-reconstructed vertex closest to the μ^- track (and analogously for μ^+)

- "Mum_zv_dist": the distance of the ZVtop-reconstructed vertex closest to the μ^- track from the primary vertex (and analogously for μ^+)

- "Mum_zv_proba": the Vf value at the position of the ZVtop-reconstructed vertex closest to the μ^- track (and analogously for μ^+)

- "Mum_zv_proba_close": the Vf value at the point of closest approach of the μ^- track to the ZVtop-reconstructed vertex closest to the μ^- track (and analogously for μ^+)

Figs.3.39-3.40 show the distributions of the previous variables for the $B_s^0 \to \mu\mu$ signal and for the generic $b\bar{b} \to \mu\mu X$ Monte Carlo background.

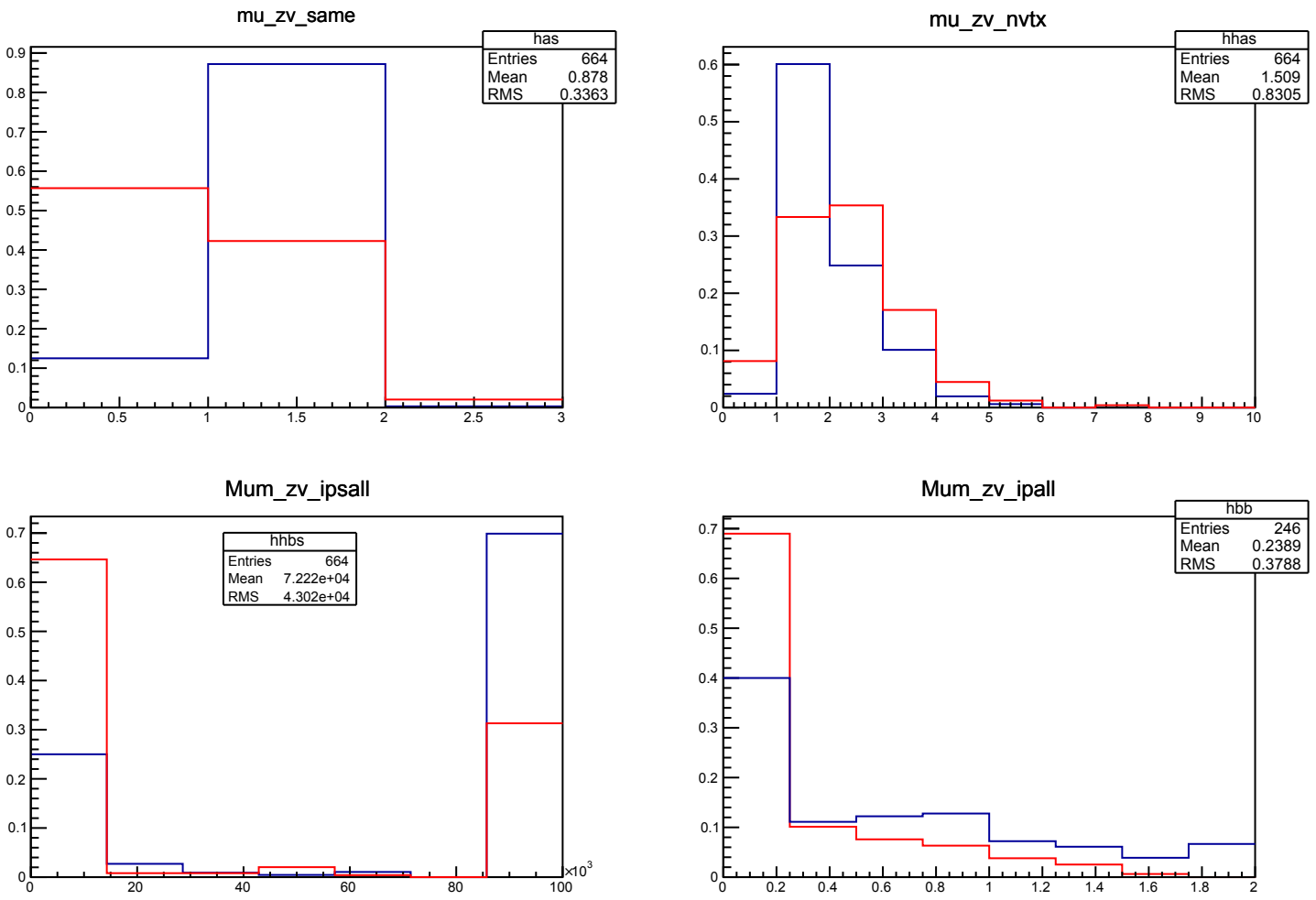


Figure 3.39: Distribution of the Type-II isolation variables for signal (blue) and generic $b\bar{b} \to \mu\mu X$ background (red). For the description of the variables the reader is referred to the text.

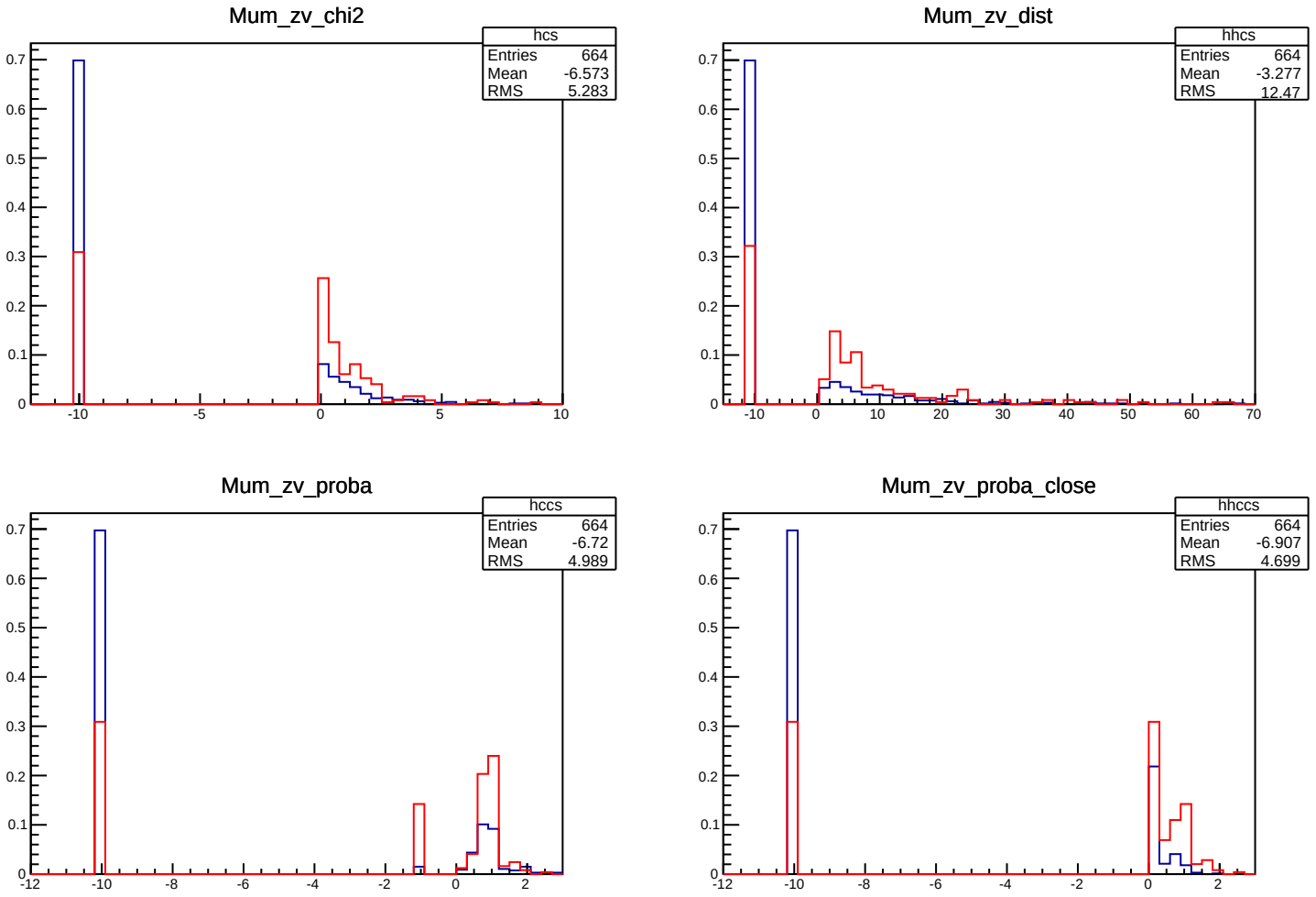


Figure 3.40: Distribution of the Type-II isolation variables for signal (blue) and generic $b\bar{b} \to \mu\mu X$ background (red). For the description of the variables the reader is referred to the text.

Combination of ZVtop related isolation variables: ZViso. A subset of the variables defined above have been combined into a unique isolation variable using a BDT operator. This variable, called "ZViso", will be tested as an extra input variable for the final BDT classifier.

Even though the BDT operator used for the classification could gain from the correlation of each of these isolation variables with the remaining twelve, the choice to combine the isolation variables defined here into a unique one before being used for the final classification appeared more suitable. This because it allows to deal with a reasonable number of isolation variables in the final BDT classification, and at the same time the signal-background discrimination of the ZViso will be much higher than the one of each single input variables used for its definition.

The ZVtop-based variables defined above are correlated with each other. The ZVtop based isolation variables used as input for the ZViso BDT training have been chosen to be as much as possible uncorrelated with each other (linear correlation less than $\sim 70\%$). Variables more strongly correlated have still been used if their correlation was significantly different in signal and background. For example, the "zv_ipsall", "zv_ipall", "zv_proba", "zv_proba_close" are linearly correlated at more than 98% level with each other and equally for signal and background hence only the first one, "zv_ipsall", has been used as input variable of the ZViso BDT. Finally the set of input variables for the ZViso BDT is the following one

- Nb_Vf_Peak_Muons
- Vf_B
- mu_zv_same
- mu_zv_nvtx
- zv_ipsall≡Mum_zv_ipsall+Mup_zv_ipsall
- zv_chi2≡Mum_zv_chi2+Mup_zv_chi2
- zv_dist≡Mum_zv_dist+Mup_zv_dist
- Vfmax_Muons
- Position_Vfmax_Muons
- Vfmax_Muons_DownStream
- Position_Vfmax_Muons_DownStream
- Vfmax_Muons_UpStream
- Position_Vfmax_Muons_UpStream

Figs.3.41,3.42 show the linear correlations between the previous variables.

3.7.3 Combination of isolation variables

In view of the improvement of the $B^0_{(s)} \to \mu\mu$ analysis, another isolation variable (called here "isoBDT") has been defined in Ref.[108], and will be briefly summarized in the following.

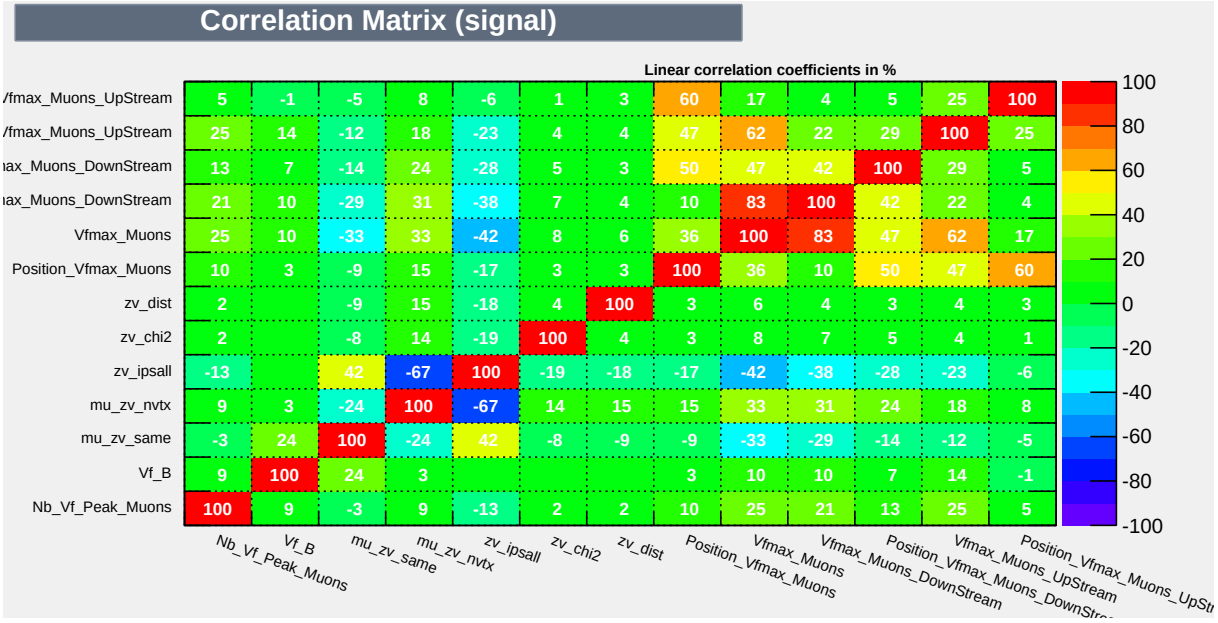


Figure 3.41: Linear correlation coefficients of the input variables of the ZViso for signal MC.

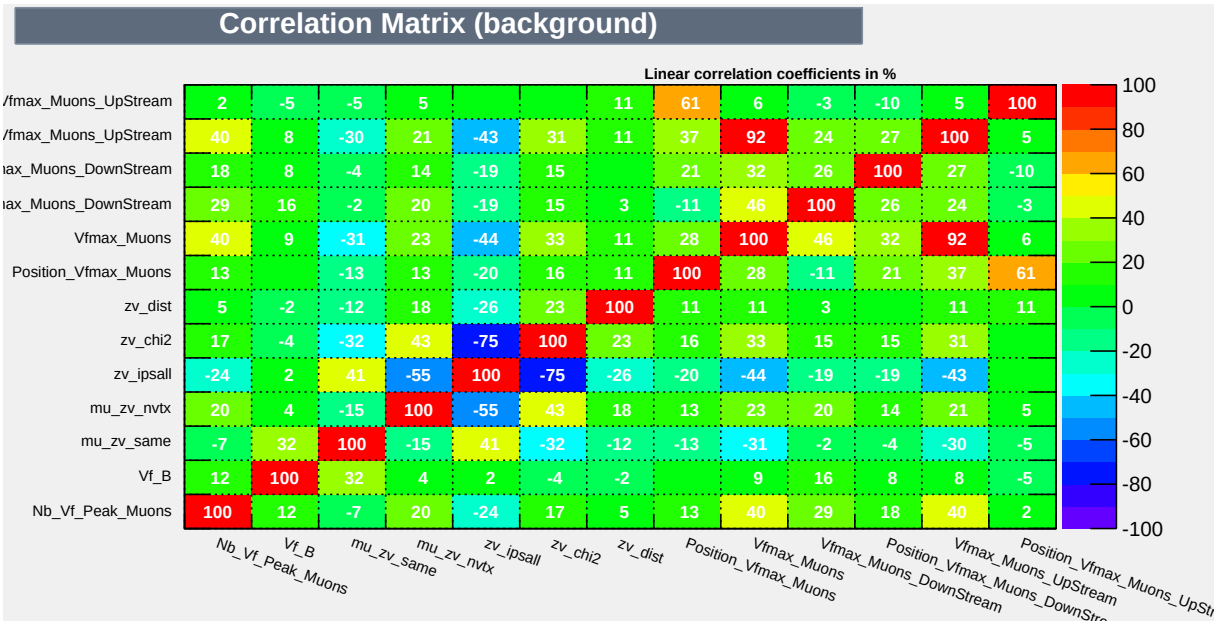


Figure 3.42: Linear correlation coefficients of the input variables of the ZViso for $b\bar{b} \to \mu\mu X$ MC generated background.

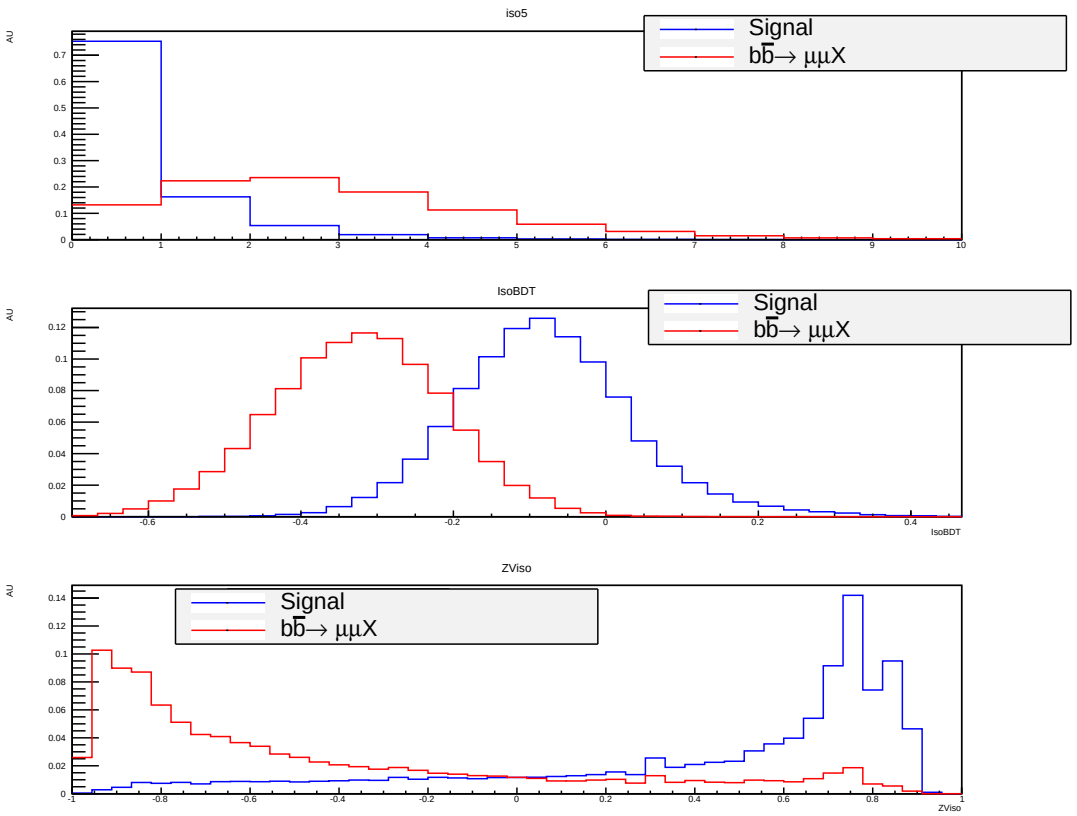


Figure 3.43: Distribution of the three isolation variables available for the $B^0_{(s)} \to \mu\mu$ analysis optimization: standard isolation (top), isoBDT (center), and ZViso (bottom). In blue is the signal, in red the $b\bar{b} \to \mu\mu X$ MC generated background.

BDT based isolation. The main idea of the isoBDT variable is to replace the cut based selection of non-isolating tracks presented in Appendix A by a BDT based selection. In particular, this new variable is defined by assigning to each track in the event the output of a BDT operator trained to discriminate the "non-isolating" tracks (considered as signal) against the "isolating" ones (considered as background) with respect to each of the two μ in the final state. In this way, two values (that are the output of the BDT classifier) are assigned to each track in the event and quantify how likely the considered track is non-isolating for the μ^+ and the μ^-. Then, the sum of the BDT values of the most likely non-isolating track with respect to the μ^+ and the μ^- is computed and used to quantify the isolation of the $B^0_s \to \mu^+\mu^-$ candidate. The BDT algorithm has been trained using the generic sample of $b\bar{b} \to \mu\mu X$ MC, by using as a proxy for signal and background the non-isolating and the isolating tracks for the $B^0_s \to \mu\mu$ candidate respectively.

The distributions of the three isolation variables are shown in Fig.3.43, while their linear correlations are shown in Fig.3.44 (for signal MC) and Fig.3.45 (for $b\bar{b} \to \mu\mu X$ MC generated background).

Isolations combinations. In order to exploit as much as possible the signal-background discrimination of these three variables, they have been combined into one unique variable through a BDT algorithm. The distribution of these variables is shown in Fig.3.46, while the comparison of the performances through the ROC curves are shown in Figs.3.47-3.48. By looking at the ROC curves, the following considerations are in order:

- the standard isolation ("iso5") and the new "ZViso" variable have similar performances;

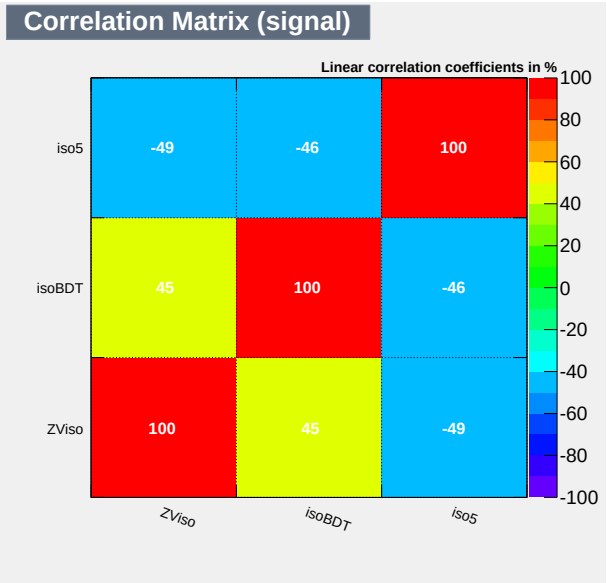


Figure 3.44: Linear correlation coefficients of the three isolation variables for signal MC.

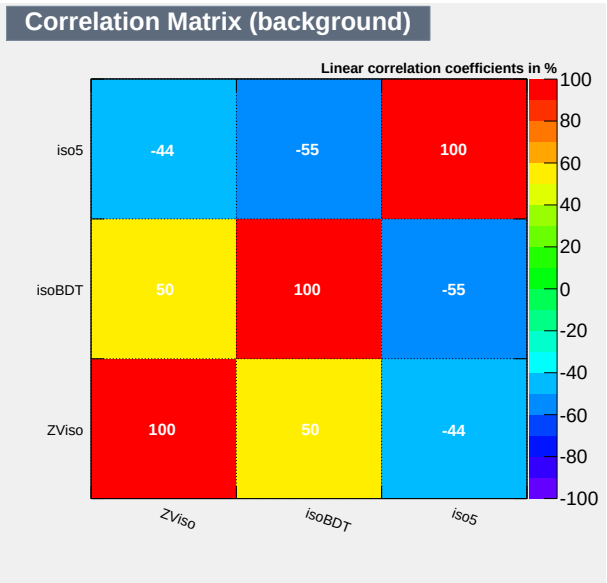


Figure 3.45: Linear correlation coefficients of the three isolation variables for $b\bar{b} \to \mu\mu X$ MC generated background.

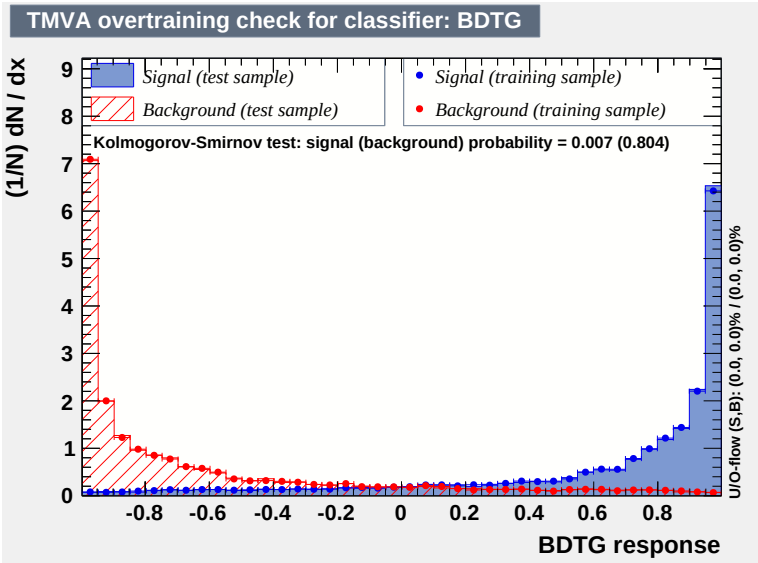


Figure 3.46: Distribution of the isolation variable obtained by combining, through a BDT algorithm, the standard isolation, the isoBDT, and the ZViso isolation variables. In blue is the signal, in red the $b\bar{b} \to \mu\mu X$ MC generated background.

- isoBDT variable performs better than the other two;
- the combination of the three is the most discriminating variable. This is a clear indication that these three variables are not fully correlated each other.

From these considerations it follows that, *a-priori*, a possible gain in performances of the final BDT classifier can be obtained by using the combination of them or all of them as input variables of the final classifier.

3.7.4 BDT classifier with new isolation variables

The effects of the previous BDT variables on the final BDT classifiers have been tested by training five BDT classifiers. Four of them use different sets of twelve input variables, differing only for the isolation variable used. The last one uses all the three isolations as separate input variables of the classifier. The comparison of the signal-background discrimination of these four algorithms are reported in Fig.3.49-Fig.3.50

These ROC curves show that the most discriminating BDT classifier is the one trained by replacing the standard muon isolation with the isoBDT one. In addition, no improvement is seen when using the combination of the three isolation variables into a unique one. The reasons of the saturation of the performances of the BDT classifier when using the isoBDT muon isolation variable can be due to the correlation of this variable with the other isolations.

To test that, the distribution of the values of the ZViso variable as a function of the isoBDT one has been studied for those combinatorial background events selected by different cuts on the output of the BDT trained with the isoBDT variable. The results are reported in Fig.3.51 and Tab.3.22. In the queue of the BDT distribution (especially the last four selected events) both the ZViso and the isoBDT variables are signal-like, being at high values of the ZViso and isoBDT variable. This correlation could justify why the use of the ZViso combined with the isoBDT does not improve the performances of the final classifier.

Another reason for the saturation of the performances can be due to the correlation of the ZViso variable with one of the remaining eleven input variables of the final BDT. In particular, being the ZViso sensitive to the overlap of the two muon tracks, it can be correlated with the

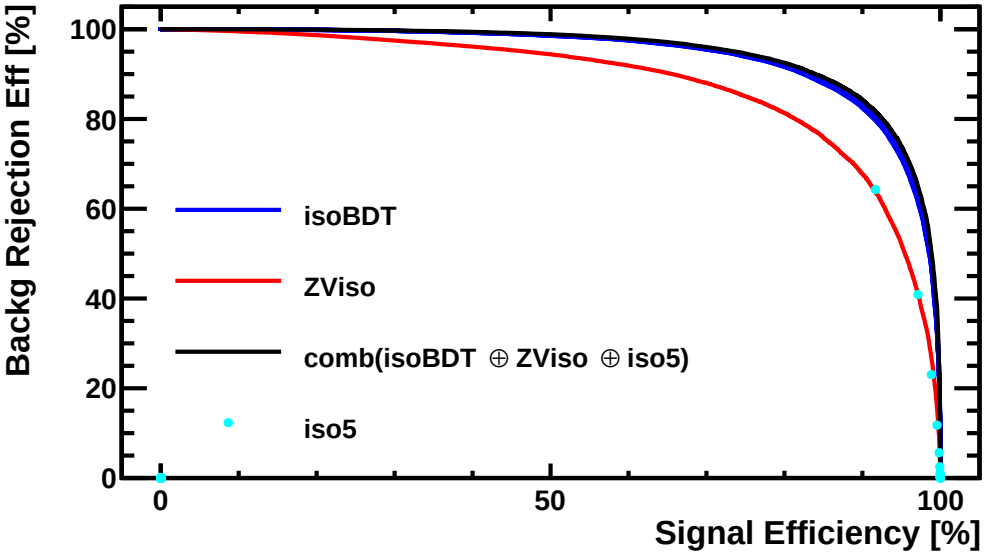


Figure 3.47: ROC curves of the four isolation variables. The standard isolation variable (here called "iso5") has isolated points being a discrete variable, while all others are continuous lines, being calculated from continuous distributions.

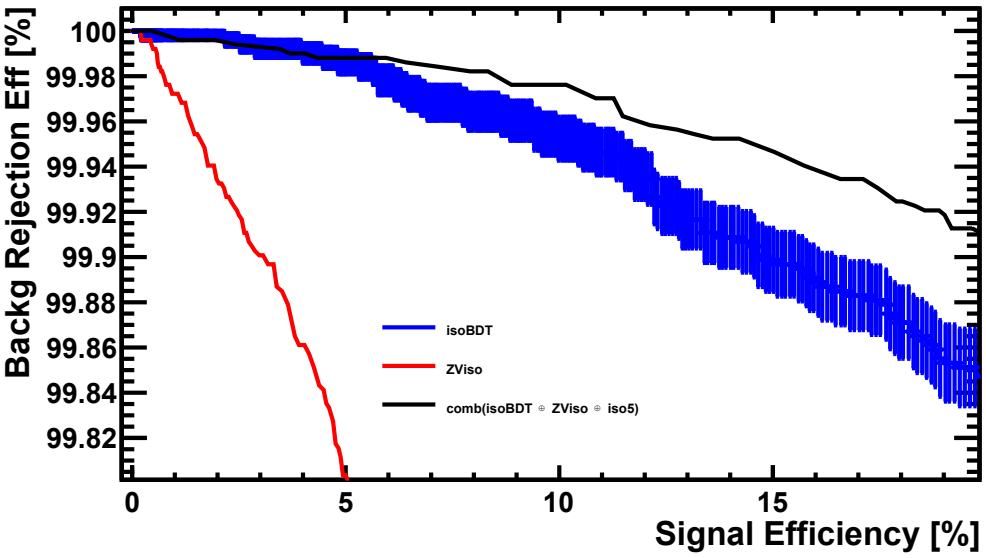


Figure 3.48: ROC curves of the four isolation variables. Zoom in the low signal efficiency region.

BDTD	# bkg evts	ϵ_{bkg} ($\times 10^4$)	ϵ_{sig} (%)
0.30	15+8+ 3+ 4	6	49.8
0.33	8+ 3+ 4	3	44.4
0.40	3+ 4	1.4	32.2
0.43	4	0.8	27.2

Table 3.22: Number of selected background events, the efficiency on background and on signal for different cuts on the output of the BDT classifier trained with the isoBDT muon isolation variable.

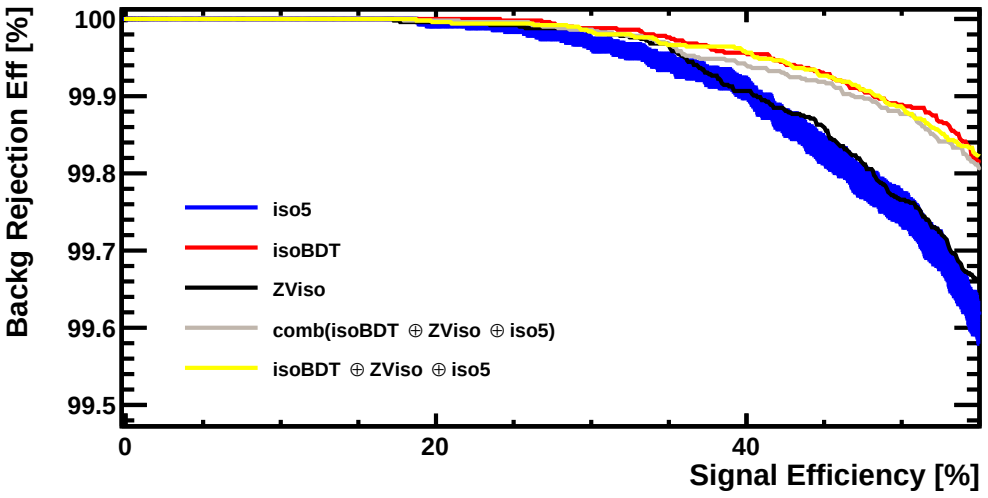


Figure 3.49: ROC curves of the four final BDT classifiers trained with different muon isolation variables. "comb(isoBDT⊕ZViso⊕iso5)" is the ROC curve of the classifier using the combination of all isolations, "isoBDT⊕ZViso⊕iso5" is the ROC curve of the classifier using the three isolations as separate input of the BDT classifier.

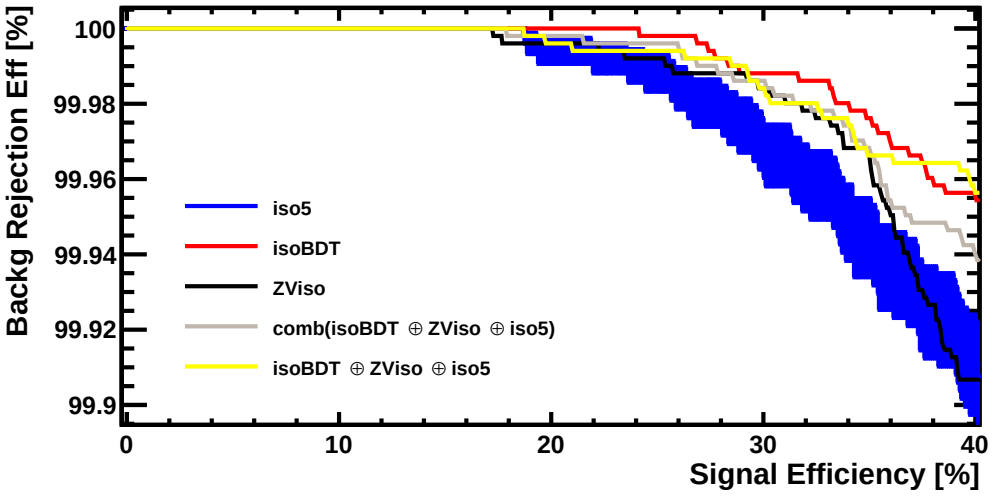


Figure 3.50: ROC curves of the four final BDT classifiers trained with different muon isolation variables. "comb(isoBDT⊕ZViso⊕iso5)" is the ROC curve of the classifier using the combination of all isolations, "isoBDT⊕ZViso⊕iso5" is the ROC curve of the classifier using the three isolations as separate input of the BDT classifier. Zoom in the low signal efficiency region.

