



**Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction
and search for the $B^0 \rightarrow \mu^+ \mu^-$ decay
with the CMS experiment**

Tesi di Dottorato di Ricerca in Fisica Sperimentale
Ph.D. Thesis in Experimental Physics

Candidato
Luca Martini

Supervisor
Dott. Fabrizio Palla

Tutor
Dott.sa Maria Agnese Ciocci

*fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza.*

Dante Alighieri, Inferno, canto XXVI, versi 119-120

Abstract

We present the first CMS measurement of the branching fraction of the decay $B_s^0 \rightarrow \mu^+ \mu^-$ and the most recent upper limit of the decay $B^0 \rightarrow \mu^+ \mu^-$. Data correspond to the full LHC 2011 and 2012 integrated luminosity for proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV, respectively. From an unbinned maximum likelihood fit to the dimuon invariant mass distribution we extract the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0_{-0.9}^{+1.0}) \times 10^{-9}$, where the uncertainty includes both statistical and systematic effects. The excess of $B_s^0 \rightarrow \mu^+ \mu^-$ events with respect to background has a significance of 4.3 standard deviations. For the decay $B^0 \rightarrow \mu^+ \mu^-$ no significant signal is measured, and an upper limit of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.1 \times 10^{-9}$ at the 95% confidence level is determined. Both results are in agreement with the expectations from the standard model and the recent measurements of the LHCb collaboration.

Acknowledgements

I wish to thank all people who made this possible. A Ph.D. thesis in physics is not a lightweight decision as it takes a not so small part of a man's life, not only mine, but of all people whose constant engagement made this a reality.

I want to thank my supervisor Fabrizio for his continuous help, assistance, guidance and support in this long journey, and for letting me have the opportunity to work in the best place of the world for a physicist, the CERN.

I would like to thank my tutor Agnese, for the proof reading of this thesis, and for her support, encouragement and blessing.

I desire to thank Urs, who made it possible to see the $B_s^0 \rightarrow \mu^+ \mu^-$.

Special thanks go to the CMS B-Physics PAG, especially Carlos, Hermine, Jack and Vincenzo.

To my friends at CERN, Alessandro, Andrea, Giovanni, Mario, Paola and Roberto. To my friends at CMS Pisa, especially Francesco and Jacopo. To my friends at PSI, Andrey, Christoph, Danek and Jose.

Last but, of course, not least!, to my mother and my father, and to Oriella, to whom I dedicate this text.

Contents

Introduction	xi
1 The standard model	1
1.1 Overview	1
1.2 The SM Lagrangian	3
1.2.1 The EW sector	4
1.2.2 The QCD sector	10
1.2.3 Supersimmetry	12
2 The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays	17
2.1 B_s^0 meson proprieties	17
2.2 Operator product expansion of the Hamiltonian	18
2.2.1 Renormalization group	19
2.3 Effective Hamiltonian for $B_s^0 \rightarrow \mu^+ \mu^-$	19
2.4 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ SM branching fraction	22
2.4.1 Flavor-averaged time-integrated branching fraction	24
2.5 $B_{(s)}^0$ rare decays and new physics	24
2.5.1 Additional Higgs bosons	25
2.5.2 MFV-MSSM and other models	26
2.5.3 The $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ constraints on SUSY phase space	26
2.5.4 New physics correlations	27
2.6 Status of experimental results	30
2.7 Production of $B_{(s)}^0$ mesons at the LHC	31
2.7.1 The factorization hypothesis	32
2.7.2 b -quark production mechanisms	34
2.7.3 Hadronization	36
2.7.4 Monte Carlo generation	38
3 The Large Hadron Collider and the CMS apparatus	40
3.1 The Large Hadron Collider	40
3.1.1 The LHC layout	41

3.1.2	The LHC luminosity	42
3.1.3	Operating conditions in the 2011 and 2012 data-taking periods	44
3.2	The Compact Muon Solenoid	47
3.2.1	CMS coordinate system	48
3.2.2	The superconducting magnet	49
3.2.3	The inner tracking system	50
3.2.4	The electromagnetic calorimeter	54
3.2.5	The hadronic calorimeter	57
3.2.6	The muon system	58
3.2.7	Track and vertex reconstruction	62
3.2.8	Trigger	73
3.2.9	Event reconstruction and processing	78
4	Observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and upper limit on $B^0 \rightarrow \mu^+ \mu^-$ decays	80
4.1	Overview	80
4.1.1	Sources of background events	82
4.2	Primary dataset and triggers	83
4.2.1	HLT paths for the signal sample	85
4.2.2	HLT path for the normalization and control samples . . .	86
4.3	Off-line signal selection	87
4.3.1	Single muon selection	87
4.3.2	Discriminating variables for $B_{(s)}^0$ -signal decay modes . . .	95
4.4	Normalization channel event selection	100
4.5	Control sample event selection	105
4.6	B_s^0 -signal selection with multivariate analysis	107
4.6.1	Preselection	107
4.6.2	Event sample splitting	107
4.6.3	BDT results	107
4.6.4	BDT pileup studies	110
4.6.5	BDT mass dependence studies	113
4.7	Background determination	115
4.8	Further systematic studies and cross-checks	118
4.8.1	Yield ratio for different run ranges	118
4.8.2	Acceptance and efficiency systematics	119
4.8.3	Production ratio f_s/f_u	123
4.8.4	Summary of systematic uncertainties	123
4.9	Statistical models	124
4.9.1	Binned analysis	124

4.9.2	Unbinned maximum likelihood fit analysis	126
4.10	Results	133
4.10.1	Results with the binned analysis	136
4.10.2	Results with the UML fit	138
4.10.3	Interpretations of the results	142
4.10.4	CMS-LHCb combination	144
4.10.5	Impact on new physics	144
4.10.6	Future prospects	147
5	Conclusions	149
	Bibliography	152
	Acronyms	164

List of Figures

1.1	Form of the Higgs field potential.	5
1.2	New boson properties.	7
1.3	Unitarity triangle constraints.	10
1.4	Example of loop in an electron-electron interaction.	11
1.5	Measurements of α_s as a function of the energy scale Q	12
1.6	One-loop quantum corrections to the Higgs squared mass	13
1.7	Summary of mSUGRA/CMSSM exclusion limits.	15
1.8	Summary of exclusion limits of CMS SUSY searches	16
2.1	LO diagrams for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay.	20
2.2	Dependence of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction on μ_t	22
2.3	MSSM LO diagrams for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay.	25
2.4	Direct and indirect constraints in CMSSM.	27
2.6	Correlation between $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ BF's.	29
2.7	Correlations between B_s^0 and B^0 , and B_s^0 and B^\pm decays.	29
2.8	Summary of reported upper limits and measurements.	31
2.9	Parton distributions.	33
2.10	LO diagrams for b -quark production.	34
2.11	Diagrams after the LO.	35
2.12	Bottom cross-sections.	36
2.13	b -quark y and p_T distributions.	37
2.14	B^0 meson pair-production asymmetry.	37
2.15	Hadronization of a parton shower.	38
3.1	Cross-section of an LHC dipole magnet.	41
3.2	Sketch of the eight sectors of the LHC.	42
3.3	SM cross-sections as a function of collider energy.	43
3.4	CMS average pileup distribution for pp collisions in 2012.	45
3.5	Integrated pp luminosity as a function of time.	45
3.6	The CERN accelerator complex.	46

3.7	Perspective view of the CMS detector.	48
3.8	CMS solenoid E/M ratio versus stored energy E	50
3.9	Tracker layers in the zr plane.	51
3.10	Tracker material budget.	51
3.11	Pixel detector layers.	52
3.12	Strip tracker points.	53
3.13	Tracker average module hit efficiencies.	54
3.14	Sketch of ECAL.	55
3.15	ECAL energy resolution as a function of energy.	56
3.16	CMS sub-detectors in the $z\eta$ plane	57
3.17	Muon system sub-detectors in the rz plane.	59
3.18	Interaction lengths as a function of $ \eta $ for the muon system.	59
3.19	Sketch of a DT cell showing drift lines and isochrones.	60
3.20	A DT chamber in position inside the iron yoke, in the $r\phi$ plane.	60
3.21	Layout of one CSC.	61
3.22	Sketch of an RPC sub-detector.	62
3.23	Helical track parametrization.	63
3.24	Track reconstruction efficiency for muons and pions.	66
3.25	Resolutions of the five track parameters for single muons.	67
3.26	Transverse momentum resolution for the three muon objects.	69
3.27	TNP tight muon identification and reconstruction efficiencies.	70
3.28	Tight muon misidentification.	71
3.29	Sketch of the adaptive vertex fitter.	72
3.30	Primary vertex resolution.	73
3.31	Sketch of the decay chain reconstruction.	74
3.32	Architecture of the L1 trigger.	75
3.33	L1 double-muon trigger rate as a function of time.	76
3.34	Data collected by early July, 2011.	77
4.1	HLT double muon trigger rate.	86
4.2	Dimuon invariant mass distribution around the J/ψ meson peak.	87
4.3	Sketch view of a decision tree.	89
4.4	Muon and kaon distributions.	92
4.5	Kaon rejection versus muon efficiency.	93
4.6	Mass difference distributions between the D^{*+} and D^0 mesons.	94
4.7	Muon misidentification rates.	94
4.8	B -candidate kinematic distributions.	96
4.9	Sketch of signal and combinatorial events.	97
4.10	Sketch of δ_{3D} , ℓ_{3D} and α_{3D} variables.	97
4.11	Normalized B -candidate distributions.	99

4.12	Isolation variable distributions.	101
4.13	$B^\pm \rightarrow J/\psi K^\pm$ invariant mass distributions.	103
4.14	Simulated partially reconstructed $B^0 \rightarrow J/\psi(\mu^+\mu^-)X$ decays. . .	104
4.15	J/ψ -constrained invariant mass distribution for B^\pm candidates. .	104
4.16	$B_s^0 \rightarrow J/\psi\phi$ invariant mass distributions.	106
4.17	MVA over-training control plot.	108
4.18	BDT ROC curves.	109
4.19	Significance versus the Boosted Decision Tree (BDT) selection. .	110
4.20	BDT distributions for the NS and CS samples.	111
4.21	Low and high pileup distributions.	112
4.22	side-band BDT mean and RMS versus pileup.	114
4.23	BDT efficiencies in data side-bands.	115
4.24	Invariant mass distributions of rare peaking backgrounds.	116
4.25	Invariant mass distributions of rare semileptonic B decays. . . .	117
4.26	Channel ratios for different data periods.	119
4.27	Isolation for the three different QCD $b\bar{b}$ production processes. . .	120
4.28	Activity distributions.	122
4.29	Mass scale and resolution uncertainties.	122
4.30	Reduced mass error distributions.	127
4.31	B_s^0 and B^0 invariant mass distributions	128
4.32	Peaking and semileptonic backgrounds invariant mass distributions.	129
4.33	BDT categories.	131
4.34	MC significance distributions	132
4.35	Dimuon invariant mass distributions.	134
4.36	BDT output versus invariant mass distributions.	135
4.37	$B_s^0 \rightarrow \mu^+\mu^-$ confidence level plots.	137
4.38	$B^0 \rightarrow \mu^+\mu^-$ confidence level plots.	137
4.39	CL_S two-sided interval confidence levels.	137
4.40	1D fit result.	139
4.41	Categorized BDT fit result.	140
4.42	Scan of the ratio of the joint likelihood.	141
4.43	Weighted invariant mass distributions.	141
4.44	Scan of the ratio of the joint likelihood for μ and λ_{ds}	143
4.45	Combined CMS-LHCb results.	145
4.46	Correlation between the branching fractions.	145
4.47	$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ versus $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ in Z and Z' scenarios . .	146

List of Tables

1.1	The four fundamental forces.	2
1.2	The three generations of leptons and quarks.	2
2.1	Main observables of the B_s^0 - \bar{B}_s^0 system.	18
2.2	Noticeable decay modes of the B_s^0 meson.	18
2.3	Relative uncertainties of the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	23
2.4	Latest upper limits on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions. . .	30
3.1	Main parameters of the LHC.	43
3.2	Main parameters of the 4 T superconducting solenoid.	49
3.3	Pitch and thickness for each different silicon strip.	53
3.4	CMSSW main data formats and size per event.	79
4.1	Kinematic selections at HLT level.	85
4.2	Muon misidentification probabilities	93
4.3	Invariant mass region definitions for the signal channel.	95
4.4	Invariant mass region definitions for the normalization channel. .	102
4.5	Invariant mass region definitions for the control channel.	106
4.6	Preselection for the BDT training.	108
4.7	Rare branching fractions included in the analysis.	118
4.8	Summary of the relative systematic uncertainties.	123
4.9	Expected number of signal and background events.	126
4.10	Categorized BDT boundaries.	130
4.11	Final efficiencies.	133
4.12	Expected and observed yields.	134
4.13	Expected uncertainties in the BF's.	148

Introduction

The standard model (SM) of particle physics, fully developed in the late '70s, is the most successful theory in explaining the electromagnetic, weak and strong interactions between elementary particles. However, the SM is not considered the final word on High Energy Physics (HEP). Among the others, in fact, it does not explain the origin of the dark matter nor of the dark energy, and it does not include the theory of gravitation. Thus, the SM is considered a sub-set of a more general theory. Many different Beyond Standard Model (BSM) constructions exist, but currently they have no empirical confirmations.

The quest of many experimental particle physicists today is to find effects which cannot be predicted by the SM, and which can help to distinguish between the different New Physics (NP) models.

The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ rare decays are among the most promising investigation channels for the search of new physics. These decays proceed through SM penguin and box diagrams and are helicity suppressed; the SM predicts very small Branching Fractions (BFs): 3.2×10^{-9} for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay and 1.1×10^{-10} for the $B^0 \rightarrow \mu^+ \mu^-$ decay. These tiny values and their small theoretical relative uncertainties (less than 10%) make these decays very sensitive to many NP scenarios.

Since the '80s, many collaborations, starting from the Argus experiment, reported upper limits on any of these two rare decays. After about thirty years of searches and upper limits, the measurements of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay have been published independently by the Large Hadron Collider beauty (LHCb) and the Compact Muon Solenoid (CMS) collaborations.

In this thesis we present the first successful reconstruction of the rare decay $B_s^0 \rightarrow \mu^+ \mu^-$, the measurement of its BF, and the most recent upper limit for the BF of the $B^0 \rightarrow \mu^+ \mu^-$ decay, at CMS. The analysis was performed on a 25 fb^{-1} sample of data collected by the CMS experiment at the Large Hadron Collider (LHC) during 2011 and 2012, using a trigger based on the selection of two reconstructed muons with a dimuon invariant mass compatible with the $B_{(s)}^0$ masses.

The main challenge of this analysis is the presence of similar backgrounds in large quantities, for instance, two muons coming from two uncorrelated B decays, whose yield is more than five orders of magnitude larger than the signal yield. Notable important background sources are also represented by two and three-body B decays, like $B_s^0 \rightarrow K^+\pi^-$ and $B_s^0 \rightarrow K^-\mu^+\nu_\mu$, which form shaped distributions close to the signal invariant mass values.

Given the expected CMS sensitivity, results are extracted with two different methods. A binned counting experiment is performed to set an upper limit on the $B^0 \rightarrow \mu^+\mu^-$ decay, with the CL_S technique. In this way a direct comparison with previous published results is possible. An unbinned maximum likelihood fit is performed on the invariant mass yield to extract the B_s^0 signal. To measure the significance of the signal, a test statistic against a background-only hypothesis is performed with a likelihood ratio test.

The branching fractions are evaluated relatively to a normalization channel: $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^+\mu^- K^\pm$. In this way, the analysis is completely independent on the uncertainties on the acquired integrated luminosity and on the $b\bar{b}$ production cross-section. Furthermore, measuring a branching fraction ratio using a decay channel with a signature similar to the signal decay has the advantage that many systematic errors cancel to first order.

Multivariate methods are used to separate the signal events from the different background contributions which could mimic the B_s^0 and B^0 topologies. We use Monte Carlo (MC) simulations to determine the acceptance and the efficiency for the signal events. Data-driven techniques are used to validate the MC simulations. Furthermore, the MC simulation of the signal is validated in data using the decay $B_s^0 \rightarrow J/\psi \phi \rightarrow \mu^+\mu^- K^+K^-$.

This thesis is organized as follows.

The theoretical framework is discussed in Chapter 1, where the SM theory is outlined.

Motivations for the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+\mu^-)$ measurements are further described in Chapter 2, where their SM expectations are examined. NP models, which predict different values of these BF's, are discussed. A brief overview of the experimental status on their measurements is also given. This analysis uses MC simulations to describe the proton-proton (pp) scatterings and the interaction of particles with the sub-detectors. Production mechanisms of the $B_{(s)}^0$ mesons at the LHC and the MC generators are discussed at the end of the same Chapter.

Chapter 3 describes the experimental apparatus used to perform the measurements. The LHC accelerator and the data-acquisition conditions are briefly illustrated. A large part of this Chapter is dedicated to the CMS detector, and especially to the inner tracker and muon stations sub-detectors, which are used to reconstruct the two muons coming from the $B_{(s)}^0$ decays. Track and

vertex algorithms are also detailed. The CMS trigger framework and the data processing are outlined at the end of this Chapter.

Chapter 4 is devoted to the detailed description of the measurements. The first part is dedicated to the overview on the procedure used to extract the BF measurements, including the description of the background sources and of the control and normalization samples. Second part describes the multivariate methods and the fitting techniques used to extract the $B_{(s)}^0$ yields. Interpretations of the results and prospects for future measurements are shown and discussed at the end of the Chapter.

The measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ BF and the upper limit on the $B^0 \rightarrow \mu^+ \mu^-$ BF have been published in *Physical Review Letters* [1] and presented for the first time at the EPS-HEP 2013 conference, together with the combination of these measurements with the LHCb results [2].

Throughout this thesis, natural units with $c = 1$ and $\hbar = 1$, where c is the speed of light and \hbar is the Planck constant, are used. For convenience, the acronyms used in this thesis are defined at the end of this text, as well as at the point of first use or when needed.

Chapter 1

The standard model

The current best understanding of nature at the high energy regime is summarized in the standard model (SM). Its description of particles and interactions has been analyzed and confirmed on a wide range of energies in numerous experiments. But there are still important open questions, therefore the role of experimentalists is to test the SM in all conceivable ways, seeking to discover whether and where something more lies beyond it.

This chapter describes important features of the SM, relevant for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ BFs, as well as an overview on NP models, which might give contributions to these two decays.

1.1 Overview

Particle physics studies the fundamental constituents of the matter (the elementary particles) and how they interact, through the force fields. There are four known forces, propagated by boson carriers: the strong, electromagnetic, weak and gravitational fields. Elementary matter is made by fermions which are divided into quark and leptons. The former interacts also through strong interactions, while the latter does not.

As of today there is no quantum theory of gravitation that can be experimentally verified. The gravitational effects are too small on single particles to be measured by detectors: for instance, the gravitational attraction between an electron and a proton is 40 orders of magnitude smaller than their electromagnetic interaction, thus gravity is not included in the SM. The electromagnetism is mediated by zero-mass spin-1 bosons (photons). Only electrically charged particles can interact with photons. The weak interaction is propagated by massive spin-1 bosons (Z and W^\pm). Every known particle but photons interacts weakly. Finally, quarks interact also strongly, through the gluon field, to form

Table 1.1: The four fundamental forces.

Interaction	Carriers	Range [m]
Gravitational	Gravitons ¹	∞
Weak	Z and W^\pm bosons	10^{-18}
Electromagnetic	Photons	∞
Strong	Gluons	10^{-15}

Table 1.2: The three generations of leptons and quarks.

Generation	Leptons		Quarks	
I	ν_e	e	u	d
II	ν_μ	μ	c	s
III	ν_τ	τ	t	b
Electric charge	0	-1	2/3	-1/3

the hadronic bound states, like protons and pions. The strong charge is called color and it has three components, commonly named red, green and blue; spin-1 gluons carry one color and one anti-color, and interact each other.

Leptons do not interact with the strong field. Since hadrons are colorless, only two combinations of quarks can exist: mesons, that are bosons formed by a quark-antiquark pair ($q_1\bar{q}_2$) and baryons, that are fermions formed by three quarks ($q_1q_2q_3$). Recently discovered resonances not belonging to the standard spectra of the c and b quarks can be interpreted as bound states of four quarks (or quark molecules) [3]. Pentaquarks, hypothetical bound states made by $q_1q_2q_3q_4\bar{q}_5$, do not have experimental evidence [4].

A summary of the force carriers is in Table 1.1. Table 1.2 shows the three observed generations of matter, which are distinguished by the masses of their constituent particles.

The SM theory describes the interactions between these fundamental particles. The foundation of the SM can be set with the development of the Quantum Electro-Dynamics (QED) theory, in the '40s [5]. Gauge invariance and symmetry motivations lie at the basis of the QED. In the '50, Yang and Mills [6] proposed a non-Abelian gauge theory for the weak interactions, while Higgs [7], Brout and Englert [8], and Guralnik, Hagen and Kibble [9] suggested a model so that vector bosons could acquire mass in a gauge invariant way. Combining these two models, Glashow [10], Salam [11] and Weinberg [12] built a unified theory of the electro-weak interactions, in 1967. The resulting predicted Z and W^\pm massive gauge bosons were discovered at CERN by the UA1 [13, 14] and UA2 [15, 16] experiments. In 1964 Gell-Mann and Zweig introduced quarks (u , d and s), as

¹Not yet discovered.

a mathematical model, to explain mass and proprieties of hadrons [17]. The fourth quark (c) was predicted in 1970 by the Glashow-Iliopoulos-Maiani (GIM) mechanism [18], noting the absence of Flavor Changing Neutral Current (FCNC) interactions. Four years later, the direct experimental observation of the J/ψ ($c\bar{c}$) meson was made during the “November Revolution” [19, 20] (1974). In 1973 the Quantum Chromo-Dynamics (QCD) theory was formulated [21]. Quarks are real particles, carrying a color charge, and gluons are massless quanta of the strong-interaction field. Still in 1973, Politzer [22], and Gross and Wilczek [23] discovered that the color theory of the strong interaction has a special property, the asymptotic freedom, for which, at high transferred momenta, quarks behave like free particles. The observation of the third generation of matter started in 1976 by Perl [24], with the discovery of the τ lepton. Experiments at Fermilab discovered the bottom quark [25] in 1977 and the top quark [26, 27] in 1995.

The SM was built by joint efforts of theoreticians and experimentalists, as it was driven by new theoretical ideas and by new experimental discoveries. Its latest confirmations are the discoveries of the τ neutrino [28] and of a Higgs-like boson [29, 30].

Despite these important successes, the SM is believed not to be the last word on high energy physics. Reasons are both experimental observations and theoretical problems. The SM contains no particles with the right properties to form the dark matter that pervades the cosmos and it does not explain all the matter-antimatter asymmetry we observe in the universe. Furthermore, it does not explain the origin of the mass values of the elementary particles. The SM contains 26 free parameters [31], many believe they are too many for a “theory of everything”. Finally, general relativity is not included in the model. For these reasons the SM is believed to be only a low-energy spontaneously-broken theory of a more symmetric model. There exist many BSM models extending it, including SuperSYmmetry (SUSY) models, technicolor or string theory, all predicting new-physics phenomena. Up to today, no BSM predictions have been observed.

The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is one of the most promising channels for the search of new physics. The reasons will be clarified further in this and in the next chapters.