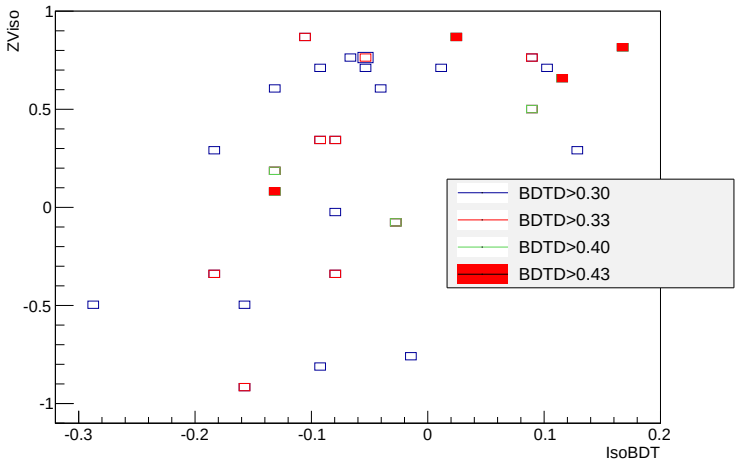


Figure 3.51: Values of the ZViso isolation variable as a function of the ones of the isoBDT for those events removed by different cuts on the output of the final BDT trained with the isoBDT muon isolation variable. For the color convention see Tab.3.22.

distance of closest approach of the two muons ($DOCA(\mu)$). This is actually what happens, as Fig.3.52 shows. This is not the case for the isoBDT variable though, as Fig.3.53 shows.

These two correlations of the ZViso with the isoBDT and the $DOCA(\mu)$ variables are the most likely reasons of the saturation of the performances of the final BDT classifier trained with the isoBDT muon isolation variable.

3.7.5 Study of the background composition of events falling in last BDT bin

In order to further improve the signal-background discrimination of the final BDT classifier, the particular composition (*i.e.* in terms of the physical decays) of the combinatorial background $b\bar{b} \to \mu\mu X$ MC generated events which fall in the very high BDT region (of the BDT used for the latest published analysis) has been studied. In particular the ten events with the higher BDT value have been examined. The following kind of events (in order of importance) are found to be the most signal like:

- 6 $B \to D(\to \mu\nu)\mu\nu$,
- 2 $B^0 \to K^0\mu\mu$,
- 1 $B^+ \to K^+\mu\mu$,
- 1 non MC truth matched event.

A first comment concerns the fact that all these events are generated by the decay of only one of the two B mesons in the event; in principle, indeed, there could also be "purely" combinatorial candidates reconstructed from the final state particles of the decays of the two b-hadrons produced in the event. In addition, the first two categories, which are the most abundant, have the common feature that the final state is characterized by the presence of the two muon tracks and several neutral unreconstructed particles. For this reason, the values of their isolation variables are signal-like and this explains why the improvement in the discrimination power of the isolation variables is not reflected in an improvement of the discriminating power of the final classifier.

The study of the combinatorial background composition of events at high BDT value, still gives some hints concerning the introduction of new input variables for the BDT classifier. In

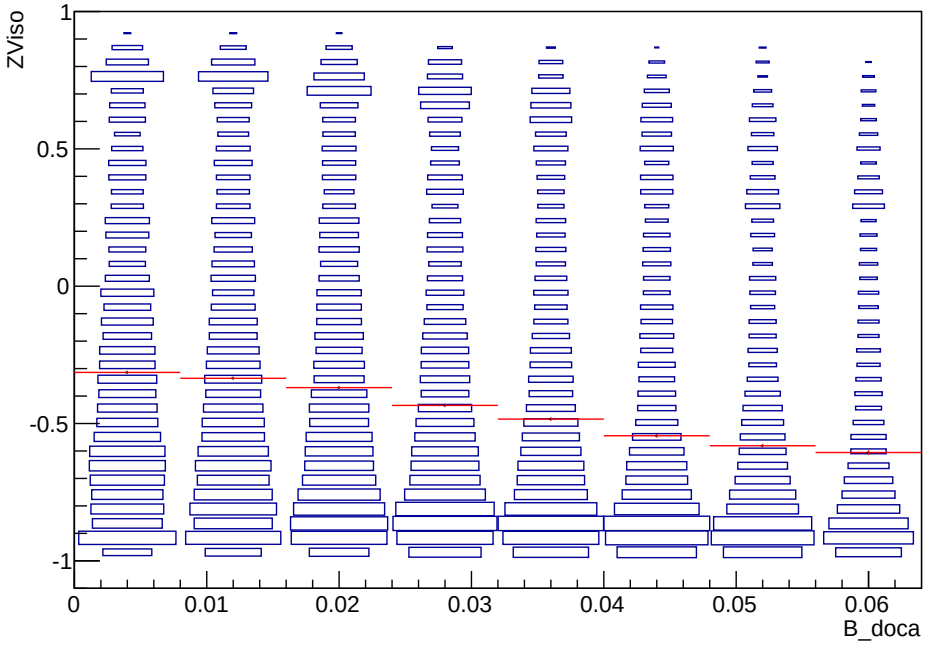


Figure 3.52: ZViso as a function of $DOCA(\mu)$ ("B_doca" in the plot) for combinatorial MC background events. In red is shown the average value of the ZViso variable as a function of $DOCA(\mu)$.

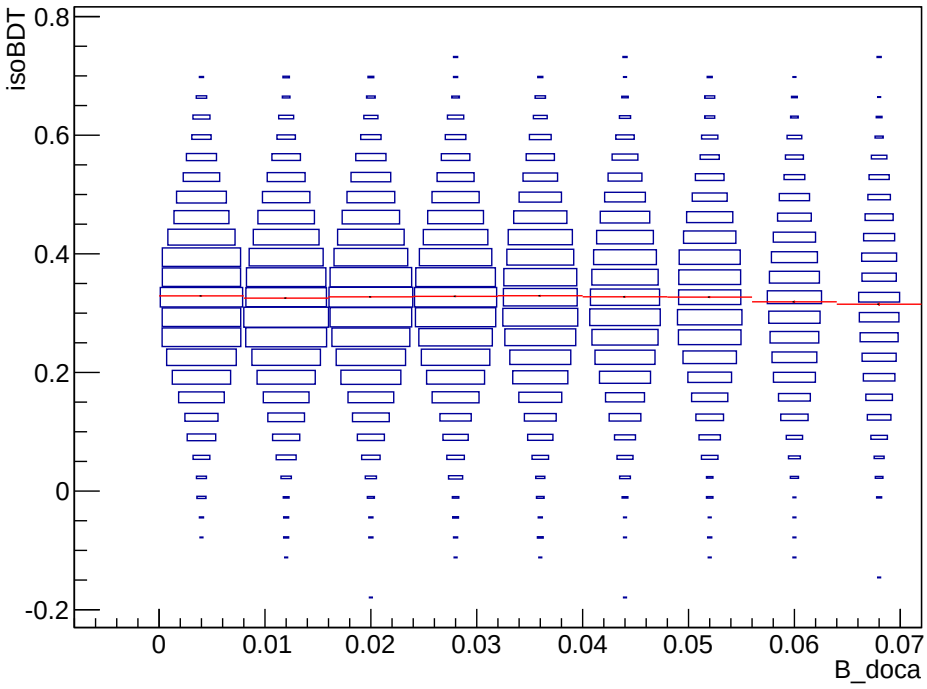


Figure 3.53: isoBDT as a function of $DOCA(\mu)$ ("B_doca" in the plot) for combinatorial MC background events. In red is shown the average value of the isoBDT variable as a function of $DOCA(\mu)$.

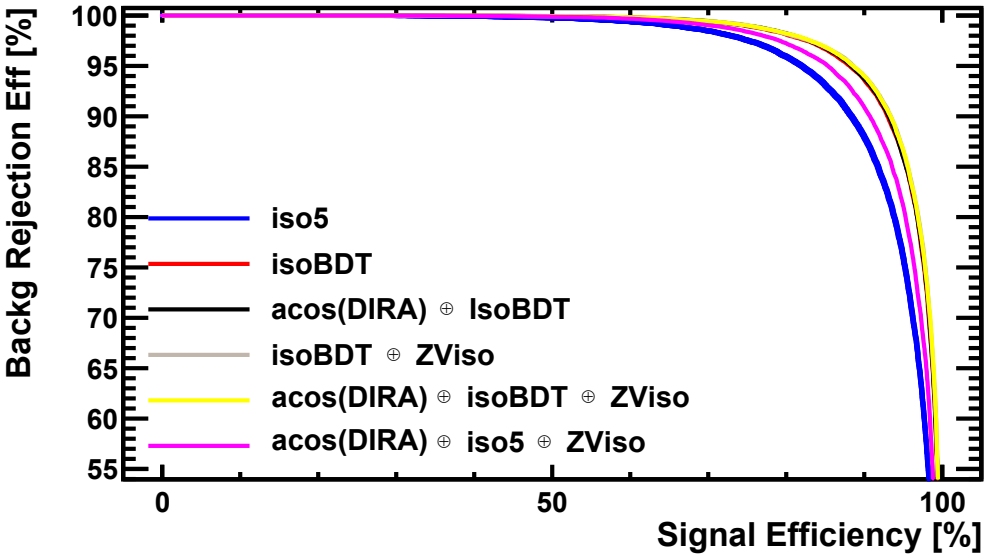


Figure 3.54: ROC curves of BDT classifiers trained using the basic set of 11 input variables plus the combinations shown in the plot. "iso5" is the standard isolation variable, "acos(DIRA)" is one of the two pointing related variable.

particular, the presence of neutral tracks in the final state will make the dimuon system total momentum not aligned with the B candidate momentum. This property can be exploited using the pointing-related variables. In the BDT with twelve variables the pointing was taken into account by the impact parameter of the B candidate. Now also the arccosine of the angle between the direction of the di-muon system momentum and the direction formed by the primary vertex and the B candidate decay vertex has been studied. This variable, here called "acos(DIRA)", has already been used in the first BDT for the selection.

3.7.6 Pointing related variables

The effects of this new input variable (when used together with the above defined isolation variables) on the performances of the classification BDT have been tested by training several BDT operators using the standard set of eleven variables plus several combinations of the isolation and pointing variables. Figs.3.54-3.55 show the comparisons of the signal-background discrimination performances of these classifiers.

From the ROC curves of these classifiers, the following conclusions can be inferred:

- the use of the ZViso and the standard muon isolation variables together with the pointing related variable improves the performances of the final BDT classifier;
- the BDT classifier trained with the isoBDT variable is still the best performant even when the pointing related variable is used.

A final attempt to improve the performances of the BDT classifier has been done by replacing the impact parameter of the B candidate with respect to the candidate origin vertex ("IP(B)") with the IP$\chi^2(B)$ variable. The ROC curves showing the performances of classifiers trained with these input variables are shown in Figs.3.56-3.57. The improvement of this classifier with respect to the one used for the latest published result (which uses the standard isolation variable and the IP(B)), and to the operator trained by replacing the standard muon isolation with the "isoBDT" one is reported in Tab.3.23.

From these ROC curves the following conclusions can be obtained:

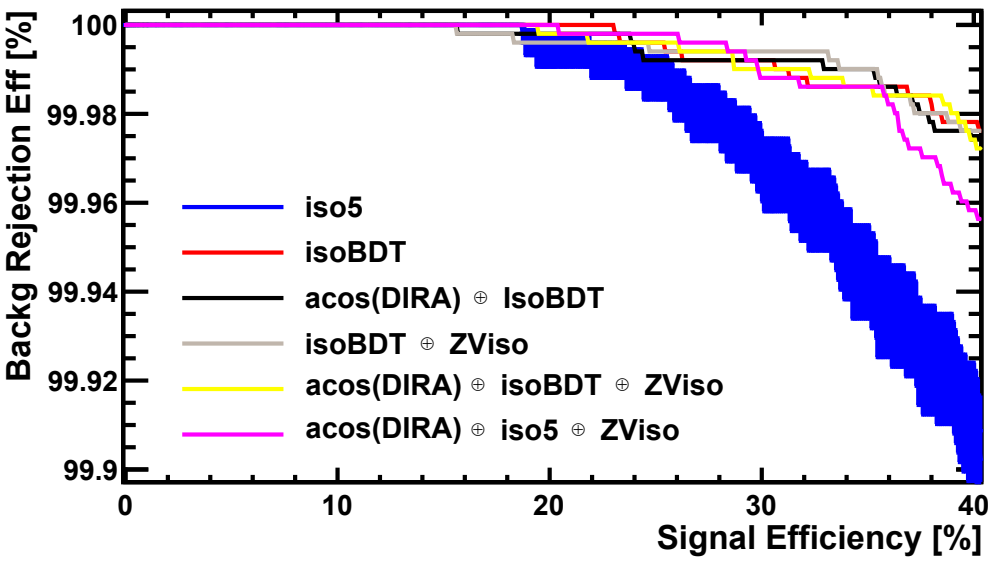


Figure 3.55: ROC curves of BDT classifiers trained using the basic set of 11 input variables plus the combinations shown in the plot. "iso5" is the standard isolation variable, "acos(DIRA)" is one of the two pointing related variables. Zoom in the low signal efficiency region.

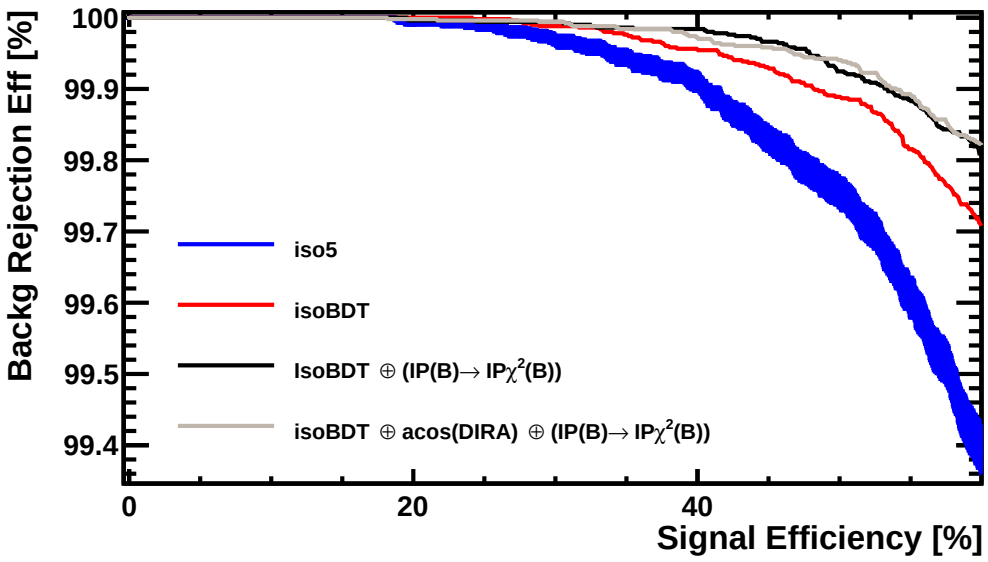


Figure 3.56: ROC curves of BDT classifiers trained using the basic set of 11 input variables plus the combinations shown in the plot. "iso5" is the standard isolation variable, "acos(DIRA)" is one of the two pointing related variables.

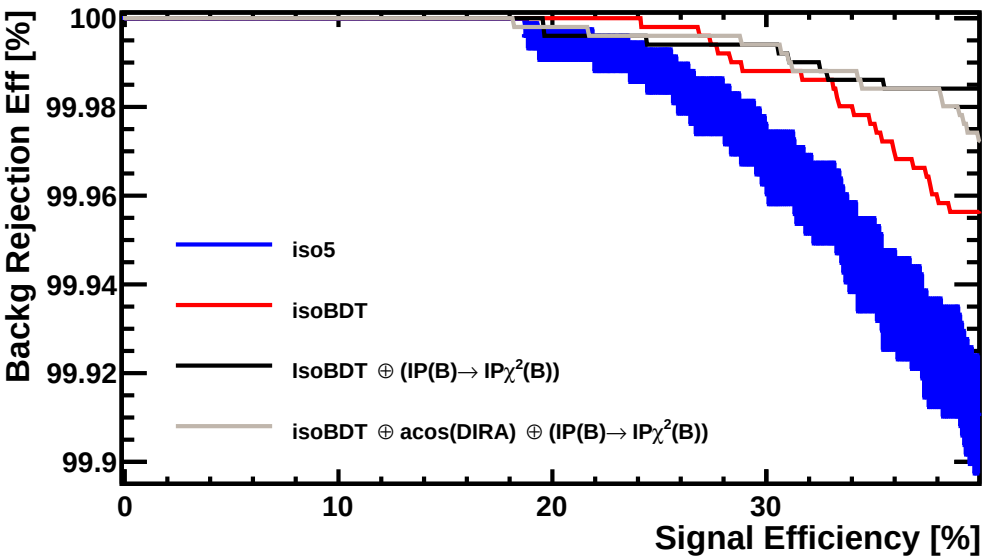


Figure 3.57: ROC curves of BDT classifiers trained using the basic set of 11 input variables plus the combinations shown in the plot. "iso5" is the standard isolation variable, "acos(DIRA)" is one of the two pointing related variables. Zoom in the low signal efficiency region.

ϵ_{Sig}	$\frac{\epsilon_{Bkg}(iso5)-\epsilon^{\star}_{Bkg}}{\epsilon_{Bkg}(iso5)}$	$\frac{\epsilon_{Bkg}(isoBDT)-\epsilon^{\star}_{Bkg}}{\epsilon_{Bkg}(isoBDT)}$
60 %	(69 ± 3)%	(40± 6)%
40 %	(83 ± 10)%	(66 ± 30)%

Table 3.23: Improvement in the background rejection of the classifier using the "isoBDT" and the "IP$\chi^2(B)$" input variables for different signal efficiencies (first column) with respect to the classifier used for the latest publication (second column), and to the classifier using only the "isoBDT" isolation variable (third column). ϵ_{Sig} is the efficiency on the signal, $\epsilon_{Bkg}(iso5, isoBDT)$ are the efficiencies on the background of the two classifiers using the two different isolation variables, $\epsilon^{\star}_{Bkg}$ is the efficiency on the background of the classifier using the "isoBDT" isolation and the "IP$\chi^2(B)$" input variables

- the replacement of the "IP(B)" with "IP$\chi^2(B)$" improves the performances of the BDT classifier, even of the one trained with the isoBDT isolation variable. At a signal efficiency of 40% the improvement in the background rejection is of 80% if compared with the classifier used for the latest analysis, and of 65% if compared with the improved version of the BDT which uses the isoBDT variable;

- the pointing related variable "acos(DIRA)" does not seems to improve the performances, even when used in the training of a classifier with the above mentioned replacement of the IP(B) variable.

Effects of data-MC agreement. All the studies shown in this section have been done using the generic $b\bar{b} \to \mu\mu X$ MC generated sample. The results shown could still depend on how well the MC reproduces the variables on data. For instance, the BDT used to classify a track as being isolating or non-isolating has been trained using the aforementioned MC sample. Now, it is known that the distribution of some of the variables related to the tracks (that are shown in Appendix B) are not well reproduced by the MC simulation; for that reason it can happen that on the true data a non-isolating track can be mis-classified as an isolating one (and *vice versa*), thus causing a worsening of performances both of the isoBDT and of the final BDT classifier variables. In order to reduce the effect of this data-MC disagreement the shape of the final BDT variable is taken from control samples in data, both for signal and for combinatorial background, as already explained in Sec.3.3. In order to further reduce the impact of the disagreement between the variables in MC and in data of the isoBDT input variables, the use of a control sample in data as a proxy for the combinatorial background in the training of the final BDT classifier is also planned to be considered for the analysis update.

3.7.7 Conclusions

The studies presented in this this chapter aimed to improve the signal background separation of the BDT classifier used for the signal search. Some hints for a further improvement of the signal-background separation of this variable have been obtained and can be summarized in the following points:

- the new isolation variable isoBDT improves the background rejection of the final BDT classifier;

- this rejection is further improved if the "IP(B)" variable is replaced with the "IP$\chi^2(B)$".

Some of the presented results still need to be investigated in more details. In particular further studies are required to fully exploit the information contained in the μ isolation variables presented in this chapter. Indeed, in principle, by looking at the combination of the three isolations (see *e.g.* Figs.3.47-3.48) an improvement of the signal-background separation of the final BDT classifier is expected when the "isoBDT" and the "ZViso" (or even their combination) are used as input. Nevertheless this is not observed and more studies are needed to better understand why this does not happen. One possible reason of this behavior can be in the correlation between the ZViso variable and the DOCA(μ), so that a part of the additional information contained in the ZViso variable is already contained by some of the other input variables of the final BDT. Likely an optimized choice of the input variables of the ZViso, as well as an optimization of the tuning parameters, would allow to reduce its correlation with the remaining input variables of the final BDT classifier and improve at the same time the signal-background separation of the classifier.

3.8 Conclusions

In this chapter the $B^0_{(s)} \to \mu\mu$ search with the $3fb^{-1}$ dataset collected in proton-proton collisions with the LHCb detector during the first run of the LHC has been presented. One of the main improvements of this analysis comes from the optimization of the BDT classifier used, together with the invariant mass of the two muons, as discriminating variables for the signal search. In particular, the optimization of the tuning parameters and the set of input variables of the BDT classifier has been done taking care of avoiding any peaking correlation between the output of the algorithm and the invariant mass of the two muons. Such a correlation would indeed create a false peak in the signal mass windows due to background. The optimization of the performances of the BDT classifier has been estimated to be the major source of improvement of the published LHCb analysis with respect to the previous ones [87].

The LHCb dataset used for the analysis presented here has also been combined with the one from the CMS collaboration. A common fit of the two datasets obtained the first observation of the $B^0_s \to \mu^+\mu^-$ and the first evidence for the $B^0 \to \mu^+\mu^-$ decays. The measured branching fraction for the two modes are in quite good agreement with the SM, and rule out huge NP effects. Nevertheless, the central value of the B^0 mode deviates a bit from the SM expectation, being slightly larger than expected ($2.2\sigma$from the SM prediction). The combined analysis has also measured the ratio $\mathcal{R} = \frac{\mathcal{BR}(B^0 \to \mu^+\mu^-)}{\mathcal{BR}(B^0_s \to \mu^+\mu^-)}$ which turns to be consistent with the SM predicted value within 2.3 standard deviations.

In view of a future analysis improvements, aiming to improve the sensitivity for the B^0 mode, new isolation variables for the muons have been defined and the improvement of the performances of the final BDT classifier trained with those variables has been tested. The variable presented in this chapter exploits an inclusive topological vertex reconstruction algorithm, called "ZVtop". A combination of these new variables into a unique more discriminating isolation has been presented. To better understand the composition of the most signal-like background events and have hints about further input variables for the final BDT classifier, a study of the physical decays composition of the combinatorial background falling in the high BDT region has been done using a sample of generic $b\bar{b} \to \mu\mu X$ MC generated events. The particular composition of these events, consisting of exclusive modes with only two muons and additional neutral particles in the final state, suggests to give more importance to variables related to the pointing of the B candidate. This has been done by introducing a new variable, the cosine between the B candidate momentum and the direction given by the PV and its decay vertex, and by replacing the IP of the B candidate (already used for the previous analysis) with the IPχ^2 variable. While the use of the first variable does not improve the signal-background discrimination the second choice does give a significant improvement in the background rejection. In particular, at a signal efficiency of 40%, an improvement in the background rejection up to 80% with respect to the classifier used for the latest published analysis seems possible using new muon isolation and pointing variables.

Chapter 4

$B^0_{(s)} \to \tau^+\tau^-$

In the present chapter a preliminary analysis strategy for the $B^0_{(s)} \to \tau^+\tau^-$ search is presented.

The main challenge of any experimental analysis studying decays containing τ leptons in the final state lies on the fact that τ cannot be directly detected and must be then reconstructed from their daughter particles. This is not a problem in itself if each τ daughter could be directly detected or indirectly reconstructed. The problem arises because of the presence of at least a ν_τ (and possibly other neutral particles) for each τ decay. Since only charged tracks are precisely detectable, the reconstruction of the τ momenta and of the whole kinematic of the decay, as well as the invariant mass $m_{\tau\tau}$ of the di-τ system, is quite challenging. Fig.4.1 shows what happens in a typical $B^0_{(s)} \to \tau^+\tau^-$ event.

For the above-mentioned reason, the analysis of the di-τ final state is quite different from the one with two muons, presented in the previous chapter. Indeed, while the μ travel all along the detector and can then be directly detected, the τ must be reconstructed from their daughter particles.

The lack of information coming from the partially reconstructed final state leads to a "degradation" of the discriminating power of each kind of variables related to the topology and kinematic of the decay, (for instance the $m_{\tau\tau}$ or the $B^0_{(s)}$ candidate decay time), and thus in a loss of sensitivity.

Nevertheless, in the final state where each of the two τ decay into two three charged pions (plus the ν_τ), it is possible, by imposing topological and mass constraints (as it will be shown in Sec.4.5), to fully reconstruct the two τ four-momenta.

In principle the di-τ final state offers the possibility to access the lepton's polarization as well by looking at the angular distributions of their daughter particles (which is not possible in the di-μ case). Nevertheless, some preliminary estimation (see Sec.4.4) gives an expected number of observed signal events (in the entire 3.1 fb^{-1} RunI dataset) smaller than one. Hence the possibility of performing an angular analysis is not realistic and the only measurable observable is (an Upper Limit on) the $\mathcal{BR}$.

The only available result so far is from the *BABAR* collaboration, and consists of an upper limit on the $\mathcal{BR}(B^0 \to \tau^+\tau^-)$ using a sample of $(232 \pm 3) \times 10^6$ $\Upsilon(4S) \to B\bar{B}$ events. The B^0_s di-τ decays have never been studied so far. As it will be shown in Sec.4.6.2, the signal search in the analysis presented here will be based on a discriminating variable not able to discriminate between the B^0 and the B^0_s modes. For this reason the measured $\mathcal{UL}$ will be on a combination of the $\mathcal{BR}$ of the two modes.

4.1 $B^0_{(s)} \to \tau^+\tau^-$ at LHCb

In the analysis performed by the *BABAR* collaboration, the two τ are reconstructed in the one prong channel with a charged (hadronic or leptonic) track.

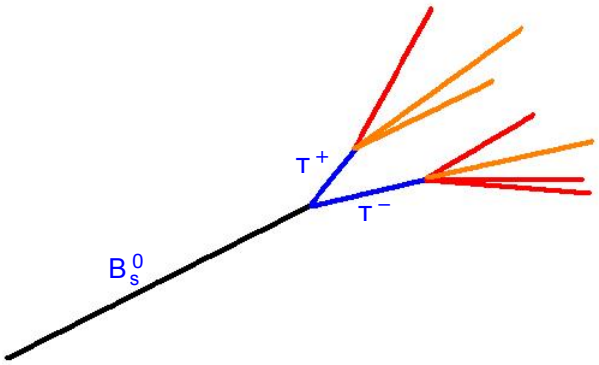


Figure 4.1: Schematic representation of a typical $B^0_{(s)} \to \tau^+\tau^-$ event: a $B^0_{(s)}$ meson is produced at the interaction point of the two proton beams. In the LHCb kinematic regime, it travels approximately for $\sim$ 1cm (black line), then it decays into two τ, which travel for $\sim$ 3mm (blue lines) before decaying into detectable (red lines) and neutral (orange lines) tracks.

At LHCb, the search for the $B^0_{(s)} \to \tau^+\tau^-$ channel is performed analyzing two different final states: the $(3\pi, 3\pi)$ and the $(3\pi, \mu)$ final state.

The "$(3\pi, 3\pi)$" final state. This corresponds to the case where each τ decays into a 3-prong vertex with 3 charged π and a ν_τ. In this final state there are only two ν, which is the minimal number in presence of two τ decays. The possibility to reconstruct the two τ decay vertices and the $B^0_{(s)}$ production vertex (PV), makes it possible to reconstruct the plane of the decay. This allows to impose kinematic constraints and to reconstruct the component of the momentum of each of the two ν orthogonal to the decay plane. In Fig.4.2 the reconstructible quantities of the $(3\pi, 3\pi)$ final state are schematically shown.

The $\mathcal{BR}$ of the three-prongs τ hadronic decay, suppresses the process by a factor [8] $(\mathcal{BR}(\tau \to 3\pi\nu))^2 = (9.32\%)^2$ with respect to the starting $\mathcal{BR}(B^0_{(s)} \to \tau^+\tau^-)$ reported in eq.(1.75) , ending up with an effective $\mathcal{BR}$ to be measured of

$$\mathcal{BR}(B^0_s \to \tau^+(\to (3\pi)^+\nu_{\bar{\tau}})\tau^-(\to (3\pi)^-\nu_\tau)) \simeq 6.7 \cdot 10^{-9}. \tag{4.1}$$

This $\mathcal{BR}$ is of the same order as the one of $B^0_s \to \mu^+\mu^-$.

However the requirement of six charged tracks in the detector acceptance as well as the lack of a dedicated trigger selection make the efficiency for the reconstruction and selection quite low (as will be shown later). The signal search is also complicated by the high level of background coming from different sources and related to the different steps of the candidate reconstruction algorithm, as will be explained in more details in Sec.4.5.

The "$(3\pi, \mu)$" final state. This corresponds to the case where one of the two τ undergoes an hadronic decay into 3 charged π (as in the previous case), while the other τ decays leptonically into a μ and two ν, *i.e.* $\tau^\pm \to \mu^\pm\nu_\tau\nu_\mu$. This channel benefits of a higher trigger and acceptance efficiency (with respect to the full hadronic final state), thanks to the requirements of ''only" four charged tracks (among which a μ) in the detector acceptance. In addition, the higher branching fraction of the purely leptonic decay of one τ ($\mathcal{BR}(\tau \to \mu\nu_\mu\nu_\tau = 17.41\%)$ enhances the total $\mathcal{BR}$ of the whole $B^0_{(s)} \to \tau^+(\to (3\pi)^+\nu_{\bar{\tau}})\tau^-(\to \mu\bar{\nu}_\mu\nu_\tau)$ decay chain with respect to the one of the $(3\pi, 3\pi)$ final state by a factor ~ 2. Indeed, the total branching fraction of the whole decay chain is:

$$\mathcal{BR}(B^0_s \to \tau^+(\to 3\pi^+\nu_{\bar{\tau}})\tau^-(\to \mu\nu_\mu\nu_\tau)) \simeq 1.25 \cdot 10^{-8}. \tag{4.2}$$

Figure 4.2: Schematic representation of how a $B^0_{(s)} \to \tau^+(\to (3\pi)^+\nu_{\bar{\tau}})\tau^-(\to (3\pi)^-\nu_\tau)$ decay is reconstructed with the LHCb detector: only the B production vertex and the two three-prongs τ decay vertices can be directly reconstructed.

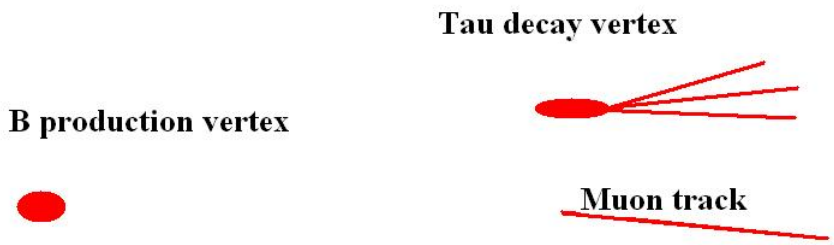


Figure 4.3: Schematic representation of how a $B^0_{(s)} \to \tau^+(\to (3\pi)^+\nu_{\bar{\tau}})\tau^-(\to \mu\bar{\nu}_\mu\nu_\tau)$ decay is reconstructed with the LHCb detector: in this case the B production vertex and the decay vertex of the τ which goes into three charged pions can be directly reconstructed, as well as the μ track coming from the leptonic decay of the other τ.

The disadvantage of this final state is that the presence of three neutrinos and the impossibility to reconstruct one of the two τ decay vertices make it impossible to impose kinematic and geometrical constraints. In Fig.4.3 the reconstructable quantities of the $(3\pi, \mu)$ final state are schematically shown.

4.1.1 The $\tau \to 3\pi\nu_\tau$ decay

The $\tau \to 3\pi\nu_\tau$ decay proceeds mainly through two hadronic resonances [111], the $a_1(1260)$ and the $\rho^0(770)$:

$$\tau^\pm \to a_1^\pm \bar{\nu}_\tau \to \rho^0\pi^\pm\bar{\nu}_\tau \to \pi^+\pi^-\pi^\pm\bar{\nu}_\tau. \tag{4.3}$$

In Tab.4.1 the properties of those two intermediate resonances are reported.

The existence of such intermediate resonances will be exploited during the selection of the candidates. Indeed the two dimensional distribution of the invariant masses $m_{\pi^+\pi^-}$ of the opposite sign pions from each τ decay (*i.e.* the Dalitz plane of the three pions) "peaks" at the

Resonance	Mass (MeV/c^2)	Γ (MeV/c^2)
$a_1(1260)$	1230 $\pm$ 40	[250-600]
$\rho^0(770)$	775.26 $\pm$ 0.25	149.1$\pm$0.8

Table 4.1: Values of the masses and widths (Γ) of the intermediate resonances for the $\tau \to 3\pi\nu_\tau$ decay [8].

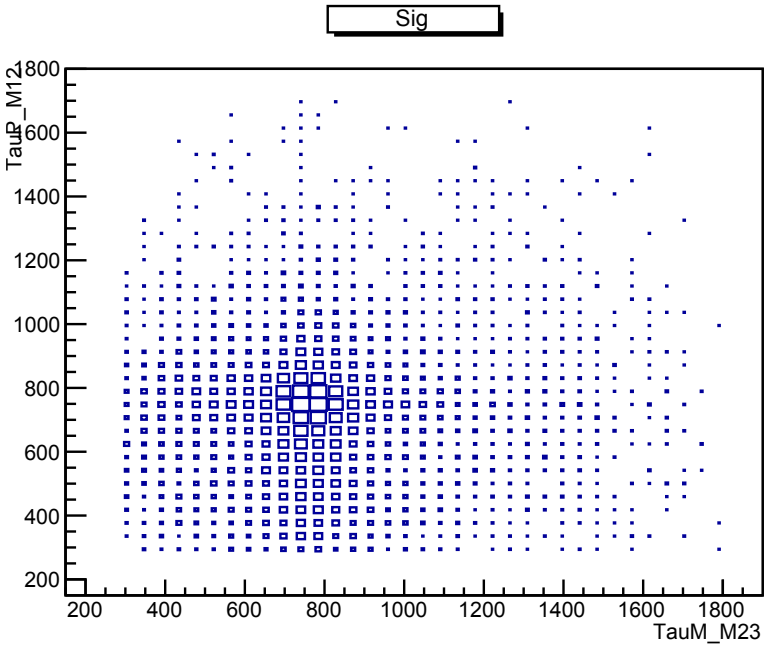


Figure 4.4: Dalitz plane of the $m_{\pi^+\pi^-}$ variables for pions coming from a $\tau \to 3\pi\nu$ decay.

values $m_{\pi^+\pi^-} = m_{\rho^0}$, while the invariant mass $m_{3\pi}$ of the 3π system "peaks" at the value of m_{a_1}. Due to the narrow width of the ρ^0 resonances the selection criteria based on the distribution of the Dalitz plane will be more efficient for τ selection with respect to selection criteria related to the invariant mass of the whole 3-π system, as it will be shown quantitatively in the following section.

The distribution of $m_{\pi^+\pi^-}$ for MC generated τ decay events is reported in Fig.4.4

4.1.2 The $(3\pi, 3\pi)$ final state analysis

In this chapter the analysis chain for the $(3\pi, 3\pi)$ final state is presented. The main issue is the understanding and characterization of the background sources polluting the sample where the signal search is performed. For this purpose, a data driven method has been defined and validated using background MC generated samples (see Sec.4.2.2).

Due to the constraints coming from the reconstruction of the decay plane, a method for the reconstruction of the two τ momenta has been studied, and the most discriminating variables obtained have been used (together with others) as input for the final MVA classifier used as the final discriminating variable for the signal search. At this stage of the analysis the knowledge of the shape of the MVA output for background events is required in order to extract the signal yield in the dataset through a fit to the output of the MVA classifier. The conversion of the signal yield into a value of the $\mathcal{BR}$ will be done using the normalization channel $B \to D^+ D^-_{(s)}$.

After an introduction about the reconstruction and preselection stages in Sec.4.2, the data driven method will be presented (in Sec.4.2.3) followed by the signal candidate selection in Sec.4.3 and the computation of the normalization factor in Sec.4.4. Then a method for a complete reconstruction of the signal will be presented in Sec.4.5, and the variable used for the signal search will be defined in Sec.4.6. Finally, two strategies currently under study for the calibration of the shape of the MVA output for background events will be presented in Sec.4.7.

4.2 Candidate reconstruction

$B^0_{(s)} \to \tau^+\tau^-$ candidates are reconstructed in events having at least six charged π tracks in the detector acceptance with a total net charge equal to zero. In particular, the reconstruction of the $B^0_s \to \tau\tau$ candidates proceeds in two steps:

- The τ candidates are built with three charged tracks identified as π coming from the same, well defined vertex (with a low χ^2 value), well displaced with respect to any other PV in the event and with an invariant mass of the 3π system ($m_{3\pi}$) compatible with the mass of the a_1 resonance. An approximate momentum, equal to the sum of the momenta of the 3π tracks, is associate to each τ candidate.

- The $B^0_{(s)} \to \tau\tau$ candidates are reconstructed using the two reconstructed τ candidate. In particular, their approximate momenta are used to reconstruct the (approximate) position of the $B^0_{(s)}$ candidate decay vertex. An approximate momentum, equal to the sum of the momenta of the six pions, is also associated to the B candidate.

Multiple candidates. The τ reconstruction algorithm looks for all possible combinations of three charged tracks forming a common vertex with a net charge $e = \pm 1$.

The topology of the selected $B^0_{(s)} \to \tau^+(\to 3\pi\bar{\nu})\tau^-(\to 3\pi\nu)$ candidates can give rise to the appearance of more than one reconstructed candidate per event (usually referred as "candidate multiplicity"). Indeed the opening angle of the decay triangle, defined by the B candidate production vertex and the two τ decay vertexes, is very narrow, hence the two τ decay vertices are usually almost collinear. Thus, a track originating at one vertex can be reconstructed as coming from the other one. For this reason, the same six tracks can be arranged in several ways in groups of three, thus making the (τ^+, τ^-) pair candidates multiplicity greater than one.

This candidate multiplicity is further increased if one or more extra charged tracks are close enough to the τ candidates' decay vertices. In this case, those tracks enter as well in the combinatorics, thus increasing the candidate multiplicity.

As it will be shown in the following, to deal with multiple candidates, only events in which only one candidate has been reconstructed are analyzed.

4.2.1 Preselection

A preselection of potentially interesting events is performed as part of the candidate reconstruction process. In particular, the following requirements are applied on tracks used to make the τ candidate:

- quality requirements: χ^2/ndf, ghost probability, and PID;

- geometrical requirements: on the impact parameter (with respect to any of the reconstructed primary vertex) expressed in units of its uncertainty $IP\chi^2$, in order to remove tracks coming directly from the collision point;

- kinematical requirements: applied on the momentum, p, and transverse momentum, p_T, of each track.

On the τ candidates the following cuts are applied:

- quality requirements: τ decay vertex χ^2;

- geometrical requirements: the Flight Distance length (FDρ), its projection along the beam axis (FDΔZ), and the FDχ^2 with respect to the approximated B decay vertex;

Variable	applied on	value
		$B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$
track χ^2/ndf	π	<3
ghost prob		< 0.3
IPχ^2		>4
p_T		> 0.25 GeV/c
p		> 2 GeV/c
ProbNNπ		>0.55
at least 1 π with p_T	τ/D	>0.8 GeV/c
DOCAMAX		<0.2 mm
Vχ^2		<16
M		[500-2000]MeV
FDχ^2		>16
FDρ		>0.1 mm
FDρ		<7 mm
FDΔZ		>5.0 mm
cos(DIRA)		>0.99
p_T		>1 GeV/c
p_T	$B_{(s)}$	> 2 GeV/c
M		[2-7] GeV/c
M_B^{corr}		<10 GeV/c
Vχ^2		<90
cos(DIRA)		>0.99
$FD\chi^2$		>225
FD		<90
max p_T of D/τ		>4 GeV/c
max IPχ^2 of D/τ		>150
min IPχ^2 of D/τ		>16
max(min IPχ^2 τ^+ ,min IPχ^2 τ^-)		>20
max p_T of K/π		>2 GeV/c
sum p_T of K/π		>7 GeV/c
max IPχ^2 of K/π		>16
Hlt2Topo		

Table 4.2: Selection criteria for $B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$; DOCA is the distance of closest approach between the two tracks, Vχ^2 is the χ^2 of the vertex, and DLL the combined PID likelihood to discriminate different particle hypotheses. The corrected mass of the B candidate M_B^{corr} is defined in sec.4.3.

- kinematical requirements: on the τ candidate approximate p_T, the p_T of its daughters and its invariant mass.

Finally on the B candidate the following requirements are applied:

- quality requirements on the B approximated vertex χ^2;
- geometrical requirements on the pointing angle, the B candidate approximated FD and FDχ^2, and on the IPχ^2 of the daughter particles;
- kinematical requirements on the B candidate approximate p, p_T, and invariant mass and transverse momentum of the daughter particles.

The complete list of selection criteria is reported in Tab.4.2.

In order to characterize the background (see Sec.4.2.2), also a sample of Same Sign (SS) events has been selected. This sample is obtained by reconstructing (unphysical) $B^{\pm\pm} \to \tau^\pm\tau^\pm$, *i.e.* by requiring two τ three prong vertices of the same electric charge and with exactly the same

Year	OS	SS (prescaled 50 %)
2011	4680	1470
2012	8550	2750

Table 4.3: Yields (per pb^{-1}) of reconstructed OS and SS $B^0_{(s)} \to \tau\tau$ events. In the current analysis only half the events selected in the SS combination are retained ("prescaled 50%").

kinematical and topological properties of the ones required to make the Opposite Sign (OS) combination.

The yields (per pb^{-1}) in data for OS and SS reconstructed events are reported in Tab.4.3. The rate of SS events is lower than the one of the OS events by a factor ~ 0.62. This can be motivated through combinatoric arguments. Indeed, using the same set of tracks, there are more combinations giving an OS than a SS candidate.

4.2.2 Background composition after preselection

Due to the $B^0_{(s)} \to \tau\tau$ candidate reconstruction algorithm, several processes can mimic the various stages of the decay chain, both for the $\tau \to 3\pi\nu_\tau$ and the $B^0_{(s)} \to \tau^+\tau^-$ decays.

To illustrate the various sources of backgrounds, the two steps of the candidate reconstruction algorithm will be followed.

$\tau \to 3\pi\nu_\tau$ fake events. The two sources of background which can mimic a $\tau \to 3\pi\nu_\tau$ decay are the following:

- three random π tracks making a common vertex, referred to in the following as "purely combinatoric" τ;

- $D^{(0,\pm)}_{(s)} \to 3\pi^\pm X$, $(X = \pi^0, ...)$ meson decays. These physical decays represent the most important source of background for the $\tau \to 3\pi\nu_\tau$ selection. Indeed $D^{(0,\pm)}_{(s)}$ mesons are the most abundant decay products of b-hadrons decays, because of the high $b \to cW^-$ transition rate. In addition, the behavior of the π^0 is very similar to the one of the ν_τ for the τ 3-prong decay. For this reason, even though the $D^{(0,\pm)}_{(s)}$ meson masses are slightly higher than that of the τ, the presence of the neutral π shifts the reconstructed invariant mass of the 3 charged pions towards lower values. In the following this kind of background will be referred to as "misidentified $D_{(s)}$".

The majority of these "purely combinatorial" τ are removed by the requirements on their vertex χ^2, on the 3π invariant mass, and on the Dalitz variables. On the other hand, the removal of the misidentified $D^{(0,\pm)}_{(s)} \to 3\pi X$ background events is more difficult. Indeed the requirements on the three π tracks common vertex χ^2 is not useful (being the three tracks originating from the same vertex). The criteria on the invariant masses of the 3π and the Dalitz variables remove most of the background from $D^\pm_s$ decays but not the one from the $D^{(0,\pm)}$, which proceed exactly through the same resonances as the $\tau \to 3\pi\nu_\tau$ decay. The "handle" to discriminate the background coming from misidentified D is the lifetime. Indeed, as shown in Tab.4.4, D meson have longer lifetimes with respect to the τ.

This is due to the fact that the s-quark in the D_s meson doesn't allow to form an a_1 and, subsequently, the ρ resonances. This is instead possible for the D meson, thanks to the presence of a first generation constituent quark.

	Mass (MeV/c^2)	τ (10^{-15} s)	Spin
τ	1776	290	1/2
D	1869	1040	0^-
D_s	1968	500	0^-

Table 4.4: Masses, lifetimes, and spins of the τ, D, and D_s mesons.

$B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$ fake events. Due to the high number of tracks and the abundance of $D_{(s)}$ mesons in the final state there are several processes that can fake the $B_s^0 \to \tau^+(3\pi)\tau^-(3\pi)$ signal. They can be categorized as follows:

- $B^0_{(s)} \to \tau\tau$ candidates made up of two purely combinatorial τ;
- genuine combinatorial background generated by two true τ coming from the semileptonic decay of the b-hadron in the event (as the combinatorial background for the $B^0_{(s)} \to \mu\mu$ mode),
- random combinations of τ and misidentified $D_{(s)}$ mesons coming from two different hadronic or semileptonic decays of the b-hadrons in the event;
- physical backgrounds generated by the semileptonic decay of a B meson either into a true τ and a misidentified $D_{(s)}$, or into two misidentified $D_{(s)}$ mesons.

Most of the background coming from purely combinatorial τ is removed by the requirements on the τ lifetime and the additional cuts on quantities related to the B reconstructed candidate. In the same way, all the events generated by combinatorics are mostly removed by the requirements on the B reconstructed candidate. The most difficult background to remove comes from by the exclusive decays of $B^{0,\pm}_{(s)}$ mesons. In this case the presence of τ and $D_{(s)}$ mesons coming from the same parent particle reduces the efficiencies of the selection on the B candidate vertex quality and of the kinematical requirements applied on masses.

4.2.3 Background characterization

The large variety of potential background sources makes a dedicated characterization using generic MC samples (as it has been done, for instance, for the $B^0_{(s)} \to \mu\mu$ analysis to characterize the combinatorial background) inefficient and unreliable. For this reason, a data driven method has been defined. The main handle to characterize the background both from combinatoric reconstruction and (as it will be shown in the following) physical decays is the above mentioned Same Sign (SS) sample.

The general composition of this sample is the following:

- purely combinatorial $B^0_{(s)} \to \tau\tau$, *i.e.* two true τ coming from two different semileptonic B meson decays
- two purely combinatorial reconstructed τ, *i.e.* six charged π tracks forming two three-prong vertices
- physical decays of $B^{0,\pm}_{(s)}$ with more than six charged tracks in the final state. Indeed the extra charged tracks can be used, while doing the combinations, to make a same sign candidate.

The fact that the SS sample can be used for the characterization of the background in the OS sample is not evident *a-priori* and a validation study of its composition in terms of physical decays must be performed.

$(3\pi, 3\pi)$	OS	SS (prescaled 50%)
evts	283	88
cand	505	165
evts matching an exclusive	241 (85%)	68 (77%)

Table 4.5: Number of selected events (first row), reconstructed candidates (second line), and (third row) selected events where a $B \to \tau\tau$ candidate has been reconstructed from tracks coming from the same parent particle (in parenthesis the fraction of events with respect to the total number of selected events) for the OS (left column) and SS (right column) combinations for the $(3\pi, 3\pi)$ final state. In the current analysis only half the events selected in the SS combination are retained ("prescaled 50%").

$(3\pi, \mu)$	OS	SS (prescaled 50%)
evts	320	55
cand	378	69
evts matching an exclusive	277 (87%)	42 (76%)

Table 4.6: Number of selected events (first row), reconstructed candidates (second line), and (third row) selected events where a $B \to \tau\tau$ candidate has been reconstructed from tracks coming from the same parent particle (in parenthesis the fraction of events with respect to the total number of selected events) for the OS (left column) and SS (right column) combinations for the $(3\pi, \mu)$ final state. In the current analysis only half the events selected in the SS combination are retained ("prescaled 50%").

The study of the composition of the SS sample is done using an inclusive sample of 4 millions MC generated $b\bar{b}$ events. This is a generic sample obtained by simulating the $b\bar{b}$ production, the hadronization of each of the two b-quarks and the subsequent b-hadron decay in any known channel with their respective $\mathcal{BR}$s.

In this sample, $B \to \tau\tau$ candidates have been reconstructed in both the OS ($\tau^+\tau^-$) and the SS ($\tau^\pm\tau^\pm$) combinations using the selection criteria presented in Sec.4.2.1. Then the selected events have been analyzed in term of the physical decays in each event, using the information from the MC truth.

The results of the reconstruction are reported in Tabs.4.5-4.6 for the $(3\pi, 3\pi)$ and the $(3\pi, \mu)$ final states respectively. The following considerations are in order:

- both the OS and the SS samples are populated mostly by physical background events, *i.e.* events where the six charged tracks used to reconstruct the $B \to \tau\tau$ candidate have a common parent particle,
- the fraction of the physical background candidates in the OS and the SS is similar,
- the $(3\pi, 3\pi)$ and the $(3\pi, \mu)$ final states have similar behaviors, as it is shown in Tab.4.6.

In addition, the fraction of exclusive decays in the SS and OS samples has been studied as a function of the number of charged tracks in their final states. The results for the $(3\pi, 3\pi)$ final state are reported in Tab.4.7. The composition of the SS sample is very similar to the OS except for channels with only six charged tracks in the final state. Indeed to have a SS combination at least seven charged tracks in the decay final state are needed.

The main hypothesis on which the analysis strategy is based is that the SS and OS sample have a similar composition in terms of background events and the differences are mostly due to the presence in the OS sample of exclusives channels with only six charged tracks in the

$n_{charged}^{tracks}$	$b\bar{b}$ OS (241 evts)	$b\bar{b}$ SS (68 evts)
6	34 (14%)	$\star$
7	60 (25%)	14 (21%)
8	88 (37%)	29 (42%)
9	32 (13%)	13 (19%)
10	19 (8%)	10 (15%)
11	4 (2%)	2 (3%)
12	4 (2%)	–

Table 4.7: Number of selected events where a $B \to \tau\tau$ candidate has been reconstructed from tracks coming from the same parent particle as a function of the number of charged tracks in the exclusive mode final state.

final state (which cannot be reconstructed in the SS combination). For this reason, the data SS sample can provide a reliable description of backgrounds (physical and combinatorial) in OS provided the final multivariate analysis reduces as much as possible the dependence on variables sensitive to the number of tracks (*e.g.* isolation and pointing variables).

4.2.4 Analysis strategy

Starting from the assumptions presented in the previous section, the signal search is organized as follows:

- after the preliminary selection, a tight selection is done to remove as much as possible exclusive background candidates with more than six charged tracks in the final state. To do that, the selection is tuned in order to select the $B_s^0 \to \tau^-\tau^+$ signal MC while rejecting the SS-like background reconstructed on data. The main goal of the tight selection (in addition to reduce the dataset at a manageable level) is to exploit as much as possible the discrimination of variables quite powerful to separate the signal against the SS-like background but which are, at the same time, too sensitive to the differences in background composition between the SS and the OS samples;

- train a BDT operator using variables that are not able to discriminate between the SS and OS samples, but are still useful to discriminate the $B_s^0 \to \tau^-\tau^+$ signal against the background contained in the SS sample;

- the signal yield (or its $\mathcal{UL}$) is extracted through a fit to the BDT output. The BDT PDF shape for signal events is taken from MC simulation, while the one for background is taken from the SS sample. Finally the contribution of the exclusives modes with only six charged tracks in the final state are added as a separate component in the fit.

4.2.5 Signal MC generated samples

The characterization of the $B_{(s)}^0 \to \tau^+\tau^-$ signal is done using MC generated samples. For the analysis presented here, the signal MC samples are

- $Sig_1 : B_s^0 \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$,

- $Sig_2 : B^0 \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$,

- $Sig_3 : B_s^0 \to \tau^+(\to 2\pi^+\pi^-\pi^0\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$.

The event selection presented in the next section has been tuned using only the $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ mode. Nevertheless, as it will be shown in Tab.4.10 the efficiencies for the first two modes are compatible with each other. Only the last mode will have a lower efficiency. Moreover, although the training of the final BDT used for the signal search is done to discriminate only the Sig_1 mode against the SS background sample, *a-posteriori* is seen that the output variable is equally distributed also for the other two modes Sig_2 and Sig_3.

4.3 Tight selection

As already mentioned, the preliminary selection described in Sec.4.2.1 is followed by a tight selection based on kinematic and isolation variables. The goal of this selection is to remove as many background events as possible, thus reducing the size of the dataset to a manageable level. The selection is tuned in order to select the signal MC while rejecting the SS-like events. The tight selection criteria are reported in Tab.4.8. Each of them helps fighting some particular background source:

- L0Hadron and Hlt1TrackAllL0 triggers;
- Dalitz plane and $m_{3\pi}$ of each τ: selection criteria based on these quantities aim to reject the events which does not proceed through the a_1 and the ρ^0 resonances;
- invariant mass of the six π in the final state ($M_{6\pi}$), of the $(3\pi)^\pm$ systems $M_{(3\pi)^\pm}$, the difference between the two $\Delta M_{B-\tau\tau} \equiv M_{6\pi} - M_{(3\pi)^+} - M_{(3\pi)^-}$, and the corrected mass of the B candidate M_B^{corr} defined as the invariant mass of the six π plus the sum of the modulus of the components of the neutrino momenta perpendicular to the decay plane. These requirements are meant to reject candidates generated by combinations of true τ and $D_{(s)}$ mesons coming from two different hadronic or semileptonic b-hadron decays;
- τ lifetimes, in order to reject unphysical candidates where the τ decay vertices are downstream with respect to the approximate reconstructed B decay vertex;
- isolations removing candidates where the 3 π tracks used to make the τ candidate come from a vertex with one or more extra charged tracks. Isolation variables can be classified into three categories:
  - candidate multiplicity (n_{cand}). This variable is the number of reconstructed $B^0_{(s)} \to \tau\tau$ candidates in the selected event. If more than six charged tracks are in the B meson final state or the two τ candidates decay vertices are very close to each other, it is likely that, due to the combinatoric search, more than one candidate per event be reconstructed; this requirement allows to select candidates where the two τ decay vertices are well separated each other;
  - vertex isolations variables. These variables are defined by the τ or $B^0_{(s)}$ candidates vertices ($\mathcal{V}_{\tau,B}$), adding to them, one by one, the other tracks in the events that are then fitted together into another vertex ($\mathcal{V}^\star_{\tau,B}$) . More precisely the following variables have been used:
    - "NumVtxWithinChi2WindowOneTrack": number of particles generating a vertex $\mathcal{V}^\star_{\tau,B}$ with a $\chi^2_{\mathcal{V}^\star_{\tau,B}} < 9$
    - "SmallestDeltaChi2MassOneTrack": invariant mass of the tracks coming from the vertex $\mathcal{V}^\star_{\tau,B}$, obtained from the one of the reconstructed τ or B candidates by adding the track giving the smallest $\Delta\chi^2$ defined as $\Delta\chi^2 \equiv \chi^2_{\mathcal{V}_{\tau,B}} - \chi^2_{\mathcal{V}^\star_{\tau,B}}$
    - "SmallestDeltaChi2MassTwoTracks": invariant mass of the tracks coming from the vertex $\mathcal{V}^{\star\star}_{\tau,B}$ obtained from $\mathcal{V}^\star_{\tau,B}$ defined above by adding the track giving the smallest $\Delta\chi^2$.

Variable	Selection
L0Hadron	true
Hlt1TrackAllL0	true
$m(\rho_{(1,2)})$	[591.6,900] MeV/c^2
$M_{3\pi}$	[1000,1350] MeV/c^2
$M_{6\pi}$	> 2400 MeV/c^2
$\Delta M_{B-\tau\tau} \equiv M_{6\pi} - m_{3\pi^+} - m_{3\pi^-}$	> 600 MeV/c^2
M_B^{corr}	> 3000 MeV/c^2
$\tau\chi^2_{B,\tau^\pm}$	> 0
$\tau_{\tau^\pm}$	$\tau_{\tau^\pm} > 0$
n_{cand}	$== 1$
B_NumVtxWithinChi2WindowOneTrack	< 1
B_SmallestDeltaChi2MassOneTrack	> 3200 MeV/c^2
Tau_NumVtxWithinChi2WindowOneTrack	≤ 3
Tau_SmallestDeltaChi2MassTwoTracks	> 2000 MeV/c^2
sum_iso_Taus	< 1
sum_iso_allPions	< 4
sum_isoBDT1_Taus_v1	< 7
sum_isoBDT1_Taus_v2	< 3
sum_isoBDT1_Taus_v3	< 3
sum_isoBDT2_Taus	$>$ -0.5
sum_isoBDT3_Taus	$>$ -0.5
sum_isoBDT1_allPions_v1	< 10
sum_isoBDT1_allPions_v2	< 6
sum_isoBDT1_allPions_v3	< 5
sum_isoBDT2_allPions	> -1
sum_isoBDT3_allPions	> -1

Table 4.8: List of trigger, kinematical, and candidate multiplicity cuts applied in the tight selection

– track isolations.

Fig.4.5 shows the Dalitz plane of the τ^- candidates for signal MC, data SS and OS events; Fig.4.6 the decay time distributions of the τ^+ candidates reconstructed in MC signal, data SS, and data OS samples. Figs.4.7-4.8 show the tracks isolation variables for the τ and π for signal MC and data SS.

The efficiency for each selection criteria for signal MC, data SS, and data OS events has been studied in details and the results are reported in Tab.4.9. The selection criteria based on Dalitz variables play a mayor role in removing background events, being a factor ten more efficient for signal events than for background. Moreover, the high suppression level of this selection (also for signal events) is due to the fact that the daughter products of both τ candidates must satisfy these requirements. The selection efficiencies for MC signal and exclusive modes are reported in the summary table 4.14.

Tab.4.11 reports the tight selection efficiencies ϵ_{sel} and the absolute yields of the selected MC generated signal events, data SS and OS, as well the selected number of SS and OS candidates in $b\bar{b}$ MC events. After the tight selection no events are left in the generic $b\bar{b}$ sample, both in the SS and in the OS combination. Hence, with the available statistics, it is not possible to characterize the background events passing the selection, as it has been done after the preliminary selection.

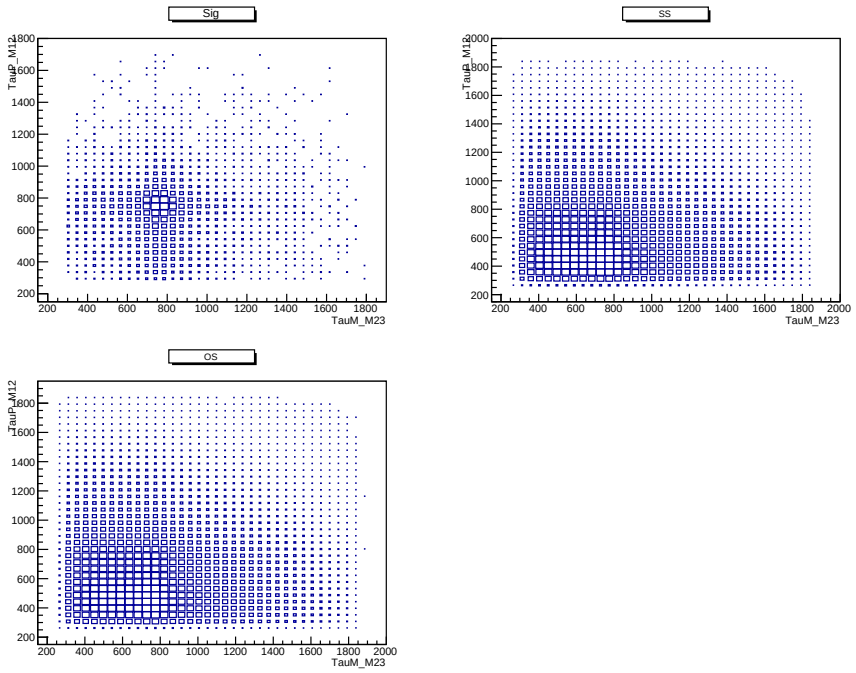


Figure 4.5: Dalitz plane of the τ^- candidates reconstructed in signal MC, data SS, and data OS samples.

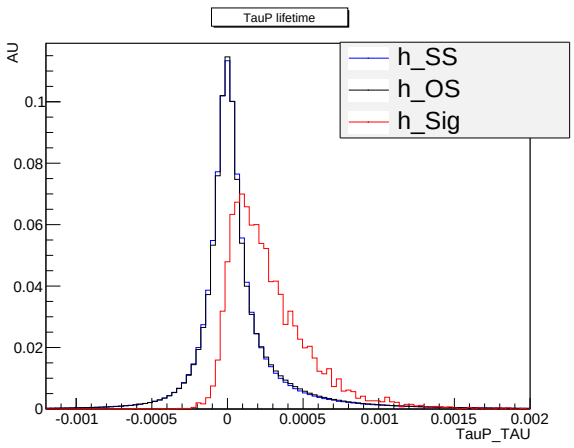


Figure 4.6: Distribution of the decay time for τ^+ candidates reconstructed in the signal MC, data SS, and data OS samples.

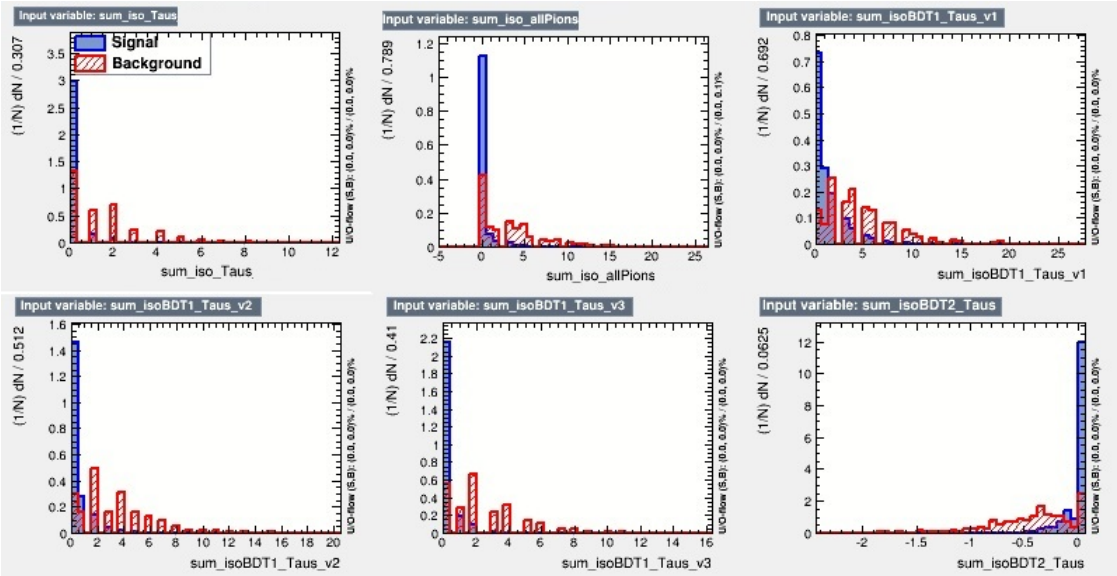


Figure 4.7: Distribution of isolation variables used in the selection for signal MC (blue), and data SS (red).

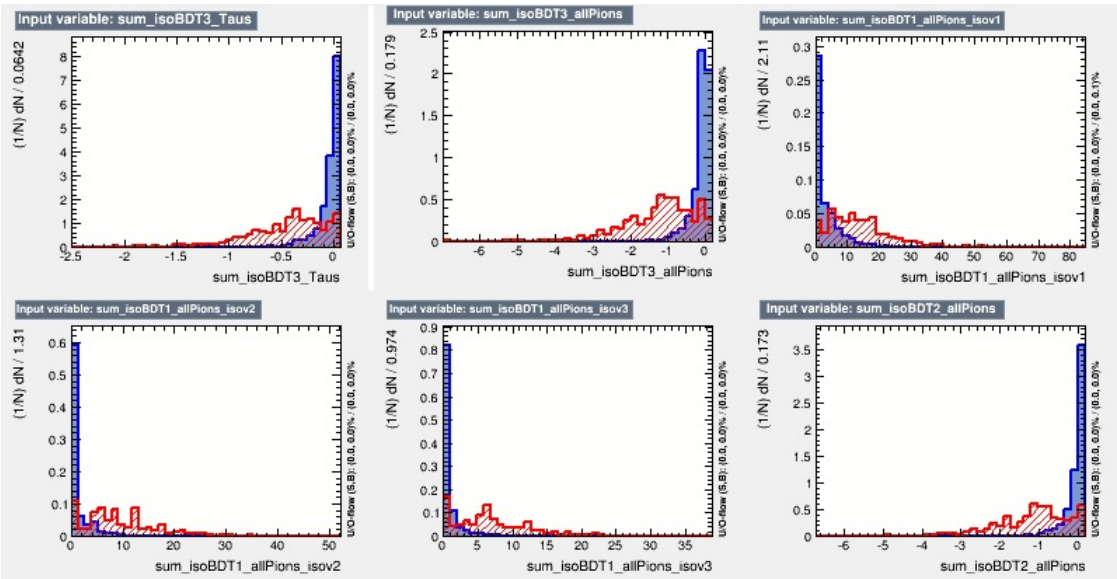


Figure 4.8: Distribution of isolation variables used in the selection for signal MC (blue), and data SS (red).

Selection efficiencies (in %)						$\frac{\epsilon_n}{\epsilon_{n-1}}$				
Selection	Sig_1	Sig_2	Sig_3	SS	OS	Sig_1	Sig_2	Sig_3	SS	OS
Trigger	87.78	87.55	86.12	81.18	80.52	—	—	—	—	—
$m(\rho_{(1,2)})$	17.60	19.18	4.53	1.27	1.51	0.20	0.19	0.04	0.015	0.02
$m_{3\pi}$	14.12	14.76	3.04	0.60	0.75	0.80	0.77	0.67	0.47	0.50
$M_{6\pi}$ & $M_{B-\tau\tau}$ & M_B^{corr}	14.08	14.72	3.01	0.27	0.40	0.99	0.99	0.98	0.45	0.53
$\tau\chi^2_{B,\tau^\pm}$	14.08	14.72	3.01	0.24	0.35	1.00	1.00	1.00	0.92	0.87
$\tau_{B,\tau^\pm}$	11.94	12.26	2.52	0.10	0.14	0.85	0.83	0.83	0.41	0.40
n_{cand}	11.81	12.08	2.47	0.03	0.06	0.99	0.98	0.97	0.30	0.43

Table 4.9: Efficiencies of the requirements on trigger, kinematic, lifetime, and candidate multiplicity variables when applied in sequence for signals MC (Sig_1: $B_s^0 \to \tau(\to 3\pi\nu_\tau)\tau(\to 3\pi\nu_\tau)$, Sig_2: $B^0 \to \tau(\to 3\pi\nu_\tau)\tau(\to 3\pi\nu_\tau)$, Sig_3: $B_s^0 \to \tau(\to 3\pi\pi^0\nu_\tau)\tau(\to 3\pi\nu_\tau)$), data SS, and data OS candidates. n_{cand} is the total number of candidates reconstructed in each event. In the right part of the table the ratio of the efficiencies of two subsequent selections are reported. The efficiencies of the selection criteria based on isolation variables are not included.