1.2 The SM Lagrangian

The SM is a quantum field theory with gauge symmetry group $SU(3) \otimes SU(2)_L \otimes U(1)_Y$, spontaneously broken to $SU(3) \otimes U(1)_Q$. $SU(3)$ is the group that generates the strong interaction; its conserved quantum number is the color (divided in the three kinds). Y and Q are the weak hypercharge and the electric

charge generators.

Given a particle ψ , the left-handed leptons and quarks

$$\psi_L = \frac{1 - \gamma^5}{2} \psi \quad (1.1)$$

form $SU(2)_L$ doublets:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L \quad (1.2)$$

while the corresponding right-handed fields

$$\psi_R = \frac{1 + \gamma^5}{2} \psi \quad (1.3)$$

transform as singlets under $SU(2)_L$. The third component of the weak hypercharge I_3 and the electric charge Q are related by the Gell-Mann–Nishijima formula [32, 33]:

$$Q = I_3 + \frac{1}{2}Y \quad (1.4)$$

The SM Lagrangian is composed of parts describing the free fields and of possible interactions between them, and it is commonly divided in two parts:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QED}} + \mathcal{L}_{\text{QCD}} \quad (1.5)$$

where \mathcal{L}_{QED} holds the QED interactions and its gauge bosons, and \mathcal{L}_{QCD} holds the gluon fields and their interactions.

1.2.1 The EW sector

The $SU(2)_L$ group is associated to three generators T_a , corresponding to three supplementary gauge bosons W_μ^a (with $a = 1, 2, 3$). The boson associated to the $U(1)_Y$ symmetry, whose generator is Y , is usually denoted as B_μ . The QED Lagrangian is the sum of kinematic and Electro-Weak (EW) interaction parts. In general the left and right-handed terms can be written as

$$\begin{aligned} \mathcal{L}_{\text{EW}} = & \bar{Q}_L (i\gamma^\mu D_\mu) Q_L + \bar{u}_R (i\gamma^\mu D_\mu) u_R + \bar{d}_R (i\gamma^\mu D_\mu) d_R \\ & - \frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \end{aligned} \quad (1.6)$$

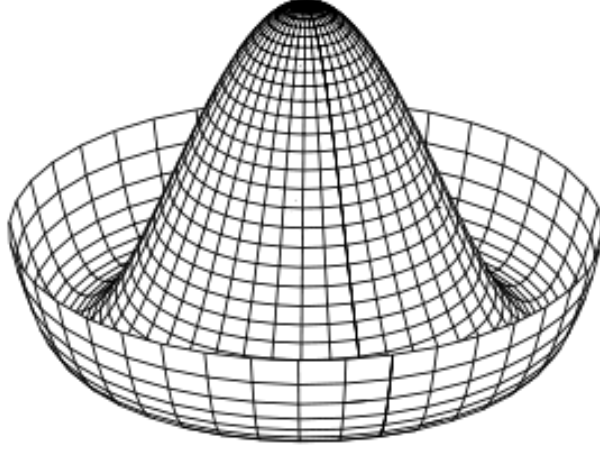


Figure 1.1: Form of the Higgs field potential.

where Q_L , u_R and d_R are, respectively, one generic left-handed doublet and two right-handed singlets (of up-type and down-type). D_μ is the covariant derivative:

$$D_\mu = \partial_\mu + igT_a W_\mu^a + i\frac{g'}{2}YB_\mu \quad (1.7)$$

The tensor fields $W_{\mu\nu}^a$ are defined as

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a \quad (1.8)$$

and similarly for $B_{\mu\nu}$.

EW gauge bosons acquire mass through the Higgs mechanism, whose Lagrangian can be written as

$$\mathcal{L}_{Higgs} = D_\mu \phi D^\mu \phi^\dagger + \mu^2 \phi^\dagger \phi - \lambda_h (\phi^\dagger \phi)^2 \quad (1.9)$$

The Higgs field is a spin-0 complex doublet of the weak isospin $SU(2)$ symmetry. If λ and μ^2 are both positive, the ground state is degenerate, with each ground state related to the other by a $SU(2)$ gauge transformation (Fig. 1.1). We can break the symmetry choosing a specific ground state such as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \quad (1.10)$$

This allows to use a perturbation theory from this ground state.

After symmetry breaking, the boson fields W_μ^a and B_μ do not correspond anymore to physical fields. Instead, they are linearly combined to form the

physical electromagnetic and weak fields:

$$A_\mu = \cos \theta_w B_\mu + \sin \theta_w W_\mu^3 \quad (1.11)$$

$$Z_\mu = -\sin \theta_w B_\mu + \cos \theta_w W_\mu^3 \quad (1.12)$$

$$W_\mu^\pm = \frac{W_\mu^1 \mp W_\mu^2}{\sqrt{2}} \quad (1.13)$$

with θ_w the Weinberg angle:

$$\cos \theta_w = \frac{g}{\sqrt{g^2 + g'^2}} \approx 0.877 \quad (1.14)$$

g and g' are the coupling constants of $SU(2)_L$ and $U(1)_Y$ respectively. $e = g \sin \theta_w$ is the positron electric charge.

The corresponding Higgs field acquires mass and the arising quadratic terms of W_μ and B_μ give masses to the weak bosons:

$$M_H = v \sqrt{2\lambda_h} \quad (1.15)$$

$$M_W = \frac{v|g|}{2} \quad (1.16)$$

$$M_Z = \frac{v\sqrt{g^2 + g'^2}}{2} = \frac{M_W}{\cos \theta_w} \quad (1.17)$$

where

$$v = |\mu|/\sqrt{\lambda_h} \approx 246 \text{ GeV} \quad (1.18)$$

is the Higgs vacuum expectation value.

The $SU(2)_L$ coupling constant g' is related to the Fermi constant G_F :

$$\frac{G_F}{\sqrt{2}} = \frac{g'^2}{8M_W^2} = 1.16639 \times 10^{-5} \text{ GeV}^{-2} \quad (1.19)$$

where $M_W \approx 80.5 \text{ GeV}$ is the mass of the W^\pm boson.

The fermion EW couplings are divided in charged and neutral current interactions: $\mathcal{L}_{\text{int}} = \mathcal{L}_{\text{CC}} + \mathcal{L}_{\text{NC}}$ where

$$\mathcal{L}_{\text{CC}} = \frac{g'}{2\sqrt{2}} (J_\mu^+ W^{+\mu} + J_\mu^- W^{-\mu}) \quad (1.20)$$

$$\mathcal{L}_{\text{NC}} = e J_\mu^{\text{em}} A^\mu + \frac{g'}{2 \cos \theta_w} J_\mu^0 Z^\mu \quad (1.21)$$

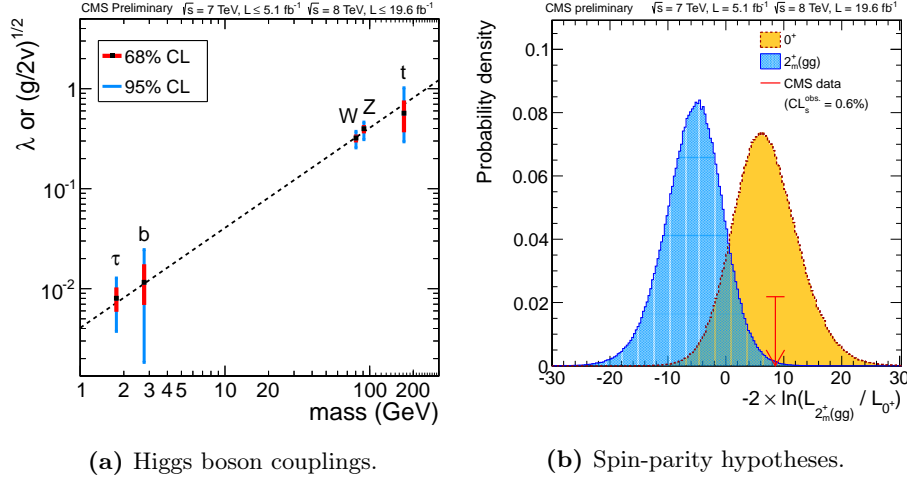


Figure 1.2: New boson couplings, expressed as function of particle masses **(a)**, and test statistic comparing the signal J^P hypotheses 0^+ and $2_m^{++}(gg)$ (graviton-like) in the best fit to the data **(b)**. The observed value is indicated by the arrow [34].

The currents are:

$$J_\mu^+ = (\bar{u}d')_{V-A} + (\bar{\nu}_e e)_{V-A} \quad (1.22)$$

$$J_\mu^{\text{em}} = \sum_f Q_f \bar{f} \gamma_\mu f \quad (1.23)$$

$$J_\mu^0 = \sum_f \bar{f} \gamma_\mu (v_f - a_f \gamma_5) f \quad (1.24)$$

where the sum runs over each fermion f . The shown currents are valid for each generation. Notation $(\bar{u}d')_{V-A}$ stands for $\bar{u}\gamma_\mu(1 - \gamma_5)d'$ and so on. The vector and axial couplings v_f and a_f follow from the Glashow-Weinberg-Salam EW theory:

$$v_f = I_3^f - 2Q_f \sin^2 \theta_w \quad (1.25)$$

$$a_f = I_3^f \quad (1.26)$$

The Higgs mass μ is a free parameter of the SM, that has to be measured experimentally. On 4 July 2012 the CMS and ATLAS collaborations announced that they independently made the same discovery of a boson of mass around 125 GeV, with about 5σ significance [29, 30]. Subsequent analyses with the full 2012 data confirmed the observation of this new boson and allowed for more precise studies of its properties such as couplings, spin and parity [34–36] (Fig. 1.2). As of today, all results are fully compatible with a SM Higgs boson.

1.2.1.1 The CKM matrix

The quark bottom parts of the vectors in Eq. (1.2) are eigenstates of the weak interaction, but they are not mass eigenstates. The relation between the two states goes through the Cabibbo-Kobayashi-Maskawa (CKM) unitary matrix [37, 38]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.27)$$

where the primed quarks are the mass eigenstates and the others are the flavor eigenstates. For convention, the V_{CKM} is applied to down-type quarks only (d , s and b).

In the quark mass eigenstate basis, the CKM matrix appears in the SM charged-current interaction Lagrangian

$$\mathcal{L}^{cc} = \frac{g}{2\sqrt{2}} \sum_{i,j} \bar{u}_i \gamma_\mu (1 - \gamma_5) (V_{\text{CKM}})_{ij} d_j W^\mu + h.c. \quad (1.28)$$

where the quark fields are $u_i = (u, c, t)$ and $d_i = (d, s, b)$, g is the weak coupling constant and W^μ is the charged vector boson field. The CKM matrix elements are the only flavor-non-diagonal couplings and CP-violating couplings (together with the neutrino mass matrix) present in the SM.

There are many parameterizations of the unitary CKM matrix in literature. In general, the CKM matrix can be parametrized using three rotation angles and one phase. Introducing the notations $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$, where the two indexes i and j are generation labels ($i, j = 1, 2, 3$) and $\theta_{ij} \in [0, \pi/2]$, the standard parametrization follows:

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -s_{23}c_{12} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \quad (1.29)$$

where $\delta \in (-\pi, \pi]$ represents all the CP violation in flavor-changing processes in the SM. The CKM matrix unitarity imposes

$$\sum_i V_{ij} V_{ik}^* = \delta_{jk}, \quad \sum_j V_{ij} V_{kj}^* = \delta_{ik} \quad (1.30)$$

The six vanishing combinations can be visualized as triangles (“unitarity triangles”) in the complex plane. One of the most useful triangles comes from Eq. (1.31):

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad (1.31)$$

divided by $V_{cd}V_{cb}^*$. Eq. (1.31) can be rewritten as

$$R_t e^{-i\beta} + R_u e^{-i\gamma} = 1 \quad (1.32)$$

with

$$R_t = \left| \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \right|, \quad R_u = \left| \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right|, \quad (1.33)$$

$$\beta = \arg \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right), \quad \gamma = \arg \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right) \quad (1.34)$$

R_u , R_t , β and γ are the non-trivial sides and angles of the normalized unitary triangle. The unitary triangle is determined by one complex number:

$$\bar{\rho} + i\bar{\eta} = R_u e^{i\gamma} \quad (1.35)$$

Thus, by the coordinates in the complex plane:

$$(0, 0), (1, 0), (\bar{\rho}, \bar{\eta}) \quad (1.36)$$

Several experimental constraints can be conveniently represented on this plane and used to determine this unitary triangle (Fig. 1.3). The latest estimates for the magnitudes of the nine CKM matrix elements, evaluated with a best fit of several measurements, are given in Eq. (1.37) [4]:

$$V_{\text{CKM}} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix} \quad (1.37)$$

The CKM matrix presents a hierarchy: $s_{13} \ll s_{23} \ll s_{12} \ll 1$. Therefore, it is convenient to do an expansion in a power series of $\lambda \approx |V_{us}| \approx 0.23$. Using the Wolfenstein [40] parameters λ , A , ρ and η , which are defined by

$$\lambda = \sin \theta_{12}, \quad A = \frac{\sin \theta_{23}}{\sin^2 \theta_{12}} \quad (1.38)$$

$$\rho = \frac{\sin \theta_{13} \cos \delta}{\sin \theta_{12} \sin \theta_{23}}, \quad \eta = \frac{\sin \theta_{13} \sin \delta}{\sin \theta_{12} \sin \theta_{23}}, \quad (1.39)$$

up to $\mathcal{O}(\lambda^4)$, the CKM matrix is

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \quad (1.40)$$

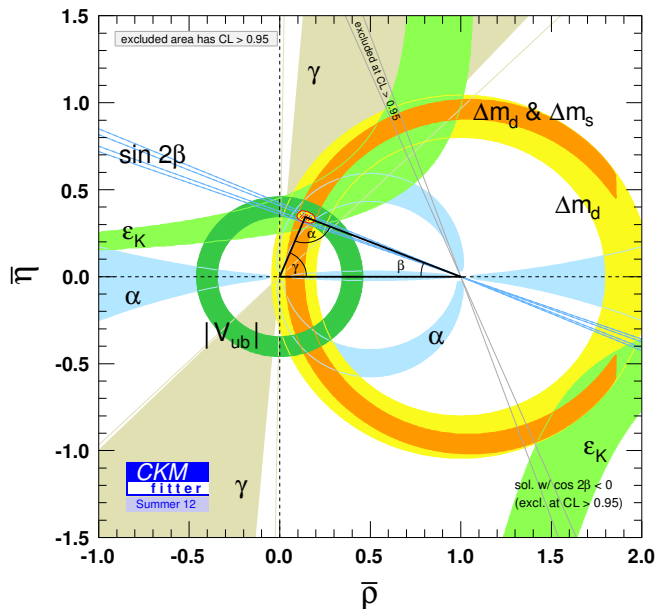


Figure 1.3: Unitarity triangle constraints. The excluded areas are at 95% Confidence Level (CL) [39].

The CKM matrix is unitary to all orders in λ . At the lowest order in λ , ρ and η coincide with $\bar{\rho}$ and $\bar{\eta}$.

1.2.2 The QCD sector

The dynamics of the quarks and gluons are controlled by the QCD Lagrangian. The gauge group is $SU(3)$, which has eight generators, called gluons, that each carries two color charges. This is an important difference with respect to the electromagnetic interaction: while photons cannot interact between themselves, since they are chargeless, gluons can self-interact.

Similarly to Eq. (1.6), the gauge-invariant QCD Lagrangian can be written in this form:

$$\mathcal{L}_{\text{QCD}} = i\bar{\psi}(\gamma^\mu D_\mu - m)\psi - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.41)$$

where ψ is the quark field and g_s is the strong coupling constant. D_μ is the following covariant derivative:

$$D_\mu = \partial_\mu + ig_s \lambda_a G_\mu^a \quad (1.42)$$

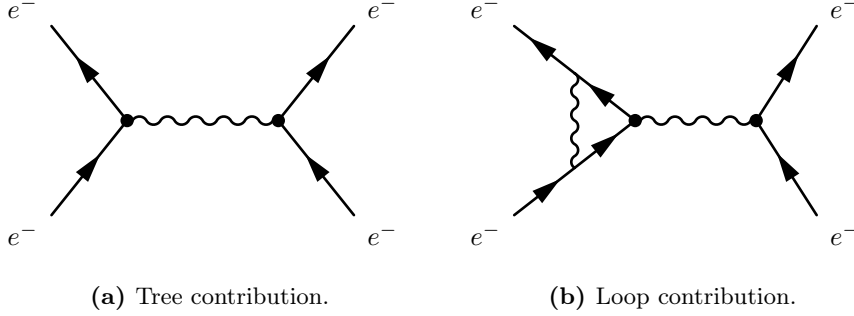


Figure 1.4: Example of leading order (a) and next-to-leading order (b) contributions in a electron-electron interaction: the right diagram contains an ultraviolet divergent loop.

and $G_{\mu\nu}^a$ is the gluon field strength tensor:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g f^{abc} G_\mu^b G_\nu^c \quad (1.43)$$

G_μ^a is the gluon field, λ^a are the $SU(3)$ generators and f^{abc} are the structure constants of $SU(3)$.

1.2.2.1 Asymptotic freedom

In the Feynman perturbative expansion of the Lagrangian, ultraviolet divergent loops can appear (Fig.1.4).

They can be eliminated through the renormalization of the fields and the parameters of the Lagrangian [41]. Given an energy scale μ , all parameters can be expressed as a function of μ times some renormalization constants. These new renormalized expressions can be then put into the Lagrangian, canceling out the divergences.

The renormalization group equations give the dependence of the renormalized parameters on the renormalization scale μ . For example, the strong coupling $\alpha_s \equiv g_s^2/(4\pi)$, at the two-loops level, obeys to [41]:

$$\frac{d\alpha_s(\mu)}{d \ln \mu} = -2\beta_0 \frac{\alpha_s^2(\mu)}{4\pi} - 2\beta_1 \frac{\alpha_s^3(\mu)}{(4\pi)^2} \quad (1.44)$$

which can be solved, giving the running of the coupling constant:

$$\alpha_s(\mu) = \frac{4\pi}{\beta_0 \ln \frac{\mu^2}{\Lambda_{\text{QCD}}^2}} \left(1 - \frac{\beta_1}{\beta_2} \frac{\ln \ln \frac{\mu^2}{\Lambda_{\text{QCD}}^2}}{\ln \frac{\mu^2}{\Lambda_{\text{QCD}}^2}} \right) \quad (1.45)$$

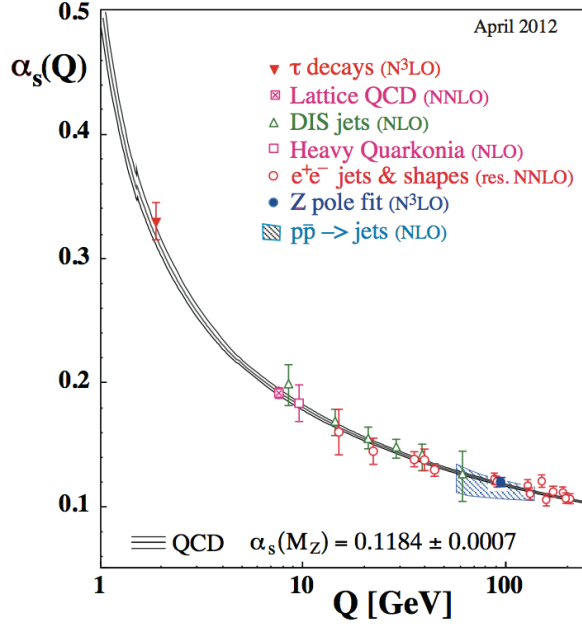


Figure 1.5: Measurements of α_s as a function of the energy scale Q , measured in different ways [43]. In the legend, the degree of QCD perturbation theory used in each extraction of α_s is indicated within the brackets.

where

$$\beta_0 = \frac{11N - 2f}{3}, \quad \beta_1 = \frac{34}{3}N^2 - \frac{10}{3}Nf - 2C_F f, \quad C_F = \frac{N^2 - 1}{2N} \quad (1.46)$$

N is the number of colors and f is the number of flavors; $\Lambda_{\text{QCD}} \approx 200 \text{ MeV}$ is the QCD scale [42]. We can yet qualitatively study the behavior of the strong interaction looking at short (high μ) and long (low μ) distances (Fig 1.5):

- when $\mu/\Lambda_{\text{QCD}} \rightarrow \infty$, $\alpha_s(\mu)$ vanishes. This shows the QCD asymptotic freedom effect: in high-energy reactions, quarks and gluons interact very weakly and a perturbative expansion can be performed;
- when the distance between two quarks is high (for low values of μ), the coupling constant takes very high values and the theory cannot be expanded perturbatively. The strong interaction amplitude at that point becomes so large that quarks cannot escape as isolated particles. This phenomenon is called “confinement”.

1.2.3 Supersimmetry

The observation of the Higgs boson (Sec. 1.2.1) is a remarkable result of both theoreticians and experimentalists. It was the last SM particle to be discovered

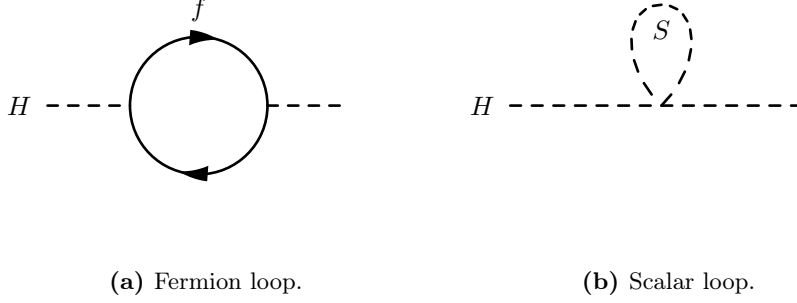


Figure 1.6: One-loop quantum corrections to the Higgs squared mass m_H^2 , due to a Dirac fermion f (a) and a scalar S (b).

and it is a confirmation of the goodness of the model, as there are still no experimental evidences of BSM physics, at least as of today.

The discovery of the Higgs boson is not, however, the end of the story. An upper limit of the validity of the SM is given, for sure, by the Planck scale $M_P = 1/\sqrt{8\pi G} = 2.4 \times 10^{18} \text{ GeV}$, where G is the Newton gravitational constant. At the Planck scale, quantum gravity effects are relevant, but they are not included in the SM.

The experimentally explored energy range reaches the TeV scale, depending on the event final state. It seems unrealistic that there exists no new physics in 15 orders of magnitude in energy [44].

This big energy leap leads also to the “hierarchy problem” of the Higgs boson. The Higgs mass receives quantum corrections from the virtual effects of every particle that can couple, directly or indirectly, to the Higgs field. For example, in a Lagrangian with a new fermion term $-\lambda_f H \bar{f} f$, the Feynman diagram in Fig. 1.6a gives the following contribution:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{UV}^2 + \dots \quad (1.47)$$

where Λ_{UV} is an ultraviolet cutoff, that can be interpreted as the energy scale of the new physics. If Λ_{UV} is of order M_P , then, this correction is more than 30 order of magnitude larger than the value of the squared Higgs mass $m_H^2 \approx (125 \text{ GeV})^2$.