Event	$\mathcal{BR}$ ($\times 10^9$)	ϵ_{gen} ($\times 10^2$)	$\epsilon_{pre-sel}$ ($\times 10^2$)	ϵ_{sel} ($\times 10^2$)	ϵ_{tot} ($\times 10^6$)	ε ($\times 10^{11}$)
Sig_1	6.70 ± 0.50	2.85	0.70 ± 0.04	9.80 ± 0.26	19.55 ± 1.53	$(1.30 \pm 0.23) \cdot 10^{-2}$
Sig_2	$(1.92 \pm 0.19) \cdot 10^{-1}$	3.12	0.71 ± 0.06	8.47 ± 0.42	19.00 ± 2.50	$(1.40 \pm 0.34) \cdot 10^{-3}$
Sig_3	3.32 ± 0.28	2.58	0.64 ± 0.06	1.57 ± 0.24	4.14 ± 0.78	$(1.37 \pm 0.37) \cdot 10^{-3}$

Table 4.10: Branching fraction, generator, pre-selection, and selection efficiencies, for the three generated MC signal samples (Sig_1: $B_s^0 \to \tau(\to 3\pi\nu_\tau)\tau(\to 3\pi\nu_\tau)$, Sig_2: $B^0 \to \tau(\to 3\pi\nu_\tau)\tau(\to 3\pi\nu_\tau)$, Sig_3: $B_s^0 \to \tau(\to 3\pi\pi^0\nu_\tau)\tau(\to 3\pi\nu_\tau)$). $\varepsilon \equiv \mathcal{BR} \cdot \epsilon_{tot} \cdot f_q/f_s$, $\epsilon_{tot} \equiv \epsilon_{gen} \cdot \epsilon_{strip} \cdot \epsilon_{sel}$; $f_d/f_s = 3.86 \pm 0.06$ [99, 100].

Type	cand after preselection	cand after selection	ϵ_{sel}
Sig_1	13960	1344	9.61%
Sig_2	4934	418	8.47 %
Sig_3	6665	105	1.57 %
$b\bar{b}$ MC SS	165	0	$< 6 \cdot 10^{-3}$
$b\bar{b}$ MC OS	505	0	$< 2 \cdot 10^{-3}$
data SS	12 261 735	610	$5.00 \cdot 10^{-5}$
data OS	33 505 876	5700	$1.67 \cdot 10^{-4}$

Table 4.11: Absolute event yields for signals MC events (Sig_1: $B_s^0 \to \tau(\to 3\pi\nu_\tau)\tau(\to 3\pi\nu_\tau)$, Sig_2: $B^0 \to \tau(\to 3\pi\nu_\tau)\tau(\to 3\pi\nu_\tau)$, Sig_3: $B_s^0 \to \tau(\to 3\pi\pi^0\nu_\tau)\tau(\to 3\pi\nu_\tau)$), SS and OS in $b\bar{b}$ MC events, SS and OS sample in data (the yields for the SS data correspond to an integrated luminosity of $2.77 fb^{-1}$, while the one for OS to $2.68 fb^{-1}$).

4.4 Normalization

The number N_{sig} of the observed $B^0_{(s)} \to \tau\tau$ signal events (or its upper limit) extracted from the dataset is converted into a value of the $\mathcal{BR}$ (or its upper limit) by multiplying it with a normalization factor α. This takes into account the whole reconstruction, trigger, and selection efficiency as well as the total number $N_{B_{(s)}}$ of $B_{(s)}$ mesons in the collected dataset. In particular, the following relation holds:

$$\mathcal{BR}(B_{(q)} \to \tau^+(3\pi\nu)\tau^-(3\pi\nu)) = \frac{N^{obs}_{sig,q}}{\epsilon_{sig} N_{B_{(s)}}} \equiv \alpha_{(q)} \cdot N^{obs}_{sig,q}. \tag{4.4}$$

By inverting the previous relation and taking into account the $\mathcal{BR}(\tau \to 3\pi\nu)$, the total number of observed $N_{sig,(s)}$ signal events is given by

$$N_{sig,(s)} = N_{B_{(s)}} \cdot \epsilon_{sig} \cdot [\mathcal{BR}(\tau \to 3\pi\nu)]^2 \cdot \mathcal{BR}(B_{(s)} \to \tau\tau) \ . \tag{4.5}$$

The total number $N_{B_{(s)}}$ of $B^0_{(s)}$ mesons in the collected dataset is computed using eq.4.4 applied to the normalization channel $B^0 \to DD_s$ with the D decaying into two charged π and a K, while the D_s goes into a pair of charged kaons and a charged π. The same number of charged tracks in the final state as well as the same geometry of the signal allow a partial cancellation of systematic uncertainties in the reconstruction, trigger, and selections when computing the ratio of those efficiencies. The value of the trigger, reconstruction, and selection efficiencies for the normalization channel have been computed on MC simulated events and a value of

$$\epsilon^{DD_s} = (4.19 \pm 0.04) \times 10^{-4} \tag{4.6}$$

has been found.

The normalization fractions are computed through the following relation:

$$\alpha_q = \frac{\epsilon^{DD_s}_{tot} \cdot \mathcal{BR}(B^0 \to DD_s) \cdot \mathcal{BR}(D \to \pi\pi K) \cdot \mathcal{BR}(D_s \to \pi KK)}{\epsilon^{sig}_{tot} \cdot N^{obs}(B^0 \to DD_s) \cdot [\mathcal{BR}(\tau \to 3\pi\nu_\tau)]^2} \cdot \frac{f_d}{f_q}. \tag{4.7}$$

Using the values of the signal efficiency in Tab.4.10 and the PDG values for the $\mathcal{BR}$ appearing in eq.4.7, the following value for the normalization factors α_s is found:

$$\alpha_s = (3.79 \pm 0.60) \times 10^{-5} \ . \tag{4.8}$$

Assuming the SM branching ratio for $B^0_s \to \tau^+\tau^-$ the expected number of signal events is $N^{exp}_{sig}(B^0_s) = 0.020 \pm 0.009$.

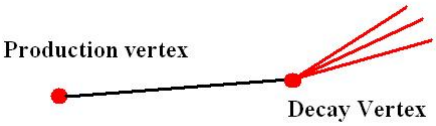


Figure 4.9: Schematic representation of the one-dimensional case.

4.5 Full reconstruction of $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ events

In this section a method for a complete reconstruction of the τ four-momenta in the $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ decay chain will be presented. In particular it will be shown that by exploiting the geometrical information from measurable quantities, and by imposing all the mass constraints (*i.e.* B, τ, and ν) it is possible to completely reconstruct the decay chain.

In Fig.4.2 a typical $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ event is shown. In particular, in the selected events the following quantities can be measured:

- the B production vertex $\vec{r}_B$ and $\tau^\pm$ decay vertexes $\vec{r}_{\tau^\pm}$ that allow to reconstruct the 3D sides of the decay triangle $\vec{w}_\pm \equiv \vec{r}_{\tau^\pm} - \vec{r}_B$,
- four-momenta $q^{\pm\mu}$ of the $(3\pi)^\pm$ system

$$q^{\pm\mu} \equiv \begin{pmatrix} E^\pm \\ \vec{q}^\pm \end{pmatrix} \equiv \begin{pmatrix} \sqrt{m^2_{3\pi^\pm} + \vec{q^\pm}^2} \\ \vec{q^\pm} \end{pmatrix} \tag{4.9}$$

  where $m_{3\pi^\pm}$ is the invariant mass of the $3\pi^\pm$ system.

Now, the relevant question is: assuming that the observed pattern is generated by the $B^0_{(s)} \to \tau^\pm\tau^\mp \to (3\pi^\pm\nu_\tau)(3\pi^\mp\nu_\tau)$ decay chain, is it possible to reconstruct the two $\tau^\pm$ candidates momenta?

4.5.1 The one-dimensional case

A case already known and studied in the literature is the one (here referred to as the "one-dimensional" case) where a τ flies between its reconstructible production and decay vertexes. There it decays into a visible system and a neutrino (Fig.4.9). This is the case of the $Z \to \tau^+\tau^-$, $\mathcal{H} \to \tau^+\tau^-$, or $B \to K^\star\tau^+\tau^-$ (schematically shown in Fig.4.10) where the τ origin vertex is reconstructible.

In this case it is possible to write down the following proportionality relation between the (unknown) four momentum of the τ and the four-vector w whose spatial component is given by the spatial separation between the τ decay and production vertexes and whose temporal component is the time interval between the τ decay and its production:

$$w = hp_\tau \tag{4.10}$$

being the Lorentz-invariant parameter h defined as

$$h \equiv \frac{\tau_\tau}{m_\tau} \ . \tag{4.11}$$

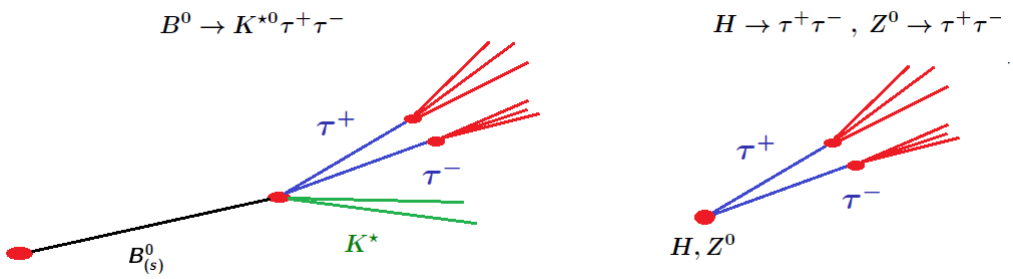


Figure 4.10: Schematic representation of the $B \to K^\star\tau^+\tau^-$ (left) and $Z^0, H \to \tau^+\tau^-$ (right) decays. In both cases the τ production vertexes are reconstructible thanks to the $K^\star$ decay products (for the first case) and to the knowledge of the H, Z production vertex (which corresponds to their decay vertexes).

In this particular case four unknowns must be fixed, which are the components of the τ momentum p^μ_τ. This can be done by exploiting the information on the τ flight direction $\hat{p}_\tau$ (which fixes two unknowns), the mass shell condition for the τ $p^2_\tau = M^2_\tau$ (one unknown) as well as the momentum conservation in the $\tau \to 3\pi + \nu$ decay, *i.e.* $(p_\tau - p_{3\pi})^2 = p^2_\nu = 0$ (one unknown), where $p_{3\pi} \equiv (E_{3\pi}, \vec{p}_{3\pi})$ and p_ν are the four-momenta of the 3π system and the ν respectively. With these constraints the modulus p of the τ space momentum is defined up to a twofold ambiguity, being the solution of the following second degree equation (see *e.g.* Ref.[112]):

$$p^2\left(1 - \frac{|\vec{p}_{3\pi}|^2}{E^2_{3\pi}}\right) - p\left(\frac{\vec{p}_{3\pi}}{E_{3\pi}} \cdot \hat{p}_\tau\right)\frac{M^2_\tau + p^2_{3\pi}}{E_{3\pi}} + M^2_\tau + \left(\frac{M^2_\tau + p^2_{3\pi}}{2E_{3\pi}}\right)^2 = 0\,. \tag{4.12}$$

4.5.2 The real case

The $B^0_s \to \tau\tau \to (3\pi\nu)(3\pi\nu)$ case is quite different, because the τ production vertexes are not known (due to the fact the B^0_s meson flies before decaying) and the two τ four-momenta cannot be reconstructed independently, as it was possible in the one-dimensional case. For this reason, the $B^0_s \to \tau\tau$ will be referred in the following as a "two-dimensional" case.

A first attempt to reconstruct the decay has been presented in [113, 114]: expressing the two τ momenta in cartesian coordinates in 3D space (which represents a non covariant approach) and imposing:

- B, τ, ν mass constraints,
- momentum conservation in $B_s \to \tau\tau$ and planarity of the decay,

the solution of the problem is (one of) the roots of a 8th degree polynomial.

This method shows that reconstruction of the τ momenta is achievable. Nevertheless this approach still presents some issues:

- there are no analytic solutions,
- numerical methods to solve the equation suffer of a certain instability,
- empirically there exists only zero, two or four real solutions but it is not clear why (being in principle up to eight),
- no variables related to the invariant mass are left to discriminate the signal against the background: only the B_s and $\tau^\pm$ decay time.

4.5.3 A new approach

In this work, a different method is proposed, differing from the previous one by the following points:

- keep manifest Lorentz covariance,
- different choice of unknown momenta.

In particular, the guiding principle of this method is to reduce the two-dimensional problem to two one-dimensional one.

Let be:

$$w^{\mu}_{\pm} \equiv (w^0_{\pm}, \vec{w}_{\pm}) \tag{4.13}$$

(being $w^0_{\pm}$ the time interval between the B production and the $\tau^{\pm}$ decay, *i.e.* the "temporal" side of the triangle) and the 2-dimensional vectors

$$W \equiv \begin{pmatrix} w^{\mu}_{+} \\ w^{\mu}_{-} \end{pmatrix} \quad , \quad P \equiv \begin{pmatrix} p^{\mu}_{+} \\ p^{\mu}_{-} \end{pmatrix} \ . \tag{4.14}$$

Using momentum conservation in $B_s \to \tau^+\tau^-$ and the definition of the four-velocity, the following relation between W and P is found:

$$W = H \cdot P \tag{4.15}$$

with

$$H \equiv \begin{pmatrix} \hat{\tau}_B + \hat{\tau}_+ & \hat{\tau}_B \\ \hat{\tau}_B & \hat{\tau}_B + \hat{\tau}_- \end{pmatrix} = \hat{\tau}_B \cdot \begin{pmatrix} 1 + t_+ & 1 \\ 1 & 1 + t_- \end{pmatrix} \tag{4.16}$$

being $\hat{\tau}_i \equiv \frac{\tau_i}{m_i}$ and $t_{\pm} \equiv \frac{\hat{\tau}_{\pm}}{\hat{\tau}_B}$.
Equation (4.15) is a (formal) relation of proportionality between four-vectors through the Lorentz scalar H, and is formally equal to the one obtained for the one-dimensional case in eq.(4.10). It must be noted that in the limit where the ratio $\frac{\tau_B}{M_B}$ goes to zero, the H matrix became diagonal and the problem is trivially equivalent to two one-dimensional cases (this is actually the case of the $H, Z^0 \to \tau^+\tau^-$).

In the ideal (though non realistic) case where the 0-th components of the 4-vectors $w^{\mu}_{\pm}$, $w^0_{\pm}$, can be measured, the four-momenta $p^{\mu}_{\pm}$ can be fixed in two steps. First the three parameters $\hat{\tau}_B$ and $t_{\pm}$ are fixed using the following equality

$$WW^T = HP(HP)^T = HPP^TH \tag{4.17}$$

and imposing the constraints $p^2_{\pm} = m^2_{\tau}$ and $p_+ \cdot p_- = \frac{M^2_B - 2m^2_{\tau}}{2}$. The three independent equations in eq.(4.17) represent quadratic constraints on the elements of H. The solution of this equation is given (in a matricial form) by

$$H = (PP^T)^{-1}[PP^TWW^T]^{\frac{1}{2}} \tag{4.18}$$

with the two additional constraints

$$\begin{cases} det(H) > 0 \\ tr(H) > 0 \end{cases} . \tag{4.19}$$

In terms of equations between the components of matrices in eq.(4.17), the relevant ones are those involving only $t_{\pm}$, being $\hat{\tau}_B$ an overall factor that can be fixed once the previous ones are found. In particular, the relevant equations for $t_{\pm}$ are the following ones:

$$m^2_{\tau}t^2_{\pm} - k_{\pm}(p_+ \cdot p_-)t_+t_- + M^2_B \left[t_{\pm} - \frac{k_{\pm}}{2}(t_+ + t_-)\right] + M^2_B(1 - k_{\pm}) = 0 \tag{4.20}$$

being $k_\pm = \frac{w_\pm^2}{w_+ \cdot w_-}$.

The system (4.20) admits only one solution for the pair (t_+, t_-) in the physical region $t_+ \geq 0$, $t_- \geq 0$. The expression for $\hat{\tau}_B$ as a function of (t_+, t_-) is

$$\hat{\tau}_B = \sqrt{\frac{w_+ \cdot w_-}{t_+ t_- (p_+ \cdot p_-) + \frac{M_B^2}{p_+ p_-}(t_+ + t_-) + M_B^2}} \ . \tag{4.21}$$

Once the parameters of the H matrix are fixed, the momenta $p_\pm^\mu$ are found through the (linear) equation (4.15). The knowledge of the three parameters (t_+,t_-,τ_B) represents the maximal information that can be extracted from the measured quantities. An estimation of the improvement of the signal-background discrimination when these three variables are used with other standard variables is reported in Appendix C. It is worth to notice that nothing about the final state of the τ is assumed, except for the fact that the τ decay vertexes must be known. Using the measured four-momenta of the two 3π systems the invariant mass of the undetected τ daughters can be reconstructed separately for each τ. In the case under study these are distributed as a delta around 0 (being such the invariant mass of the neutrino), giving thus two peaking variables for the signal.

In the real case, nevertheless, it is not possible to measure the $w_\pm^0$ components and the four-momenta $p_\pm$ must be fixed through a different procedure, replacing the constraints from $w_\pm^0$ with the information from the subsequent τ decays, and by using linear and quadratic constraints, as is shown in the following.

H is a real 2×2 symmetric matrix diagonalized in the form:

$$H = R(\theta) \cdot D_\lambda \cdot R^{-1}(\theta) \tag{4.22}$$

being

$$D_\lambda \equiv \hat{\tau}_B \cdot \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \quad , \quad R(\theta) \equiv \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{4.23}$$

We find for the eigenvalues $\lambda_\pm$ and for the rotation parameter $\cos\theta$

$$\lambda_\pm = \frac{2 + \Delta \pm \sqrt{4 + \delta^2}}{2} = 1 + \frac{\Delta}{2} \pm \sqrt{1 + \frac{\delta^2}{4}} \tag{4.24}$$

$$\cos\theta = \left[1 + \left(\sqrt{1 + \frac{\delta^2}{4}} - \frac{\delta}{2} \right)^2 \right]^{-\frac{1}{2}} \tag{4.25}$$

being

$$\Delta \equiv t_+ + t_- \quad , \quad \delta \equiv t_+ - t_- \tag{4.26}$$

It is interesting considering some particular cases, in order to understand what is the physical "meaning" of the three Lorentz invariant parameters $(\theta, \lambda_+, \lambda_-)$:

- if $t_+ = t_- \equiv t$ the following set of values $(\theta = \frac{\pi}{4}, \lambda_+ = 2 + t, \lambda_- = t)$ is found. A particular case is obtained when $t = 0$: in this case H is not invertible: both τ are created and decay *immediately* (because the triangle degenerates to a line)

- if $t_+ \neq t_-$ but still $t_+ \simeq t_-$ (*i.e.* $\delta << 1$) the following approximation for $\cos(\theta)$ is found:

$$\cos(\theta) = \cos(\frac{\pi}{4}) + \frac{1}{\sqrt{32}}(t_+ - t_-) + \mathcal{O}\left((t_+ - t_-)^2\right) \ . \tag{4.27}$$

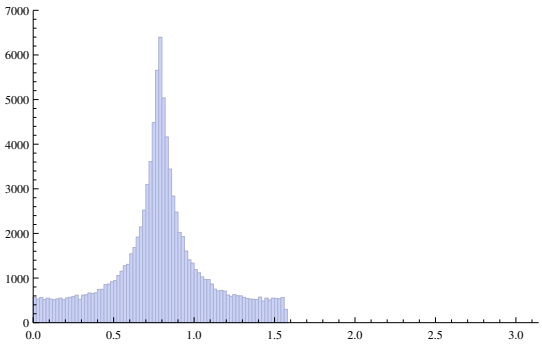


Figure 4.11: Distribution of the θ angle for MC generated signal events.

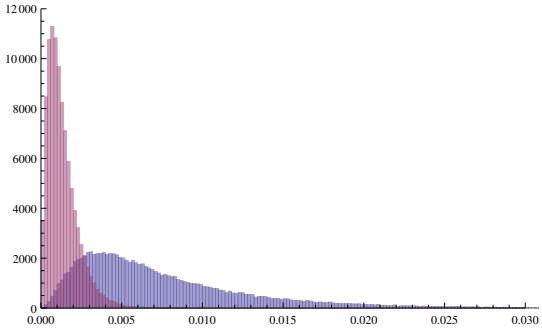


Figure 4.12: Distribution of the λ_+ (violet) and λ_- (purple) eigenvalues for MC generated signal events.

The distributions of θ and the two eigenvalues $\lambda_\pm$ for Monte Carlo generated signal events are shown respectively in Fig.4.11 and 4.12.

$\cos\theta$ is sensitive only to the difference $t_+ - t_-$, hence it measures the asymmetry of the triangle in the proper time of each of the two τs.

Fig.4.13 shows the distribution of $\frac{\delta}{2}$ (which is the relevant parameter to linearize the expression (4.24) for $\lambda_\pm$) and $\cos(\theta) - \cos(\frac{\pi}{4})$ as a function of δ and $|\delta|$ (for the convention on the sign see the figure caption).

In the "rotated" basis:

$$\tilde{W}(\theta) \equiv R^{-1}(\theta) \cdot W = \begin{pmatrix} \tilde{w}^\mu_+ \\ \tilde{w}^\mu_- \end{pmatrix} = \begin{pmatrix} \cos\theta w^\mu_+ + \sin\theta w^\mu_- \\ -\sin\theta w^\mu_+ + \cos\theta w^\mu_- \end{pmatrix} \tag{4.28}$$

$$\tilde{P}(\theta) \equiv R^{-1}(\theta) \cdot P = \begin{pmatrix} \tilde{p}^\mu_+ \\ \tilde{p}^\mu_- \end{pmatrix} = \begin{pmatrix} \cos\theta p^\mu_+ + \sin\theta p^\mu_- \\ -\sin\theta p^\mu_+ + \cos\theta p^\mu_- \end{pmatrix} , \tag{4.29}$$

the following relation holds (using eqs.(4.15), (4.22)):

$$\tilde{W}(\theta) = D_\lambda \cdot \tilde{P}(\theta) \tag{4.30}$$

With the choice (4.28) of the unknown momenta there are no more mixed terms $\tau^+ - \tau^-$, being

$$\left\{ \begin{array}{lcl} \tilde{w}^\mu_+ & = & \frac{1}{\sigma_+}\tilde{p}^\mu_+ \\ \tilde{w}^\mu_- & = & \frac{1}{\sigma_-}\tilde{p}^\mu_- \end{array} \right. \Rightarrow \left\{ \begin{array}{lcl} \vec{\tilde{w}}_+ & = & \frac{1}{\sigma_+}\vec{\tilde{p}}_+ \\ \vec{\tilde{w}}_- & = & \frac{1}{\sigma_-}\vec{\tilde{p}}_- \end{array} \right. . \tag{4.31}$$

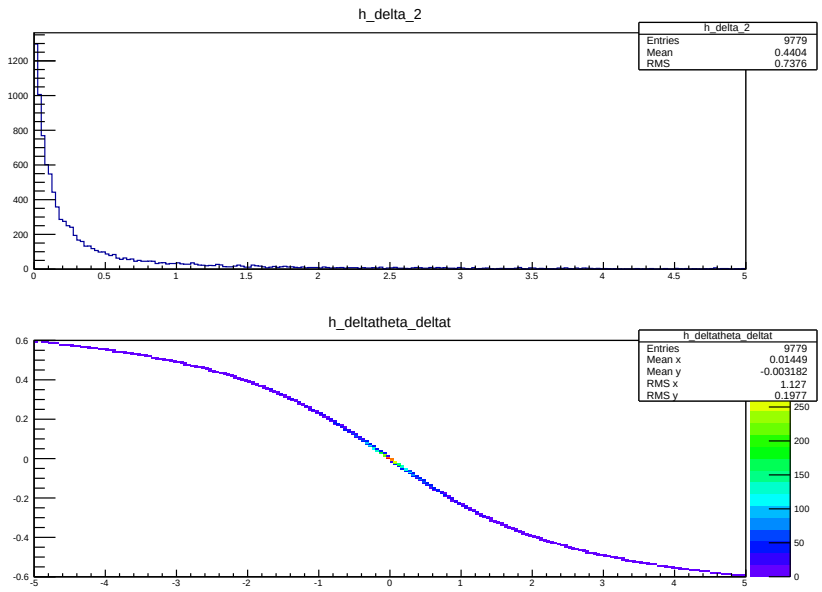


Figure 4.13: Top: distribution of $\frac{\delta}{2}$. Bottom: dependence of $\theta - \frac{\pi}{4}$ as a function of δ.

where, from eq.(4.31), the Lorentz invariant unknown $\sigma_\pm$ are defined as:

$$\sigma_\pm \equiv \frac{1}{\hat{\tau}_B \lambda_\pm} \ . \tag{4.32}$$

By imposing the two (θ-dependent)

- mass-shell condition for $\tilde{p}_\pm^2$:

$$\tilde{p}_\pm^2 \equiv M_\pm^2 = m_\tau^2 \pm \sin(2\theta) s \ , \tag{4.33}$$

  where m_τ^2 is the τ mass and $s \equiv p_+ \cdot p_-$ (for signal events $s = \frac{M_B^2 - 2m_\tau^2}{2}$);

- constraint on the direction of $\vec{\tilde{p}}_\pm$:

$$\vec{\tilde{p}}_\pm \propto \hat{\tilde{w}}_\pm \ ; \tag{4.34}$$

and the (θ independent)

- constraint on $p_\pm^\mu \cdot q_\mu^\pm$ (assuming $p_\nu^2 = 0$):

$$p_\pm \cdot q^\pm \equiv m_\pm^2 = \frac{m_\tau^2 + m_{3\pi^\pm}^2}{2} \ , \tag{4.35}$$

  being $m_{3\pi^\pm}^2$ the invariant mass of the $3\pi^\pm$ system

a system of two equations of second degree in two unknowns $\sigma_\pm$ is found (see next section for the proof and for the explicit expression of the equations):

$$\mathcal{P}_\pm^{(2)}(\sigma_\pm, \theta) = 0 \tag{4.36}$$

It's important to stress the fact that these two equations *linearly* depend on the four product $s = p_+ \cdot p_-$, because of the constraint in eq.(4.33).

The coefficients of the eq.(4.36) depend only on the parameter θ, and, once it is fixed, exact solutions are obtained. In principle, this parameter can be determined by using the additional constraint

$$p_+ \cdot p_- = \frac{M_B^2 - 2m_\tau^2}{2} \ . \tag{4.37}$$

Once this constraint is imposed it is possible to completely reconstruct the event by solving the system of three equations in three unknowns. Nevertheless, this strategy requires the resolution of a highly complex trigonometric equation to fix θ, and has been left aside so far. Instead some approximation of θ, that will be presented in Sec.4.5.6, have been used to solve the system of equations (4.36). With this strategy several discriminating variables have been sorted out and used for the signal search (see Sec.4.5.7).

At present, a different strategy aiming to fully exploit the kinematic equations (4.36),(4.37) is under study. This new approach will allow to leave the s quantity a free two-body peaking variable (for signal) and more details will be presented in Sec.**??**.

4.5.4 Calculations

In this section the relevant equations to solve for the τ unknown momenta will be obtained.

Let be

$$q^{\pm\mu} \equiv \begin{pmatrix} E^\pm \\ \vec{q}^\pm \end{pmatrix} \tag{4.38}$$

the four-momenta of the two 3π system from the $\tau^\pm$ decay;

by using the constraints in eqs.(4.33,4.34), the four-vector $\tilde{p}_\pm$ can be written as

$$\tilde{p}_\pm^\mu \equiv \begin{pmatrix} \sqrt{M_\pm^2 + |\vec{\tilde{w}}_\pm|^2\sigma_\pm^2} \\ \sigma_\pm \vec{\tilde{w}}^\pm \end{pmatrix} \ . \tag{4.39}$$

By using the following intermediate quantities:

$$C_{3\pi} \equiv \begin{pmatrix} \vec{q}_+ \cdot \vec{w}_+ & \vec{q}_+ \cdot \vec{w}_- \\ \vec{q}_- \cdot \vec{w}_+ & \vec{q}_- \cdot \vec{w}_- \end{pmatrix} \ , \tag{4.40}$$

$$\begin{pmatrix} a_+ \\ a_- \end{pmatrix} \equiv R(-\theta) \begin{pmatrix} m_+^2/E^+ \\ m_-^2/E^- \end{pmatrix} \ , \tag{4.41}$$

$$\begin{pmatrix} b_+ \\ b_- \end{pmatrix} \equiv \left[\begin{pmatrix} \cos^2\theta & \sin^2\theta \\ \sin^2\theta & \cos^2\theta \end{pmatrix} \circ [C_{3\pi} R(\theta)]^T \right] \begin{pmatrix} 1/E^+ \\ 1/E^- \end{pmatrix} \ , \tag{4.42}$$

$$\begin{pmatrix} c_+ \\ c_- \end{pmatrix} \equiv \sin\theta\cos\theta\sigma_1 [C_{3\pi} R(\theta)]^T \begin{pmatrix} -1/E^+ \\ 1/E^- \end{pmatrix} \ , \tag{4.43}$$

being $\sigma_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, the constraints in eq.(4.5.3) give the following equation

$$\sqrt{M_\pm^2 + |\vec{\tilde{w}}_\pm|^2\sigma_\pm^2} = a_\pm + b_\pm\sigma_\pm + c_\pm\sigma_\mp \ . \tag{4.44}$$

Taking the square of the previous two equations the following equalities are obtained

$$(b_\pm^2 - |\vec{\tilde{w}}_\pm|^2)\sigma_\pm^2 + 2c_\pm b_\pm \sigma_+\sigma_- + c_\pm^2\sigma_\mp^2 + 2a_\pm b_\pm\sigma_\pm + 2a_\pm c_\pm\sigma_\mp + a_\pm^2 - M_\pm^2 = 0 \ . \tag{4.45}$$

The $\circ$ product between matrices denotes the Hadamard product. It is defined, for two generic $n \times m$ matrices A and B as $(A \circ B)_{ij} \equiv A_{ij}B_{ij}$.

The determinant of the quadratic part of eq.(4.45) is equal to $-|\vec{\tilde{w}}_{\pm}|^2 c_{\pm}^2$, which is always negative. Thus, the two equations (4.45) describe two hyperbolas in the (σ_+, σ_-) plane. It is important to stress the fact that this result does not depend on the value of θ; nevertheless, the existence, the number, and the position of the solutions of the system do depend on θ. Being the possible number of common points of intersection between two conics equal to 0,2,4 this justifies what was found in [113, 114].

Finally, the solution of the problem is equivalent to finding the roots of a 4th degree polynomial:

$$\mathcal{P}^{(4)}(\xi) = \sum_{i=0}^{4} a^{(i)}(\theta)\xi^i = 0 \ . \tag{4.46}$$

The explicit expressions for the coefficients $a^{(i)}(\theta)$ as well as the definition of the unknown ξ are shown in the Appendix D.

4.5.5 Choice of the "right" solution

The solutions of the fourth degree equation are in the form

$$\xi_1 = a - ib \ , \ \xi_2 = a + ib \ , \ \xi_3 = c - id \ , \ \xi_4 = c + id \tag{4.47}$$

with $a, b, c, d \in \mathcal{C}$.

For each of these solutions a pair of four-momenta $\tilde{p}_{\pm}(\theta, \xi_i)$ is found and the modulus of their spatial components are given by $\vec{\tilde{p}}_{\pm}(\theta, \xi_i) = \sigma_{\pm}(\theta, \xi_i)\vec{\tilde{w}}_{\pm}(\theta)$.

Being these solutions related to the physical observables $\sigma_{\pm}$, at least two of them must be real for signal events. In order to study the behavior of these solutions, eq.(4.46) has been solved for a sample of Monte Carlo generated signal events, using the true value of θ. Results are summarized in the following and in Fig.4.14:

- the first two solutions are real for 95.58% of events (meaning that $b = i\tilde{b}$, $a, \tilde{b} \in \Re$),
- for 1.24% of events the last two are real as well,
- due to numerical resolution, complex solutions can appear also when using the true value of the θ parameter; nevertheless complex solutions are considered "good" if their imaginary part is smaller than 5% the real one. The appearance of this small imaginary part (even though the true values for the θ parameters is used) could be due to the fact that the equations are non linear, so that a small initial uncertainty (due essentially to approximations in the input quantities like momenta and lengths) can propagate non linearly, ending up to a sizable final effect.
- apparently there is no way to choose *a-priori* the correct one among them. Nevertheless, as shown in Fig.4.15 the distribution of the first two solutions well reproduce the one of the true value (in agreement with what was already observed in [113, 114]), while the last two are very differently distributed with respect to the true one.

In the "real-life" case, the true value of the angle θ is not known and it must be estimated through measurable quantities (see following sections). For this reason, and due also to

- final state radiation emitted by charged particles,
- detector resolution,

In the sense that the imaginary part is only due to numerical approximation.

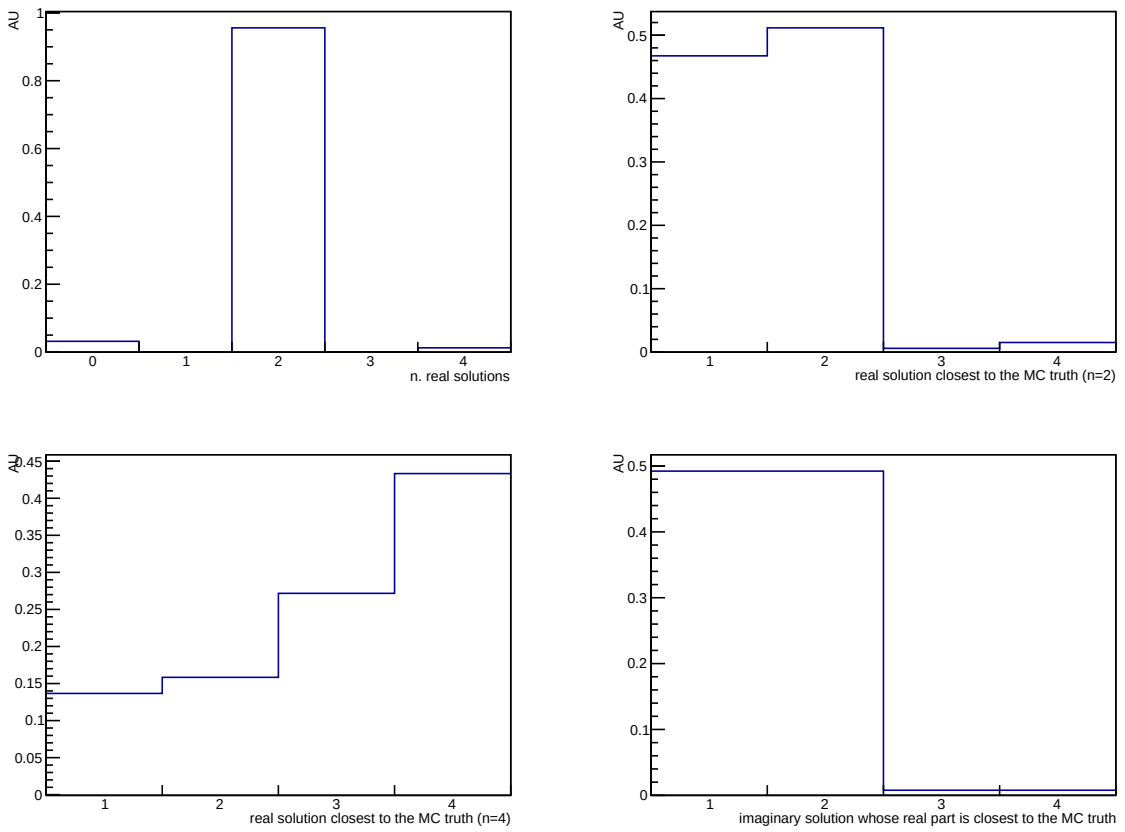


Figure 4.14: Behavior of the solutions of eq.(4.46) solved using the true value of θ. Top-left: fraction of events as a function of the number of real solutions for eq.(4.46); top-right: for events for which eq.(4.46) admits only two real solutions, fraction of events as a function of the index (according to the ordering in eq.(4.47)) of the closest on the MC truth; bottom-left: for events for which eq.(4.46) admits 4 real solutions, fraction of events as a function of the index (according to the ordering in eq.(4.47)) of the closest on the MC truth; bottom-right: for events for which eq.(4.46) admits only imaginary solutions, fraction of events as a function of the index (according to the ordering in eq.(4.47)) of the one whose real part is closest on the MC truth value.

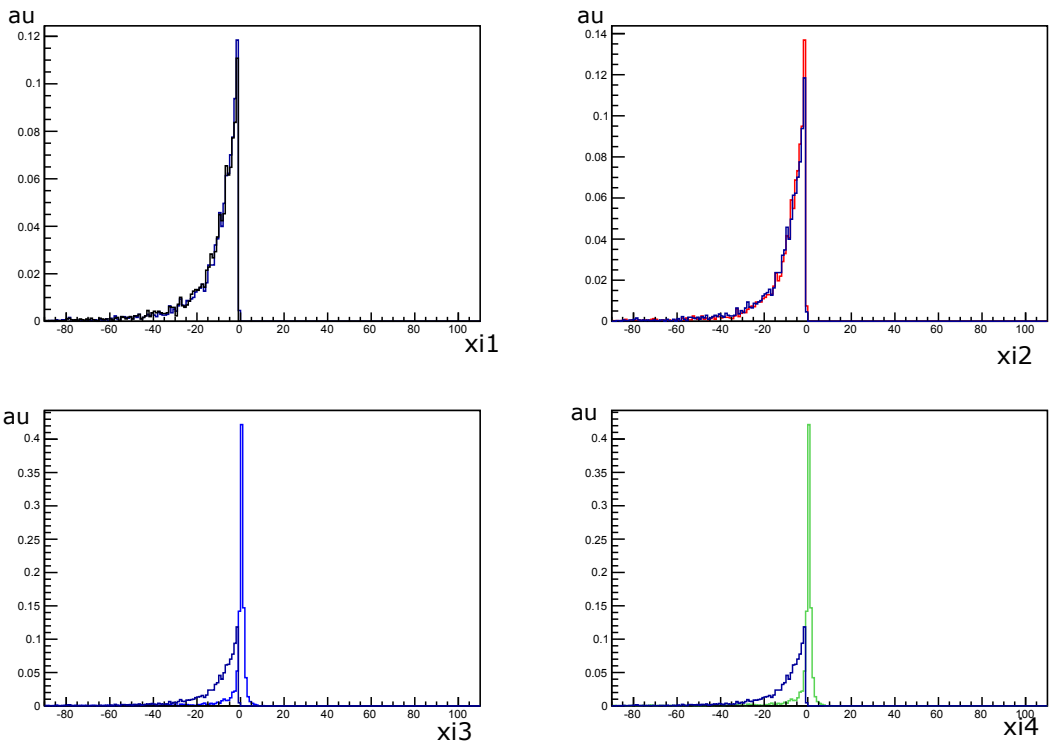


Figure 4.15: Distribution of $Re(\xi_i)$ for the four roots of eq.(4.46) and of the true value ξ_{true} superimposed (in blue). According to the classification of the solutions of eq.(4.46) in eq.(4.47), ξ_1 is in black, ξ_2 is in red, ξ_3 is in light blue, ξ_4 is in green.

a "non-trivial" imaginary part for the true solution appears.

The appearance of an imaginary part for the solution is due to the violation of some constraints assumed while deriving the equations. It is, in some sense, a measure of how much the starting hypothesis that the observed pattern is generated by the decay chain $B_s \to \tau^+\tau^-$ is incorrect. For this reason, even if its meaning is not evident (depending on the various determinants which the solutions of the equation depend on) it nevertheless carries an information which is physical. For this reason, these imaginary parts can also be used as discriminating variables as it will be discussed in Sec.4.5.7.

4.5.6 θ approximation

To deal with the dependence on θ of the coefficients of eq.(4.45), the following options have been considered:

- substitute θ with its average value $\overline{\theta} = \frac{\pi}{4}$ (*i.e.* assuming the same decay time of the two τ);
- find a better approximation of θ which relies on measurable quantities.

The first option relies on the fact that even if this approximation is not optimal for the signal, it will be still less correct for the various resonant backgrounds, where there are usually D particles (that have lifetime quite different from τ).

Nevertheless, finding a method to get for each event a more reliable approximation of θ seemed to be the best way to be pursued. Indeed, being θ a "measure" of the asymmetry of the triangle, it might be possible to express it as a function of the difference between $\vec{w}_+$ and $\vec{w}_-$. In particular, a function of (spatial) quantities measured in the laboratory frame which is Lorentz invariant, being such θ, must be found.

The method presented here to approximate $\bar{\theta}$, consists in finding a matrix $\bar{H}$ whose elements are functions of measurable quantities and which feature the same "structure" of the matrix H in eq.(4.16)

$$\overline{H} = \begin{pmatrix} 1+\bar{t}_+ & 1 \\ 1 & 1+\bar{t}_- \end{pmatrix} \tag{4.48}$$

being $\bar{t}_\pm$ functions of $(\vec{w}_+, \vec{w}_-)$. In this way, the $\overline{\theta}$ angle parameterizing the rotation matrix $R(\overline{\theta})$ which diagonalizes $\overline{H}$ can be a reliable approximation of the true θ angle.

In analogy with the H matrix defined in eq.(4.16), the matrix $\bar{H}$ must satisfy the following properties :

- $\overline{H}$ must transform as H for exchange $(+ \leftrightarrow -)$, *i.e.* $\overline{H}_{11} \leftrightarrow \overline{H}_{22}$,
- the functions $\bar{t}_\pm(\vec{w}_+, \vec{w}_-)$ must be adimensional and
- Lorentz invariant (at least at first order in a power expansion of the boost γ_B of the B_s)

In the following, two possible parameterizations for the $\overline{H}$ matrix will be presented.

Approximation through the triangle sides. The functions

$$\bar{t}_\pm \equiv \frac{|\vec{w}_\pm|}{\alpha|\vec{w}_+ + \vec{w}_-|} \tag{4.49}$$

satisfy all these conditions (for $\gamma_B \gg 1$).

The reason for the introduction of the α parameter lies on the fact that for the H matrix the following relation holds: $H_{11} - H_{22} = \delta$, while $\overline{H}_{11} - \overline{H}_{22} = \frac{1}{\alpha}\frac{|\vec{w}_+|-|\vec{w}_-|}{|\vec{w}_++\vec{w}_-|} = \frac{1}{\alpha}\cdot\delta_W$, where $\delta_W \equiv \bar{H}_{11} - \bar{H}_{22}$. Now, to make these two expressions similar, α must be set equal to the

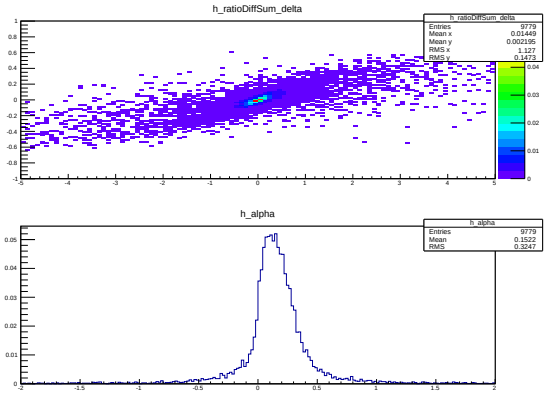


Figure 4.16: Top: δ_W as a function of δ; bottom: distribution of the ratio between δ_W and δ whose mean can be used as a rough estimation of their ratio.

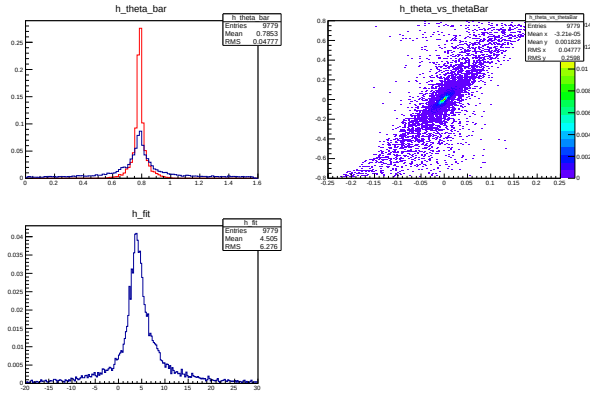


Figure 4.17: Top-left: for $\alpha = 1$ distribution of $\overline{\theta}$ (red) and θ (blue); top-right: $\theta - \frac{\pi}{4}$ (y axis) *vs* $\overline{\theta} - \frac{\pi}{4}$ (x axis); bottom-left: distribution of $(\theta - \frac{\pi}{4})/(\overline{\theta} - \frac{\pi}{4})$, whose mean can be used as a rough estimation of the coefficient between θ and $\overline{\theta}$.

coefficient between δ_W and δ. This parameter depends on the masses and the emission angle of the τs in the B rest frame and thus it must be estimated empirically (see *e.g.* Fig. 4.16).

By diagonalizing the $\bar{H}$ matrix a rotation angle $\overline{\theta}$ is found, whose distribution and correlation with the θ angle are shown for two different values of α in Figs.4.17-4.18.

In Tab.4.12 are reported the efficiencies for finding a real solution with different approximations of θ.

The estimation of θ through $\bar{\theta}$ improves the efficiencies by a factor greater than 50% with respect to the case where the approximation $\theta = \frac{\pi}{4}$ is used. In Figs.4.19(a),4.19(b) the quantity $2\frac{(|\xi_{best}|-|\xi_{true}|)}{|\xi_{best}|+|\xi_{true}|}$ is reported for different approximations of θ. This shows that the precision of the closest solution as well is improved using the estimation of θ through $\bar{\theta}$. In particular the residual is distributed symmetrically with respect to zero in all the α-dependent $\bar{\theta}$ approximations, while it is manifestly asymmetric and biased towards negative values if the approximation $\theta = \frac{\pi}{4}$ is used.

As far as the choice of the value of the α parameter is concerned, $\alpha = 1$ has been taken. Indeed, even if the approximation of the θ angle is worse than in the case of $\alpha = 0.2$ (see Figs.4.17-4.18), the efficiency to find a real solution of eq.(4.46) is higher (see Tab.4.12) and, at the same time, the precision on the reconstructed solution closest on the the true one are compatible with each other, as it is shown in Fig.4.19(a).

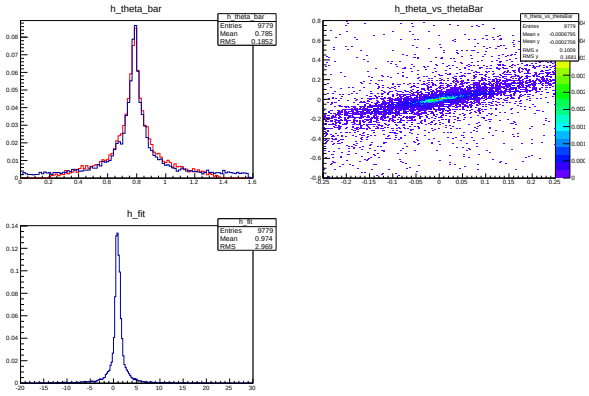
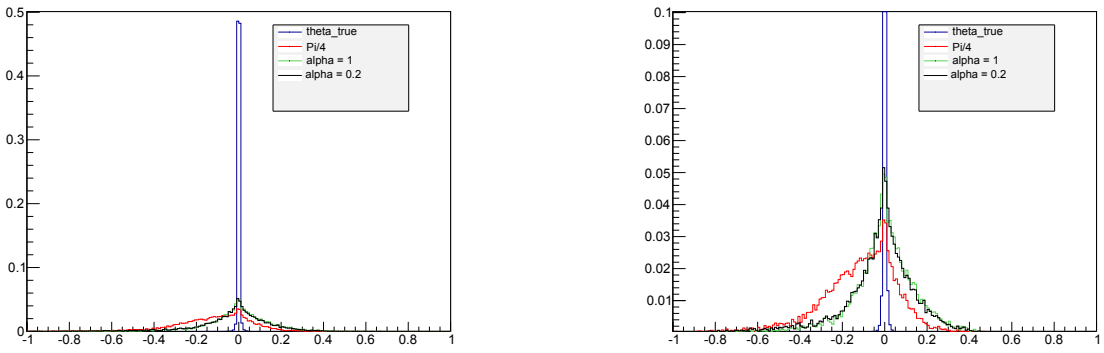


Figure 4.18: Top-left: for $\alpha = 0.2$ distribution of $\overline{\theta}$ (red) and θ (blue); top-right: $\theta - \frac{\pi}{4}$ (y axis) *vs* $\overline{\theta} - \frac{\pi}{4}$ (x axis); bottom-left: distribution of $(\theta - \frac{\pi}{4})/(\overline{\theta} - \frac{\pi}{4})$, whose mean can be used as a rough estimation of the coefficient between θ and $\overline{\theta}$.

θ_{true}	$\theta = \frac{\pi}{4}$	$\theta(\overline{\theta}_{\alpha=1})$	$\theta(\overline{\theta}_{\alpha=0.2})$
99.90	21.40	31.00	29.30

Table 4.12: Efficiencies (in %) to find a real solution (including complex solutions with $\frac{Im(\overline{\xi})}{Re(\overline{\xi})} < 2\%$) with different approximations of θ.



(a) (b)

Figure 4.19: Precision on the reconstructed solution whose value is closest on the true one ($i.e.\ 2\frac{(|\xi_{best}|-|\xi_{true}|)}{|\xi_{best}|+|\xi_{true}|}$) found for different approximations of θ.

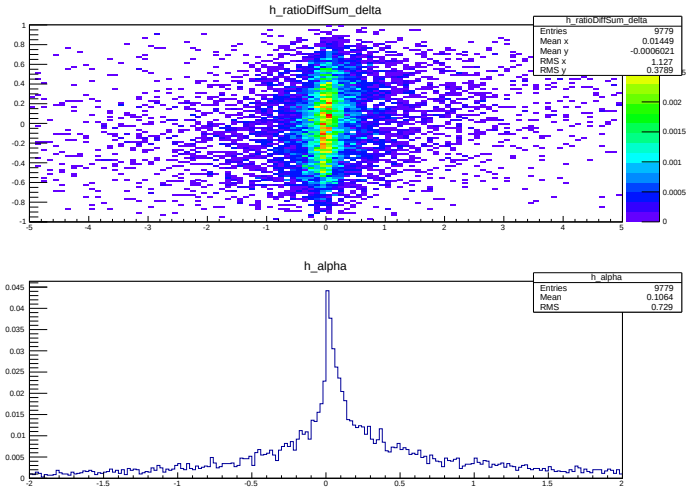


Figure 4.20: Top: $\delta_{IP} \equiv \frac{b_+ - b_-}{b_+ + b_-}$ as a function of δ; bottom: distribution of the ratio between δ_{IP} and δ whose mean can be used as a rough estimation of their ratio.

Approximation through the $(3\pi)^{\pm}$ system impact parameters. Another possible form for the functions $\bar{t}_{\pm}$ has been tested

$$\bar{t}_{\pm} \equiv \frac{b_{\pm}}{\alpha(b_+ + b_-)} \ , \tag{4.50}$$

being $b_{\pm}$ the impact parameter of the $3\pi^{\pm}$ system with respect to the B_s PV.

In this case, as well, it is easy to see that the functions in eq.(4.50) satisfy the requirements on the $\bar{H}$ matrix elements.

Nevertheless, in this case the result is not as good as the one obtained with the previous approximation. This is due to the fact that, when using the impact parameters, an information which is not related to the decay triangle structure only (*i.e.* the direction of the 3π system) is introduced. Instead, in the previous case, the $\bar{t}_{\pm}$ function are quantities which "genuinely" describe the asymmetry of the triangle. The way to exploit the information about the impact parameter as well is not as straightforward as in the other case.

Ideas about an improved recursive approximation. In this section a recursive method aiming to improve the approximation of θ together with the ''quality" of the solution is presented.

The θ angle can always be expressed in the following form

$$\theta = \theta^{\star} + \delta\theta \tag{4.51}$$

being $\theta^{\star}$ any approximation for θ, for example through $\bar{\theta}$ or $\theta^{\star} = \frac{\pi}{4}$ and $|\frac{\delta\theta}{\theta^{\star}}| << 1$.

To fix θ the general relation between $\tilde{s}$ and s:

$$\tilde{s}(\theta) = s\cos 2\theta \tag{4.52}$$

can be exploited.

By developing both sides of the previous equation around $\theta^{\star}$ the following relation is found:

$$\tilde{s}(\theta^{\star} + \delta\theta) = \tilde{s}(\theta^{\star}) + \delta\theta \frac{d\tilde{s}}{d\theta}\|_{\theta^{\star}} = s(\cos^2\theta^{\star} - \sin^2\theta^{\star} - 4\delta\theta\cos\theta^{\star}\sin\theta^{\star}) \ , \tag{4.53}$$

where the quantities on the left hand side are meant to be computed by solving eq.(4.46). The value of $\frac{d\tilde{s}}{d\theta}\|_{\theta^{\star}}$ must be evaluated numerically, by solving the fundamental equation with

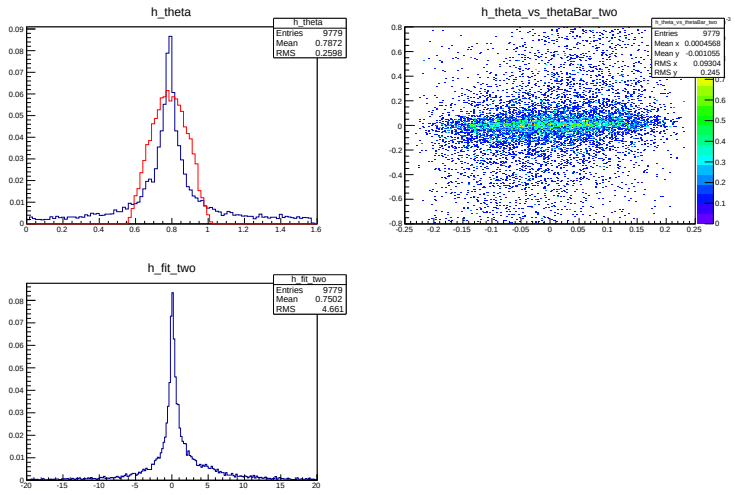


Figure 4.21: Top-left: distribution of $\overline{\theta}$ (red) and θ (blue); top-right: $\theta - \frac{\pi}{4}$ (y axis) *vs* $\overline{\theta} - \frac{\pi}{4}$ (x axis); bottom-left: Distribution of $(\theta - \frac{\pi}{4})/(\overline{\theta} - \frac{\pi}{4})$, whose mean can be used as a rough estimation of the coefficient between θ and $\overline{\theta}$.

$\theta = \theta^\star \pm d\theta$ and for each of these two values computing the value $\tilde{s}(\theta^\star \pm d\theta)$ and then the incremental ratio $\frac{\tilde{s}(\theta^\star + d\theta) - \tilde{s}(\theta^\star - d\theta)}{2d\theta}$.

From the previous equation, it follows that

$$\delta\theta = \frac{s(\cos^2\theta^\star - \sin^2\theta^\star) - \tilde{s}(\theta^\star)}{4s\cos\theta^\star\sin\theta^\star + \frac{d\tilde{s}}{d\theta}\|_{\theta^\star}} \ . \tag{4.54}$$

While doing this computation the value of $d\theta$ must be chosen smaller than the expected value of the correction $\delta\theta$, and once a value is chosen for $d\theta$ and $\delta\theta$ has been evaluated, this condition must be checked a-posteriori.

The procedure can be iterated n times up to a chosen convergence criterium is satisfied. In particular, the maximum number of iteration has been fixed to 20. The $d\theta$ parameter has been taken of the form $d\theta_n = \frac{0.01}{2^n}$, and the convergence criteria are:

- $|\delta\theta_{n+1}| > 0.66 \cdot |\bar{\theta}_n|$,
- $|\delta\theta| < 10^{-4}$,
- $||\theta_n - \theta_{n-1}| - |\theta_n - \theta_{n-2}|| < 10^{-3}$.

The improvement in the approximation of the θ angle and, as consequence, of the solution ξ is summarized in Fig.4.22 and in Tab.4.13.

4.5.7 Discussion about discriminating variables

The approximated value of the θ parameter is used to compute the coefficients of the polynomial in eq.(4.46) and to solve for the unknown ξ. Once the correct solution is chosen among the possible four, the two (approximate) τ momenta can be reconstructed as well as the topology of the decay. The knowledge of the τ momenta allows to reconstruct their common origin vertex as well as the B candidate momentum. Thus, from a strictly conceptual point of view, the only variables that can be computed to discriminate signal against background are the decay times of the two τ and of the B candidate (which are the elements of the matrix H). Indeed, all the kinematic variables such as the masses of the τ and the B, as well the ones of the two ν have already been imposed as external constraint in the derivation of the equations.

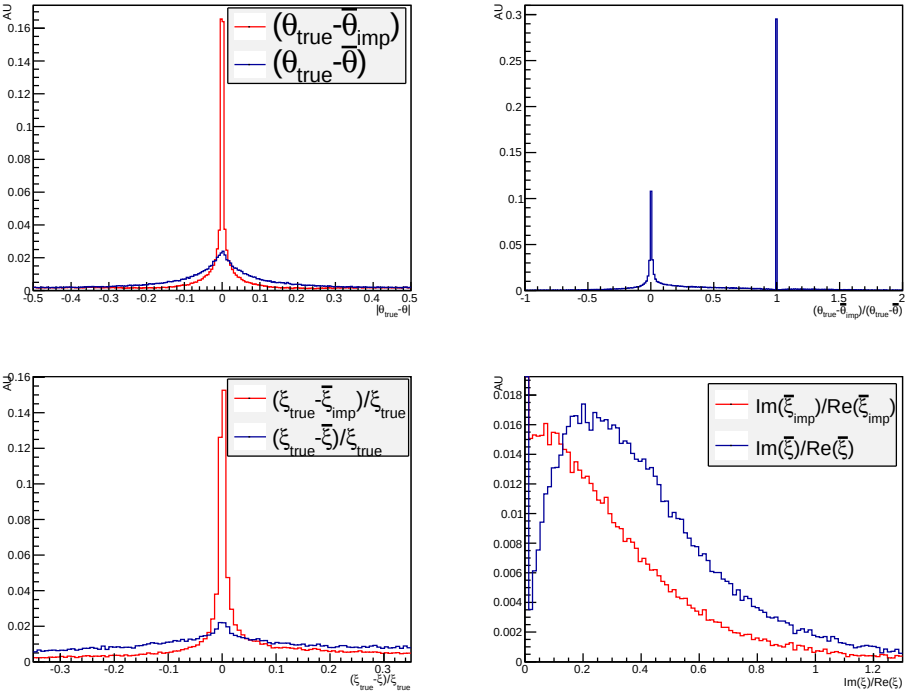


Figure 4.22: Top-left: fraction of events as a function of the difference between the values of the true angle θ and the first approximation $\bar{\theta}$ (blue) and the improved one $\bar{\theta}_{improved}$ (red); top-right: fraction of events as a function of the ratio $\frac{\theta-\bar{\theta}_{improved}}{\theta-\bar{\theta}}$. The peak at 1 corresponds to all those events for which the iteration fails at the first step, so that the zero order approximation $\bar{\theta}$ is used to approximate θ; bottom-left: fraction of events as a function of the difference between the first (blue) and improved (red) approximated and true solution with respect to this last one; bottom-right: fraction of events as a function of the ratio $\frac{Im(\bar{\xi})}{Re(\bar{\xi})}$ for the first (in blue) and improved (in red) approximation of θ. The distribution in the first bin is out of the y-axis scale.

eff. in %	solutions	θ_{true}	$\bar{\theta}$	$\bar{\theta}_{imp}$
Real	2	95.6 (98.7)	26.60 (100.00)	48.35 (99.97)
	4	1.2 (1.3)	0.00 (0.00)	0.01 (0.03)
	$2\oplus 4$	96.8	26.60	48.36
Complex	good	3.06 (96.7)	0.30 (0.40)	1.50 (2.86)
	bad	0.10 (3.35)	73.10 (66.60)	50.15 (97.14)
	good $\oplus$ bad	3.17	73.4	51.64

Table 4.13: Efficiencies (in%) for finding real or imaginary solutions for eq.(4.46) using the true value of θ_{true}, its zero order approximation $\bar{\theta}$, and the improved approximation using the iterative method $\bar{\theta}_{imp}$. In the first block the efficiencies to find two (first row) or four (second row) real solutions are reported; last row is the sum of the previous two. In the second block the efficiencies to find complex solutions are reported; complex solutions are considered "good" (first row) if the imaginary part smaller than 2% the real one (*i.e.* $Im(\xi)/Re(\xi) < 2\%$). The values in parenthesis refers to the fraction of events with respect to the Real or Complex categories only; for each block ("Real" or "Complex") last row is the sum of the previous two.

Nevertheless, in the various steps of the reconstruction algorithm, several variables which are functions of observables quantities can be sorted out and used to discriminate signal against background. These variables do not have an evident physical meaning, being much more related to the mathematics needed to derive the relevant equations. In more details, these variables are:

- the coefficients of eqs.(4.45-4.46);
- the four complex solutions of eq.(4.46) (no matter which the true one is);
- the four possible complex values for the s and $\tilde{s}$ variables;
- the approximation of θ, $\lambda_{\pm}$.

The justification behind their use for the signal-background discrimination lies in the facts that they are functions of measurable quantities in the event, and of the approximated value of the θ angle as well, which turn out to have different distribution for signal and background events. In particular, different approximations of θ can result in slightly different distributions of the same variable and the above-mentioned quantities have been computed with three approximations of θ: $\theta = \frac{\pi}{4}$, $\theta = \bar{\theta}$ and the estimation $\theta^{\star}$ obtained with the iterative method.

Among the above-mentioned variables, the most discriminating are found to be the following ones:

1. Re_x_3: $Re(|\vec{\tilde{p}}_{+}(\theta = \bar{\theta}, \xi_3)|)$;
2. Im_x_1_ar: $Im(|\vec{\tilde{p}}_{+}(\theta = \theta^{\star}, \xi_1)|)$;
3. Im_stildepm_1_ar: $Im(\tilde{p}_{+}(\theta = \theta^{\star}, \xi_1)\tilde{p}_{-}(\theta = \theta^{\star}, \xi_1))$;
4. Im_stildepm_3_ar: $Im(\tilde{p}_{+}(\theta = \theta^{\star}, \xi_3)\tilde{p}_{-}(\theta = \theta^{\star}, \xi_3))$;
5. Re_stildepm_1_Pi4: $Im(\tilde{p}_{+}(\theta = \pi/4, \xi_1)\tilde{p}_{-}(\theta = \pi/4, \xi_1))$;
6. Re_xi_1_ar: $Re(\xi_1)$;
7. theta_bar_W: $\bar{\theta}$;

and their distributions are shown in Fig.4.23 for signal MC events and the SS data sample after the analysis selection presented in Sec.4.3.

These variables are analog, for instance, to the output of a MVA classifier. In this case the output variable is a function of the input quantities that does not have an evident physical meaning but it is differently distributed for signal and background events.

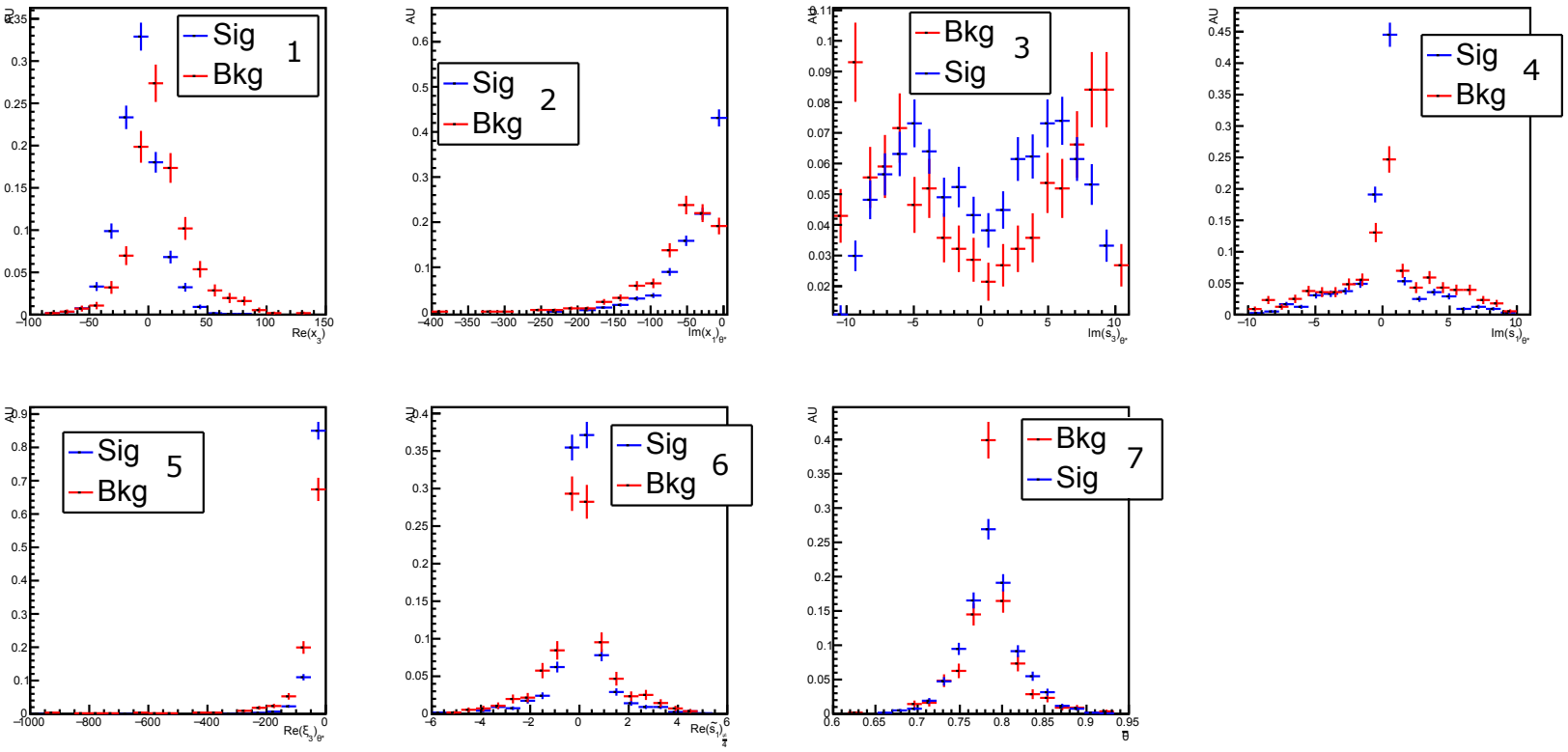


Figure 4.23: The most discriminating variables from the τ reconstruction algorithm presented in this section. The number of each variable refers to the ordering defined on page 144.

4.5.8 Conclusion and prospects

In this section, a method for a full reconstruction of the τ four-momenta in the decay chain $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ has been presented. It has been shown, in particular, that a full reconstruction of the event is possible if:

- the four-momenta of the two three-pions systems and the lengths of the two sides of the decay triangles (given by the B candidate production vertex and the two τ decay vertexes) are measured,
- kinematical constraints using the masses of the B, τ, and ν are imposed.

The relevant equations depend on only one Lorentz invariant event-dependent parameter θ, which is proportional to the differences of the decay times of the two τ. Even though it is possible, in principle, to exactly determine this parameter, a different approach has been used so far, because of the high complexity of the trigonometric equations that fix it. In particular some approximation of the true value of this parameter have been presented and have been used to (approximately) reconstruct the momenta of the two τs and to define some discriminating variables that will be used for the signal search.

The studies presented in this section set the basis for further development of the reconstruction method, some of them presented in Appendix E. In particular, other estimations of the true value of the θ parameter are being studied, as well as the possibility of combining the information from the different approximations to obtain an improved estimation of the true value of θ. In addition, also a different use of the kinematical constraints is being explored. This new strategy can allow, in principle, to leave the invariant mass of the B candidate a free variable (which peaks for signal events) that can be then used for the signal search.