If, instead of a fermion, there exists a massive complex scalar particle S , with mass m_S , which couples with the Higgs boson with $-\lambda_S |H|^2 |S|^2$, then the correction is (Fig. 1.6b):

$$\Delta m_H^2 = \frac{\lambda_S}{16\pi^2} (\Lambda_{UV}^2 - 2m_S^2 \ln(\Lambda_{UV}/m_S) + \dots) \quad (1.48)$$

Therefore, m_H^2 is sensitive to the masses of all particles to which H couples. To keep low the Higgs mass, then, it is required an extreme “fine tuning” among the different terms (at level of about 10^{-32}), raising a problem of “naturalness” known as the “hierarchy problem”.

Comparing Eq. (1.47) with Eq. (1.48), we can see that the fermion and boson terms raise opposite-sign contributions to the Higgs mass. It is possible to have a systematic cancellation of these contributions to Δm_H^2 if we introduce a new symmetry. This symmetry between fermions and bosons is called SuperSYmmetry (SUSY). The SUSY operator Q is an anticommuting spinor which transforms a fermion into a boson and vice versa:

$$Q |Boson\rangle = |Fermion\rangle \quad (1.49)$$

$$Q |Fermion\rangle = |Boson\rangle \quad (1.50)$$

In SUSY, for each SM particle, exists a supersymmetric partner with opposite-statistics behavior. SUSY models predict the existence of a plethora of new particles, one new particle for each elementary SM particle, forming “supermultiplets”, in which the two particles differ by $1/2$ of angular moment.

Since superpartners remain unobserved experimentally, their masses must be at a higher scale than those of SM particles, therefore, SUSY is a broken symmetry.

Supersymmetry predicts the existence of weakly interacting massive particles that can be candidates for dark matter and offers, in the context of the Grand Unified Theories (GUTs), the convergence of the electroweak and strong coupling constants at high energies. Therefore, searches for SUSY are an experimental priority for particle physics.

Searches for SUSY proceed through the search of supersymmetric partners (direct search) or measuring its effects on observables of known particles, like in the branching fractions (indirect search). Up to today, at the end of the LHC run I, there are no direct evidences for SUSY effects and stringent limits on SUSY parameters are set by the CMS and ATLAS collaborations [45, 46]. Some exclusion limits for SUSY predicted superparticles are shown in Fig. 1.7 and Fig. 1.8. The shown parameters come for the Minimal Supersymmetric Standard Model (MSSM) flavor of SUSY, in which the soft SUSY-breaking mass parameters are assumed to be universal at a very high energy scale where theories unify. m_0 is the universal mass for scalars, $m_{1/2}$ is the universal gaugino and higgsino mass and A_0 is the trilinear soft breaking parameter at the GUT scale [44].

Even with such strong negative search results, these direct limits do not allow to exclude SUSY or to point to specific directions. Given the large number of

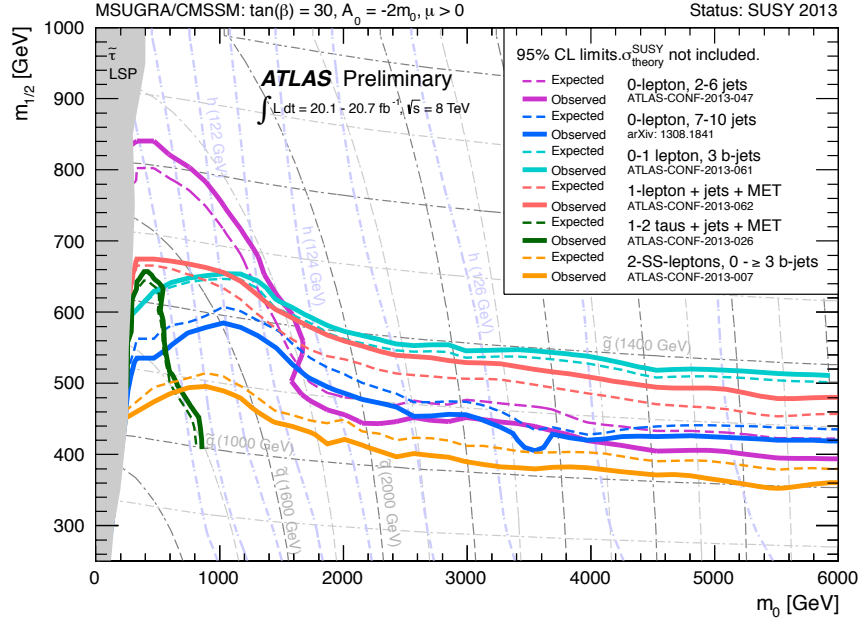


Figure 1.7: Exclusion limits at 95% CL for 8 TeV analyses in the $(m_0, m_{1/2})$ plane for the mSUGRA/CMSSM. Part of the model plane accommodates a lightest neutral scalar Higgs boson mass of 125 GeV [46].

parameters involved, there is no reason that SUSY should manifest itself in its simplest configuration. The SUSY phase space is large and some configurations could simply be experimentally more challenging. Thus, even the present strong exclusion limits allow to possible low energy supersymmetry.

On the other hand, indirect searches could provide complementary information that, together with the direct search results, could point to specific and testable scenarios. The complementarity between direct and indirect explorations is fundamental today for the searches of new physics. For instance, the discovery of the Higgs boson has provided very important information on the SUSY predictions.

In the recent years the focus of the B -physics experiments moved from the measurement and the evaluation of the CKM parameters, toward the indirect searches for new physics. The $B_s^0 \rightarrow \mu^+ \mu^-$ BF is an ideal candidate. As it will be shown in Chapter 2, thanks to its tiny value and the clean precision of its SM prediction, it is very sensible to the presence of new possible particles via loop effects.

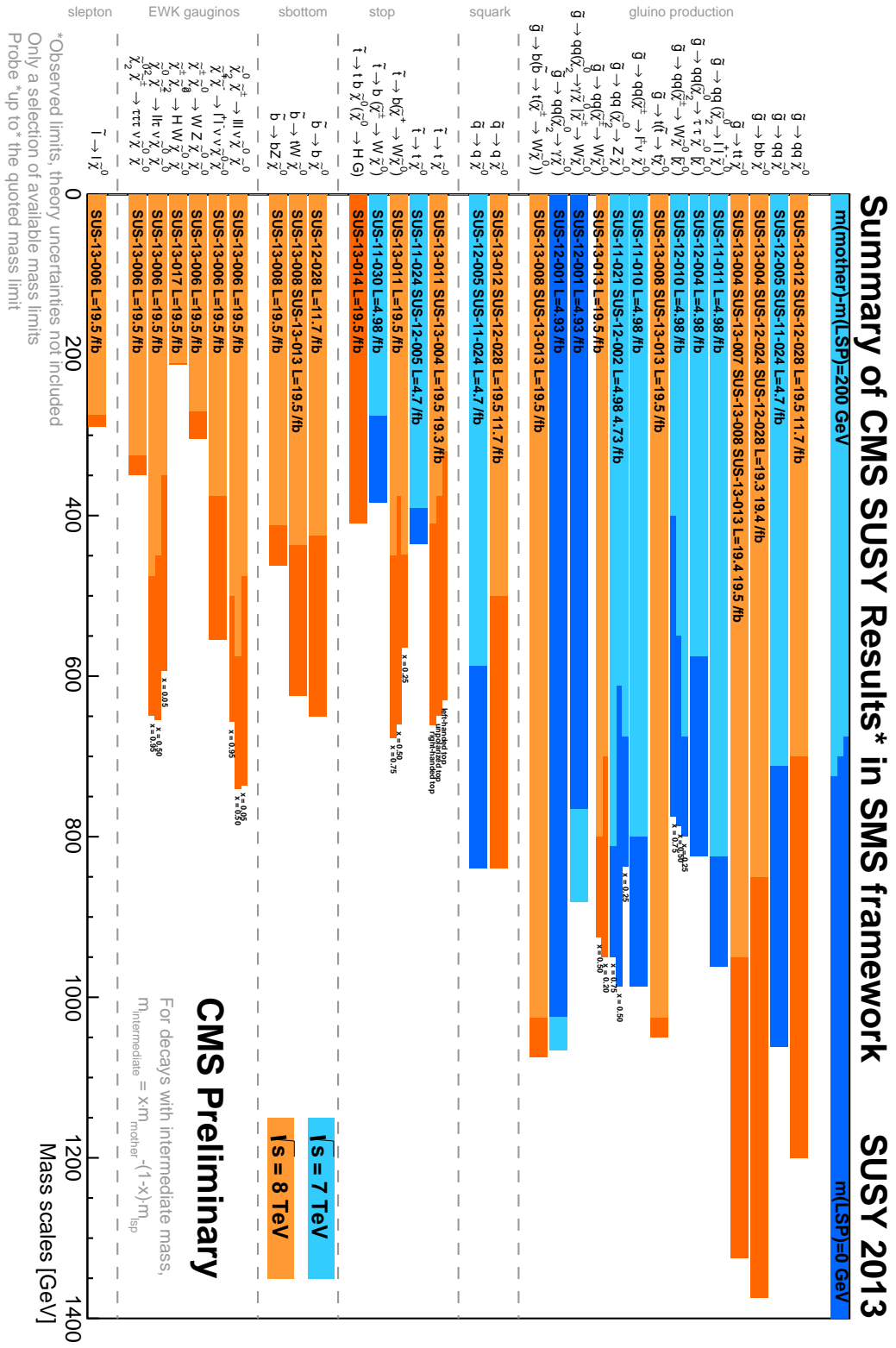


Figure 1.8: Summary of exclusion limits of CMS SUSY searches [45].

Chapter 2

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays

This Chapter examines the theoretical motivations for the search of the decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, as well as the SM branching fraction expectations. Main predictions of BSM theories are briefly discussed.

An accurate simulation of the B production in the LHC environment is fundamental to trust the MC simulations used by this analysis. The $B_{(s)}^0$ meson generation and simulation at the LHC is the topic of the last part of this Chapter.

Within this Chapter, the focus will be on the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. The SM $B^0 \rightarrow \mu^+ \mu^-$ decay behavior is similar, and differences will be stressed when needed.

2.1 B_s^0 meson proprieties

The B_s^0 meson is a neutral $\bar{b}s$ bound ground state, with zero spin and negative parity. Together with its anti-particle (\bar{B}_s^0), they form a two-component system, with a light and a heavy mass eigenstates. This means that the created flavor eigenstate meson oscillates between B_s^0 and \bar{B}_s^0 before decaying. Some noticeable B_s^0 meson parameters are summarized in Table 2.1.

The B_s^0 decay modes are dominated by those of the b quark, $b \rightarrow cW^{*-}$ (see Table 2.2); this is an example of “spectator decay”. The virtual W^{*-} boson can then decay into a pair of leptons (semileptonic decay) or of hadronizing quarks. The lifetime is quite long ($c\tau = 449 \mu\text{m}$), since the b -quark decay is highly suppressed by the CKM matrix elements (see Eq. (1.37)).

In the rest of this section the evaluation of the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ within the SM will be discussed. Any deviation from the SM value implies contributions from new physics.

Table 2.1: Main observables of the B_s^0 - \bar{B}_s^0 system [4].

Observable	Value	Unit
Mass	5366.77 ± 0.24	MeV
Mean life	$(1.466 \pm 0.031) \times 10^{-12}$	s
Width difference	$(0.100 \pm 0.013) \times 10^{-12}$	s
$B_s^0 - \bar{B}_s^0$ Mass difference	$(116.4 \pm 0.5) \times 10^{-10}$	MeV

Table 2.2: Noticeable decay modes of the B_s^0 meson [4].

Decay mode	Branching fraction \mathcal{B}
$D_s^- X$	$(9.3 \pm 2.5) \times 10^{-1}$
$l \nu_l X, l = e \mu$	$(9.5 \pm 2.7) \times 10^{-2}$
$J/\psi \phi$	$(1.09^{+0.28}_{-0.23}) \times 10^{-3}$

2.2 Operator product expansion of the Hamiltonian

Hadron weak decays, like in $B_s^0 \rightarrow \mu^+ \mu^-$, are mediated by the weak interactions of their quark constituents [41]. The initial states are bound by strong interactions whose typical scale energy is about 1 GeV. In the Operator Product Expansion (OPE) method, these decays are evaluated through an effective low energy theory in which the momentum transfer k through the weak boson vectors is very smaller than their mass. Therefore, terms of order $\mathcal{O}(k^2/M_W^2)$ can be neglected and the product of charged current operators can be expanded into a series of local long-distance operators $\langle Q \rangle$, multiplied by effective short-distance coupling constants, called the Wilson coefficients C .

In the OPE formalism the effects of a short-range exchange force mediated by a heavy boson correspond to a point interaction, as it was in the old Fermi theory of the weak interactions.

The quark strong interactions at short distances can be evaluated in perturbative series, thanks to the QCD asymptotic freedom effect (Sec. 1.2.2.1). The Wilson coefficients for QCD and weak interactions are functions of the strong coupling α_s , the weak boson mass M_W and the renormalization scale μ . They can be determined requiring that the amplitude in the full theory is reproduced by the corresponding amplitude in this effective theory, at the same order in α_s .

Therefore, the Hamiltonian operator, describing the weak decay amplitude, can be formally written as:

$$\langle \mathcal{H}_{\text{eff}} \rangle = \frac{G_F}{\sqrt{2}} V_{\text{CKM}} \langle \vec{Q}^T(\mu) \rangle \vec{C}(\mu) \quad (2.1)$$

where V_{CKM} contains the relative CKM matrix elements of the interaction under study.

2.2.1 Renormalization group

The Wilson coefficients contain terms of order [41]:

$$C_i \propto \left(\frac{\alpha_s}{4\pi} \ln \frac{M_W^2}{\mu^2} \right)^i \quad (2.2)$$

where $\mu = \mathcal{O}(m_b)$. At such low energy scales ($\approx 1 \text{ GeV}$), with respect to the weak scale M_W , the logarithm term of Eq. (2.2) becomes large, therefore the convergence of the series is slow. The use of the Renormalization Group (RG), that is the group of transformations between different choices of the renormalization scale μ , allows to change the argument of the power series expansion. The RG equations describe how the renormalized quantities evolve as a function of μ , allowing to sum up all the terms $(\alpha_s \ln(M_W/\mu))^n$, with $n = 0, \dots, \infty$, of the perturbation theory.

The sub-leading term extends the summation to all terms $\alpha_s(\alpha_s \ln(M_W/\mu))^n$, that represent an $\mathcal{O}(\alpha_s)$ correction to the leading term. Since we are interested in the short-distance QCD terms, α_s is small and this series converges quickly.

In summary, the procedure to evaluate the weak decay amplitudes of hadrons in a low energy effective theory, Eq. (2.1), can be divided in three steps [41]:

1. perturbation theory: calculation of the Wilson coefficients $\vec{C}(\mu_W)$ at order α_s , at $\mu_W \approx M_W$, so that contributions of the form $\ln(\mu_W/M_W)$ are not large and the standard perturbative series can be done;
2. RG improvement of the perturbation theory: evaluation of the coefficient evolution to the low energy scale μ , with the RG equations: $\vec{C}(\mu) = U(\mu, \mu_W) \vec{C}(\mu_W)$, where $U(\mu, \mu_W)$ is the μ -evolution operator;
3. calculation of the hadronic matrix elements $\langle \vec{Q}(\mu) \rangle$ normalized at the appropriated low energy scale μ .

2.3 Effective Hamiltonian for $B_s^0 \rightarrow \mu^+ \mu^-$

For the $B_s^0 \rightarrow \mu^+ \mu^-$ decay, the leading order Feynman diagrams are shown in Fig. 2.1. Within the SM, this decay goes through a Flavor Changing Neutral Current (FCNC) process determined by Z -penguin and box diagrams. There are no contributions from the photon propagator due to vector current conservation: a lepton-anti-lepton pair with zero angular momentum has charge conjugation

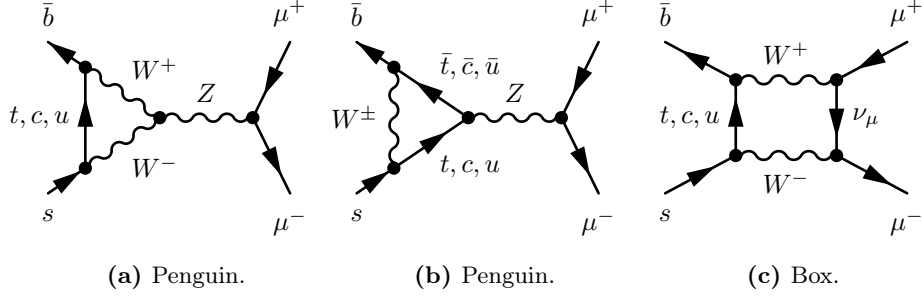


Figure 2.1: Leading Order (LO) diagrams for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. The box diagram (c) is suppressed by a factor m_W^2/m_t^2 with respect to the Z-penguin diagrams (a) and (b).

quantum number $C = 1$, while the photon has $C = -1$. Therefore this decay presents a hard quadratic GIM suppression [18] in small internal quark masses, instead of logarithmic only.

The effective Hamiltonian can be written in this general form [41]:

$$\mathcal{H}_{\text{eff}} = \sqrt{G_F} \sqrt{2} \frac{\alpha}{2\pi \sin^2 \theta_w} (\lambda_u F(x_u) + \lambda_c F(x_c) + \lambda_t F(x_t)) (\bar{b}s)_{V-A} (\bar{\mu}\mu)_{V-A} \quad (2.3)$$

where $x_i = m_i^2/M_W^2$ and $\lambda_i = V_{ib}^* V_{is}$. The functions $F(x_i)$ describe the dependence on the internal up-type quark masses m_i , and include also all QCD corrections. They are increasing functions of the quark masses, a property that is particularly important for the top quark contribution. Imposing the unitarity of the CKM matrix, the dependence of the Hamiltonian on the internal quarks comes in the form

$$\sum_{i=u,c,t} \lambda_i F(x_i) = \lambda_c (F(x_c) - F(x_u)) + \lambda_t (F(x_t) - F(x_u)) \quad (2.4)$$

Due to the GIM mechanism, $F(x_i)$ behaves quadratically in quark masses. Thus, since $x \ll 1$, then $F(x) \approx x \ln x$ and the u quark contribution in Eq. (2.3) can be neglected.

Passing to the charm sector, $F(x_c)/F(x_t) \approx \mathcal{O}(10^{-3}) \ll 1$. This contribution has to be considered together with the CKM factors involved. The result is the relative importance of the c -quark versus the t -quark. Looking at the Wolfenstein parametrization Eq. (1.40), we can see that the CKM factors λ_c and λ_t have the same order of magnitude (they are both proportional to λ^2), thus also the charm contribution is negligible.

The top and charm sectors show the same structure, the only difference coming from the quark masses. Therefore, this decay is completely determined by the top sector and allows to study the high energy scales of the SM: especially

the top quark mass m_t and the CKM coupling V_{ts} .

A theoretically clean prediction for this process can be obtained since long distance contributions are negligible and the low energy hadronic matrix elements are those of quark currents between hadron states, which can be obtained from other non rare semileptonic decays.

The SM process contains only the operator

$$Q(\bar{\mu}\mu) = (\bar{b}s)_{V-A}(\bar{\mu}\mu)_{V-A} \quad (2.5)$$

and the Hamiltonian is

$$\mathcal{H}_{\text{eff}} = -\frac{G_F}{\sqrt{2}} \frac{\alpha}{2\pi \sin^2 \theta_w} V_{tb}^* V_{ts} Y(x_t) (\bar{b}s)_{V-A} (\bar{\mu}\mu)_{V-A} + h.c. \quad (2.6)$$

where Y is the loop function consisting of Z -penguin and box diagram contributions (Fig. 2.1), and also QCD and leading electroweak corrections in the large m_t limit:

$$Y(x) = Y_0(x) + \frac{\alpha_s}{4\pi} Y_1(x) \quad (2.7)$$

Without such corrections, $Y(x)$ reduces to the LO Inami-Lim function $Y_0(x)$ [47]:

$$Y_0(x) = \frac{x}{8} \left[\frac{4-x}{1-x} + \frac{3x}{(1-x)^2} \ln x \right] \quad (2.8)$$

The Next to Leading Log (NLL) contributions can be expressed as [48]:

$$\begin{aligned} Y_1(x) = & \frac{4x + 16x^2 + 4x^3}{3(1-x)^2} - \frac{4x - 10x^2 - x^3 - x^4}{(1-x)^3} \ln x \\ & + \frac{2x - 14x^2 + x^3 - x^4}{2(1-x)^3} \ln^2 x + \frac{x + x^3}{(1-x)^2} L_2(1-x) \\ & + 8x \frac{\partial Y_0(x)}{\partial x} \ln x_\mu \end{aligned} \quad (2.9)$$

where $x_\mu = \mu^2/M_W^2$ ($\mu = \mathcal{O}(m_t)$) and

$$L_2(1-x) = \int_1^x \frac{\ln t}{1-t} dt \quad (2.10)$$

Figure 2.2 shows the dependence of the evaluated branching fraction on the renormalization scale of the top sector μ_t , in the range $100 \text{ GeV} \leq \mu_t \leq 300 \text{ GeV}$, for a top mass $m_t = 170 \text{ GeV}$. An important aspect is that the branching fraction including α_s corrections is less sensitive to the renormalization scale uncertainties than the leading order.

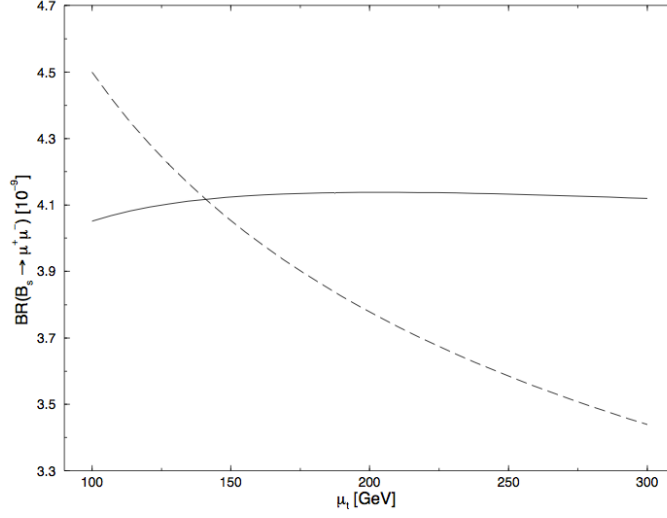


Figure 2.2: Dependence of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction on the renormalization scale μ_t with (solid curve) or without (dashed curve) $\mathcal{O}(\alpha_s)$ corrections [41].

2.4 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ SM branching fraction

The corresponding $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is given by [48]:

$$\mathcal{B}_{SM}^0(B_s^0 \rightarrow \mu^+ \mu^-) = \tau(B_s^0) \frac{G_F^2}{\pi} \left(\frac{\alpha}{4\pi \sin^2 \theta_w} \right)^2 F_{B_s^0}^2 m_\mu^2 m_{B_s^0} \sqrt{1 - 4 \frac{m_\mu^2}{m_{B_s^0}^2}} |V_{tb}^* V_{ts}|^2 Y^2(x_t) \quad (2.11)$$

where $\tau(B_s^0)$ is the B_s^0 mean life and F_{B_s} is the B_s^0 decay constant. Some noteworthy elements of Eq. (2.11) are the following:

$\frac{G_F^2}{\pi} \left(\frac{\alpha}{4\pi \sin^2 \theta_w} \right)^2 |V_{tb}^* V_{ts}|^2$: contains the gauge and CKM constants;

$F_{B_s^0}^2$: comes from the hadronic matrix element:

$$\langle 0 | \bar{b} \gamma^\alpha \gamma_5 s | B_s^0(p) \rangle = -i F_{B_s^0} p^\alpha \quad (2.12)$$

$F_{B_s^0}$ is among the simplest quantities for lattice QCD, thus a high-precision calculation is possible;

m_μ^2 : is a consequence of the chiral suppression;

$Y^2(x_t)$: is the short-distance function (the Wilson coefficient).