4.6 BDT classification

In order to discriminate the signal against the background in the selected sample, the events are classified through a BDT algorithm. This variable is the one used for the signal search. As already explained in Sec.4.2.4, the BDT classifier is trained using variables similarly distributed between the SS and the OS samples, but that are still discriminating between signal and SS sample. In particular, from the background characterization studies presented in Sec.4.2.3, these are the variables with a weak dependence on the number of charged tracks in the final state. The input variables used for the training of the BDT are the seven introduced in Sec.4.5.7 plus the following ones:

- Min(Tau_DistZ): minimum of the projection along the beam axis of the flight distance of the τ candidates with respect to the approximate B candidate decay vertex ;
- Min(Tau_TAU) : minimum of the decay time of the two τ candidates;
- M_B^{corr}: corrected mass of the B candidate;
- Min(Tau_NumVtxWithinChi2WindowOneTrack): minimum of the "NumVtxWithinChi2WindowOneTrack" variable for the two τ;
- Max(Tau_ENDVERTEX_CHI2): the maximum of the χ^2 of the τ decay vertices;
- $I_{CDF}(B)$: CDF isolation for the B candidate (see definition on page 52).

Fig.4.24 shows the distributions of these variables for signal MC and data SS candidates.

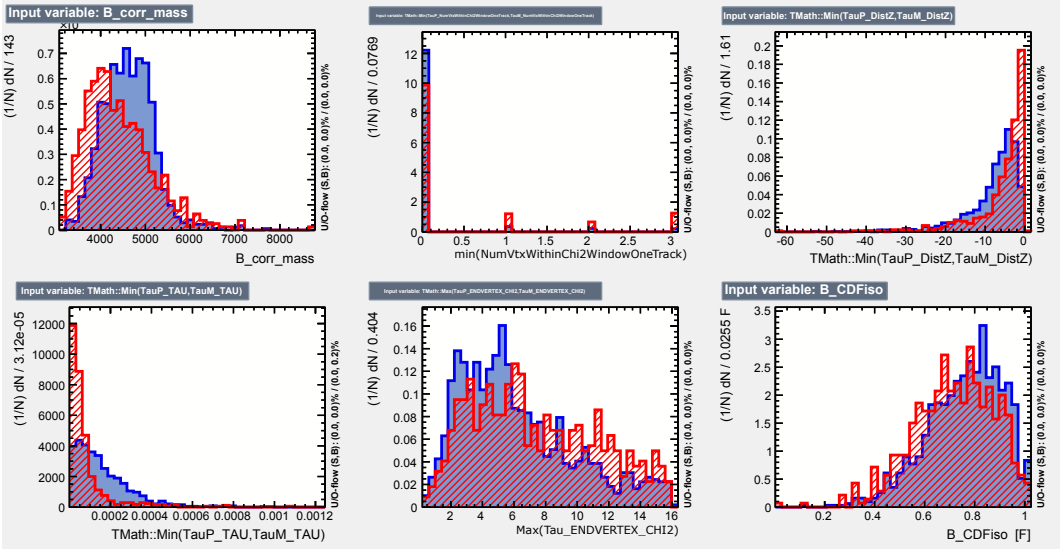


Figure 4.24: Input variables of the BDT classifier (used together with the ones shown in Fig4.23) used for the signal search.

BDT training strategy. The available number of candidate events that can be used to train and test the BDT classifier is of 610 for background (data SS) and 1344 for the $B_s^0 \to \tau\tau$ signal (as reported in Tab.4.11). Because of the limited numbers of events, a simple minded twofold division of the samples to get a test and training samples is not suitable. With so few events the training process is indeed dominated by statistical fluctuations. To overcome this problem and increasing the dimension of the training sample the following strategy is adopted (known also as "*k-folding*" technique [115]):

- both signal and SS sample are split into ten parts,
- for each of the two categories (signal and background) the training samples are made of nine of these ten subsamples; the resulting BDT operator is then applied to the remaining one,
- a cyclical permutation is done in order to apply the BDT classifier to each of the ten subsamples.

The limited statistics available for the training of the BDT also affects the signal-background discrimination because of the low value of the depth reachable in the decision trees. This is a consequence of the small number of cuts that can be applied before the minimum number of events in each leaf is reached. For this reason, the discriminating power of the input variables of the classifier is not fully exploited, because some of them could not be used at all in the decision process even though they are still able to discriminate the signal against the background. An improvement in the signal-background discrimination is thus expected once a larger training sample will be available (see *e.g.* Sec.4.7.2).

Application of the BDT operator. The "*k-folding*" technique leads to ten BDT operators, one for each pair of signal and background subsample not used for the training. The application of these classifier to the signal and data SS subsamples is unambiguous, in the sense that each of those classifiers can be applied to the unique pair of signal and SS subsamples that have not been used in the training of the BDT.

The distributions of the ten classifiers on the respective signal and data SS subsample are reported in Figs.4.25-4.26 respectively.

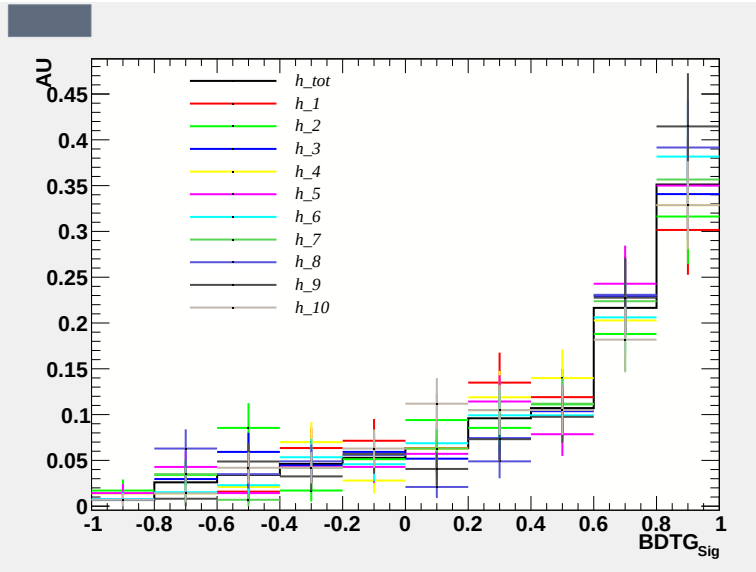


Figure 4.25: Distribution of the BDT values of each of the ten classifiers and their average (continuous histogram, "h_{tot}") on the respective test subsample for signal.

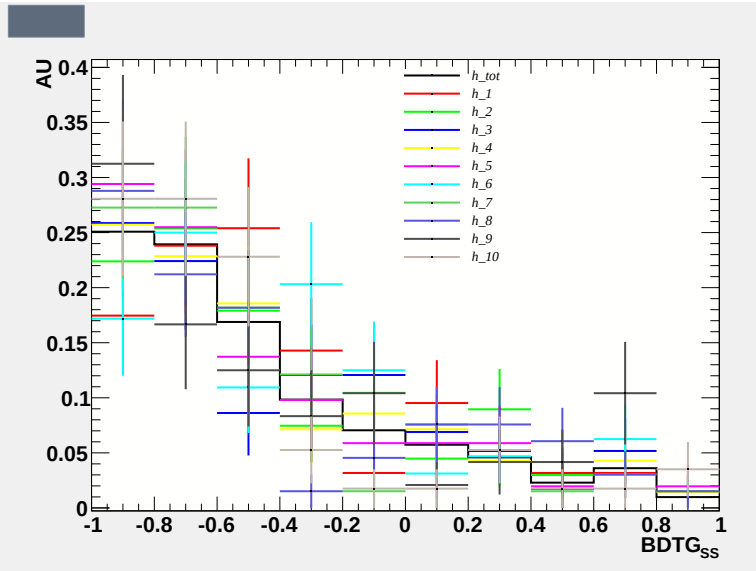


Figure 4.26: Distribution of the BDT values of each of the ten classifiers and their average (continuous histogram, "h_{tot}") on the respective test subsample for SS events.

To obtain the BDT distribution of the whole signal and SS samples the BDT distributions of each subsample must be added together. Nevertheless, a simple minded merging of the ten subsamples to get the distribution of the whole sample is not suitable. The reason is the following: the shape of the output variable of a BDT does not have an absolute meaning. Indeed, even though the same tuning parameters and input variables are used in the training, the output shape of the BDT for signal and background events can vary depending on the training samples (essentially due to statistical fluctuations). Nevertheless what is really significant in the BDT output distributions is the overall separation between the signal and the background samples. This separation should not vary too much, no matter the particular shapes of the signal and background samples. In other words, the statistical fluctuations in the training samples affect not only the overall signal-background separation of the output variable (which is a "natural" effect of statistical fluctuations), but also the shapes of the output variable for the two samples (which is, instead, an artifact of the learning machine). For that reason the output distributions for the ten signal and background subsamples are "normalized" to the same shape before being added together.

Such "normalization" of the BDT output shape can be achieved, for instance, through the flattening transformation

$$x \to \frac{\int_{x_{min}}^{x} f_{sig}(y)dy}{\int_{x_{min}}^{x_{max}} f_{sig}(y)dy} \,, \tag{4.55}$$

that has already been used for the $B^0_{(s)} \to \mu^+\mu^-$ search. Such transformation does not change the overall signal-background separation and, at the same time, makes each signal distribution flat, no matter of the prior BDT output shape. Once the BDT shapes for each of the signal and SS subsamples are compatible with each other they can be added together and their sum can be used to describe the BDT output shape for the whole signal and data SS samples.

The evaluation of the BDT operator on the OS sample is not so straightforward as for the signal and SS sample. This is just because there are ten possible classifiers that can be used and a choice on how to combine them must be made. The computation of the BDT variable on the OS sample is done as follows:

- for each OS event all ten classifiers are evaluated,
- for each of the ten classifiers, the BDT value is transformed according to the flattening with respect to the corresponding signal distribution,
- the average of the 10 BDT values after flattening is computed and assigned to the candidate.

4.6.1 BDT distributions

The distribution of the BDT output for the SS data and the three generated MC signal samples is reported in Fig.4.27.

The following considerations are in order:

- the shapes of three MC generated signal samples are compatible within statistical uncertainties, even though the BDT has been trained to discriminate between the $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ and the data SS, (see Sec.4.2.5). For this reason the analysis presented here is equally sensitive to the sum of the $B^0_s \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$ branching fractions. This last point will be discussed in more details in Sec.4.6.2;
- the region at high values of the BDT has a high sensitivity to the signal. In terms of absolute event yields only one SS candidate falls in this region.

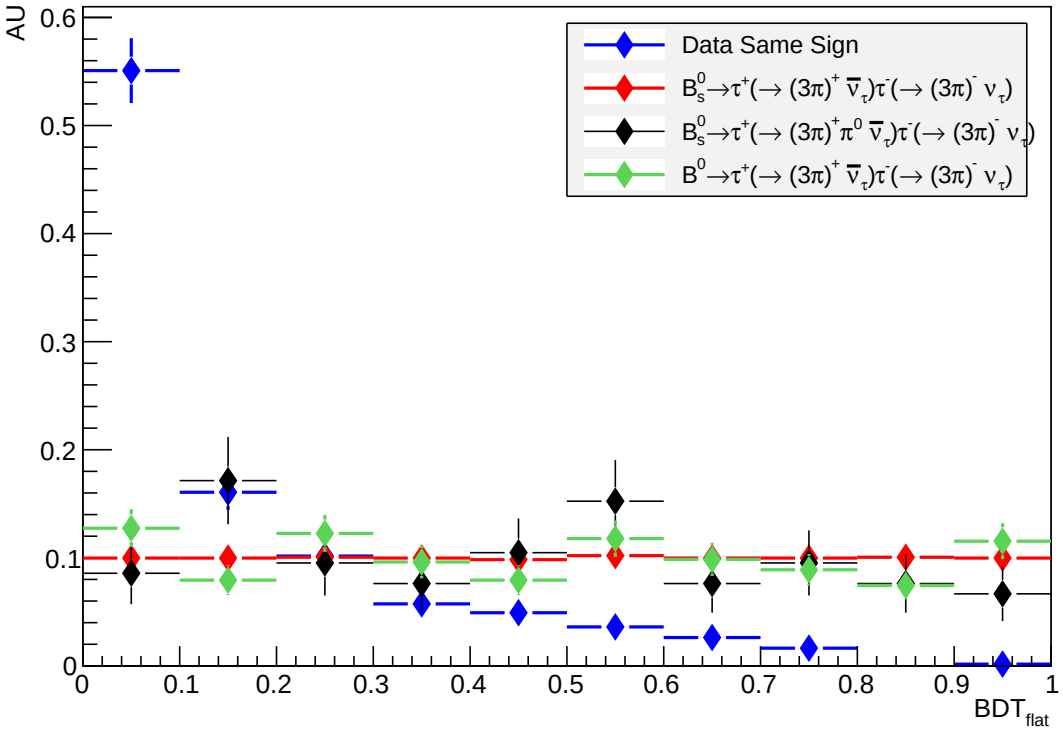


Figure 4.27: Distribution of the BDT variable for $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ MC signal events (red), $B^0 \to \tau^+(\to 2\pi^+\pi^-\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ MC signal events (green), $B^0_{(s)} \to \tau^+(\to 2\pi^+\pi^-\pi^0\bar{\nu}_\tau)\tau^-(\to 2\pi^-\pi^+\nu_\tau)$ signal MC events(black), and SS data (blue). All the distributions are normalized such that their integral all over the BDT range is equal to 1.

The comparison between the distribution of the BDT outputs for the SS and the OS data samples is shown in Fig.4.28(a)-4.28(b). The distribution of the OS sample has been blinded in the most signal sensitive region, *i.e.* for BDT>0.7, in order to avoid any bias in the analysis. The distribution of the SS dataset seems to reproduce quite well the shape of the OS sample. Nevertheless, due to the limited statistics, the knowledge of the SS shape is affected by a relatively large error, if compared to the one on the OS shape. The largest discrepancy between the two shapes is found in the first BDT bin, *i.e.* for BDT<0.1, which is also the one less sensitive to signal.

The discrepancies between the SS and OS BDT shapes are mostly due to the presence, in the latter, of the exclusives modes with only six charged tracks in the final states. How to deal with such modes is still under study at the present and some ideas will be presented in Sec.4.7.

4.6.2 Discussion about the observables

The shapes of the BDT outputs are similarly distributed and compatible within uncertainties for the $B_s^0 \to \tau\tau$ and $B^0 \to \tau\tau$ signals. For this reason the only quantity that can be measured using this variable is the sum N_{tot} of the yields $N_{B_s^0}$ and N_{B^0} for the $B_s^0 \to \tau\tau$ and $B^0 \to \tau\tau$ signals respectively (from now on, with the expression "observed yield of signal events", the sum of the two signals yields will be meant). This total yield N_{tot} is given by a linear combination of the branching fractions of the two modes. In particular, using the definition in eq.4.4 the observed yield of signal events as a function of their branching fraction is given by the following expression

$$N_{tot} \equiv N_{B_s^0} + N_{B^0} = \frac{\mathcal{BR}(B_s^0)}{\alpha_s} + \frac{\mathcal{BR}(B^0)}{\alpha_d} \ . \tag{4.56}$$

The previous equation can be arranged in several ways depending on the quantity that has to be measured. If the quantity of interest is the $\mathcal{BR}(B_s^0)$, the following relation holds:

$$\mathcal{BR}(B_s^0) = \alpha_s \cdot N_{tot} - \frac{\alpha_s}{\alpha_d}\mathcal{BR}(B^0). \tag{4.57}$$

In this way, measuring the total signal event yield and using the available experimental result on the $\mathcal{BR}(B^0 \to \tau\tau)$ it is possible to constrain $\mathcal{BR}(B_s^0 \to \tau\tau)$.

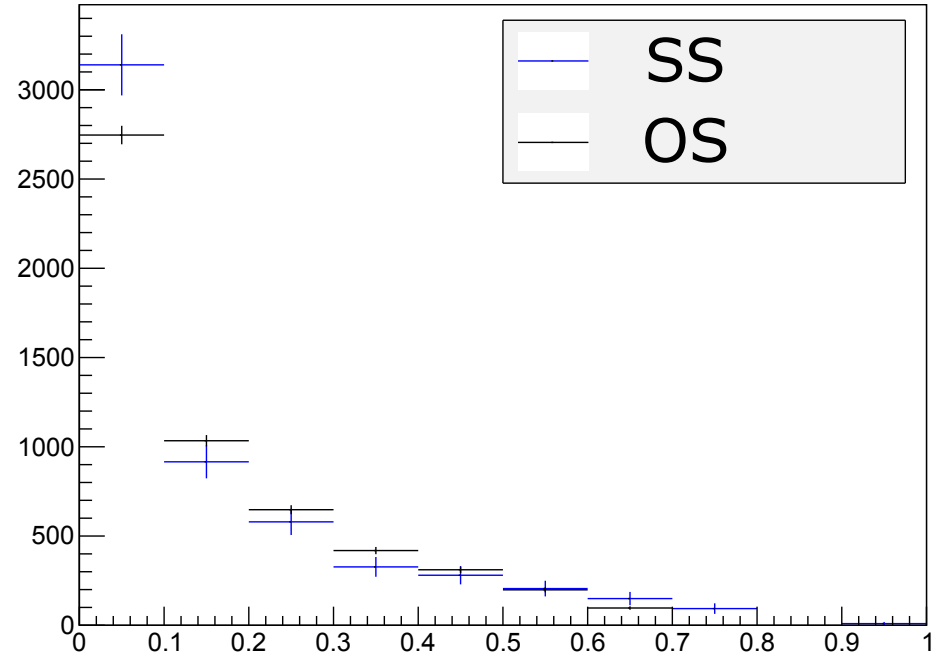
The ratio $\frac{\alpha_s}{\alpha_d}$ is approximately equal to the ratio $\frac{f_d}{f_s}$, and corrections arise from differences in the efficiencies for the B_s^0 and B^0 signals.

4.7 BDT PDF calibration

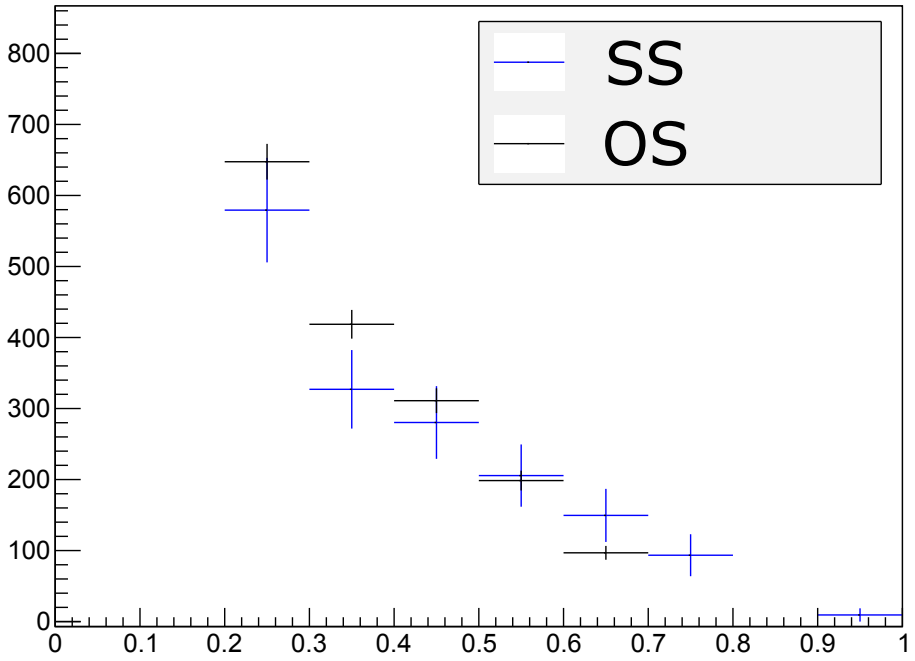
The flattening transformation of the output of the final BDT makes the signal BDT PDF shape being constant over the whole BDT range and the SS background BDT PDF peaking at zero. While the BDT PDF for signal is taken from the MC sample, the calibration of the BDT PDF shape for the background in the OS is less trivial and requires careful understanding of the background sources in the final selected dataset. In the following sections two methods for the background PDF calibration, still under study at the present, will be presented. The first one relies on the use of the data SS sample and MC generated samples accounting for the presence of exclusive decays with six charged tracks in the final state, that are not present in the former; the second method relies instead on the use of control regions in the OS sample to have an inclusive description of the background.

4.7.1 Calibration with the SS sample

Thanks to the choice of the input variables of the BDT classifier, the shape of the SS and OS data samples are by construction quite similar, hence the BDT PDF shape of the SS sample can be used to describe the background in the OS sample, modulo the presence, in the latter, of



(a)



(b)

Figure 4.28: Distribution of the BDT variable for SS (blue) and OS (black) data. The two distributions have the same normalization. The distribution of the OS sample has been blinded in the most signal sensitive region, *i.e.* for BDT>0.7, in order to avoid any bias in the analysis. Fig.4.28(b) is a zoom of Fig.4.28(a).

physical decays with six charged tracks in the final state. The effect of these six charged tracks exclusive modes must then be estimated using MC generated samples.

Monte Carlo generated exclusive background samples. Tab.4.14 lists the branching fractions and the selection efficiencies for the most relevant modes (*i.e.* the ones with the highest rates) generated, mostly with six charged tracks in the final state. This list of modes is not exhaustive and other decays can still contribute to the six tracks final state background in the OS. In Fig.4.29 the BDT distributions for those modes are shown together with the one of the SS sample for comparison. From these distributions it is evident that the data SS cannot efficiently take into account such modes. Indeed, even though they peak at zero as the SS does, their shapes are different.

The expected yields of these generated modes in the whole $3.1 fb^{-1}$ OS dataset, according to the values in Tab.4.14, is (730 ± 250) events. Nevertheless, this estimation does not take into account the presence of other modes, and some more work is still required to correctly estimate the yields of these modes.

In order to include the effects of the six charged tracks modes in the background BDT PDF, the weighted sum of them (the weight of each mode being given by the total efficiency ε reported in Tab.4.14) is added as an independent component in the fit. The OS dataset will be modeled by three PDFs, one for signals, one from the SS dataset, and another for the exclusive modes with six charged tracks. Then the total yield of these modes can be fixed to its expectation or can be extracted from the fit (as a nuisance parameter) by constraining it to its expectation and allowing it to vary within its uncertainty.

No matter which of the previous strategies are used, the use of the MC generated samples for the BDT PDF calibration present some limitations:

- first of all, the list of available MC generated modes is not exhaustive of all the possible dangerous modes that can pollute the sample. Even though each of these decays can be representative of a class of similar modes with respect to topology or kinematics properties, a full characterization of the exclusive modes with six charged tracks through MC simulated samples is quite challenging and some extrapolation and assumption will have to be made;

- systematic errors due to the specific MC decay model have to be taken into account.

Event	$\mathcal{BR}$ ($\times 10^5$)	ϵ_{gen} ($\times 10^2$)	$\epsilon_{pre-sel}$ ($\times 10^2$)	ϵ_{sel} ($\times 10^2$)	ϵ_{tot} ($\times 10^6$)	ε ($\times 10^{11}$)
$B_s^0 \to D_s^{(\star)+}(\to 3\pi\pi^0, \tau\nu) D_s^{(\star)-}(\to 3\pi\pi^0, \tau\nu)$	14.00 ± 6.32	0.29	0.610 ± 0.003	3.09 ± 0.10	0.54 ± 0.02	7.56 ±3.70
$B_s^0 \to D_s^-(\to \tau^-\bar{\nu})\tau^+\nu$	11.24 ± 10.8	2.33	0.300 ± 0.005	4.36 ± 0.38	3.04 ± 0.31	34.16 ± 36.31
$B_s^0 \to D_s^-(\to \tau^-\bar{\nu})3\pi$	3.25 ± 0.72	3.31	0.550 ± 0.007	2.76 ± 0.21	5.02 ± 0.44	16.31 ± 5.05
$B_s^0 \to D_s(\to 3\pi\pi^0(\pi^0))3\pi$	35.10 ± 10.6	3.08	0.720 ± 0.008	0.00 ± 0.01	0.00 ± 0.02	0.00 ± 0.80
$B^0 \to D^{(\star)-}\tau^+\nu$	1.67 ± 0.34	0.38	0.610 ± 0.004	0.70 ± 0.05	0.16 ± 0.01	1.03 ± 0.30
$B^0 \to D^{\star-}(\to D^-\pi^0)\omega(\to \pi^+\pi^-\pi^0)\pi^+$	0.89 ± 0.17	2.54	0.440 ± 0.007	0.09 ± 0.04	0.10 ± 0.05	0.34 ± 0.24
$B^0 \to D^{\star-}(\to D^-\pi^0)3\pi\pi^0$	6.10 ± 1.46	2.60	0.420 ± 0.005	0.26 ± 0.05	0.28 ± 0.06	6.59 ± 3.10
$B^0 \to D^{\star-}(\to D^-\pi^0)3\pi$	2.43 ± 0.48	3.14	0.570 ± 0.005	1.10 ± 0.10	1.90 ± 0.20	17.82 ± 5.67
$B^0 \to D^-3\pi$	7.23 ± 1.30	3.20	0.600 ± 0.005	1.15 ± 0.10	2.20 ± 0.20	61.40 ± 17.57
$B^0 \to D^-(\to (3\pi)^-K^0)3\pi$	19.96 ± 2.88	2.43	0.670 ± 0.008	1.70 ± 0.16	2.76 ± 0.30	212.65 ± 57.10
$B^+ \to D^{\star 0}(\to D^0(\to K^-3\pi^+)\pi^0)3\pi$	51.53 ± 9.70	3.23	0.690 ± 0.008	0.26 ± 0.01	0.57 ± 0.03	113.37 ± 29.07
$B^+ \to D^{\star 0}(\to D^0(\to K^-3\pi^+\pi^0)\pi^0)3\pi$	26.82 ± 7.00	2.72	0.660 ± 0.007	0.01 ± 0.01	0.02 ± 0.02	2.07 ± 2.64
$B^+ \to D^{\star\star}(\to D^{\star 0}(\to D^+\pi^-)\pi^0)\tau^+\nu$	0.27 ± 0.35	2.64	0.330 ± 0.004	0.45 ± 0.08	0.39 ± 0.07	0.40 ± 0.61

Table 4.14: Branching fraction, generator, pre-selection, and selection efficiencies, and expected number of events, for the MC generated exclusives decay modes. $\varepsilon \equiv \mathcal{BR} \cdot \epsilon_{tot} \cdot f_q/f_s$, $\epsilon_{tot} \equiv \epsilon_{gen} \cdot \epsilon_{strip} \cdot \epsilon_{sel}$; $f_d/f_s = 3.86 \pm 0.06$ [99, 100]. D mesons decay as $D \to 3\pi\pi^0$, τ leptons as $\tau \to 3\pi\nu_\tau$.

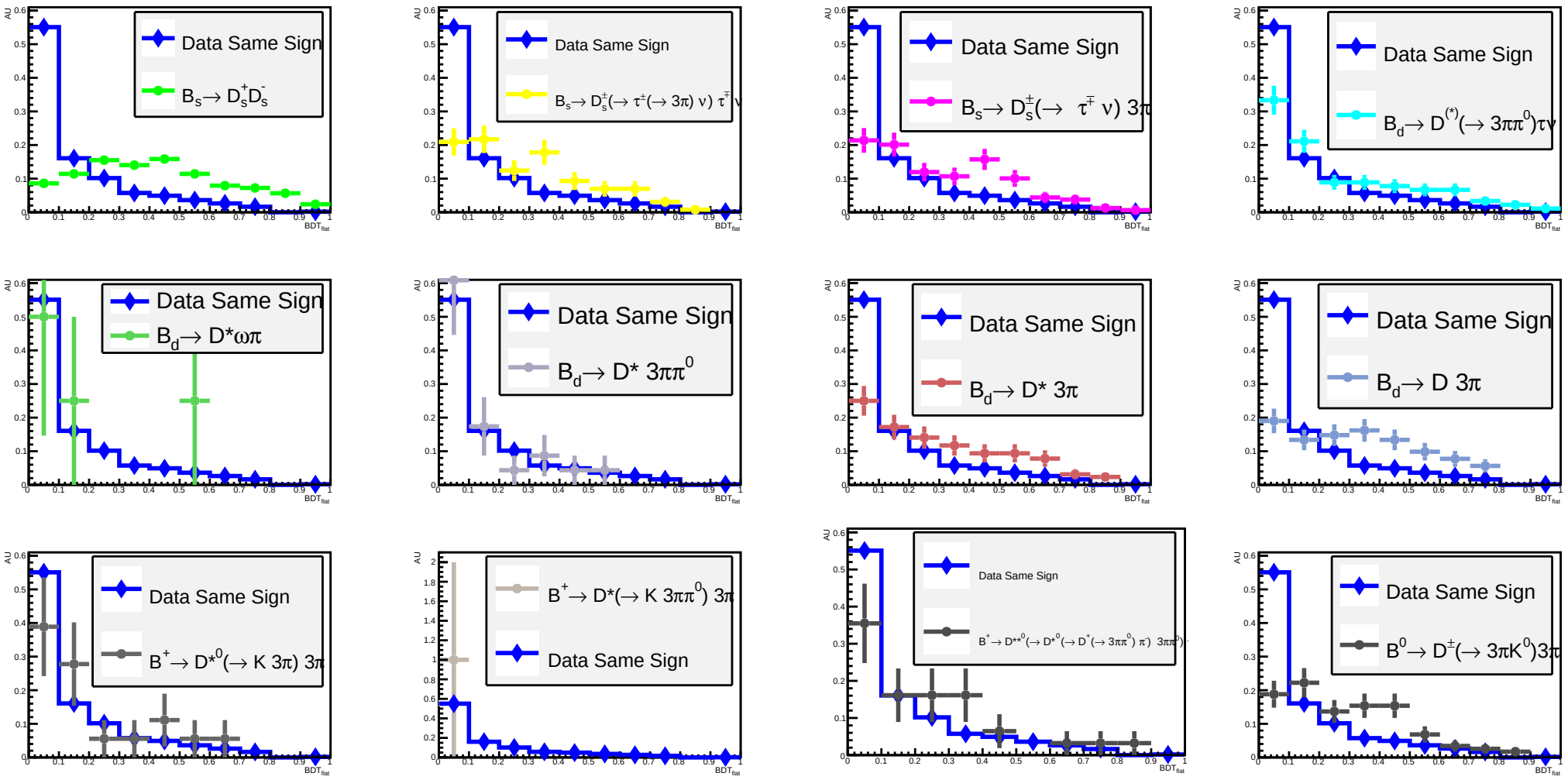


Figure 4.29: Distribution of the BDT values for the twelve generated exclusive decays for which at least one event passes the whole selection chain. All distributions are normalized such that their integral all over the whole BDT range be equal to 1.

4.7.2 Control regions in OS

An alternative method under study is based on the definition of control regions in the OS sample which are reasonably free of signal and can thus be used to calibrate the BDT PDF. The advantage of this method is that it allows an inclusive characterization of the background, without the distinction between modes with six or more charged tracks that is required when using the SS sample together with the MC generated modes. In addition, more events can be used to calibrate the background PDF, thanks to higher statistics in the OS sample with respect to the SS one; this allows reducing the statistical uncertainty on the BDT shape of the background, and also better modeling of the queue of the background distribution in the most sensitive signal region, which is poorly described by the SS sample due to the lack of statistics.

The main point of such method is the definition of the control region in the OS sample. This can be done by inverting some of the selection requirements presented in Sec.4.3. The one which seems most promising is the inversion of the Dalitz plane requirements. In this case the "signal" region for one τ is defined by the requirements reported in Sec.4.3. The control region is defined by requiring one τ candidate inside and the other one outside the signal region in the Dalitz plane. Fig.4.30 shows the definition of the signal and control regions in the Dalitz plane of one τ.

The distribution of the BDT output for OS events in these two regions is shown in Fig.4.31. The distribution of events in the signal region in the Dalitz plane of the τ has been blinded in the high BDT region. From this comparison it seems that the BDT shape of the events in the control region in the Dalitz plane well reproduces the one of the events in the signal region, at least for events with a BDT smaller than 0.8; as a consequence the BDT distribution of events in the control region can be used as a reliable description of the background in the dataset selected for the signal search.

As a prospect, the use of the control region in the OS data sample could allow to define a more discriminating BDT variable by using it as background sample in the training of the MVA operator. Indeed, in this case, the available statistics is higher than the one in the SS, thus allowing to exploit the information of variables that, because of the lack of events, cannot be used when training on the SS sample. In addition, there is no need to use only those variables similarly distributed between SS and OS, so that new inputs, that have not been used so far because too sensitive to the different number of tracks in the final state of events selected in the two samples, can be used for the training of the BDT.

.

4.8 Conclusions and prospects

In this chapter an analysis strategy for the $B^0_{(s)} \to \tau^+\tau^-$ search with the LHCb detector has been presented. In this chapter the analysis of the final state where each τ decays into three charged π and a neutrino has been presented. Another possible final state, where one of the τ goes into a μ and two neutrinos, has been considered also for the signal search. Each of these modes presents some advantages and disadvantages. The $(3\pi, 3\pi)$ final state allows to reconstruct the two τ decay vertices and thus to impose more kinematical constraints but suffers of the lower acceptance of the detector (because of the presence of six charged tracks) and of a smaller $\mathcal{BR}$; the $(3\pi, \mu)$ final state has a higher detector acceptance thanks to the presence of only four tracks among whom there is a muons which increase also the trigger efficiency but the impossibility of reconstructing one τ decay vertex does not allow to impose kinematical constraints.

A new method for the τ four-momenta reconstruction has been presented. It has been shown that a complete reconstruction of the two τ four-momenta is possible by i) imposing kinematical constraints on the masses of the B, τ, and ν, and ii) by exploiting the geometrical and kinematical information obtainable from the quantities that can be measured directly, *i.e.* the four-momenta of the two three-pions systems and the length of the sides of the decay triangle.