Eq. (2.11) shows the FCNC-related suppression times a chiral factor m_μ^2/M_W^2 , bringing an additional 10^{-6} factor. The chiral suppression and the purely leptonic nature of this process yields to an extremely rare but also very clean decay. Thus,

Table 2.3: Main relative uncertainties of Eq. (2.13).

Parameter	Value	Unit	Relative error	Ref.
$\tau_{B_s^0}$	1.503 ± 0.010	ps	0.7%	[51]
$ V_{tb}^* V_{ts} $	$(4.05 \pm 0.08) \times 10^{-2}$		4%	[39, 52]
$F_{B_s^0}$	225 ± 3	MeV	1.3%	[53]
M_t	173.4 ± 1.0	GeV	0.6%	[54]

this decay channel is one of the best available probes of new physics at the LHC reach.

Putting the latest values in Eq. (2.11), the branching fraction results [49, 50]:

$$\mathcal{B}_{\text{SM}}^0(B_s^0 \rightarrow \mu^+ \mu^-) = (3.25 \pm 0.17) \times 10^{-9} \quad (2.13)$$

with a relative uncertainty of about 5%. Eq. (2.13) can be decomposed to get a deeper insight of the principal contributing parts. A phenomenological expression of the BF is:

$$\begin{aligned} \mathcal{B}_{\text{SM}}^0(B_s^0 \rightarrow \mu^+ \mu^-) = 3.25 \times 10^{-9} \times \left(\frac{\tau_{B_s^0}}{1.503 \text{ ps}} \right) \times \left| \frac{V_{tb}^* V_{ts}}{4.05 \times 10^{-2}} \right|^2 \times \\ \times \left(\frac{F_{B_s^0}}{225 \text{ MeV}} \right)^2 \times \left(\frac{M_t}{173.4 \text{ GeV}} \right)^{3.07} \end{aligned} \quad (2.14)$$

The relative uncertainties on the measured quantities are shown in Table 2.3.

The same procedure can be applied to the rare decay $B^0 \rightarrow \mu^+ \mu^-$, switching all variables involving the strange quark with those of the down quark. The result of the calculation is

$$\mathcal{B}_{\text{SM}}^0(B^0 \rightarrow \mu^+ \mu^-) = (1.07 \pm 0.10) \times 10^{-10} \quad (2.15)$$

An important quantity is also the branching fraction ratio

$$\frac{\mathcal{B}_{\text{SM}}^0(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}_{\text{SM}}^0(B_s^0 \rightarrow \mu^+ \mu^-)} \quad (2.16)$$

that is a crucial test of the SM and a possible handle to distinguish between different new physics models (as will be shown in Sec. 2.5.4).

2.4.1 Flavor-averaged time-integrated branching fraction

What is measured by the LHC experiments is the flavor-averaged time-integrated BF:

$$\langle \mathcal{B}(B_s^0 \rightarrow f) \rangle_{[t]} = \frac{1}{2} \int_0^t [\Gamma(B_s^0(t') \rightarrow f) + \Gamma(\bar{B}_s^0(t') \rightarrow f)] dt' \quad (2.17)$$

where $\Gamma(B_s^0(t') \rightarrow f)$ is the decay distribution as a function of the proper time t' , at initial time. To evaluate it, given the total decay widths of the two mass eigenstates, $\Gamma_S^{H,L}$, one can take their semi-difference times the lifetime (the most recent values are from [51]):

$$y_s = \tau_{B_s^0} \frac{\Gamma_s^L - \Gamma_s^H}{2} = 0.062 \pm 0.009 \quad (2.18)$$

For $t \gg \tau_{B_s^0}$, the SM correction factor $\kappa_{\text{SM}}^{\mu\mu}(t, y_s)$ for the $\mathcal{B}_{\text{SM}}^0(B_s^0 \rightarrow \mu^+ \mu^-)$ branching fraction can be shown to behave like [49, 55]:

$$\kappa_{\text{SM}}^{\mu\mu}(t, y_s) \xrightarrow{t \gg \tau_{B_s^0}} \frac{1}{1 - y_s} \quad (2.19)$$

increasing the experimentally expected SM BF to:

$$\mathcal{B}_{\text{SM}}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.46 \pm 0.18) \times 10^{-9} \quad (2.20)$$

2.5 $B_{(s)}^0$ rare decays and new physics

Given that the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are highly suppressed in the SM, and have also relatively small theoretical uncertainties, they are potentially sensitive probes for physics beyond the SM. These purely leptonic decays depend on a small number of operators, so they are accessible to model-independent studies of new physics. While, in the SM, all six $B_q^0 \rightarrow \ell^+ \ell^-$ decays (with $q = d, s$ and $\ell = e, \mu, \tau$) are related to each other in a simple way, this is not necessarily true in NP models.

The SM $B_s^0 \rightarrow \mu^+ \mu^-$ effective Hamiltonian of Eq. (2.6), that contains only the operator of Eq. (2.5), may be extended to include new operators O_i , predicted by NP models [56]:

$$H = \frac{G_F}{\sqrt{2}} \frac{\alpha}{\pi \sin^2 \theta_w} V_{tb}^* V_{tq} \sum_i (C_i O_i + C'_i O'_i) + h.c. \quad (2.21)$$

where C_i is the Wilson coefficient for the operator O_i and the prime operators are chirality-flipped with respect to the non-primed operators, and they are

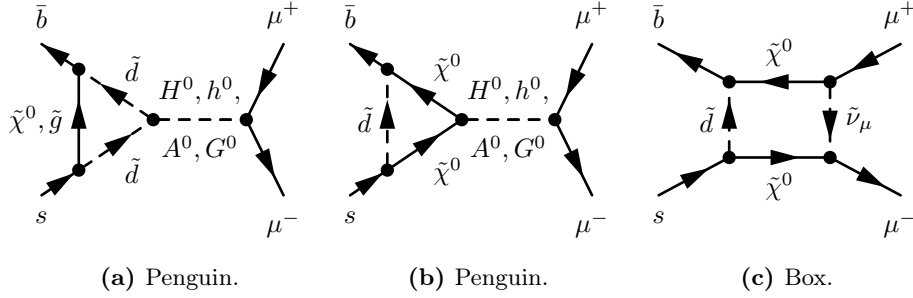


Figure 2.3: MSSM LO diagrams for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay.

highly-suppressed in the SM.

In the general MSSM, the first contributing diagrams are those shown in Fig. 2.3, where the decays can go through a profusion of SUSY different particles: H^0 , h^0 , A^0 and G^0 are the neutral Higgs and would-be Goldstone bosons; $\tilde{\nu}_\mu$ is the sneutrino, that is the supersymmetric partner of the SM neutrino; \tilde{d} denotes the down-type superpartners of the quarks (squarks); $\tilde{\chi}^0$ are the neutralinos, that are the mass eigenstates of the neutral Higgs and electroweak superpartners; and, finally, \tilde{g} is the spin-1/2 superpartner of the gluon, the gluino [57, 58].

2.5.1 Additional Higgs bosons

The SM effective Hamiltonian of Eq. (2.6) is helicity suppressed by the factor $m_\ell/M_{B_q^0}$. This makes these BFs sensible to NP models with an extended Higgs sector [59].

Higgs bosons couple to fermions with Yukawa couplings y_f . In the SM, y_b and y_ℓ , which are proportional to m_b/M_W and to m_ℓ/M_W , respectively, are so small that the Higgs penguin diagrams are negligible.

NP models with two or more Higgs multiplets generally contain tree-level FCNC couplings of neutral Higgs bosons. In one type of these models (the type-II model), one Higgs doublet, H_u , couples only to up-type fermions and the other one, H_d , couples only to down-type fermions [60]. The parameter controlling the size of the down-type Yukawa coupling is $\tan \beta = v_u/v_d$, the ratio of the vacuum expectation values acquired by H_u and H_d . The Yukawa coupling y_f of H_d to the fermion f satisfies $y_f \sin \beta = m_f \tan \beta / v$, with $v = \sqrt{v_u^2 + v_d^2} = 246 \text{ GeV}$ (see also Eq. 1.18).

The dominant contributions can be expressed in terms of $\tan \beta$ and the charged Higgs boson mass M_{H^\pm} [61]:

$$C_S = C_P = \frac{m_\ell}{4M_W^2} \tan^2 \beta \frac{\ln r}{r-1} \quad \text{with} \quad r = \frac{M_{H^\pm}^2}{\mu^2} \quad (2.22)$$

where C_S and C_P are the Wilson coefficients of the scalar and pseudoscalar operators

$$Q_S = m_b \bar{b}_R q_L \bar{\ell} \ell \quad (2.23)$$

$$Q_P = m_b \bar{b}_R q_L \bar{\ell} \gamma_5 \ell \quad (2.24)$$

respectively.

The branching fraction can be enhanced for very large values of $\tan \beta / M_{H^+}$, but it could also be reduced: for example, for $\tan \beta = 60$ and $M_{H^+} = 500 \text{ GeV}$, the $B_q^0 \rightarrow \ell^+ \ell^-$ BFs would be reduced by 50%.

2.5.2 MFV-MSSM and other models

The generic MSSM contains many new sources of flavor violation in addition to the Yukawa couplings, that could easily exceed those of the CKM mechanism. Because of the very good CKM description of flavor changing transitions one may add to the MSSM the hypothesis of Minimal Flavour Violation (MFV) [62]. In the MFV-MSSM model, the only sources of flavor violation are still the Yukawa couplings, as in the SM.

The MSSM has two Higgs doublets which, at the one-loop level, couple to all fermions.

In MFV-MSSM, the Wilson coefficients C_S and C_P differ from Eq. (2.22) because they involve the third power of $\tan \beta$ and they depend on the neutral mass $M_{A^0} \approx M_{H^0}$, instead of the charged Higgs boson mass. In this case the branching fractions scale as [63]:

$$\mathcal{B}(B_q^0 \rightarrow \ell^+ \ell^-)_{\text{SUSY}} \propto \frac{m_b^2 m_\ell^2 \tan^6 \beta}{M_{A^0}^4} \quad (2.25)$$

The $\tan^6 \beta$ factor makes these branching fractions very sensible to SUSY contributions.

Other variants of the MSSM incorporate the parameter constraints from GUTs. The $B_s^0 \rightarrow \mu^+ \mu^-$ BF is interesting in GUTs based on the symmetry group $SO(10)$ because, in these theories, the top and bottom Yukawa couplings y_b and y_t unify at a high scale, implying that $\tan \beta$ is of order 50. Thus, any experimental information, such as in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, on the deviation of y_b / y_t , would probe the Higgs sectors of GUT theories.

2.5.3 The $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ constraints on SUSY phase space

Flavor observables play an important role nowadays on the allowed SUSY phase space. In fact, the $B_s^0 \rightarrow \mu^+ \mu^-$ BF measurement contribution to the SUSY

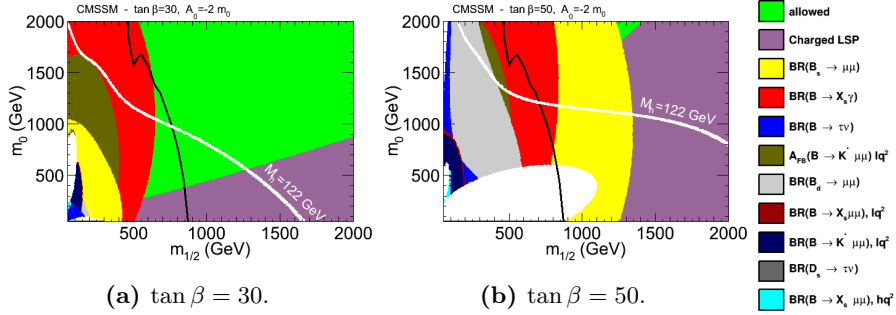


Figure 2.4: Flavor constraints in the CMSSM, in the $(m_{1/2}, m_0)$ parameter plane with $A_0 = -2m_0$, for $\tan \beta = 30$ (a) and $\tan \beta = 50$ (b). The black lines delimit the ATLAS SUSY direct search limits with 20.3 fb^{-1} of data and the white lines show where the Higgs mass can reach a value of 122 GeV [64].

search can be as strong as the current direct searches, depending on the supposed model.

As examples of constraints, Fig. 2.4 shows the allowed values of the parameters $(m_{1/2}, m_0)$ in the Constrained Minimal Supersymmetric Standard Model (CMSSM), due to direct and indirect searches. At large $\tan \beta$ the B -physics constraints, especially from the current $B_s^0 \rightarrow \mu^+ \mu^-$ BF measurement, in yellow, are stronger than the direct searches.

2.5.4 New physics correlations

An important and promising aspect of the $B_s^0 \rightarrow \mu^+ \mu^-$ BF is that, in NP models, its possible deviation from the SM is correlated to other observables that can be measured with a good precision in modern experiments.

For instance, in the CMSSM model, the observation of the Higgs-like boson shrinks the allowed $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ range. Fig. 2.5 shows the permitted $B_s^0 \rightarrow \mu^+ \mu^-$ BF as a function of $\tan \beta$ and m_0 (gray points). The constraint of a light Higgs boson h with mass in $[123, 129] \text{ GeV}$ (green points) does not allow to have $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ lower than the SM prediction.

Correlations are relevant also within the flavor sector. As an example, the ratio of the two B rare decays:

$$\frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = \frac{\tau(B^0)}{\tau(B_s)} \frac{m_{B^0}}{m_{B_s}} \frac{F_{B^0}^2}{F_{B_s}^2} \left| \frac{V_{td}}{V_{ts}} \right|^2 \quad (2.26)$$

is of particular interest because the theoretical errors cancel to a large extent.

These correlations can play an important role in distinguishing different models of NP with similar LHC signatures, and in probing the flavor structure of NP theories [66].

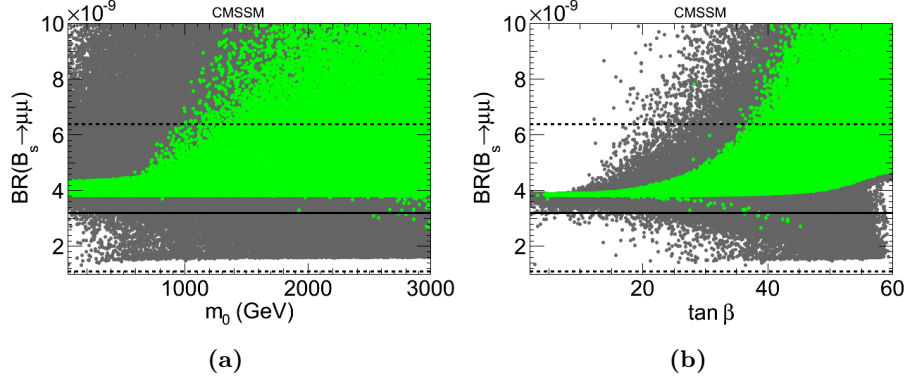


Figure 2.5: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ as a function of m_0 (a) and $\tan \beta$ (b). Solid lines correspond to the central value of the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ measurement, and dashed lines to the 2σ experimental deviations from [65]. The green points are those in agreement with the Higgs mass constraint [64].

To obtain correlations between the predictions for various observables, assumptions have to be made on a NP theory, like using the flavor symmetry. In the MFV, flavor violation is allowed only through the CKM matrix (Sec. 1.2.1.1). Thus, all Wilson coefficients are flavor independent:

$$C_i^{bs} = C_i^{bd} = C_i^{sd} \quad (2.27)$$

Allowed $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ BF's for different NP models are shown in Fig. 2.6. The orange MFV line is obtained with the flavor independence of Eq. (2.27). The AKM and RVV2 models predict values which are close to the ones of MFV; the AC model predicts a higher $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ than the MFV. The fourth-generation model (SM4) predicts a possible enhancement of only one of the two rare decays. In a MFV model with left-handed currents only (LL), there is no correlation. Other models, like the RSc [67] and LHT [68] ones, predict effects below today and near-future experimental sensitivities.

Other correlations can also be obtained. For example, models with partially composite quarks (like composite Higgs models or models with a warped extra dimension) predict interesting correlations (as they may be experimentally measurable with precision in a near future) also between the $B_s^0 \rightarrow \mu^+ \mu^-$ and the $B^0 \rightarrow \mu^+ \mu^- K^{*0}$ BF's [75] (Fig. 2.7).

Therefore, the correlations between flavor observables are a powerful tool to distinguish between various NP models and to select the ones that are consistent with data.

To conclude, the $B_s^0 \rightarrow \mu^+ \mu^-$ decay is considered a “golden channel” for NP models, as it has a low theoretical error (Sec. 2.4) and it is accessible to experiments with a relatively simple signature. Thus, precise experimental

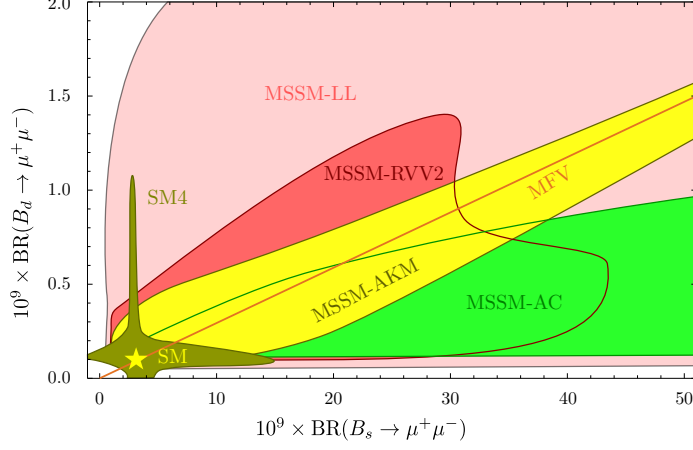


Figure 2.6: Correlation between the branching fractions of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ in MFV [69], SM4 [70] and in four MSSM flavor models: AC [71]), RVV2 [72]), AKM [73] and LL [74]. The SM point is marked by a star [66].

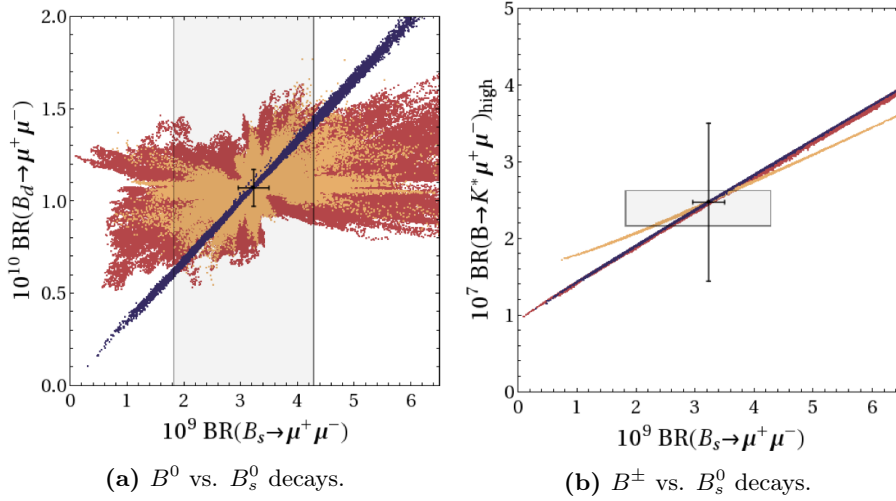


Figure 2.7: Correlations between branching fractions in B_s^0 and B^0 decays (a) and in B_s^0 and B^\pm (in the high q^2 region) decays (b), in the anarchic triplet model (yellow/light gray), in the anarchic bidoublet model (red/medium gray) and in the bidoublet model with $U(2)^3$ left-handed compositeness (blue/dark gray). The shaded regions indicate experimental 1σ ranges as of 2012. The black cross shows the SM predictions with theory uncertainties [75].

Table 2.4: Latest upper limits on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions, for each collaboration.

Collaboration	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	Ref.
ARGUS (1987)		5.0×10^{-5} (90% CL)	[77]
UA1 (1991)		8.3×10^{-6} (90% CL)	[78]
CLEO (2000)		6.1×10^{-7} (90% CL)	[79]
Belle (2003)		1.6×10^{-7} (90% CL)	[80]
BABAR (2008)		5.2×10^{-8} (90% CL)	[76]
ATLAS (2012)	2.2×10^{-8} (95% CL)		[81]
CMS (2012)	7.7×10^{-9} (95% CL)	1.8×10^{-9} (95% CL)	[82]
LHCb (2012)	4.5×10^{-9} (95% CL)	1.1×10^{-9} (95% CL)	[83]
CDF (2013)	3.1×10^{-8} (95% CL)	4.6×10^{-9} (95% CL)	[84]
D0 (2013)	1.5×10^{-8} (95% CL)		[85]

measurements of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ are very important and promising for BSM theories.

2.6 Status of experimental results

Many modern HEP detectors have published results on the search for $B_s^0 \rightarrow \mu^+ \mu^-$ or $B^0 \rightarrow \mu^+ \mu^-$ decays. Over the past 30 years, significant progress in sensitivity has been made, with exclusion limits on the branching fractions improving by five orders of magnitude. The main reason of the improvement is the collected statistics, that, especially for the B factories, has been too low for approaching the SM predictions.

For instance, at BABAR the latest published results [76] use an integrated luminosity of 347 fb^{-1} , that corresponds to 3.8×10^8 B^0 - \bar{B}^0 pairs, produced from the $\Upsilon(4S)$ decay. Considering the SM expectation from Eq. (2.15), the full statistics, without even taking into account the efficiency, contains less than 0.1 expected $B^0 \rightarrow \mu^+ \mu^-$ decays.

The hadron colliders like the Tevatron and the LHC, due to the higher energy involved, produce many B^0 mesons, and of course also B_s^0 mesons. This, together with their relatively high instantaneous luminosities, make possible to approach the SM expectations.

Table 2.4 shows the latest results on the upper limits for each collaboration.

In 2012 the CDF collaboration, using a neural network analysis, observed an excess of B_s^0 candidates, with the data sample collected at the Tevatron, at $\sqrt{s} = 1.96 \text{ TeV}$. Their measurement with the full data sample of 10 fb^{-1} is $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (1.3_{-0.7}^{+0.9}) \times 10^{-8}$, with a significance of 2.3σ [84].

The most recent results from the LHCb collaboration were obtained with the

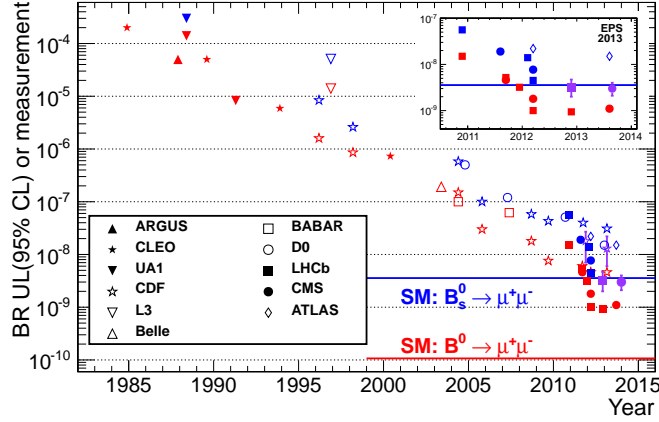


Figure 2.8: Summary of the reported upper limits and measurements on the branching fractions $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ as a function of time [87], for each experiment.

two years of LHC data-taking: a total integrated luminosity of 3 fb^{-1} [86]. A boosted decision tree analysis in eight independent bins reports a time-integrated measured BF of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.9_{-1.0}^{+1.1}) \times 10^{-9}$ with a significance of 4σ , and a $B^0 \rightarrow \mu^+ \mu^-$ upper limit of $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 7.4 \times 10^{-10}$ at 95% CL.

The CMS collaboration, at the same time as the latest results from the LHCb collaboration, reported their results with the total acquired integrated luminosity of 2011 and 2012 runs (about 25 fb^{-1}) [1]. The $B^0 \rightarrow \mu^+ \mu^-$ upper limit is set to $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.1 \times 10^{-9}$. The $B_s^0 \rightarrow \mu^+ \mu^-$ BF is measured to be $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0_{-0.9}^{+1.0}) \times 10^{-9}$. The excess has a significance of 4.3σ .

An illustrative summary of the upper limits and measurements on these two branching fractions is shown in Fig. 2.8.

The way in which the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction has been measured by the CMS collaboration is the topic of this thesis.

2.7 Production of $B_{(s)}^0$ mesons at the LHC

Free quarks cannot be observed: a quark never exists on its own for a time longer than $\approx 1/\Lambda_{\text{QCD}}$: up, down, strange, charm, and bottom quarks all hadronize, becoming part of a meson or baryon. The top quark, due to the high value of its mass, decays before it has time to hadronize. The hadronization is caused by the color confinement of QCD: color charged particles cannot be observed singularly. This couples the hadronization of a parton to the flavor content and momentum of the other parton, providing an explicit dependence on the production environment. When two quarks become separated, at some point it is more energetically favorable for a new quark-antiquark pair to spontaneously

appear, than to extend the starting pair further. As a result of this, when quarks are produced in particle accelerators, instead of the individual quarks in detectors, jets of many color-neutral particles (mesons and baryons) appear, clustered together.

The hadronization occurs at a significantly longer time scale than the production process. Based on this factorization, one can subdivide the hadron formation process in two distinct phases, the perturbative one, when the quarks are produced, and the nonperturbative one, when these quarks hadronize.

2.7.1 The factorization hypothesis

Due to the QCD asymptotic freedom the interaction between quarks and gluons becomes arbitrarily weak at short distances. Thus, at large transferred momenta, hadrons can be seen as collections of almost non-interacting partons: a scattering between two hadrons is actually a scattering between two free partons within the hadrons and a perturbative method can be used to describe their interaction.

If a parton carries the momentum fraction x_i of a hadron i with momentum P_i , the virtuality Q^2 of a parton-parton interaction is defined as

$$Q^2 = \hat{s} = x_1 x_2 s \quad (2.28)$$

where $s = (P_1 + P_2)^2$ is the square of the center-of-mass energy of the total system¹.

In a hadron-collider, as at the LHC, the center-of-mass of the parton interaction is unknown and it may be normally boosted with respect to the laboratory frame. For this reason it is useful to describe the process with variables that are invariant under longitudinal boosts:

- the transverse momentum of the partons \hat{p}_T in the center-of-mass frame of the collision;
- the rapidity

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (2.29)$$

where E is the energy and p_z is the momentum along the longitudinal, colliding axis, which reflects the momentum imbalance of the partons. At relativistic energies its value is close to the spatial coordinate η , called pseudorapidity:

$$\eta = \frac{1}{2} \ln \frac{p + p_z}{p - p_z} \quad (2.30)$$

¹Here we assume that the parton mass values are negligible.

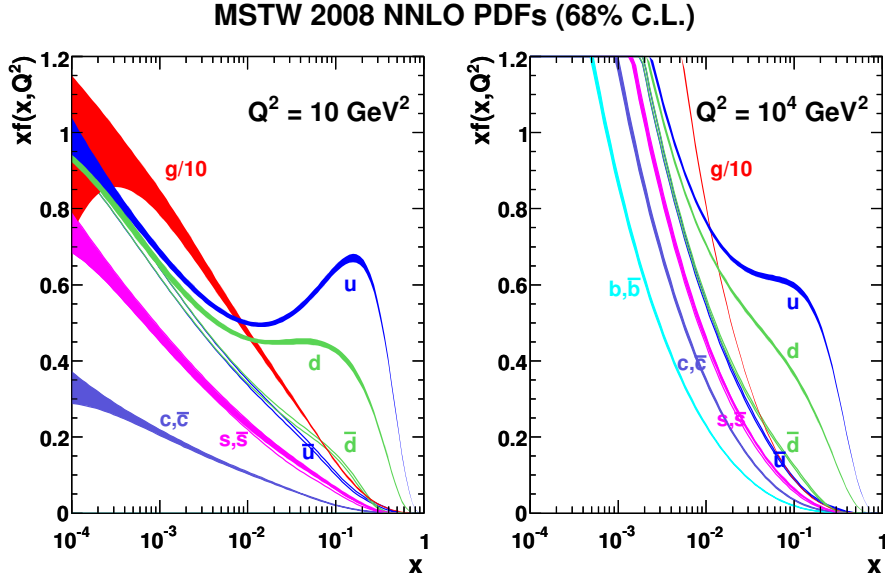


Figure 2.9: Distribution of x times the unpolarized parton distributions $f(x)$ (where $f = u_v, d_v, \bar{u}, \bar{d}, s, c, b, g$), at a scale $\mu^2 = 10 \text{ GeV}^2$ (left) and $\mu^2 = 10\,000 \text{ GeV}^2$ (right), using the NNLO MSTW2008 parametrization [89].

where p is the total momentum. In scattering collisions, particle production is roughly constant as a function of η .

The short-distance part covers the hard process, calculable in perturbative QCD (pQCD), while the long-distance part includes collinear and soft divergences which are not accessible to perturbative calculations.

The factorization theorem [88] states that the cross-section for a hard scattering between two hadrons with momenta P_1 and P_2 can be put in this form:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i^1(x_1, Q^2) f_j^2(x_2, Q^2) \hat{\sigma}_{i,j}(x_1 P_1, x_2 P_2, Q^2) \quad (2.31)$$

where $\hat{\sigma}_{i,j}$ is the scattering cross-section for partons i and j , evaluable by pQCD. $f_i^j(x, Q^2)$ is the Parton Distribution Function (PDF) for parton i in hadron j , that describes the probability density to find the parton with a certain momentum fraction x of P_i , at virtuality Q^2 . PDFs, due to their non-perturbative nature, cannot be obtained by pQCD, but can instead be measured from lattice QCD calculations and experimental data (Fig. 2.9).

The factorization theorem implies that the probability of extracting the parton can be treated independently from the parton undergoing an interaction. This assumption was successfully verified at the deep inelastic lepton-hadron scattering experiments [90, 91].

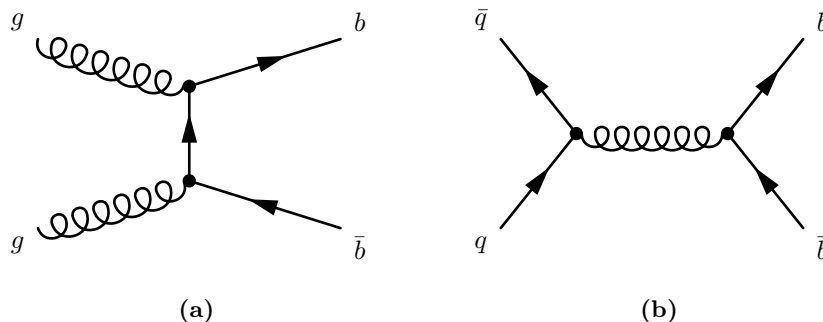


Figure 2.10: LO diagrams for b -quark production.

2.7.2 b -quark production mechanisms

Many mechanisms can contribute to the b -production, at a given energy scale Q^2 . For QCD processes, the starting LO diagrams are the $\mathcal{O}(\alpha_s^2)$ quark-antiquark annihilation and gluon-gluon fusion interactions shown in Fig. 2.10:

$$q\bar{q} \rightarrow Q\bar{Q}, \quad gg \rightarrow Q\bar{Q}. \quad (2.32)$$

If also Next to Leading Order (NLO) diagrams are considered, the $b\bar{b}$ generation mechanism usually may be divided in three categories [92] (see Fig. 2.11):

1. Pair Creation (PC), which includes $b\bar{b}$ production through $q\bar{q}$ annihilation and gluon fusion, and higher-order correction to these processes. The LO diagrams of Fig. 2.10 belong to this category. The higher orders do not change the production cross-sections, but they modify the kinematics (Fig. 2.11a). In fact, at LO, the b and \bar{b} quarks have to emerge back to back, to conserve the momentum, instead, at NLO a net recoil can be taken by one or more further partons;
2. Flavor Excitation (FE) production, where a $b\bar{b}$ couple from the quark sea inside the proton is excited into the final state and only one b -quark undergoes a hard QCD interaction (Fig. 2.11b);
3. Gluon Splitting (GS), where the $b\bar{b}$ is generated from a gluon splitting, in the initial or final state, and the heavy flavors do not enter the hard process (Fig. 2.11c).

Thus, these three mechanisms above are distinguished by having 2, 1 or 0, respectively, heavy flavors in the final state of the hard scattering. b quarks can be produced in decays of heavier resonances, such as $Z \rightarrow b\bar{b}$, $H \rightarrow b\bar{b}$, $\Upsilon \rightarrow b\bar{b}$ and $t \rightarrow bW^+$. Their contribution is, however, negligible with respect to the QCD production in a pp or $p\bar{p}$ collider.

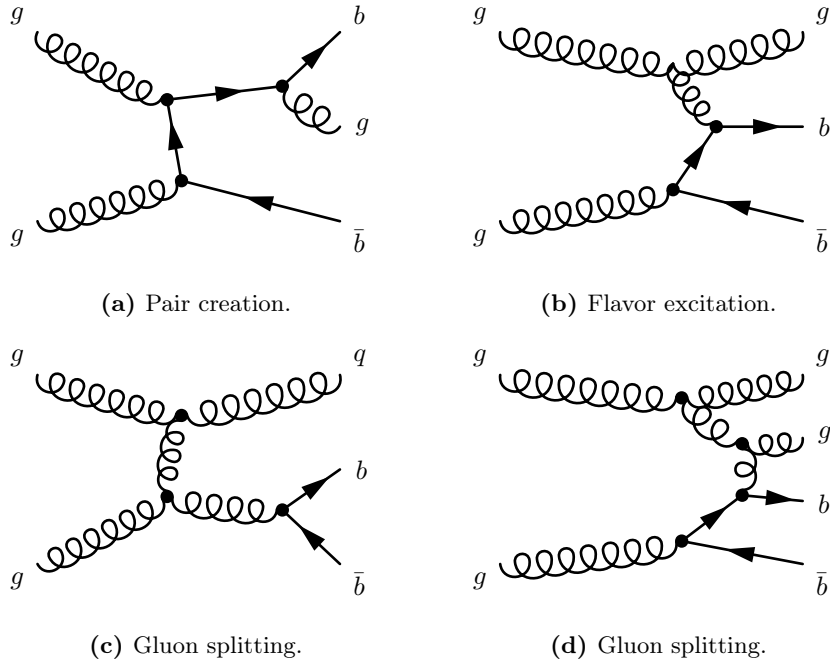


Figure 2.11: Examples of diagrams after the LO: pair creation with gluon emission (a), flavor excitation (b), gluon splitting (c) and events classified as gluon splitting but of flavor-excitation nature (d) [92].

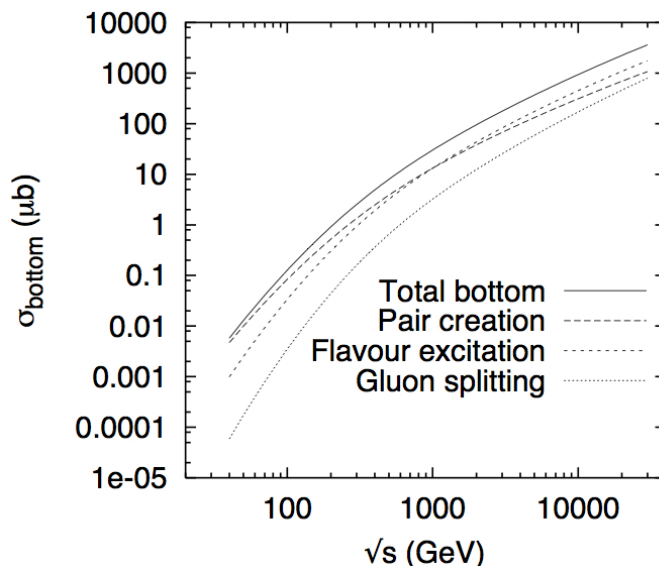


Figure 2.12: Bottom cross-sections for pp collisions as a function of \sqrt{s} . The contributions from pair creation, flavor excitation and gluon splitting are shown separately.

In Fig. 2.12 the predicted b -quark cross-section for pp collisions is shown as a function of \sqrt{s} . For small energies, the production is dominated by the pair-creation mechanism, followed by a significant fraction of flavor excitation, and the gluon splitting contribution is negligible, due to the fact that the flavor excitation mechanism requires not strong kinematics. As we increase the energy, flavor excitation overtakes the pair production and the gluon splitting mechanism starts to approach the other two. This behavior is understood noticing that any partonic process only contains one $2 \rightarrow 2$ scattering; instead, due to the higher gluon virtuality, the number of branchings in the associated initial and final state showers increases with energy. The three generation mechanisms also make slightly different kinematical distributions for the signal b -quark (Fig. 2.13).

At the LHC an interesting feature is that the colliding particles are two protons, so the total valence quark-antiquark content differs from zero. This leads to an asymmetric production of B and \bar{B} mesons (Fig. 2.14).

2.7.3 Hadronization

The partonic state, that is going to hadronize, consists of the outgoing partons from the hard scattering and of the beam remnant partons. Since the value of the coupling α_s rises strongly at large distances (Fig. 1.5), the fragmentation process cannot be calculated in pQCD and phenomenological models have to be applied.

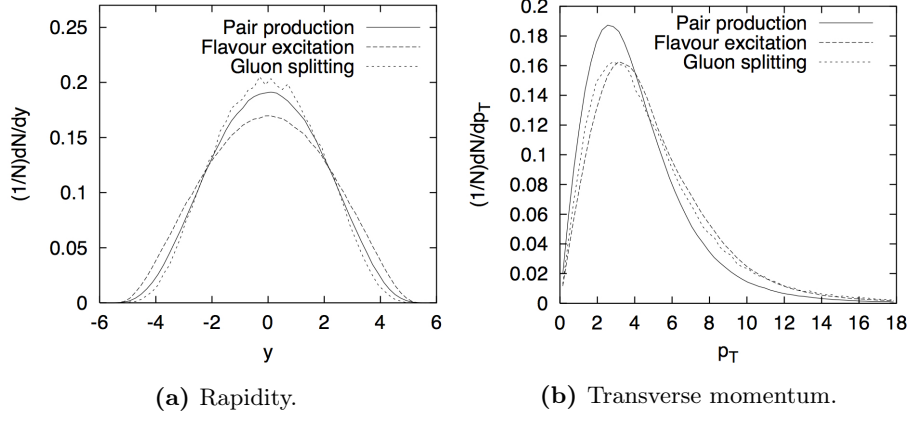


Figure 2.13: Normalized y (a) and p_T (b) distributions for single b quarks at a 2 TeV $p\bar{p}$ collider, divided into the three NLO processes.

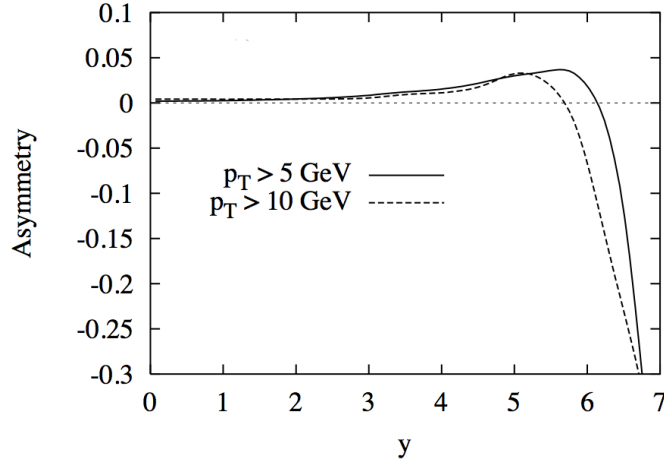


Figure 2.14: B^0 meson pair-production asymmetry $A = \frac{\sigma(B^0) - \sigma(\bar{B}^0)}{\sigma(B^0) + \sigma(\bar{B}^0)}$ at the LHC, as a function of rapidity, for two different p_T selections.

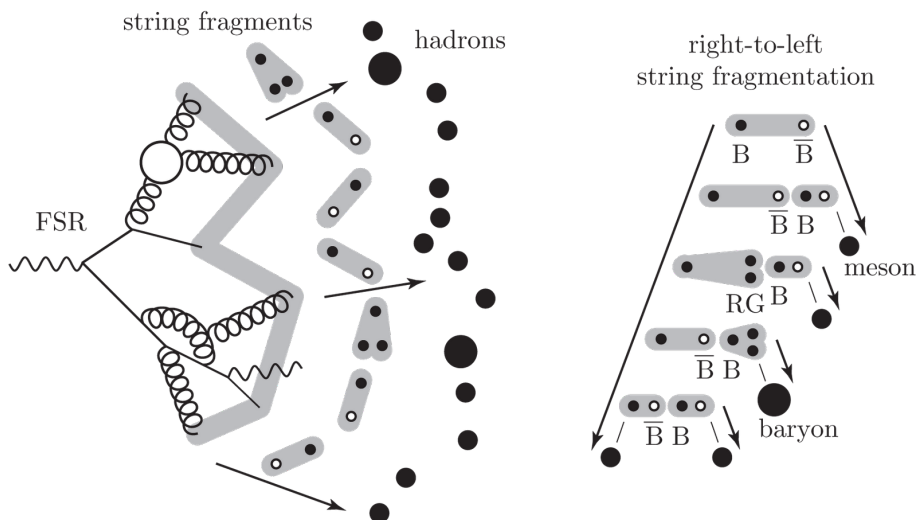


Figure 2.15: Sketches of a hadronization of a parton shower from final state radiation. Grey lines denote the color strings.

The Lund string fragmentation is a typical model for studying the non-perturbative hadronization [93], implemented in many popular Monte Carlo simulation softwares (like PYTHIA [94] and UCLA [95]). In this model the gluon color field is modeled as a narrow tube (or string) of color field between the quark-antiquark pair. These confinement strings correspond to a Lorentz-invariant description of a linear potential $V(r) = \kappa r$, with string tension $\kappa \approx 1 \text{ GeV/fm}$. Each string has a color charge at one end and its anticolor at the other, made of quarks; the double color charge of the gluon corresponds to the string (Fig. 2.15). This approach may be used for all strings above a cut-off mass of a few GeV.

The hadronization of a parton is presumed to be universal, not depending on the environment of the production. The function describing the hadronization is called the fragmentation function $D_i^h(z, \mu^2)$, that represents the probability that the parton i fragments into a hadron h carrying a fraction z of the parton momentum. If the fragmentation functions are combined with the cross-sections for the inclusive production of each parton type in the given physical process, predictions can be made for the momentum spectra of final state hadrons.

2.7.4 Monte Carlo generation

This analysis uses MC simulation samples to determine the signal and normalization efficiencies, to estimate rare background contributions and to parametrize the invariant mass shapes of signals and backgrounds.

In the presence of a hard process, characterized by particles, or clusters of particles, with a large transverse momentum with respect to the beam direction,

the final state of hadron-hadron interactions can be described as the superposition of several contributions:

1. products of the partonic hard scattering with the highest p_T , including initial and final state radiation;
2. hadrons produced in additional multiple parton interactions;
3. beam-beam remnants resulting from the hadronization of the partonic constituents that did not participate in other scatterings.

The last two contributions form the underlying event.

The event generation starts from minimum-bias processes, that are needed for a faithful description of the physical environment. This environment is especially required to study the isolation variables (see also Sec. 4.3.2.1), which are among the most important ones to separate the signal from the background.

The generation of each sample is done through the PYTHIA6 MC simulator [96] and the Z2* tune [97], which uses the CTEQ6L parton distribution functions [98]. Tunes differ in the treatment of parton radiation and hadronization, and in the choice of underlying event parameters including those regulating parton showers, color reconnections, and cutoff values for the multiple parton interaction mechanism. Values of these parameters were chosen to provide a reasonable description of existing LHC pp differential data measured in minimum-bias and hard QCD processes.

The unstable particles are then decayed via the EVTGEN simulation [99].

The PYTHIA simulation generates a couple of b -quarks. This $b\bar{b}$ couple tends to behave differently, depending on the type of generation mechanism. A pair coming from a PC mechanism tends to be back-to-back in the azimuthal angle ϕ , and have a balance in the transverse momentum. Instead, a FE pair will be flat in $\Delta\phi$ and asymmetric in p_T . Finally, GS b -quarks tend to be close to each other, resulting in peaking at small values of $\Delta\phi$.

For these reasons, MC generators must take into account all these production mechanisms, in the right proportions, when generating B mesons, and a possible systematic uncertainty has to be evaluated.

Chapter 3

The Large Hadron Collider and the CMS apparatus

The measurements described in this thesis are based on a data sample collected by the CMS detector during 2011 and 2012 at the LHC. This Chapter provides a general description of the experimental apparatus, both collider and detector, focusing on the more relevant elements for this analysis: the tracker and muon sub-detectors, the tracking algorithms and the data acquisition.

3.1 The Large Hadron Collider

The LHC is a two-ring, superconducting accelerator and collider installed at the Conseil Européenne pour la Recherche Nucléaire (CERN) site in the 27 km long tunnel, that once was containing the Large Electron-Positron Collider (LEP) [100]. Its main purposes are the discovery of the Higgs boson and the search for new physics.

The LHC at design can generate pp scattering events with a center-of-mass energy of 14 TeV. Furthermore, it can accelerate and collide heavy ions, for studies regarding dense matter and quark-gluon plasma.

Seven experiments are currently installed around the ring. Two of them, AToroidal LHC ApparatuS (ATLAS) [101] and CMS [102], are general purposes detectors especially designed for the Higgs boson and high energy, rare new physics searches. The LHCb experiment [103] aims specifically for b -physics and the TOTal Elastic and diffractive cross section Measurement (TOTEM) experiment [104] for the measurement of the proton elastic scattering at small angles. The A Large Ion Collider Experiment (ALICE) experiment [105] has been especially designed to study heavy ions (Pb-Pb) collisions. The Large Hadron

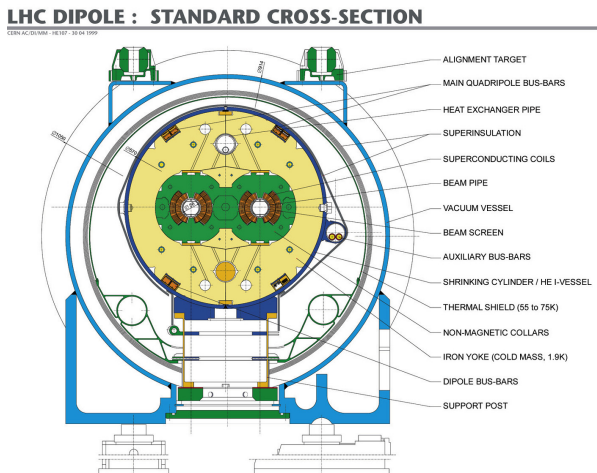


Figure 3.1: Cross-section of an LHC dipole magnet with cold mass and vacuum chamber.