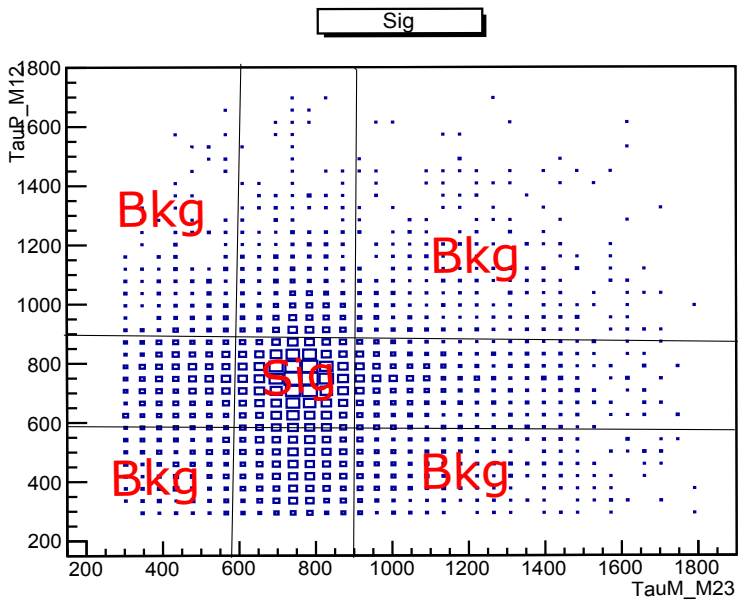


Figure 4.30: Definition of the control region (indicated as "Bkg") using the Dalitz plane of one τ. The signal region is labeled as "Sig".

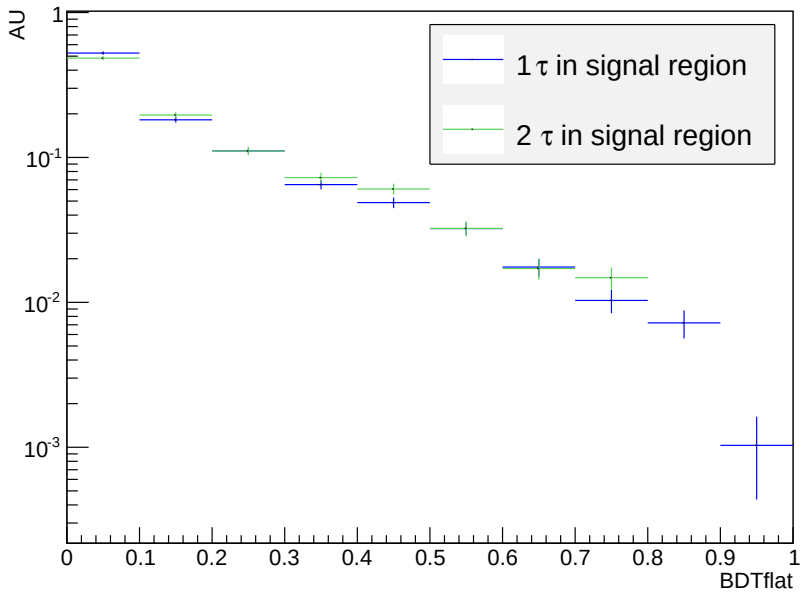


Figure 4.31: Distribution of the BDT variable for OS events in the signal (in green) and the control (blue) regions in the τ Dalitz plane. The last two bins of the distribution of the signal region events have been blinded.

The relevant equations depend on a the Lorentz invariant parameter θ, which is a function of the difference of the decay times of the two τ. Even if this parameter is, at least in principle, exactly determinable as the solution of a trigonometric equation, so far its true value has been estimated from observable quantities, and has been used to find the approximate momenta of the two τ. Different approximations for θ have been presented, and some discriminating variables have been sorted out and used in the experimental search. Currently new ideas are being explored in order to improve the approximation of the θ parameter and to define a different strategy for the resolution of the kinematical equations. This last point allows, in principle, to leave the invariant mass of the two τ (so far used as an external kinematical constraint) as a free peaking variable that can be used for the signal search.

Concerning the experimental search, the main challenge of the analysis is the control of the background. For its description and characterization a sample of unphysical $B \to \tau^{\pm}\tau^{\pm}$ candidates has been selected in data. The use of such sample for the description of the backgrounds (coming from B decays with at least seven charged tracks in the final state) is done using a MC generated sample. After the candidate reconstruction, the analysis strategy is based on a selection aiming to reject as much as possible background events, followed by a classification of the selected events through a BDT operator whose output is used for the signal search. In particular, the definition of the BDT variable is strictly related to the BDT PDF calibration strategy, and two possible strategies are under study at present.

Even though some work is still required before a result can be obtained, the ideas presented in this chapter constitute the basis of the current analysis as well as the starting point for further improvements.

Chapter 5

Conclusions and Prospects

The Standard Model of elementary particle physics provides the most accurate description of nature available so far. Nevertheless, this model is believed to be only a low-energy approximation of a more fundamental underlying theory.

As in the past, flavor observables can provide useful hints to understand the features of possible New Physics Scenarios. In particular the comparison between the measured values of flavor observables with their Standard Model predictions, can shed light on the pattern of the New Physics.

One of the "golden" modes are the rare dileptonic $B^0_{(S)} \to \ell\bar{\ell}$ decays. These modes are generated by a Flavor Changing Neural Current and are thus loop suppressed. In addition they receive, in the Standard Model, an additional suppression due to helicity conservation. They offer a wide range of observables, such as the branching ratio or the effective lifetime $\tau_{\ell\bar{\ell}}$. The theoretical prediction for these processes includes electromagnetic and QCD corrections up to the order $\mathcal{O}(\alpha_{em})$ and $\mathcal{O}(\alpha_s^2)$ respectively. The major source of uncertainty in these predictions is given by the knowledge of the $B^0_{(s)}$ decay constant and of the V_{CKM} matrix elements.

$B^0_{(s)} \to \mu^+\mu^-$. Experimentally the modes with two muons are very clear and thanks to a dedicated detector design, the first evidence for the $B^0_s \to \mu\mu$ has been obtained in 2012 (after a search last thirty years) by the LHCb Collaboration analyzing a dataset of $2fb^{-1}$ integrated luminosity. This analysis has been updated in 2013 with the whole $3fb^{-1}$ collected by LHCb during the first LHC Run, and the dataset has been used, together with the one collected by the CMS collaboration, for a combined analysis, resulting in the first observation of the $B^0_s \to \mu\mu$ mode and the first evidence of $B^0 \to \mu\mu$ in 2015. The measured branching fraction are:

$$\mathcal{BR}(B^0_s \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9} \,, \tag{5.1}$$

$$\mathcal{BR}(B^0 \to \mu^+\mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10} \,. \tag{5.2}$$

The branching ratio of the B^0_s mode is slightly below the Standard Model prediction (1.2σ) while the one of the B^0 is above of a factor ~ 3 (2.2σ). The combined LHCb and CMS analysis has also provided the first measurement of the ratio of the two branching fractions $\mathcal{R}$:

$$\mathcal{R} = 0.14^{+0.08}_{-0.06};. \tag{5.3}$$

This parameter is relevant to set constraints on MFV scenarios and is compatible with the Standard Model prediction at the 2.3σ level.

The measured branching fractions for the two modes, as well as their ratio, allow to exclude huge New Physics enhancements to those quantities, even if they sill leave some room for non-Standard Model contributions. In particular some deviations from the predicted values are observed even though none of them is significant enough to be interpreted as a clear signal of

New Physics. In this contest, the new data recorded in the coming years, during the second LHC Run, will allow to strengthen or alleviate the above-mentioned discrepancies. Surely, the study of the $B^0_{(s)} \to \mu\mu$ modes is moving from a simple discovery regime, up to few years ago, to a precision measurement era. In particular the next analysis which will be performed with the new datasets available in the future will focus on the improvement of the sensitivity for the B^0 mode, and the measurement of the effective lifetime $\tau_{\mu\mu}$. This last observable can still show up New Physics effects even though the branching fraction is in agreement with the Standard Model prediction.

$\boldsymbol{B^0_{(s)} \to \tau^+\tau^-}$**.** The modes with two electrons or two taus in the final states are much more challenging experimentally. The former is out of the experimental reach in the near future, because of its small predicted branching fraction. The modes with two τ have a higher branching fraction, because of a smaller helicity suppression. Nevertheless they are quite challenging under the experimental point of view, because of the missing energy carried away by the ν_τ in the secondary τ decays, which does not allow for a full reconstruction of their momenta. In addition, the higher level of background with respect to channels with the two μ in the final states increase the complexity of the search. For this reason these modes are largely unexplored, made exception for an upper limit from the BABAR collaboration on the $B^0 \to \tau\tau$ mode, still four orders of magnitudes above the predicted value in the Standard Model. Despite the experimental difficulties, these modes are triggering an increasing interest in connection to hints of Lepton Flavor Universality New Physics scenarios. In addition, correlations between the branching ratio and other flavor observables related to the $B^0_{(s)} - \bar{B}^0_{(s)}$ mixing, notably the ratio $\Delta\Gamma_{(d,s)}/\Gamma_{(d,s)}$, have been pointed out in the past; deviation from the Standard Model in these branching ratios could also help to accommodate the anomalous like-sign dimuon asymmetry observed by the D‰ and CDF collaboration at Tevatron.

The work presented in this thesis represents the first attempt to set up an analysis chain for the $B^0_s \to \tau^+\tau^-$ search. This is performed with the LHCb detector, by looking at the final state where each of the two τ goes into three charged pions and a neutrino. The main issues of the analysis are the reconstruction of the signal, made challenging by the presence of two undetectable neutrinos in the final state, and the understanding and characterization of the sources of background. On both sides a huge effort has been done and is still going on to optimize the signal selection and reconstruction. A discriminating variable has already been defined for the signal search and can already be used to have at least a preliminary result. Even though some work is still required to understand and model the most dangerous sources of background, a measurement of a non trivial value of the upper limit (*i.e.* smaller than the ones obtained from indirect constraints) is expected. In addition new ideas concerning the signal reconstruction method as well as the characterization of the background leave room for further improvements in the future.

A final remark concerns the fact that the $B^0(s) \to \tau^+\tau^-$ decays are considered among the most interesting modes to be studied in the future high luminosity e^+e^- machines currently under study, *e.g.* the Future Circular Collider e^+e^- (FCC-ee). The high statistics sample collected with such facilities together with the particularly clean environment of a leptonic machine will allow the observation of the signal and the measurement of angular observables related to the final state particles coming from the decay of the two τ. A measurement of the τ polarization, for example, is then feasible allowing the study of several additional observables not accessible with the di-muon or di-electron final states.

Appendices

Appendix A

Muons isolation variables

Among the variables most discriminating between signal and background there are the so-called *isolation* variables. These variables measure the activity in terms of other particles in the vicinity of the two muons or their displaced vertices. In a signal event, indeed, there shouldn't be any track close to one μ, except for the other μ used to make the $B^0_{(s)} \to \mu\mu$ candidate; instead, in a combinatorial background event, where the two μ's used to make the $B^0_s \to \mu\mu$ candidate originate from two different b-hadron decays, is more likely that other tracks, in addition to the other μ, are close to each of the two μ's.

In more details, the standard isolation of the two muons used in the analysis presented here, is defined as the number of long tracks making a "good" vertex with each of the two muons from the $B^0_{(s)} \to \mu\mu$ candidate [85]. To compute it, for each long track in the event the vertex TV that it forms with the $\mu^\pm$ tracks is found. The TV is defined as the point of equal and minimal distance between the two tracks. Then, for each of the two muons $\mu^\pm$ from the $B^0_{(s)} \to \mu\mu$ candidate, the number $n^\pm$ of tracks selected by a set of rectangular cuts on the following quantities:

- $pvdis \equiv (\vec{r}_{PV} - \vec{r}_{TV})$,
- $svdis \equiv (\vec{r}_{SV} - \vec{r}_{TV})$,
- DOCA between track and μ,
- the $IP\chi^2$ of the track with respect to the PV,
- angle between the track and the μ,
- fc defined, for each pair of a muon and a generic tracks, as

$$fc = \frac{|\vec{P}_\mu + \vec{P}_{tr}| \cdot \alpha^{\mu+tr,PV}}{|\vec{P}_\mu + \vec{P}_{tr}| \cdot \alpha^{\mu+tr,PV} + P_{T\mu} + P_{Ttr}} \tag{A.1}$$

  where

  - $\vec{P}_\mu$, $\vec{P}_{tr}$ are the spatial momenta with respect to the beam axis of the μ and the track,
  - $P_{T\mu}$, P_{Ttr} are the transverse momenta with respect to the beam axis of the μ and the track,
  - $\alpha^{\mu+tr,PV}$ is the angle between the sum of the momentum of the μ and the track and the direction defined by the PV and the common vertex between the μ and the track.

  By definition $fc \to 0$ if $\vec{P}(track + \mu)$ points toward the PV,
- *ghost prob* of the track (the reader is referred to pag.33),
- track χ^2

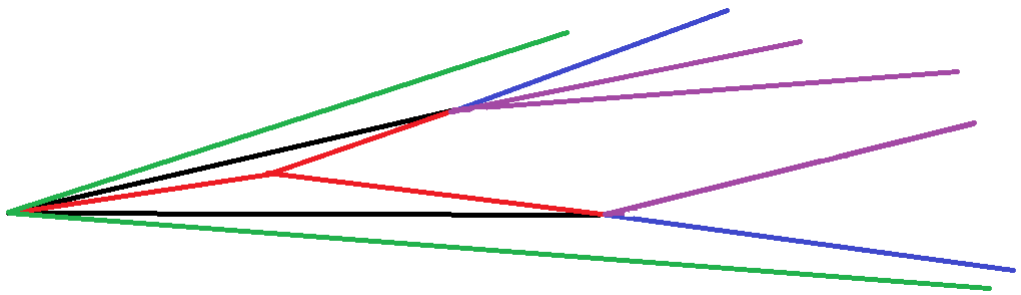


Figure A.1: A combinatorial background event: the two b-hadrons (black lines) produced in the event, together with other isolating tracks (in green), decay semileptonically giving two μ's (in blue), which are reconstructed as a $B^0_{(s)} \to \mu\mu$ candidate (in red), and other non-isolating tracks (in violet).

is found.

The requirements are tuned on generic $b\bar{b}$ MC in order to discriminate two kinds of tracks (see also Fig.A.1):

- *non-isolating tracks* which share an ancestor with the $\mu^\pm$ from the $B^0_{(s)} \to \mu\mu$
- *isolating tracks* defined as the all "non-isolating"

Computing the muon isolation consist in selecting and counting the number $n^\pm$ of non-isolating tracks for each μ.

Appendix B

Data-MC comparison for isolation inputs

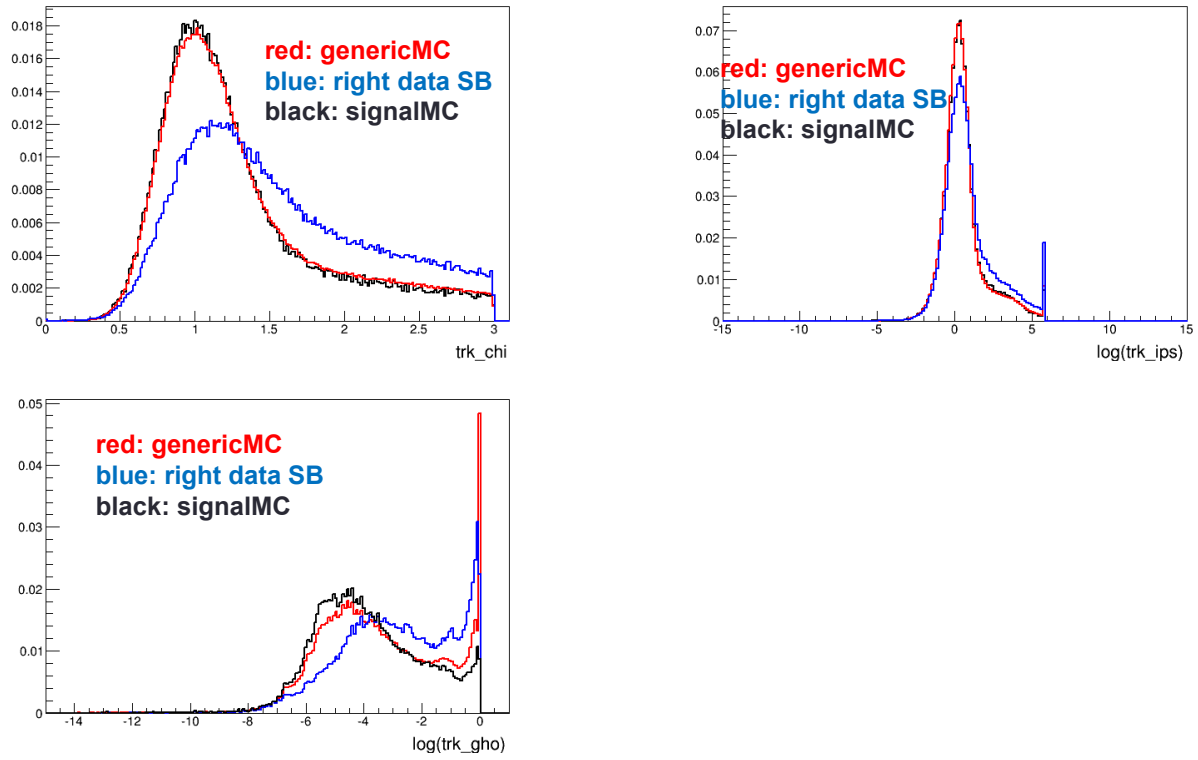


Figure B.1: Distribution of input variables used to train the isoBDT on which the MC simulations (both for signal and for the generic $b\bar{b} \to \mu^+\mu^- X$) does not well reproduce the distribution in data (taken from the high di-muon invariant mass sideband). Top-left: track χ^2; top-right: the logarithm of the IPχ^2 of the tracks; bottom-left: the logarithm of the ghost probability of the track.

Appendix C

Impact of H related variables on signal-background discrimination.

In this appendix the improvement in the signal background discrimination when the maximal information coming from the knowledge of the parameters of the H matrix introduced in eq.(4.15) is shown.

A MC generated sample of $B \to D^{\star}\tau\nu$ is considered as background for the $B \to \tau\tau$ signal. The detector acceptance and resolution effects have not been taken into account, and no selection requirements have been applied.

The signal-background discrimination is evaluated looking at the output of several BDT algorithms using different sets of input variables. In particular the following three sets of variables have been used:

- Geometrical (Vert $\oplus$ IP):
  - the maximum and minimum lengths of the sides of the triangle
  - maximum and minimum of the IP of the 2 $(3\pi)^{\pm}$ systems with respects to the B^0 origin production vertex
- Kinematic (Masses):
  - maximum and minimum invariant mass of the two $(3\pi)^{\pm}$ systems
  - maximum and minimum of the corrected mass of the two $\tau^{\pm}$ candidates
  - invariant mass of the 6π system
  - corrected mass of the 6π system
  - sum and difference of the ν momentum components orthogonal to the decay plane
- H variables (H):
  - $\hat{\tau}_B$, $\hat{\tau}_{\pm}$
  - $t_{\pm}$
  - $\lambda_{\pm}$

Several BDT algorithms have been trained on a sample of 50000 events both for signal and for background (and tested on an equivalent set), using different sets of input variables:

- Vert $\oplus$ IP
- H
- Masses

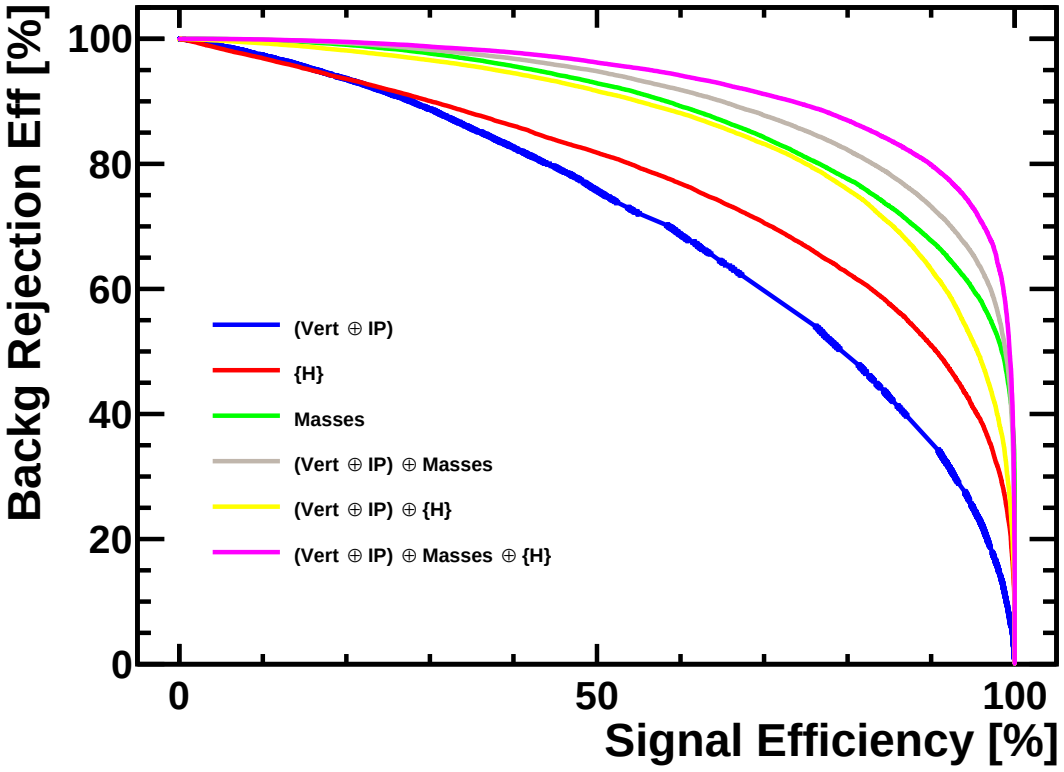


Figure C.1: ROC curves obtained with the output variables of several BDT classifiers trained with different sets of input variables. A statistically significant improvement is observed (pink line) when the elements of the H matrix are added to the set of the input variables of the classifier.

- (Vert ⊕ IP) ⊕ Masses
- (Vert ⊕ IP) ⊕ H
- (Vert ⊕ IP) ⊕ Masses ⊕ H

and the resulting ROC curves are reported in Fig.C.1-C.2

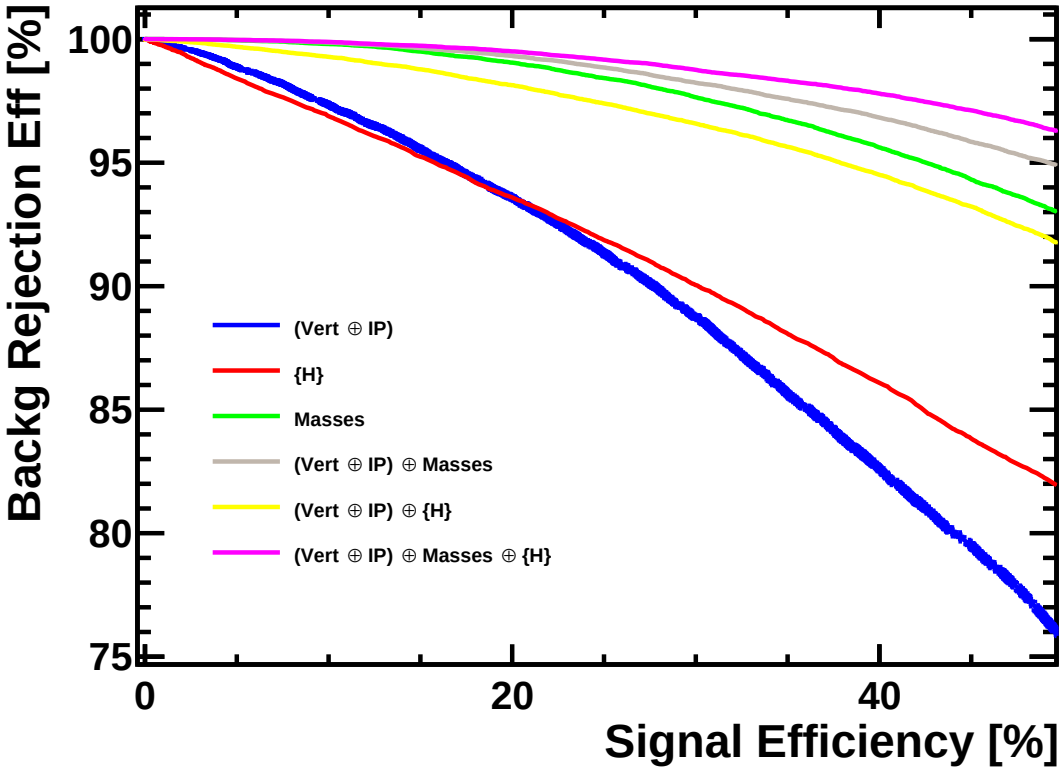


Figure C.2: ROC curves obtained with the output variables of several BDT classifiers trained with different sets of input variables. A statistically significant improvement is observed (pink line) when the elements of the H matrix are added to the set of the input variables of the classifier. Zoom of Fig.C.1 in the low signal efficiency region.

Appendix D

Solution of the system for $B^0_{(s)} \to \tau^+\tau^-$ signal reconstruction

With the following definitions:

$$\begin{cases} A \equiv \bar{b}_+^2 - 1 \\ B \equiv \bar{b}_+\bar{c}_+ \\ C \equiv \bar{c}_+^2 \\ D \equiv \bar{a}_+\bar{b}_+ \\ E \equiv \bar{a}_+\bar{c}_+ \\ F \equiv \bar{a}_+^2 - M_+^2 \\ \\ P \equiv \bar{b}_-^2 - 1 \\ Q \equiv \bar{b}_-c_- \\ R \equiv \bar{c}_-^2 \\ S \equiv \bar{a}_-\bar{b}_- \\ T \equiv \bar{a}_-\bar{c}_- \\ U \equiv \bar{a}_-^2 - M_-^2 \end{cases} \tag{D.1}$$

where $\bar{a}_\pm \equiv \frac{a_\pm}{|\vec{\bar{w}}_\pm|}$, $\bar{b}_\pm \equiv \frac{b_\pm}{|\vec{\bar{w}}_\pm|}$, and $\bar{c}_\pm \equiv \frac{c_\pm}{|\vec{\bar{w}}_\pm|}$ ($a\pm$, $b_\pm$, and $c_\pm$ defined in eqs.(4.41-4.43), and

$$x \equiv \sigma_+|\vec{\bar{w}}_+| \quad , \quad y \equiv \sigma_-|\vec{\bar{w}}_-| \tag{D.2}$$

the eq.(4.45) can be written as:

$$\begin{cases} Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F &= 0 \\ Px^2 + 2Qxy + Ry^2 + 2Sx + 2Ty + U &= 0 \end{cases} \tag{D.3}$$

with $AC - B^2 < 0$ and $PR - Q^2 < 0$.

To solve the system in eq.D.3 the coordinates of a point on the first hyperbola are parametrized as a function of a parameter ξ as follow:

$$\begin{cases} Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0 \\ y = y_0 + \xi(x - x_0) \end{cases} \tag{D.4}$$

being $Ax_0^2 + 2Bx_0y_0 + Cy_0^2 + 2Dx_0 + 2Ey_0 + F = 0$ (*i.e.* $(x_0, y_0) \in \mathcal{C}_1$).

Is found that

$$
\begin{aligned}
&Ax^2 + 2Bxy_0 + 2Bx\xi(x - x_0) + Cy_0^2 + 2Cy_0\xi(x - x_0) \\
&+ C\xi^2(x - x_0)^2 + 2Dx + 2Ey_0 + 2E\xi(x - x_0) + F = 0 \\
&Ax^2 - Ax_0^2 + 2Bxy_0 - 2Bx_0y_0 + 2Bx\xi(x - x_0) + 2Cy_0\xi(x - x_0) \\
&+ C\xi^2(x - x_0)^2 + 2Dx - 2Dx_0 + 2E\xi(x - x_0) = 0 \\
&A(x^2 - x_0^2) + 2By_0(x - x_0) + 2Bx\xi(x - x_0) + 2Cy_0\xi(x - x_0) \\
&+ C\xi^2(x - x_0)^2 + 2D(x - x_0) + 2E\xi(x - x_0) = 0 \\
&A(x + x_0) + 2By_0 + 2Bx\xi + 2Cy_0\xi + C\xi^2(x - x_0) + 2D + 2E\xi = 0
\end{aligned}
\tag{D.5}
$$

giving the following expressions:

$$
\begin{cases}
x(\xi) = \frac{C\xi^2 x_0 - 2Cy_0\xi - 2E\xi - 2D - Ax_0 - 2By_0}{p_2(\xi)} \\
y(\xi) = -\frac{C\xi^2 y_0 + 2Bx_0\xi^2 + 2E\xi^2 + 2D\xi + 2Ax_0\xi - Ay_0}{p_2(\xi)}
\end{cases}
\tag{D.6}
$$

being

$$
p_2(\xi) \equiv C\xi^2 + 2B\xi + A \; . \tag{D.7}
$$

Now, a starting point (x_0, y_0) must be chosen as a function of the parameters of the conic $\mathcal{C}_1$. To do that, a generic point (x_0, y_0) satisfying the equation defining $\mathcal{C}_1$ is considered:

$$
Cy_0^2 + 2Bx_0y_0 + 2Ey_0 + Ax_0^2 + 2Dx_0 + F = 0 \tag{D.8}
$$

from which it follows, solving for y_0,

$$
y_0 = \frac{-(Bx_0 + E) \pm \sqrt{B^2x_0^2 + 2BEx_0 + E^2 - ACx_0^2 - 2CDx_0 - CF}}{C} \; . \tag{D.9}
$$

To make this y_0 meaningful, x_0 must be chosen in such a way that the argument of the square root be positive, *i.e.*

$$
\mathcal{P}(x_0) \equiv (B^2 - AC)x_0^2 + 2(BE - CD)x_0 + E^2 - CF > 0 \; . \tag{D.10}
$$

$\mathcal{P}(x_0)$ is the equation of a parabola in x_0 with second derivative positive (being such the coefficient of x_0^2). Moreover, using eqs.D.1 it can be shown that the following relations hold

$$
BE - CD = 0 \; ,
$$

$$
E^2 - CF = c_+^2 M_+^2 > 0
$$

so that the parabola $\mathcal{P}(x_0)$ is even for exchange $x_0 \to -x_0$ and has its minimum (which is positive!) for $x_0 = 0$.

As starting point (x_0, y_0)

$$
(0, \frac{\sqrt{E^2 - CF} - E}{C}) \tag{D.11}
$$

can be chosen.

Setting

$$\begin{cases} a \equiv Cx_0 \\ b \equiv -(2Cy_0 + 2E) \\ c \equiv -(2D + Ax_0 + 2By_0) \\ \\ d \equiv -(Cy_0 + 2Bx_0 + 2E) \\ e \equiv -(2D + 2Ax_0) \\ f \equiv Ay_0 \\ \\ l \equiv C \\ m \equiv 2B \\ n \equiv A \end{cases} \tag{D.12}$$

it is found that

$$x(\xi) = \frac{a\xi^2 + b\xi + c}{l\xi^2 + m\xi + n} \quad , \quad y(\xi) = \frac{d\xi^2 + e\xi + f}{l\xi^2 + m\xi + n} \quad . \tag{D.13}$$

Inserting these expressions in the second equation of system D.3 the following equation for ξ is obtained:

$$\Gamma\xi^4 + \Phi\xi^3 + \Psi\xi^2 + \Omega\xi + i = 0 \tag{D.14}$$

being

$$\begin{gathered} \Gamma \equiv Pa^2 + 2Qad + Rd^2 + 2Sal + 2Tdl + Ul^2 \ , \\ \Phi \equiv 2Pab + 2Qae + 2Qbd + 2Rde + 2Sam + 2Sbl + 2Tdm + 2Tel + 2Ulm \ , \\ \Psi \equiv Pb^2 + 2Pac + 2Qaf + 2Qbe + 2Qcd + Re^2 + 2Rdf + 2San + \ , \\ 2Sbm + 2Scl + 2Tdn + 2Tem + 2Tfl + Um^2 + 2Uln \ , \\ \Omega \equiv 2Pbc + 2Qbf + 2Qce + 2Ref + 2Sbn + 2Scm + 2Ten + 2Tfm + 2Umn \ , \\ i \equiv Pc^2 + 2Qcf + Rf^2 + 2Scn + 2Tfn + Un^2. \end{gathered} \tag{D.15}$$

Appendix E

Latest developments of the full reconstruction of the $B^0_{(s)} \to \tau^+\tau^-$ events

In this section the latest developments of the method for the signal reconstruction will be presented. In particular two main "directions" have been pursued:

- improve the approximation of the θ parameter,
- exploit in a more efficient way the kinematic equations (4.36), (4.37).

It will be clear in the following that these two aspects are correlated each other.

θ approximation. An improved approximation of the θ parameter is based on eq.(4.18). In particular that relation has been used by:

- using the spatial components of the $w_\pm$ and $p_{3\pi^\pm}$ four-vectors (being the $w^0_\pm$ components not known),
- approximating the four-momenta of the $\tau^\pm$, $p_\pm$, with the four-momenta $p_{3\pi^\pm}$ of the $3\pi^\pm$ system.

In this way a matrix $H_{3\pi}$ is found and the rotation angle $\theta_{3\pi}$ that diagonalizes it can be used as an approximation of the true value of θ. The comparison between the distribution of the true value of θ and the approximation $\theta_{3\pi}$ for MC generated signal events is shown in Fig.E.1, while Fig.E.2 shows the correlation between the two quantities.

Kinematic equations. In order to fully exploit the available information encoded in the three kinematic equations (4.36), (4.37), these equations have been rearranged as follows ($s \equiv p_+ \cdot p_-$):

- two s independent equations: $\mathcal{S}_\pm(\sigma_\pm, \theta) = 0$,
- one s linearly dependent equations: $s = \mathcal{S}_0(\sigma_\pm, \theta)$.