Collider forward (LHCf) [106] aims to study the production of particles in the forward region. Finally, the goal of the Antihydrogen Laser PHysics Apparatus (ALPHA) [107] is the matter-antimatter analysis, through the production of antihydrogen atoms.

3.1.1 The LHC layout

The LHC uses twin bore magnets, formed by two sets of coils and beam channels, inside the same mechanical structure and cryostat (see Fig. 3.1 for the dipole section).

The LHC is made of eight arcs, each containing 154 dipole bending magnets, and eight 528 m long insertions, consisting of a long straight section plus two transition regions (Fig. 3.2). A sector is the region between two insertion points. ATLAS and CMS are located at diametrically opposite straight sections: point 1 and point 5, respectively. Point 2 and point 8 contain ALICE and LHCb experiments, respectively, and also the injection systems for the two beams. Two collimation systems reside at Point 3 and 7 while point 4 contains two Radio Frequency (RF) systems. Finally, Point 6 contains the beam dump insertion where the two beams are extracted from the machine.

Protons that will collide are extracted from a hydrogen source, and then tunneled into many different accelerators, before entering the LHC (see Fig. 3.6 at the end of this section). The first one is the LINear particle ACcelerator (LINAC)-2, generating 50 MeV protons, which are then injected into the Proton Synchrotron Booster (PSB). There, protons are accelerated to 1.4 GeV and then fed into the Proton Synchrotron (PS), where they reach the energy of 26 GeV.

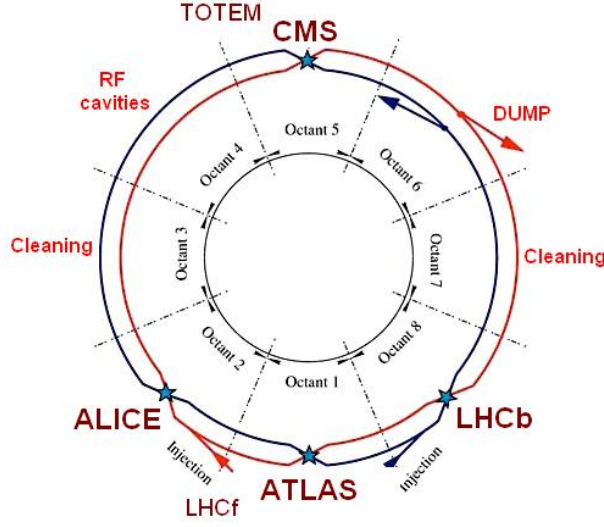


Figure 3.2: Sketch of the eight sectors of the LHC. The four main experiments are highlighted in bold.

Finally, the Super Proton Synchrotron (SPS) is used to further increase their energy up to 450 GeV, before the injection enters into the LHC ring.

The 1232 superconducting dipoles in the LHC work at the nominal magnetic field of 8.33 T; 392 focusing quadrupole magnets complete the steering of the beams. The induced RF electrical field runs at 400 MHz and provides an accelerating field of 5 MV/m. To achieve this huge value of the magnetic field, the structure operates in superfluid helium at 1.9 K. The large thermal conductivity and low viscosity of He allow for the dissipation of the generated heat.

The main LHC design parameters are summarized in Table 3.1.

3.1.2 The LHC luminosity

For each physics process, the event rate n_{event} is given by the product of its production cross-section σ_{event} , at that particular energy (Fig. 3.3), and of the luminosity \mathcal{L} , that depends only on the beam parameters:

$$n_{\text{event}} = \mathcal{L} \sigma_{\text{event}} \quad (3.1)$$

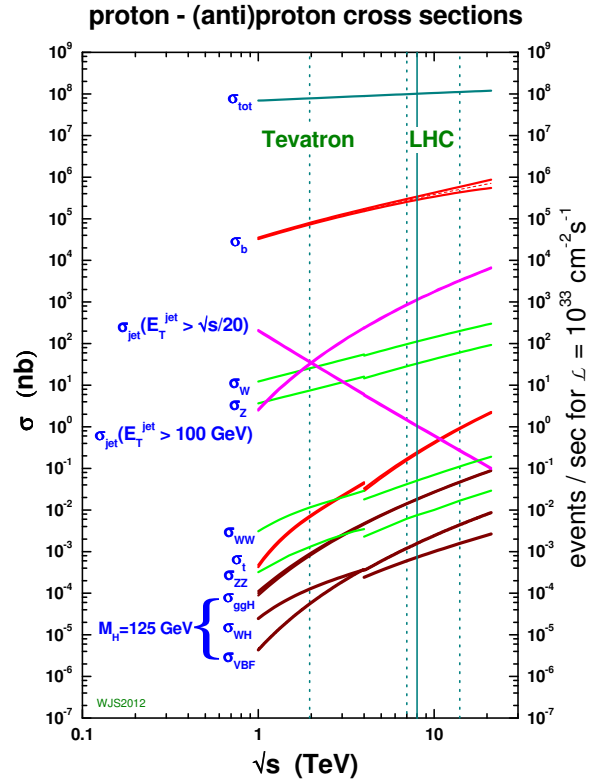
In case of equal beam parameters and Gaussian shapes, the luminosity can be expressed as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (3.2)$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor,

Table 3.1: Main parameters of the LHC, at design and for $\sqrt{s} = 14$ TeV.

Parameters	Value	Unit
Proton energy	7	TeV
Relativistic gamma	7461	
Number of particles per bunch	1.15×10^{11}	
Number of bunches	2808	
Transverse normalized emittance ϵ_n	3.75	$\mu\text{m rad}$
Circulating beam current	362	A
Stored energy per beam	362	MJ
RMS bunch length	7.55	cm
RMS beam size at point 5	16.7	μm
β function at point 5	0.55	m
Peak luminosity at point 5	1.0×10^{34}	$\text{cm}^{-2} \text{s}^{-1}$
Inelastic cross-section	60	mb
Total cross-section	100	mb
Events per bunch crossing	19	
Luminosity lifetime	14.9	h

**Figure 3.3:** SM cross-sections as a function of collider energy [108].

ϵ_n is the normalized transverse emittance, β^* is the betatron function at the Interaction Point (IP) and F is the geometric luminosity reduction factor for the crossing angle:

$$F = 1 / \sqrt{1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2} \quad (3.3)$$

where θ_c is the full crossing angle at the IP, σ_z is the Root Mean Square (RMS) bunch length and σ^* is the transverse RMS beam size at the IP. The betatron function $\beta(s)$ and the emittance ϵ_n are correlated [109] by:

$$y(s) = \sqrt{\beta(s) \frac{\epsilon_n}{\beta\gamma}} \cos(\nu\phi(s) + \delta) \quad (3.4)$$

where $y(s)$ is the x or y transverse coordinate along the arc length s , ν is the number of betatron oscillations in one revolution, $\phi(s)$ is a phase function and δ is determined from initial conditions; $\beta\gamma$ is the Lorentz relativistic factor.

The design peak luminosity of the LHC is $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, obtained at the beginning of each run. The main cause for the decay of the luminosity are beam collisions. Furthermore, Touschek scattering [110] and particle losses due to emittance blow-up contribute to the decay constant. The design net luminosity lifetime is 15 hours.

3.1.3 Operating conditions in the 2011 and 2012 data-taking periods

During the 2011 data-taking period the center-of-mass energy for pp runs was 7 TeV. The instantaneous luminosity of the LHC was raised from low luminosities up to $\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, thanks to the increasing of the number of bunches and to the optimization of β^* .

An important feature of the LHC collisions is the number of primary vertices per bunch crossing: the so-called pileup. Given that the time spacing between each bunch crossing was 25 ns, from Eq. 3.1 results that the mean pileup was $\langle N_{PV} \rangle = 8$, with an RMS equal to 5.6 cm. The pileup is a challenge for the detectors because of the multiplicities of the produced tracks coming from uncorrelated processes, and care must be taken in the physics analyses to treat this effect.

For the 2012 data-taking period, the center-of-mass energy for pp runs was raised to 8 TeV. Thanks to the LHC excellent performance, the instantaneous luminosity could reach $8 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ during the second part of the year. The mean 2012 pileup was $\langle N_{PV} \rangle = 21$ (Fig. 3.4).

Figure 3.5 shows the 2010, 2011 and 2012 integrated luminosities delivered to CMS. The overall CMS recording efficiency was higher than 90%.

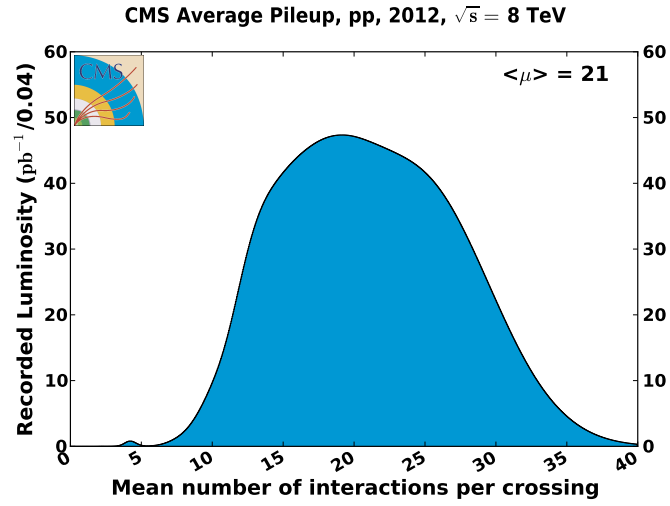


Figure 3.4: CMS average pileup distribution for pp collisions in 2012.

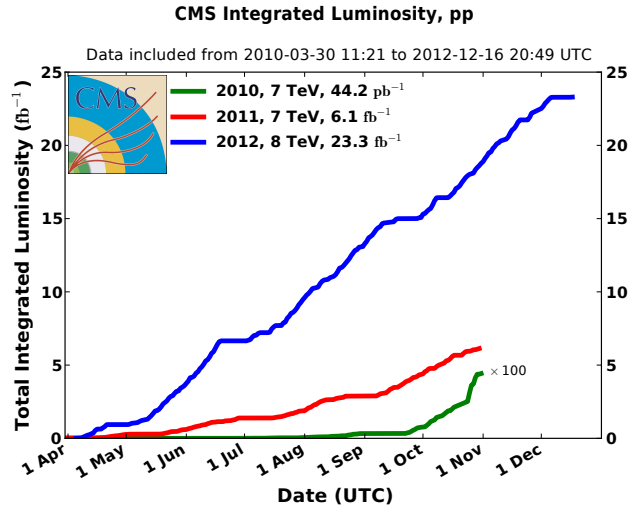


Figure 3.5: Integrated pp luminosity as a function of time, for the first three years of LHC run, delivered to CMS.

3.1. THE LARGE HADRON COLLIDER

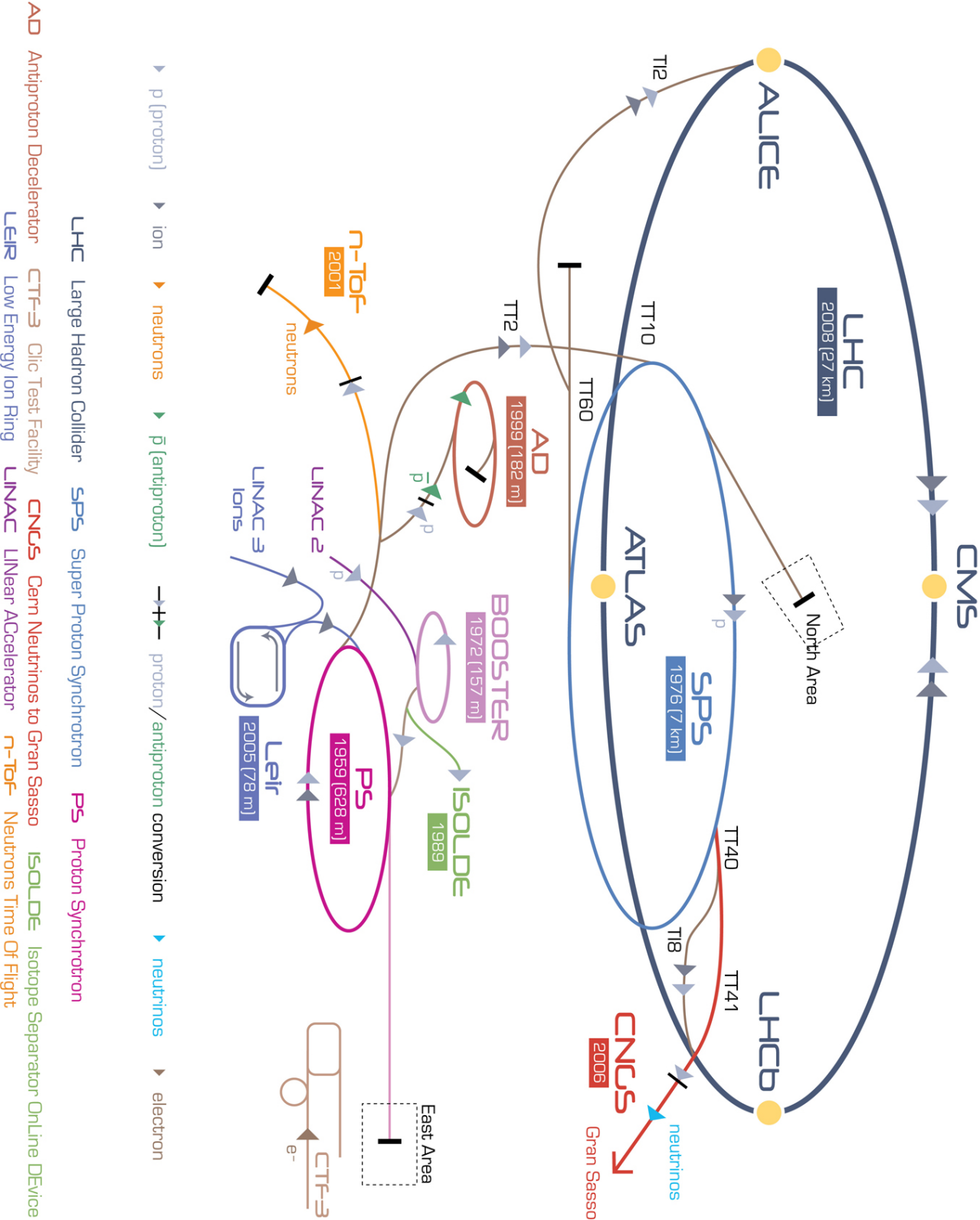


Figure 3.6: The CERN accelerator complex.

3.2 The Compact Muon Solenoid

In this section the main features of the Compact Muon Solenoid (CMS) [111] are described; in particular, the tracker and muon sub-systems, which are the main sub-detectors utilized by this analysis.

CMS is designed to study the proton-proton and nucleon-nucleon scattering physics at the energies and luminosities of the LHC. CMS is situated at Point 5 of the LHC ring, about 100 m underground near Cessy in France, between the Jura mountains and Lake Geneva. It has a cylinder geometry, with two “endcaps” closing the “barrel”. CMS is 21.6 m long, it has a diameter of 14.6 m, for a total weight of 12 500 t.

The main features of CMS that allow to meet the requirements of the LHC physics programme are the following:

- good muon identification and transverse momentum (p_T) resolution. CMS can reconstruct muons over a wide range, from $p_T \approx 1$ GeV in the forward region, up to $p_T \approx 1$ TeV. This is achieved by the inner tracker and the muon station systems;
- good charged-particle p_T resolution and reconstruction efficiency in the tracker. Pixel detectors very close to the IP are required to distinguish different Primary Vertices (PVs) in a high luminosity environment and to distinguish the Secondary Vertices (SVs) present in b and τ jet events;
- good electromagnetic energy resolution and good electron and photon coverage. This is performed by the electromagnetic calorimeter;
- good missing transverse energy resolution. A hadron calorimeter system with a large geometric coverage and fine segmentation is required, extending the coverage down to $|\eta| \approx 5$.

Going from the beam collision point outward, the main sub-detectors are the inner silicon tracker, the electromagnetic calorimeter, the hadronic calorimeter and the muon stations (see Fig. 3.7).

One of the most important requests for the concept of CMS is the magnetic field bending power for charged particles. At the heart of CMS is the superconducting solenoid, 13 m long, 6 m large, capable of generating a homogeneous 4 T magnetic field along the beam direction. The silicon inner tracker and the two calorimeters are all within the solenoid. The muon stations are placed outside the solenoid, in the return yoke, which also contains overall 1.5 m of iron along the radial axis; the bending power for a muon reaching the muon stations is 12 T m.

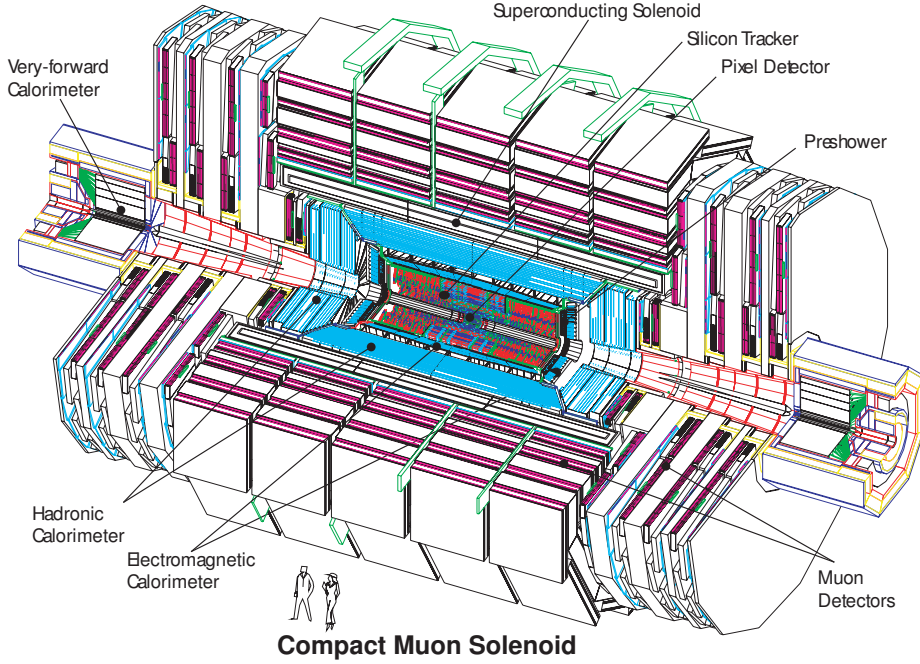


Figure 3.7: Perspective view of the CMS detector, showing the main sub-detectors.

At the center-of-mass energy of $\sqrt{s} = 14$ TeV, the total pp cross-section is roughly 100 mb (see also Fig. 3.3 in the previous section). With the LHC design luminosity, this implies a rate of 10^9 inelastic events/s. This huge amount of data is a challenge for the storage and the computing. Discerning the interesting events for physics searches from all produced events is of capital importance. CMS makes use of hardware triggers installed into the sub-detectors for reducing the event rate down to 10^5 Hz. After that, the events must pass software trigger selections, that reduce further the event rate down to few hundreds per second. Finally, the accepted events can be recorded, processed and stored in sites around the world, ready for physics analyses.

All components of CMS will be described in the following sections, focusing mainly on those used by this analysis.

3.2.1 CMS coordinate system

CMS uses a right-handed coordinate system centered at the nominal IP:

- the y -axis points vertically upward, orthogonal to the LHC plane;
- the x -axis points radially inward toward the center of the LHC;
- the z -axis points along the beam direction toward the Jura mountains.

Table 3.2: Main parameters of the 4 T superconducting solenoid.

Parameter	Value	Unit
Magnetic length	12.5	m
Operating temperature	4.5	K
Cold bore diameter	6.3	m
Central magnetic induction	4	T
Total Ampere-turns	41.7	MA turns
Nominal current	19.14	kA
Inductance	14.2	H
Stored energy	2.6	GJ
Radial thickness of cold mass	312	mm
Radiation thickness of cold mass	3.9	X_0
Weight of cold mass	220	t
Total mass of iron in return yoke	10 000	t

The radial distance from the IP is defined as $r = \sqrt{x^2 + y^2}$. The azimuthal angle ϕ is measured from the x -axis in the xy plane. The polar angle θ starts from the z -axis, and the pseudorapidity variable is related to θ as $\eta = -\ln \tan(\theta/2)$.

3.2.2 The superconducting magnet

The trajectories of the charged tracks can be measured thanks to the large CMS magnetic field produced by the superconducting coil. The charged particle transverse momentum p_T is measured in the xy plane. Since the transverse momentum resolution Δp_T of charged particle tracks depends on the magnetic field and the solenoid radius as

$$\frac{\Delta p_T}{p_T} \propto \frac{p_T}{BL^2} \quad (3.5)$$

where B is the magnetic field intensity and L is the path length, CMS design aimed to maximize both quantities. The 6 m diameter solenoid is designed to produce a magnetic field of 4 T within it. However, for longevity reasons, it was decided to be set at 3.8 T for the LHC 2011 and 2012 runs.

Outside the magnet, the magnetic flux returns through a 10 000 t iron yoke, divided in 11 large components, in which the magnetic field intensity is about 2 T.

The 64 mm \times 22 mm magnet winding is composed of four layers made of stabilized reinforced NbTi conductor. For physics reasons, the radial extent of the coil had to be kept small: the structural thickness is only 170 mm.

The main parameters of the CMS magnet are summarized in Table 3.2.

The CMS solenoid is very different from other detector magnets when looking at the stored energy and its ratio to the cold mass: 11.6 kJ/kg (see Fig. 3.8).

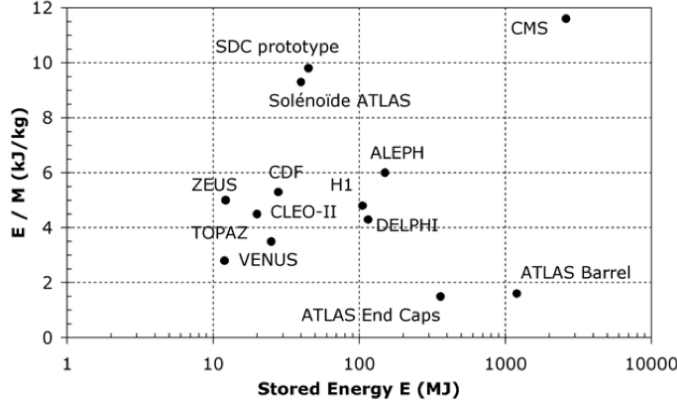


Figure 3.8: CMS solenoid energy versus mass ratio E/M versus stored energy, together with other detectors.

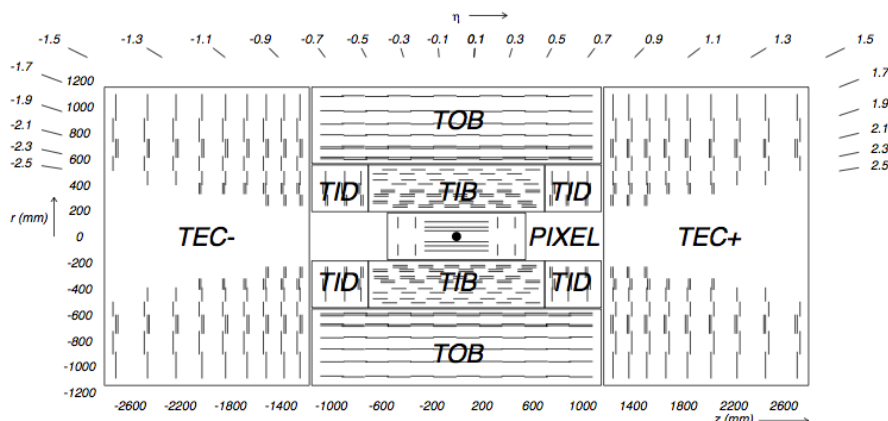
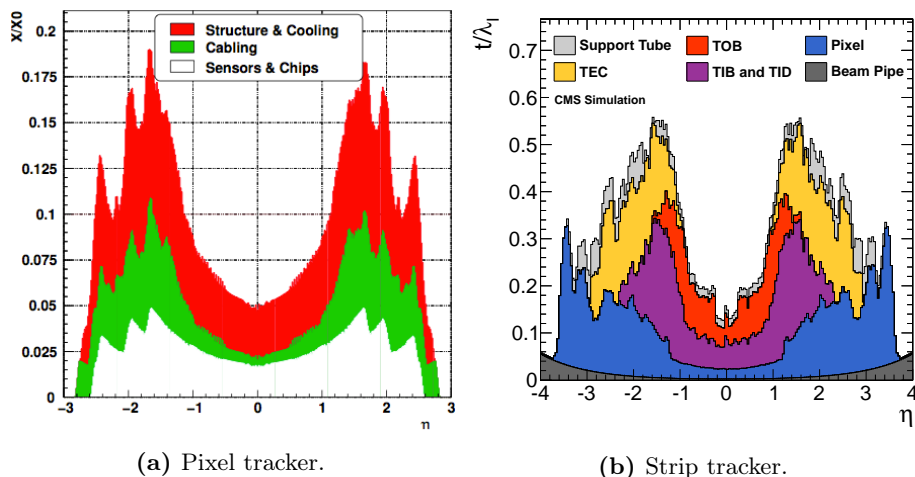
This high ratio can cause relatively large mechanical deformations (0.15%) when energized. To limit the mechanical stress, CMS uses an innovative design of self-supporting conductor, that includes the structural material. In this way, the 130 MPa magnetic hoop stress is shared between the layers and the cylindrical mandrel, rather than by the mandrel only as it was in previous thin detector solenoids.

The operating temperature is 4.5 K. The cooling is made by the helium refrigeration plant, which has a cooling capacity of 800 W at operating temperature. The cool-down of the solenoid from room temperature to 4.5 K takes around 24 days. The current-sharing temperature T_g , defined as the maximum temperature for which the current can flow without dissipation in the superconducting part, is $T_g = 6.44$ K; thus the temperature margin is 1.94 K.

The iron yoke is made by five wheels in the barrel and three disks in each endcap. Their weights range from 400 t up to 1920 t for the central wheel, that includes the coil and its cryostat. All the elements are aligned with an accuracy of 2 mm.

3.2.3 The inner tracking system

The innermost sub-detector of CMS is the silicon tracker. Its aim is the accurate, robust and efficient tracking of the charged particles produced by the collisions. Of fundamental importance is the precise measurement of secondary vertices coming from long-lived hadrons and τ -leptons. At the LHC design, each bunch-crossing produces about one thousand particles crossing the tracker every 25 ns. This large number of tracks, coming from different primary vertices, needs a fast, high-granularity and radiation-resistant detector. For all these reasons and to

Figure 3.9: Tracker layers in the zr plane.Figure 3.10: Pixel (a) and strip (b) tracker material budget as a function of η .

keep high redundancy and low occupancy, CMS uses a full silicon technology for its inner tracking system. The CMS silicon tracker is capable of high efficiency reconstruction for tracks with $p_T > 1$ GeV and $|\eta| < 2.5$.

In the barrel, from $r = 4.4$ cm to 10.2 cm, the tracker consists of three layers of pixel detectors. Farther, in a lower occupancy region, ten layers of silicon strips complete the silicon tracker at a distance of 1.1 m. Each tracker endcap consists of three pixel disks and twelve disks of silicon strips. Overall, there are 1440 pixel modules and 15 148 strip modules, forming around 200 m² of active silicon area, making the CMS tracker the largest silicon tracker ever built. Figure 3.9 shows the arrangement of all the tracker layers and Fig. 3.10 shows the pixel and strip material budget crossed by the particles.

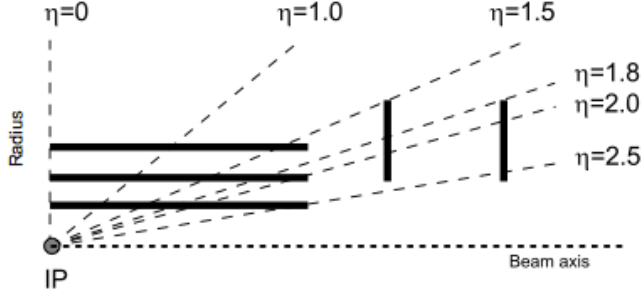


Figure 3.11: Pixel detector layers in the zr plane.

3.2.3.1 The pixel detector

The pixel layers are the closest ones to the beam spot. The pixel detector contributes to the measurement of the impact parameter, that is important for good secondary vertex reconstruction.

The arrangement of the pixel modules is shown in Fig. 3.11. There are three pixel layers in the barrel, at $r = 4.4$ cm, 7.3 cm and 10.2 cm, forming the Pixel Barrel (BPix) sub-detector and two disks forming the Pixel Forward (FPix) detector, at $z = \pm 34.5$ cm and ± 46.5 cm. In this way each particle crosses three layers for almost all the acceptance region, $|\eta| < 2.5$.

The size of each pixel is $100\text{ }\mu\text{m} \times 150\text{ }\mu\text{m}$ such to have about the same resolution in the $r\phi$ and z directions. Overall, the BPix contains 48 million pixels and the FPix 18 million. Thanks to these dimensions, at a typical charged particle flux of $10^8\text{ cm}^{-2}\text{s}^{-1}$, the occupancy of each pixel is only about 0.01%.

The position resolution is improved thanks to charge sharing, that helps to separate signal and noise hits and also to disentangle large hit clusters from overlapping tracks. In the presence of a magnetic field, due to the Lorentz drift, the charge of the collected signals is spread over more than one pixel and the interpolation of the analog pulse permits a spatial resolution of about $15\text{--}20\text{ }\mu\text{m}$ for the BPix. To improve the Lorentz drift also in the endcaps and to have a similar resolution, the FPix forward modules are tilted of 20° in a turbine-like configuration. The pixel sub-detector has a zero-suppressed read-out scheme with analog pulse height read-out.

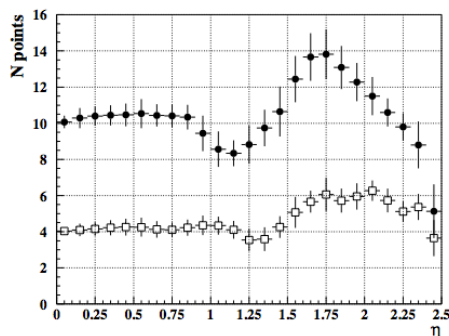
3.2.3.2 Silicon strip tracker

Beyond the pixel detector, the particle flux is lower and silicon micro-strip detectors can be utilized.

The micro-strip sensors are of single sided p -on- n type. They are disposed on a total of ten layers in the barrel and twelve disks per endcap, as yet shown

Table 3.3: Pitch and thickness for each different silicon strip.

Strip tracker sub-detector	Pitch (μm)	Thickness (μm)
TIB	80–120	320
TOB	122–183	500
TID	100–141	320
TEC	97–184	320–500

**Figure 3.12:** Number of points in the strip tracker versus $|\eta|$. Open squares show the number of stereo layers, filled circles show the total number [112].

in Fig. 3.9. The Tracker Inner Barrel (TIB) and Tracker Inner Disk (TID) are placed close to the pixel detector, and reach a radius of 55 cm. They contain four barrel layers and three disks for each endcap. The single-point resolution for TIB and TID is around 23–35 μm .

Outside, there is the Tracker Outer Barrel (TOB), containing six layers with a higher thick than the TIB. The point resolution for the TOB varies from 35 to 53 μm .

At the z ends of the TOB ($z = \pm 118\text{ cm}$) there are the nine disks of each Tracker EndCap (TEC), that extend up to $|z| = 282\text{ cm}$ and $r = 113.5\text{ cm}$. Each disk carries up to seven rings of silicon strip detectors.

Table 3.3 shows the main geometrical parameters of the silicon strips per each sub-detector.

Furthermore, many modules of the silicon strip detector contain a second module, back to back mounted to the first, with a stereo angle of 100 mrad. In this way it is possible to measure the z coordinate in the barrel and the r coordinate in the endcaps.

Thus, a particle in the tracker acceptance region crosses at least nine modules, ensuring high efficiency and good resolution (see Fig. 3.12).

The tracker average module hit efficiencies are shown in Fig. 3.13 for each layer and disk. The left plot (Fig. 3.13a) shows a flat 97% pixel efficiency except

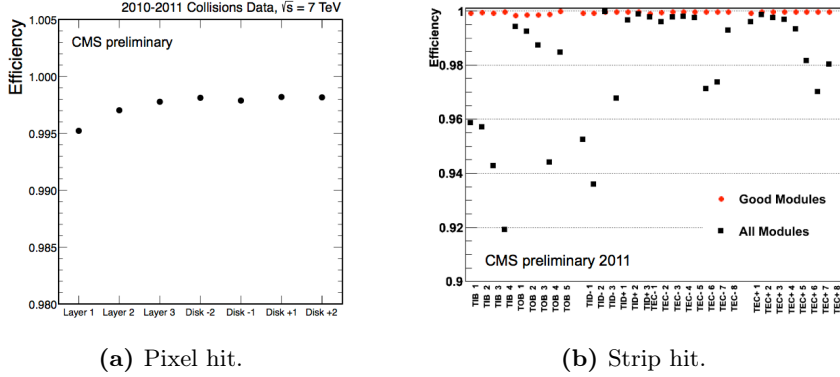


Figure 3.13: Tracker average module hit efficiencies for the pixel (a) and the strip (b) tracker, for each layer and disk [112].

for the first two layers, due to read-out losses. The right plot (Fig. 3.13b) shows the same for the strip tracker. Few modules are known to be inefficient and removing them from the read-out allows the average hit efficiency to be almost 100% (red points).

3.2.4 The electromagnetic calorimeter

The main aim of the CMS Electromagnetic CALorimeter (ECAL) is photon and electron detection and the measurement of their positions and energies. ECAL is a homogeneous and hermetic calorimeter made by lead tungstate crystals (PbWO_4 , density $\rho = 8.28 \text{ g/cm}^3$). It is formed by 61 200 crystals in the barrel and 7324 crystals in each endcap. The choice for the lead tungstate is driven by:

1. fast response: the 80% of the light is emitted in only 25 ns;
2. granularity: it has short radiation length ($X_0 = 0.89 \text{ cm}$) and Molière radius (2.2 cm), allowing best resolutions for jet positions and separations;
3. resistance to radiations, up to 10 Mrad.

The emitted light has a maximum for 420–430 nm (green-blue) and the refraction index for light around this peak is $n = 2.29$. On the other hand, PbWO_4 releases only about 30 photons/MeV and this implies the need to use intrinsic gain photodiodes, able to work in high magnetic fields. Thus, in association with the crystals, silicon avalanche photodiodes in the barrel, and vacuum phototriodes in the endcaps, are used.

A sketch of the CMS electromagnetic calorimeter is shown in Fig. 3.14.

The Barrel calorimeter (EB) is divided in 36 supermodules, each spanning 20° in ϕ , until $|\eta| < 1.479$, at $z = 2.68 \text{ cm}$. Each barrel crystal has a truncated

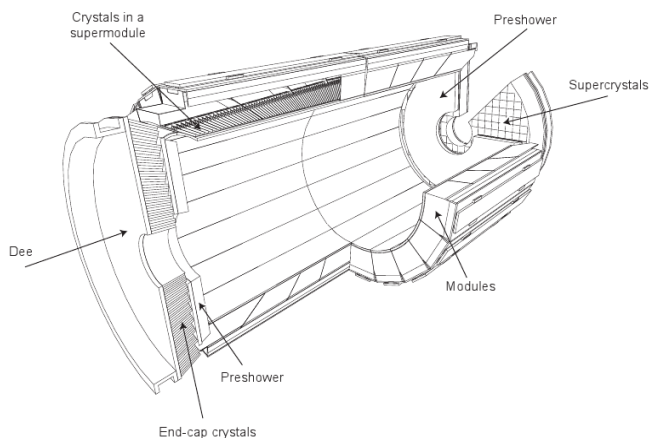


Figure 3.14: Sketch of ECAL.

pyramid shape, whose smallest base is large $22 \text{ mm} \times 22 \text{ mm}$ and it is 1.29 m distant from the beam direction; the other base is large $26 \text{ mm} \times 26 \text{ mm}$. The height is 230 mm , corresponding to $25.8 X_0$.

The Endcap calorimeters (EEs) are 3144 mm far from the origin. Each EE is formed by 5×5 crystals, grouped in mechanical units, called supercrystals. The two halves of each EE, called “dees”, range in $|\eta| \in [1.479, 3.0]$. The endcap crystals have the front base equal to $28.6 \text{ mm} \times 28.6 \text{ mm}$, the rear base $30 \text{ mm} \times 30 \text{ mm}$ and they are 220 mm deep, corresponding to $24.7 X_0$.

In front of each EE is placed a “preshower” detector, for the detection and localization of photon conversions with high pseudorapidity. It is made by two silicon strip layers, 1.9 mm high, separated by two lead disks, $2 X_0$ and $3 X_0$ deep. The two preshowers can detect in the pseudorapidity range $|\eta| \in [1.653, 2.6]$.

The calorimeter energy resolution ΔE can be parametrized as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.6)$$

where S , N and C are the stochastic, noise and constant terms, respectively. The stochastic term is mainly due to:

1. fluctuations in the lateral shower containment, event per event; the expected contribution is about 1.5% when summing 5×5 supercrystals;
2. a 2.1% photostatistics contribution, given by $\sqrt{\frac{F}{N_{pe}}}$, where N_{pe} is the number of primary photoelectrons per unit of energy and F takes into account the fluctuations in the gain process. This factor is about $2\text{--}2.5$, depending on the region;

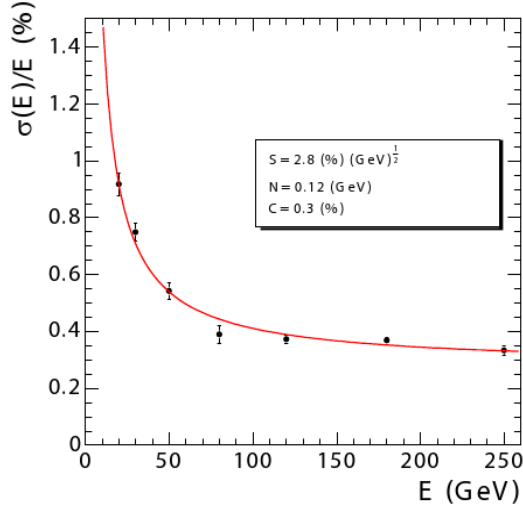


Figure 3.15: ECAL energy resolution as a function of energy. The parametrization of eq. 3.7 is used for a best fit.

3. (where present) fluctuations in the deposited energy in the preshower absorber, with value roughly $5\%/\sqrt{E}$.

The principal contributions to the noise term are

1. electronic and digitization noises from the instrumentation, leakage currents from neutron irradiation, and the output of the pre-amplifiers;
2. pileup noise, measured taking into account also the remnant energy from previous bunch crossings.

Finally, the constant term comes from

1. non-uniformity of the longitudinal light collection;
2. energy leakage from the rear of the crystal;
3. intercalibration uncertainty.

Test beam results indicate that by measuring energy in a 3×3 crystal lattice, a typical resolution of the CMS ECAL (Fig. 3.15) is

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{0.028}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.0030)^2 \quad (3.7)$$

where E is expressed in GeV.

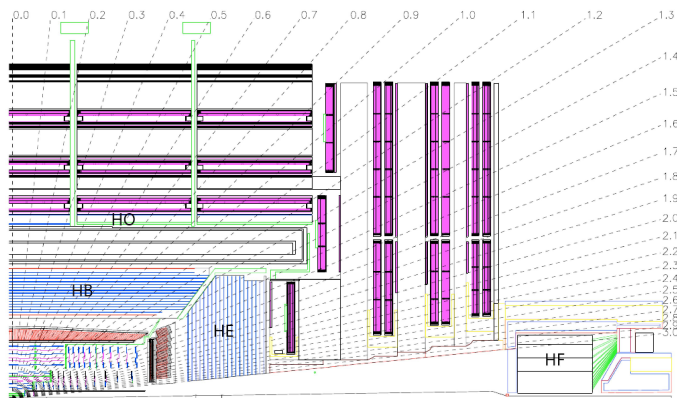


Figure 3.16: CMS sub-detectors in the $z\eta$ plane.

3.2.5 The hadronic calorimeter

The Hadronic CALorimeter (HCAL) is located beyond ECAL and lies almost entirely inside the magnet. It is designed to have a high geometrical acceptance and to measure precisely the hadronic jet energies and the event transverse missing energy, needed for the detection of non-interacting particles like the neutrinos, or predicted supersymmetric particles like neutralinos.

HCAL is divided in four parts: Hadron Barrel (HB), Hadron Outer (HO), Hadron Endcaps (HE), and Hadron Forward (HF). A longitudinal view of HCAL is drawn in Fig. 3.16.

The HCAL active part is made by layers of Kuraray SCSN81 active plastic scintillator, chosen for its time stability and relatively high radiation resistance. In the first fifteen layers the scintillator thickness is 3.7 mm, while in the last one (HO) is 9 mm, to detect late developing showers at the back of HB.

Light is collected by 0.94 mm large green wavelength-shifting fibers and it is brought to a hybrid photodiode, with a gain of about 2000.

The layers are separated by brass (70% copper and 30% zinc) absorbers or, for structural reasons, by stainless steel. Brass is chosen for its low X_0 (1.49 cm), that maximizes the radiation lengths, it is easy to manipulate and it is not ferromagnetic.

A special place is taken by the HF detector, starting at 11.2 m from the IP and with 1650 mm of absorber.

Overall, there are 41 detector towers, that are scintillator piles, segmented in pseudorapidity.

At high pseudorapidities ($3.0 < |\eta| < 5.0$) the particle flux is very higher than on the rest of the CMS acceptance region. On average the HF absorbs 760 GeV per pp interaction, compared to 100 GeV for the rest. After ten years of

operation, the HF will absorb about 10 MGy of radiations. The HF is designed to be stable even inside this hostile environment. Quartz fibers (fused-silica core and polymer hard-cladding) are the chosen active medium, while the absorber is made of 5 mm thick steel plates. The HF detects charged shower particles above their Cherenkov thresholds, through the generated Cherenkov light.

3.2.6 The muon system

Muon tracking is one of the most important features of CMS. This system performs three functions on muon particles: identification, momentum measurement and triggering. This sub-detector is placed outside of the hadron calorimeter and extends up to about 750 cm from the beam direction.

Three different gas detectors are used (see Fig. 3.17). On the barrel, at low pseudorapidities ($|\eta| < 1.2$), Drift Tubes (DTs) are used. Four muon stations are separated by iron layers, that form the return yoke of the solenoid magnetic field and that serve also as hadron absorbers.

Each station contains eight chambers, divided in two groups of four, to measure the $r\phi$ plane coordinates, and, only in the first three stations, four chambers which can measure the z direction. To have no blind spots in the detection, each drift cell is shifted by half-cell width with respect to the drift cell in the other layer.

In the endcaps ($1.4 < |\eta| < 2.4$), in a higher particle flux environment, Cathode Strip Chambers (CSCs) have been chosen, for their high time response, segmentation and radiation resistance. Each endcap contains four layers of CSC stations, with every chamber placed perpendicular to the beam direction, thus, measuring the coordinates on the $r\phi$ plane.

CMS utilizes also a third muon sub-detector, both in the barrel and in the endcaps (up to $|\eta| < 1.6$), made of Resistive Plate Chambers (RPCs). They offer an independent and complementary way for the muon detection. RPCs have a lower spatial resolution than DTs or CSCs, but they have a better timing resolution, allowing to identify the correct bunch crossing. The barrel contains six RPC layers: two for the first two stations each, and one in the third and fourth station. In the endcaps, RPCs are contained in the first three stations, to trigger and to improve time and p_T resolution. Figure 3.17 shows the three different muon sub-detectors in the rz plane.

The cylinder is longitudinally cut in five wheels. In the endcaps, CSCs and RPCs are mounted in four stations, placed on disks, perpendicular to the beam axis. Every disk is divided in two concentric rings, except the first one, divided in three. The interaction lengths are shown in Fig. 3.18 as a function of the pseudorapidity.

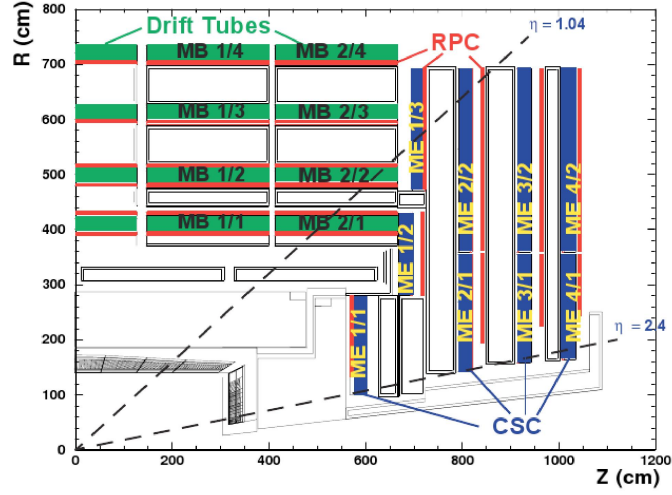


Figure 3.17: Muon system sub-detectors in the rz plane.

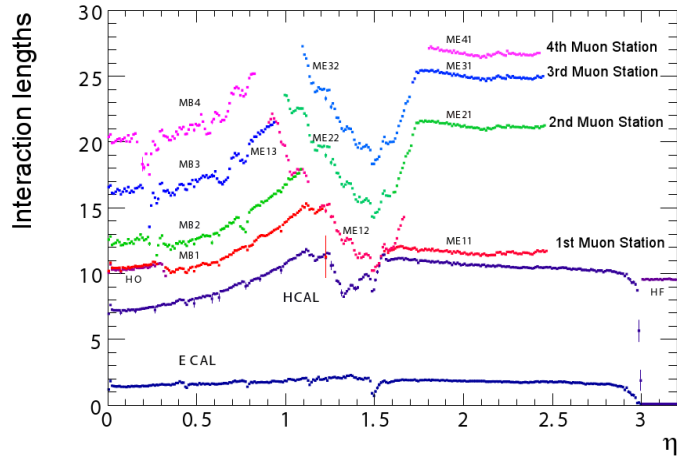


Figure 3.18: Interaction lengths as a function of $|\eta|$ for the muon system.