Also in this case the three equations $\mathcal{S}_{\pm,0}(\sigma_\pm, \theta) = 0$ represent the equations of hyperbola, independently of the value of θ.

Once the equations are re-arranged in this form, it is possible to find $\sigma_\pm$ by solving the system of the two s-independent equations $\mathcal{S}_\pm(\sigma_\pm, \theta) = 0$ where θ is replaced by an approximation (*e.g.* $\theta_{3\pi}$ or $\bar{\theta}$); then the s quantity can be expressed as a function of θ and $\sigma_\pm$. The main advantage of this strategy lies in the fact that s is distributed for signal as a Dirac peak (for the

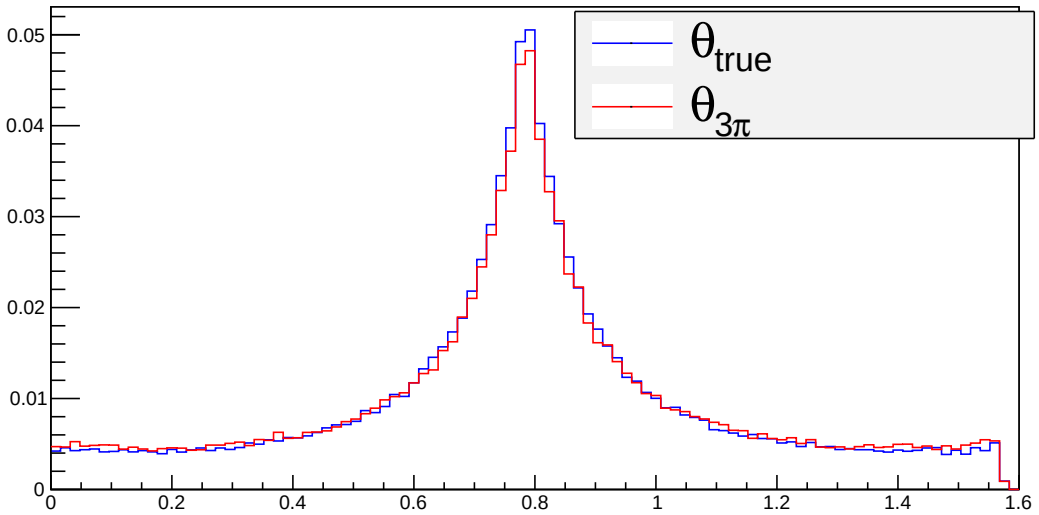


Figure E.1: Distribution of the θ angle (blue) and the $\theta_{3\pi}$ approximation (red) for MC generated signal events.

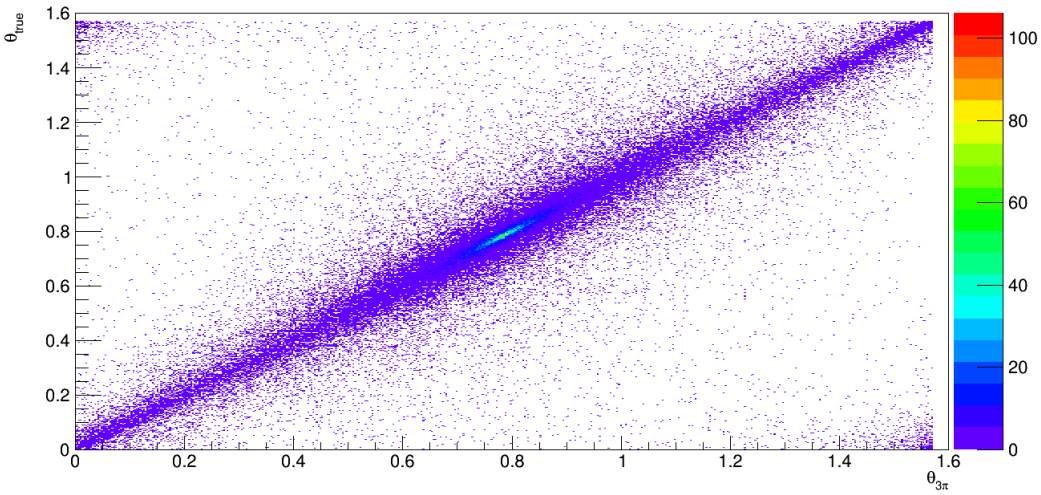


Figure E.2: θ angle (on the y-axis) as a function of $\theta_{3\pi}$ (x-axis) for MC generated signal events.

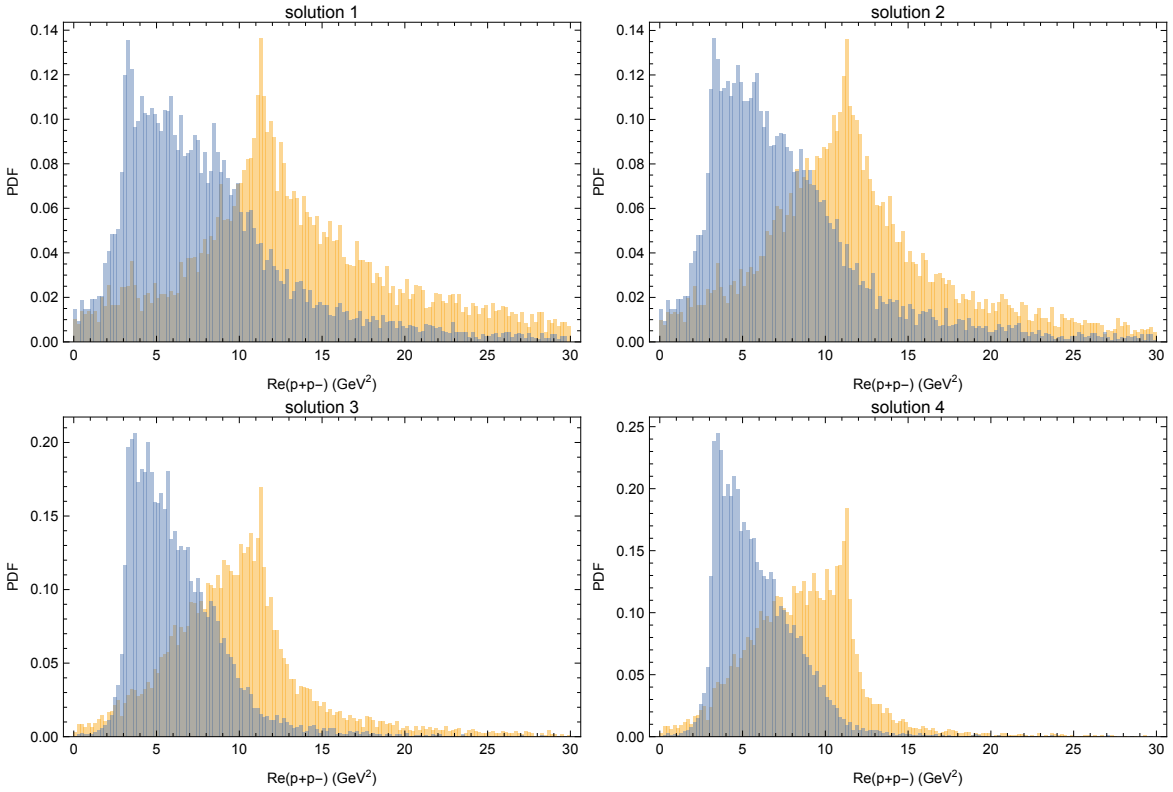


Figure E.3: Distributions of the four possible values of s, approximated through $\theta_{3\pi}$, for signal MC generated events (orange) and $B^0 \to \bar{D}\tau^+\nu_\tau$ background MC generated events (light blue).

B_s^0 at the value $11.24\frac{GeV^4}{c^2}$) and for that reason is more discriminating than the decay times of the B_s^0 and $\tau^\pm$ that are accessible after the full reconstruction of the event.

Being the $\sigma_\pm$ the solutions of a system of second degree in two unknowns, four possible solutions can exist and so four possible values for s, whose distributions, for signal MC and the $B^0 \to \bar{D}\tau^+\nu_\tau$ background is shown in Fig.E.3.

A criteria to choose the correct solutions among the possible four is not known at present and further studies are needed to understand if it is possible to determine it *a-priori*. In addition, because of the approximation of θ, imaginary solution can appear. In the studies presented here only those events for which at least a real solution exists are considered and, if more than one real solution exists, the one which gives the value of s closest to the one of the signal ($11.24\frac{GeV^4}{c^2}$) has been used to reconstruct to invariant mass of the B_s^0.

All the results presented in the following have been obtained by approximating the θ parameter through $\theta_{3\pi}$.

Fig.E.4 shows the distribution of the reconstructed invariant mass of the B_s^0 candidate for signal and the $B^0 \to \bar{D}\tau^+\nu_\tau$ background. Remarkably, the choice of the solution corresponding to the value of s closest to the one of signal does not create a false peak of the background at the signal value.

The effect of the detector resolution and of the emission of radiation from particles in the final state does not change the position of the peak for the signal, while it just spreads the distribution of the reconstructed invariant mass of the B_s^0 candidate around the peak position, as it is shown in Fig.E.5.

Finally, Fig.E.6 shows the comparison between the distribution of the reconstructed invariant mass of the B_s^0 candidate for signal MC simulated signal events (where the simulation includes also the detector resolution and the radiation emission from charged particles in the final state of the decay) and for data SS.

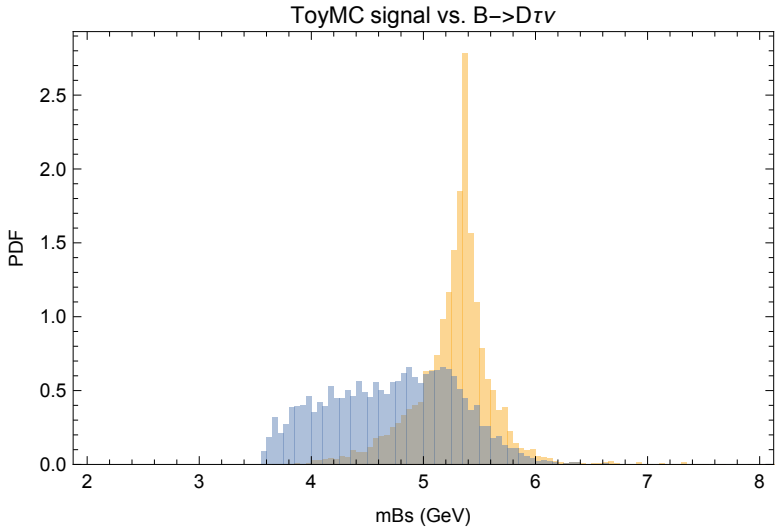


Figure E.4: Distribution of the "best" reconstructed invariant mass of the B_s^0 candidate (*i.e.* whose value is closest to the nominal one) for MC generated signal events (orange) and the $B^0 \to \bar{D}\tau^+\nu_\tau$ background (light blue).

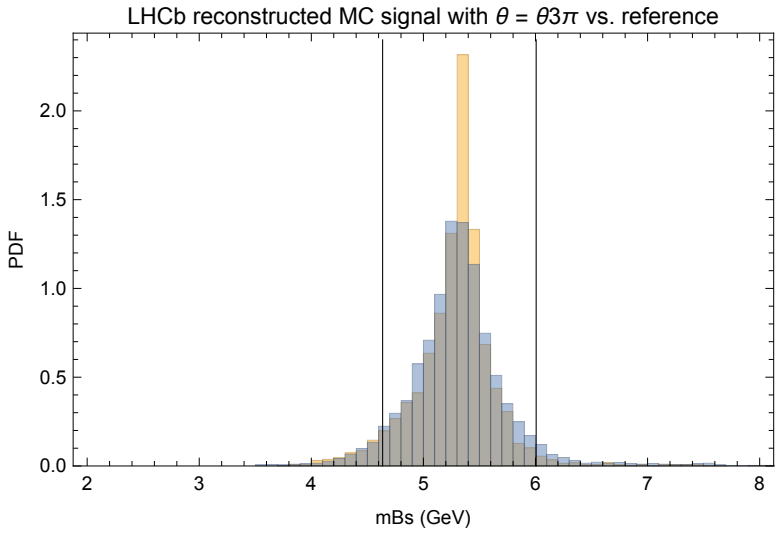


Figure E.5: Distribution of the "best" reconstructed invariant mass of the B_s^0 candidate (*i.e.* whose value is closest to the nominal one) for MC simulated signal events (orange) and MC simulated signal events which take into account the effects of the resolution of the detector and of the radiation emission from charged particles in the final state (light blue).

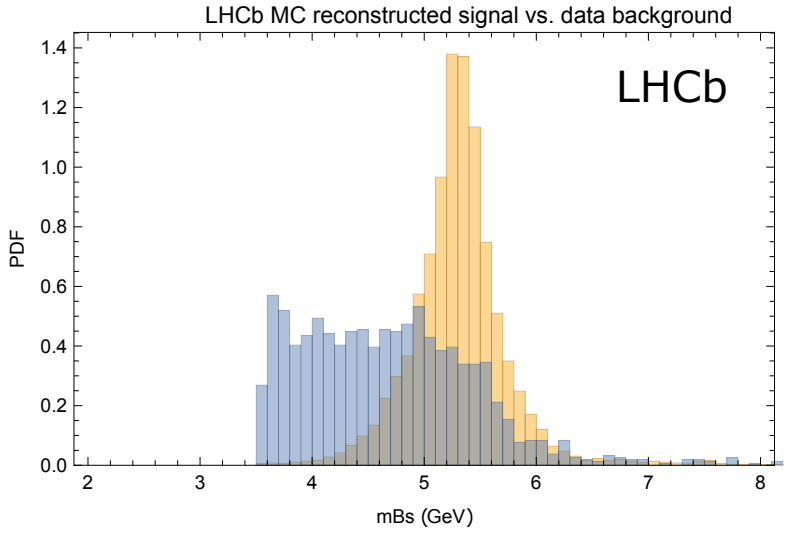


Figure E.6: Distribution of the "best" reconstructed invariant mass of the B_s^0 candidate (*i.e.* whose value is closest to the nominal one) for MC generated signal events (orange) and the background from the data SS sample (light blue).

Conclusions. In this appendix the latest developments of the signal reconstruction method have been presented. The new ideas and methods will allow, in prospect, to define more discriminating variables to be used for the signal search.

The main idea of this new approach is to use the approximation of the θ parameter, together with the kinematic constraints (in the form of the eqs.(4.36), (4.37)), to estimate the invariant mass of the B_s^0 candidate, that is left as a free variable. In this framework the approximation of this angle became more and more important and a new way has been studied. Further improvements, not yet studied, can be obtained by combining the different approximation of θ (*e.g.* $\bar{\theta}$ and $\theta_{3\pi}$) using Multivariate Analysis Algorithms, such as *Regression Trees* (for a detailed introduction to this method the reader is referred to ref. [116]).

Still some aspects of the approach presented in this part needs to be investigated in more details. The most important concerns:

- choice of the correct solution in case of multiple solutions,
- how to deal with the events for which only imaginary solutions exist.

The possibility to approximate the invariant mass of the B_s^0 candidate, with a width of ~ 200 GeV/c^2, will allow to optimize and redesign some of the steps of the experimental analysis and to increase its sensitivity.

Bibliography

[1] S.Weinberg, *The quantum theory of fields, Vol.I,II*, Cambridge University Press, 1995

[2] M.Maggiore, *A Modern Introduction to Quantum Field Theory*, Oxford University Press, 2005

[3] M.P.Peskin, D.V. Schroeder, *An Introduction To Quantum Field Theory*, Addison-Wesley Publishing Company, 1995

[4] E. Noether, *Invariante Variationsprobleme*, Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse (1918): 235–257.

[5] P.W.Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett., vol. 16, n.16, (1964)

[6] F. Englert, R. Brout, *Broken Symmetries and the mass of a gauge vector mesons*, Phys. Rev. Lett., vol. 13, N. 9, (1964)

[7] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** 1945 (1983)

[8] K.A. Olive *et al.*, *Particle Data Group*, Chin. Phys. C, 38, 090001 (2014).

[9] C.S. Wu *et al.*, *Experimental Test of Parity Conservation in Beta Decay*, Phs. Rev. **105**, 4, 1413-1415

[10] J. H. Christenson *et al.*, *Evidence for the 2π Decay of the K_2^0 Meson* Phys. Rev. Lett. **13**, 138

[11] B. Aubert *et al.* (BaBar Collaboration), *Measurement of CP-Violating Asymmetries in B^0 Decays to CP Eigenstates*, Phys. Rev. Lett. **86**, 2515

[12] K. Abe *et al.* (Belle Collaboration), *Observation of Large CP Violation in the Neutral B Meson System*, Phys. Rev. Lett. **87**, 091802

[13] S.M.Bilenky *et al.*, *Lepton Mixing, $\mu \to e\gamma$ Decay and Neutrino Oscillation*, Phys. Rev. Lett. B *67* 309 (1977)

[14] MEG Collaboration, *New Constraint on the Existence of the $\mu^+ \to e^+\gamma$ Decay'*, Phys. Rev. Lett. **110**, 201801 (2013)

[15] S.L.Glashow, D.Guadagnoli, K.Lane, *Lepton Flavor Violation in B Decays?*, arXiv:14110565v2

[16] R.Alonso *et al.*, *Lepton Universality violation and lepton flavor conservation in B-meson decay*, arXiv:1505.05164v1

[17] S.L. Adler *et al.*, *Absence of high order correction in the anomalous axial vector divergence*, Phys. Rev. **182** 1517 (1969)

[18] Super-Kamiokande Collaboration, *Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days*, Phys.Rev.Lett, **81** (1998), 1158-1162

[19] A.Giuliani *et al.*, *Neutrinoless Double-Beta Decay*, Advances in High Energy Physics, Vol. 2012 (2012),

[20] F.Zwicky, *Spectral displacement of extra galactic nebulae*, Helvetica Physica acta, **6**, 110-127 (1933)

[21] M.Roberts *et al.*, *Comparison of rotation curves of different galaxy types* Astronomy and Astrophysics, **26** 483-485 (1973)

[22] WMAP Collaboration, *Nine-Year Wilkinson Microwave Anisotropy Probe (WMPA) Observation: Cosmological Parameter Results*, arXiv:1212.5226, (2912)

[23] P.Sikivie *et al.*, *Isospin breaking in technicolor models*, Nucl. Phys. B 173, 189 (1980)

[24] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at* $\sqrt{s} = 7$ *and* 8 *TeV with the ATLAS and CMS Experiments*, Phys. Rev. Lett. **114**, 191803 (2015)

[25] R.Peccei *et al.*, *CP Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. **38**, 1440 (1977)

[26] S.L.Glashow *et al.*, *Weak Interactions with Lepton–Hadron Symmetry*, Phys. Rev. D **2**, 1285 (1970)

[27] M. K. Gaillard *et al.*, *Rare decay modes of the* K *mesons in gauge theories* Phys. Rev. D **10**, 897 (1974)

[28] J.-E. Augustin *et al.*, *Discovery of a Narrow Resonance in* e^+e^- *Annihilation*, Phys. Rev. Lett. **33** 1406 (1974)

[29] J.J. Aubert *et al.*, *Experimental Observation of a Heavy Particle J*, Phys. Rev. Lett. **33** 1404 (1974)

[30] M. Kobayashi *et al.*, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (2): 652-657 (1973)

[31] S. W. Herb *et al.*, *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39**, 252 (1977)

[32] F. Abe *et al.* (CDF collaboration), *Observation of Top Quark Production in* $p\bar{p}$ *Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74**, 2626 (1995)

[33] W.Altmannshofer, *The Flavor Puzzle*, Colloquium, Aspen Center for Physics, 26 June 2014

[34] A.J.Buras, *Weak Hamiltonian, CP violation and rare decays*, hep-ph/9806471

[35] E.Fermi, *Versuch einer Theorie der beta-Strahlen. I*, Zeitschrift für Physik **88** 161, (1934); F.L.Wilson, *Fermi's theory of Beta Decay*, American Journal of Physics **36** 12, (1968)

[36] R.Alonso *et al.*, $SU(2) \times U(1)$ *gauge invariance and the shape of new physics in rare B decays*, Phys. Rev. Lett. **113**, 241802 (2014)

[37] K. De Bruyn *et al.*, *Probing New Physics via the* $B_s^0 \to \mu^+\mu^-$ *Effective Lifetime*, Phys. Rev. Lett. 109, 041801

[38] G.Hiller *et al.*, *R_K and future $b \to s\ell\ell$ BSM opportunities* Phys. Rev. D **90** 054014 (2014)

[39] C.Bobeth *et al.*, *$B_{s,d} \to \ell\bar{\ell}$ in the Standard Model with Reduced Theoretical Uncertainty*, Phys. Rev. Lett. **112**, 101801 (2014)

[40] A.Greljo, G.Isidori, D.Marzocca, *On the breaking of Lepton Flavor Universality in B decay*, arXiv:1506.01705v1

[41] A.J.Buras, *Relation between $\Delta M_{s,d}$ and $B_{s,d} \to \mu\bar{\mu}$ in models with minimal flavor violation*, Phys. Lett. B **566**, 115-119, (2003)

[42] G.D'Ambrosio *et al.*, *A Minimal flavor violation: an effective field theory approach*, Nucl. Phys. B **645** 155-187 (2002)

[43] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. 102, 201801

[44] R.Aaij *et al.*, *First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. 110, 021801

[45] BABAR Collaboration, *A Search for the rare decay $B^0 \to \tau^+\tau^-$ at BABAR*, Phys.Rev.Lett. **96** 241802, (2006)

[46] Y.Grossman *et al.*, *$B \to \tau^+\tau^-(X)$ decays: First constraints and phenomenological implications*, Phys. Rev. D **55**, 2760 (1997)

[47] A.Dighe *et al.*, *Enhanced $B_s - \bar{B}_s$ lifetime difference and anomalous like-sign dimuon charge asymmetry from new physics in $B_s \to \tau^+\tau^-$*, Phys. Rev. D **82**, 031502(R) (2010)

[48] C.Bobeth *et al.*, *New Physics in Γ_{12}^2: $(\bar{s}b)(\bar{\tau}\tau)$ Operators*, Acta Phys. Polon. **B44**, 127 (2013)

[49] CMS and LHCb collaborations, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522**, 68-72 (04 June 2015)

[50] A.Dighe *et al.*, *Possibility of large lifetime differences in neutral B meson system*, Phys. Rev. D **76**, 054005 (2007)

[51] A.Dighe *et al.*, *How large can the branching ratio of $B_s \to \tau^+\tau^-$ be?*, Phys. Rev. D **86**, 054023 (2012)

[52] V-M-Abazov *et al.* (DØ Collaboration), *Evidence for an anomalous like-sign dimuon charge asymmetry*, Phys. Rev. D **82**, 032001 (2010)

[53] V-M-Abazov *et al.* (DØ Collaboration), *Evidence for an Anomalous Like-Sign Dimuon Charge Asymmetry*, Phys. Rev. Lett. **105**, 081801 (2010)

[54] V-M-Abazov *et al.* (DØ Collaboration), *Measurement of the anomalous like-sign dimuon charge asymmetry with 9 fb^{-1} of $p\bar{p}$ collisions*, Phys. Rev. D **84**, 052007 (2011)

[55] CDF Collaboration, *CP Asymmetry in semileptonic B decays*, CDF Note Report No.9015 (2007)

[56] A.Lenz *et al.*, *Anatomy of new physics in $B - \bar{B}$ mixing*, Phys. Rev. D **83**, 036004 (2011); A.Lenz *et al.*, *Numerical updates of lifetimes and mixing parameters of B mesons*, arXiv:1102.4274; A.Lenz, *Theoretical update of B-Mixing and Lifetimes*, arXiv:1205.1444

[57] C.W. Bauer *et al.*, *Comment on new physics contributions to Γ_{12}^s*, Phys. Lett. B **696**, 362-366 (2011)

[58] R.Aaij *et al.*, *Test of Lepton Universality Using $B \to K^+\ell^+\ell^-$ Decays*, Phys. Rev. Lett. **113** 151601 (2014)

[59] S. Fajfer *et al.*, *$B \to D^\star\tau\bar{\nu}_\tau$ sensitivity to new physics*, Phys. Rev. D **85**, 094025 (2012)

[60] D.Becirevic *et al.*, *$\bar{B} \to D\tau\bar{\nu}_\tau$ vs. $\bar{B} \to D\mu\bar{\nu}_\mu$* , Phys. Lett. B **716**, 208 (2012)

[61] J. Lees *et al.* (The BaBar Collaboration), *Measurement of an excess of $\bar{B} \to D^{(\star)}\tau^-\bar{\nu}_\tau$ decays and implications for charged Higgs Bosons*, Phys. Rev. D **88**, 072012 (2913)

[62] A. Matyja *et al.* (Belle Collaboration), *Observation of $B^0 \to D^{\star-}\tau^+\nu_\tau$ Decay at Belle*, Phys. Rev. Lett.**99** 191807 (2007)

[63] A. Bozek *et al.* (Belle Collaboration), *Observation of $B^+ \to \bar{D}^{\star 0}\tau^+\nu_\tau$ and evidence for $B^+ \to \bar{D}^0\tau^+\nu\tau$ at Belle*, Phys. Rev. D **82** 072005 (2010)

[64] M. Huschle *et al.* (Belle Collaboration), *Measurement of the branching ratio of $\bar{B} \to D^{(\star)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(\star)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, arXiv:1507.03233 (to appear in Phys. Rev. D)

[65] R. Aaij *et al.* (LHCb Collaboration), *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{\star+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{\star+}\mu^-\bar{\nu}_\mu)$*, arXiv:1506.08614

[66] Heavy Flavor Average Group, `http://www.slac.stanford.edu/xorg/hfag/semi/eps15/eps15_dtaunu.html`

[67] A.Aad *et al.*, *Luminosity determination in pp collision at $\sqrt{s} = 7$ TeV using the ATLS detector at the LHC*, European Physical Journal, **C71** 4, 1-37 (2011)

[68] G.Antchev *et al.*, *First measurement of the total proton-proton cross-section at the LHC energy of $\sqrt{s} = 7$ TeV*, European Physics Letters, **96** 2, 21002 (2011)

[69] G.Papotti *et al.*, *Experience with offset collisions in the LHC*, TUPZ025, Proc. IPAC, (2011)

[70] V. Andreev, *Heavy Quark Production at CMS and ATLAS*, (2009)

[71] T. Sjostrand *et al.*, *PYTHIA 6.4 Physics and Manual*, Journal of High Energy Physics, **0605** (2006), 026

[72] T. Sjostrand *et al.*, *A Brief Introduction to PYTHIA 8.1*, Computer Physics Communications, **178** (2008), 852-867

[73] LHCb Collaboration, *Tracking strategies used in LHCb*, `https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingStrategies#Long_track_reconstruction`

[74] M. Needham, *Performance of the LHCb Track Reconstruction Software*, CERN-LHCb-2008-026 (2008)

[75] M.Williams *et al.*, *The HLT2 Topological Lines*, CERN-LHCb-PUB-2011-002 (2011)

[76] C.Adrover *et al.*, *Search for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with 3 fb^{-1} at LHCb*, LHCb-ANA-2013-032 (2013)

[77] R.Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ Branching Fraction and Search for $B^0 \to \mu^+\mu^-$ Decay at the LHCb experiment*, Phys. Rev. Lett. 111, 101805

[78] E. Rodrigues, *Dealing with clones in the tracking* CERN-LHCb-2006-057 (2006)

[79] M.Needham, *Identification of Ghost Tracks using a Likelihood Method* CERN-LHCb-2008-026 (2008)

[80] L. Breiman, *Classification and Regression Trees*, Wadsworth, Stamford, (1984)

[81] Y. Aditya *et al.*, *Faking* $B_s^0 \to \mu^+\mu^-$, Phys. Rev. **D87** (2013), 074028

[82] LHCb Collaboration, *Central Exclusive Dimuon Production at* $\sqrt{s} = 7$ *TeV*, LHCb-CONF-2011-002 (2011)

[83] G. Lanfranchi *et al.*, *The Muon Identication Procedure of the LHCb Experiment for the First Data.*, LHCb-PUB-2009-013, CERN-LHCb-PUB-2009-013 (2009)

[84] Toolkit for Multi Variate Analysis with ROOT, http://tmva.sourceforge.net

[85] G. Mancinelli *et al.*, *Study of Muon Isolation in the* $B_s^0 \to \mu^+\mu^-$ *Channel*, LHCb-INT-2010-011 (2010)

[86] C. Adrover Pacheco, *Towards the first* $B(B^0_{(s)} \to \mu^+\mu^-)$ *measurements with the LHCb detector*, CERN-THESIS-2012-123 (2012)

[87] C.Adrover *et al.*, *Search for the* $B_s^0 \to \mu^+\mu^-$ *and* $B^0 \to \mu^+\mu^-$ *decays with* $3fb^{-1}$ *at LHCb*, LHCb-ANA-2013-032; CERN-LHCb-ANA-2013-032 (2013)

[88] CDF collaboration, *Search for* $B_s^0 \to \mu^+\mu^-$ *and* $B^0 \to \mu^+\mu^-$ *decays in* $p\bar{p}$ *collisions with CDFII*, Phys. Rev. Lett., **95** (2005), 221805

[89] M.Perrin-Terrin, *Optimisation of the binning of the discriminating variables used in the computation of* $B_s^0 \to \mu^+\mu^-$ *upper limits with the modified frequentist approach*, LHCb-INT-2012-003 (2012)

[90] T. Skwarnicki, *A study of the radiative cascade transitions between the upsilon-prime and upsilon resonances*, Ph.D. thesis, Cracow TU, Hamburg (1986),

[91] Perrin-Terrin *et al.*, *Towards a model independent branching ratio measurement of* $B_s^0 \to \mu^+\mu^-$, LHCb-INT-2013-012 ; CERN-LHCb-INT-2013-012 (2012)

[92] B. Adeva *et al.*, *Roadmap for selected key measurement of LHCb*, LHCb-PUB-2009-029 (2009)

[93] Particle Data Group, J.Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001

[94] R. Aaij *et al.* (LHCb Collaboration), *Determination of the quark coupling strength* $|V_{ub}|$ *using baryonic decays*, Nature Physics 11, 743–747 (2015)

[95] W.F. Wang *et al.*, *The semileptonic decays* $B/B_s \to (\pi, K)(l^+l^-, l\nu, \nu\bar{\nu})$ *in the perturbative QCD approach beyond the leading-order.*, arXiv:1207.0265

[96] Y.M.Wang *et al.*, *Lambda(b), Lambda transition form factors from QCD light-cone sum rules*, Phys. Rev.**D80** (2009) 074012

[97] R. Aaij *et al.*, *First observation of the decay* $B^+ \to \pi^+\mu^-\mu^-$, JHEP 1212 (2012) 125

[98] A.Morata *et al.*, *Measurement of trigger efficiencies and biases*, CERN-LHCb-2008-073

[99] R. Aaij *et al.*, *Average* f_s/f_d *b-hadron production fraction ratio for 7 TeV pp collisions*, LHCb-CONF-2013-011

[100] R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, J. High Energy Phys. 04 (2013) 01

[101] T. Junk, *Confidence level computation for combining searches with small statistics*, Nuclear Instruments and Methods in Physics Research, **A434** (1999), 435-443

[102] A.L.Read, *Presentation of search result: The CL_s technique*, Journal of Physics, **G28** (2002), 2693-1704

[103] CMS collaboration, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decay with the CMS experiment* , Phys. Rev. Lett. **111**, 101804 (2013)

[104] F. Beaujean *et al.*, *Constraints on tensor and scalar couplings from $B \to K\bar{\mu}\mu$ and $B \to \bar{\mu}\mu$* To appear in Eur. Phys. J. **C**, ArXiv 1508.01526

[105] T. Hurt *et al.*, *Global fits to $b \to s\ell\ell$ data and signs for lepton non-universality* JHEP12, 053 (2014)

[106] T. Altmannshoffer *et al.*, , arXiv:1503.06199

[107] H. Cliff *et al.*, *$B_s^0 \to \mu^+\mu^-$ Effective Lifetime Studies*, (`https://indico.cern.ch/event/369545/session/1/contribution/2/attachments/735092/1008486/lifetime_presentation.pdf`), Talk presented at "Workshop on the optimization of the analysis for $B \to \mu\mu$ decays", Cambridge, 13-15 April 2015

[108] F. Archilli *et al.*, *Background studies for $B^0 \to \mu^+\mu^-$ analysis optimization*, LHCb-INT-2014-047 ; CERN-LHCb-INT-2014-047

[109] D.Jackson, *A topological vertex reconstruction algorithm for hadronic jets*, Nuclear Instrument and Methods in Physics Research. **A388**, 1-2, 247-253 (1997)

[110] M.Perrin-Terrin, *Searches for B meson decays to purely leptonic final states*, CERN-THESIS-2013-249 (2013)

[111] ALEPH Collaboration, *Branching ratios and spectral functions of τ decays: Final ALEPH measurements and physics implications*, Phys. Reports **421** 191-284 (2005)

[112] B. Gripaios *et al.*, *Reconstruction of Higgs bosons in the di-tau channel via 3-prong decay*, arXiv:1210.1938

[113] A. Keune, *Reconstruction of the Tau Lepton and the Study of $B^0 \to D^\star\tau\nu$ at LHCb*, PhD Thesis, CERN-THESIS-2012-321 (2012)

[114] A. Keune *et al.*, *Ideas on $B_s^0 \to \tau^\pm\tau^\mp$ reconstruction*, LHCb-INT-2011-039 (2011)

[115] A.Blum *et al.*, *Beating the hold-out: Bounds for k-fold and progressive cross-validation*, Preceedings of the Twelfth Annual Conference on Computational Learning Theory (COLT '99), New York, NY, USA, pp.203-208, ACM (1999), doi:10.1145/307400.307439

[116] Rokach, Lior; Maimon, O. *Data mining with decision trees: theory and applications*, World Scientific Pub Co Inc, (2008).