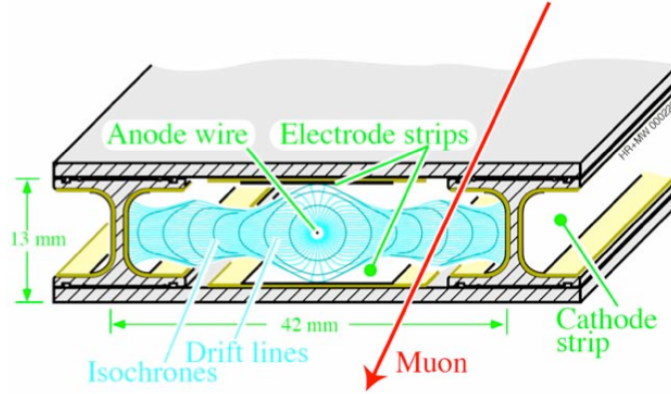


Figure 3.19: Sketch of a DT cell showing drift lines and isochrones.

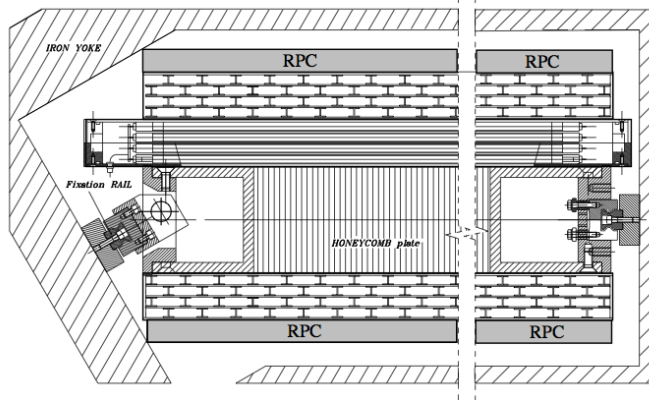


Figure 3.20: A DT chamber in position inside the iron yoke, in the $r\phi$ plane. One can see the two super layers with wires along the beam direction and the other perpendicular to it. In between is a honeycomb plate with supports attached to the iron yoke. Not shown are the RPCs, which are attached to the DT chambers via support plates glued to the bottom or top faces, depending on chamber type.

3.2.6.1 The drift tubes

The four concentric cylinders of the muon barrel are divided in twelve sectors. The first three cylinders contain twelve muon stations, while the fourth one contains fourteen of them. The DTs (see Fig.3.19 for a sketch) are placed at 4.0, 4.9, 5.9, and 7.0 m from the beam direction. Every chamber contains twelve DTs, divided in three groups of four layers, called super layers (Fig. 3.20). The transverse dimension of every cell is 21 mm, corresponding to a drift time of about 380 ns. This drift time makes the occupancy of each cell negligible. On the other hand, the cell dimensions are large enough to limit the number of the active electronic channels. The plates at the top and bottom of each cell are at ground potential. The voltages applied to the electrodes are +3.6 kV for wires,

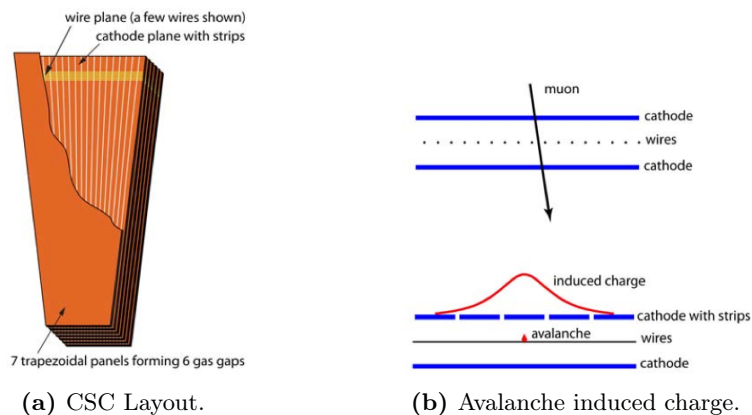


Figure 3.21: (a) Layout of one CSC, consisting of seven trapezoidal panels. (b) The avalanche induced charge is interpolated to obtain a more precise position of the avalanche along the wire direction.

+1.8 kV for strips, and -1.2 kV for cathodes. Near the wire the field increases by several orders of magnitude and electron multiplication occurs. About 10^5 electrons arrive at the wire and are read out.

The gas is a mixture of argon (85%) and CO_2 (15%). The chamber global resolution in $r\phi$ is $100\ \mu\text{m}$.

3.2.6.2 The cathode strip chambers

In the muon endcap region there are 468 CSCs. Every CSC has a trapezoidal shape and it is composed by six gas detector layers. Each particle crosses at least two CSCs along the ϕ direction (except for those in ME1/3, Fig. 3.17 shows also the muon station terminology). Except for ME2/1, ME3/1 and ME4/1, which contain 18 muon chambers, the rest has 36 chambers each.

CSCs are multiwire proportional chambers, composed by six anode wire planes interleaved by seven cathode panels (Fig. 3.21). Wires are placed azimuthally and define the track r coordinate. The ϕ coordinate is extracted interpolating the charges induced on the strips, that have a constant $\Delta\phi$ width. The overall CSC sensitive region is about $5000\ \text{m}^2$, the gas volume is more than $50\ \text{m}^3$ and the number of the wires is around 2 million. The ME1/1 spatial resolution is about $100\ \mu\text{m}$; for the others, it is $200\ \mu\text{m}$. The angular resolution is $\Delta\phi \approx 10\ \text{mrad}$.

In the intermediate region ($0.9 < |\eta| < 1.2$) muons are detected both by DTs and CSCs.

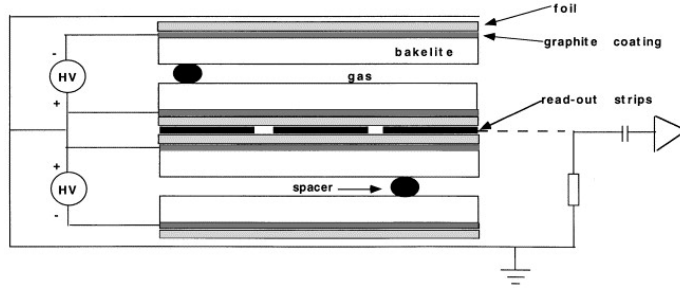


Figure 3.22: Sketch of an RPC sub-detector.

3.2.6.3 The resistive plate chambers

RPCs are gaseous parallel plate detectors with a very good tile resolution: an RPC can tag the time much more faster than the interval between two LHCs bunch crossings. For this reason a dedicated muon trigger, based on RPCs, is very important in the presence of a high background rate, that can go up to 1 kHz/cm^2 .

Six RPC layers are placed in the barrel, for a total of 480 chambers. The redundancy of the first layers allows the trigger algorithms to operate with at least four layers, also for low energy muons, that can stop after the first iron layers. An RPC module consists of two bakelite gaps, 2 mm high, operating in avalanche mode. The read-out electronic is placed in the middle (Fig. 3.22). The induced signal is given by the sum of the signals of the two gaps. The gas is a mixture of $\text{C}_2\text{H}_2\text{F}_4$ (96.2%), iC_4H_{10} (3.5%), and SF_6 (0.3%).

3.2.7 Track and vertex reconstruction

The trajectory of a charged particle in a constant homogeneous magnetic field is a helix, which can be described by five parameters, depending on the charge, momentum and the chosen pivot.

If the coordinate system is set such that the magnetic field points along the z axis (as it is in CMS), the helix can be parametrized as

$$\begin{cases} x = x_0 + d_\rho \cos \phi_0 + \frac{\alpha}{\kappa} (\cos \phi_0 - \cos (\phi_0 + \phi)) \\ y = y_0 + d_\rho \sin \phi_0 + \frac{\alpha}{\kappa} (\sin \phi_0 - \sin (\phi_0 + \phi)) \\ z = z_0 + d_z - \frac{\alpha}{\kappa} \tan \lambda \cdot \phi \end{cases} \quad (3.8)$$

where d_ρ , ϕ_0 , κ , d_z and $\tan \lambda$ are the five helix parameters. The position vector $\mathbf{x}_0 = (x_0, y_0, z_0)$ is an arbitrary reference point on which d_ρ , ϕ_0 and d_z depend. The constant α is defined such that $\alpha \equiv 1/B$ with B the magnetic field. Given

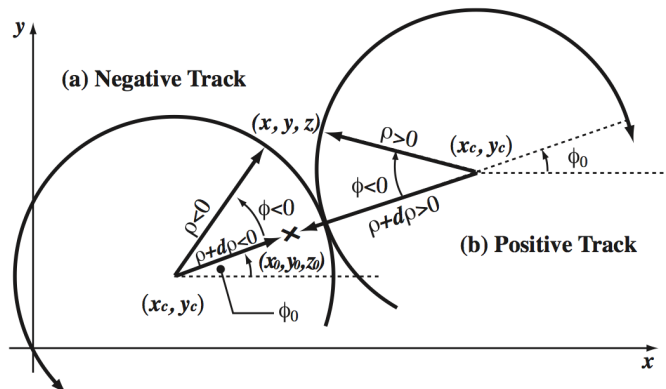


Figure 3.23: Helical track parametrization [113] for a negative (a) and a positive (b) particle. For an explanation of the parameters, see Sec. 3.2.7.

that the reference point \mathbf{x}_0 is arbitrary, it can be taken as the hit point, called pivot. The angle ϕ measures the deflection with respect to this point. The ratio $\alpha/\kappa \equiv \rho$ is the radius of the helix, while the parameter $\kappa \equiv Q/p_T$ with Q being the charge. See Fig. 3.23 for a picture of all referenced parameters.

In the following section, track and vertex reconstructions in CMS are described.

3.2.7.1 Tracker track reconstruction

Trajectory reconstruction has to take into account magnetic field inhomogeneities, energy losses and multiple scattering through the detector material budget. In CMS the track reconstruction algorithms begin taking the nucleon-nucleon interaction region (the beamspot) as an estimate of the primary interaction point in the transverse plane [114]. The RMS on the fitted beamspot coordinates can be measured and correspond to $\Delta x, y = 0.05$ mm and $\Delta z = 2$ cm.

After the beamspot reconstruction, track (and also vertex) reconstruction starts using the pixel hits. CMS standard track reconstruction is the Combinatorial Track Finder (CTF). Starting from the reconstructed hits, track reconstruction can be decomposed in four steps [115, 116]:

1. seed generation;
2. pattern recognition, or trajectory building;
3. ambiguity resolution;
4. final track fit.

3.2.7.1.1 Seed generation Since for the determination of the transverse momentum at least three points are needed, the starting point is a “track seed”, that can be either a triplet of hits, or only two hits plus a vertex constraint. All hits must belong to different detector layers. There are allowed ranges for the reciprocal hit positions, that are determined by the momentum direction, the transverse momentum and the primary vertex constraint.

The seeding from pixel triplets has some differences with respect to the seeding from pairs: the number of seeds is much smaller and the purity of the seeds is higher, but their efficiency is significantly lower, due to the pixel geometrical and readout inefficiencies. For these reasons the standard track reconstruction uses seeds from hit pairs and seeding from triplets is used in special cases, such as reconstruction of events from heavy-ion collisions.

3.2.7.1.2 Pattern recognition The pattern recognition is based on a combinatorial Kalman filter method. This is an iterative algorithm, that takes into account the information of the successive detection layers and updates the track parameters at each step. In this way, with every new added hit, the track parameters gain in precision.

Given the initial seed, the trajectory is extrapolated to the next compatible layer. The compatibility criterion is given by the charged particle equation of motion, corrected by energy losses and multiple scattering. For each compatible hit in the new layer, a new trajectory candidate is built. Also, another trajectory is built, with no measured hit (as known as “invalid hit”), to take into account the possibility that the particle did not leave a hit on that layer, due to inefficiencies. Once a new trajectory is created, new hits are added and the errors updated, until the tracker boundaries are reached. This procedure is iterated for each layer.

3.2.7.1.3 Ambiguity resolution Pattern recognition may create a lot of different trajectories per each seed, and, furthermore, a given track may be reconstructed from different seeds. To avoid double counting the tracks, these ambiguities must be removed. The ambiguity is defined as the fraction of hits shared between two trajectories:

$$f_{\text{shared}} = \frac{N_{\text{shared}}^{\text{hits}}}{\min(N_1^{\text{hits}}, N_2^{\text{hits}})} \quad (3.9)$$

where $N_{\text{shared}}^{\text{hits}}$ is the number of shared hits and N_i^{hits} is the total number of hits of the i^{th} track. The seed is then propagated to search for other compatible hits. If $f_{\text{shared}} > 0.5$, the track with the least number of hits or with the highest χ^2 is removed.

3.2.7.1.4 Final track fit At the final stage, the trajectory is rebuilt again with all the associated hits. The trajectory is firstly refitted with a Kalman filter. Then, a second filter is initialized with the result of the first one, and it is run backwards, towards the beam line. This smoothing procedure provides optimal estimates of the trajectory parameters at each measured point.

At the end of a CTF iteration, poor reconstructed or fake tracks are removed. Many of the fake tracks can be removed by choosing only tracks that have a good fit χ^2 and are compatible with the beamline or the primary vertices of the event. This selection can be refined further by varying the cuts for the track p_T , η , and the number of layers with hits. Short tracks with few hits are far more likely to be fakes, so a tighter selection is applied. In contrast, tracks with hits on every layer are unlikely to be fakes and a much looser selection can be used. Tracks are selected on the normalized χ^2 , the transverse (d_{xy}) and longitudinal (d_z) impact parameters and their significances, the number of tracker layers with or without a hit on the track. Tracks failing the loosest selection are rejected, while those that pass the tightest selection are labeled *highPurity* [117]. The impact parameter resolutions can be parametrized as

$$\sigma_{d_{xy}} = \sqrt{a^2 + (b/p_T)^2} \quad (3.10)$$

$$\sigma_{d_z} = \cosh(\eta) \cdot \sigma_{d_{xy}} \quad (3.11)$$

For *highPurity* tracks, $a = 30 \mu\text{m}$ and $b = 10 \mu\text{m GeV}$.

Every time a CTF iteration ends, the used tracker hits are removed to avoid track duplication and a new iteration begins. The first iterations use only the pixel seeds to reconstruct prompt tracks, then the algorithm looks for pixel and strip layers as seeds, to reconstruct displaced tracks (that may come from b decays, for example). Finally, strip-only seeds are used for tracks reconstructed in inefficient pixel regions.

Selection cuts are optimized on simulated data samples, considering processes of vary multiplicity, leading to the following formulas used for the selection on the above parameters:

$$\begin{aligned} \chi^2/d.o.f &< \alpha_0 n_{\text{layers}} \\ |d_{xy}| &< (\alpha_1 n_{\text{layers}})^\beta \cdot \sigma_{d_{xy}} \\ |d_z| &< (\alpha_2 n_{\text{layers}})^\beta \cdot \sigma_{d_z} \\ d_{xy}/\delta_{d_{xy}} &< (\alpha_3 n_{\text{layers}})^\beta \\ d_z/\delta_{d_z} &< (\alpha_4 n_{\text{layers}})^\beta \end{aligned} \quad (3.12)$$

where $\delta_{d_{xy}}$ and δ_{d_z} are the transverse and longitudinal impact parameter uncer-

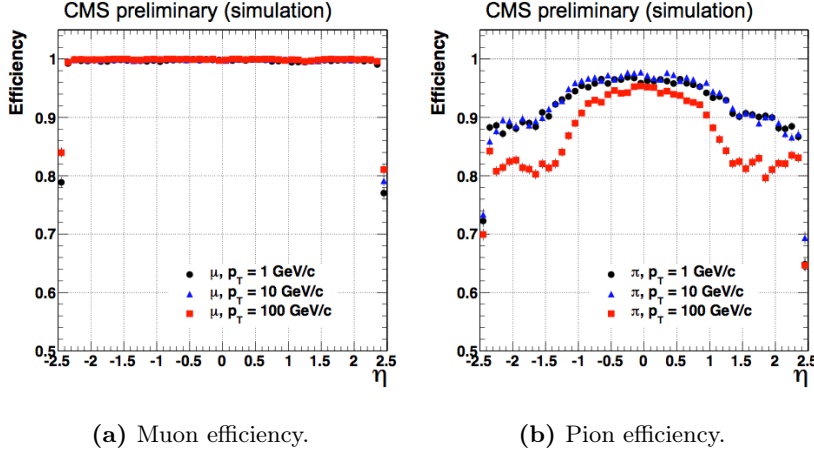


Figure 3.24: Track reconstruction efficiency for muons (a) and pions (b) of transverse momenta of 1 GeV, 10 GeV and 100 GeV, as a function of η [118].

tainties, respectively, and n_{layers} is the number of tracker layers with a hit on the track. The parameters α_i and β are configurable and different values are used for each tracking iteration.

3.2.7.1.5 Track performances The track reconstruction efficiencies for MC single muons and pions of different p_T are shown in Fig. 3.24. Plots show the efficiency of correctly reconstructed tracks produced inside the beam pipe. This efficiency includes the algorithm, acceptance and hit efficiencies. Tracks are reconstructed using seeds in the pixel detector and are required to have a minimum of eight hits, with a hit missing in at most one layer, and $p_T > 0.8$ GeV. The lower efficiency for pions is due to the fact that hadrons interact with the tracker material.

The track resolutions for the different helix parameters are shown in Fig. 3.25 as a function of $|\eta|$, for different transverse momenta. They are measured from the σ parameter of a Gaussian probability density function (pdf) fitted to the residuals:

$$x_{\text{res}} = x_{\text{reco}} - x_{\text{gen}} \quad (3.13)$$

where x_{reco} is the reconstructed quantity and x_{gen} is the generated quantity, for each parameter x . The d_0 and z_0 resolutions at low p_T are degraded by the multiple scattering. At high momenta the resolution is dominated by the hit resolution of the first pixel hit. At high transverse momentum the resolution is about 1-2%, up to $|\eta| < 1.6$, then the reduced lever arm worsens the resolution. Figure 3.25 shows also the p_T mean residual. The small bias for $|\eta| > 1.5$ is caused by inhomogeneities of the magnetic field at large radius.

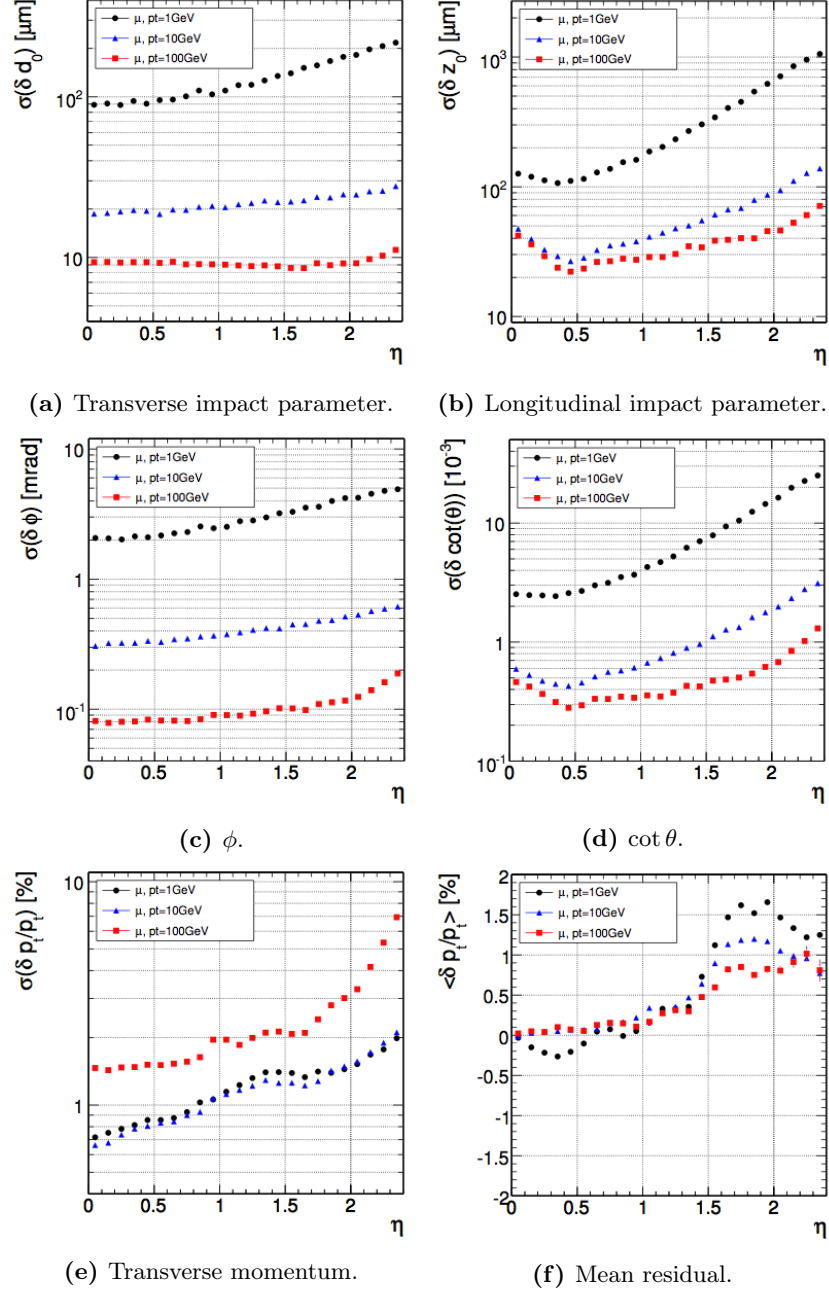


Figure 3.25: Resolutions of the five track parameters for single muons with transverse momenta of 1 GeV, 10 GeV and 100 GeV, versus $|\eta|$ [116]. From top to bottom and left to right: (a) transverse, (b) longitudinal impact parameter, (c) ϕ , (d) $\cot \theta$, and (e) transverse momentum. Plot (f) shows the mean residual for the transverse momentum.

3.2.7.2 Muon tracks

In CMS, muon tracks can be reconstructed by the inner tracker and by the muon system [119]. There are two different reconstruction algorithms: “Tracker Muon” and “Global Muon” algorithms.

The Tracker Muon reconstruction uses an “inside-out” approach. All tracker tracks with $p_T > 0.5 \text{ GeV}$ and $p > 2.5 \text{ GeV}$ can be reconstructible muon candidates. Every track is extrapolated to the muon system, taking into account magnetic field, energy losses and Coulomb multiple scattering. The track must match one muon segment, that is a track stub, formed by DT or CSC hits. The matching is valid if the distance between the extrapolated track and the muon segment is less than 3 cm, or if it is less than four times the combined uncertainties. If this happens, the tracker track becomes a Tracker Muon.

On the other hand, the Global Muon reconstruction uses an “outside-in” approach [120]. Starting from a standalone reconstructed muon, the muon trajectory is extrapolated from the innermost muon station to the outer tracker surface, taking into account the muon energy loss in the material and the effect of multiple scattering. Silicon layers compatible with the muon trajectory are then determined, and a region of interest within them is defined in which to perform regional track reconstruction. The determination of the region of interest is based on the track parameters and their corresponding uncertainties of the extrapolated muon trajectory, obtained with the assumption that the muon originates from the interaction point. If the matching is valid, the combined track is a Global Muon. The global track is determined by a Kalman filter technique, adding together the hits from the two originating tracks.

The p_T resolutions for the different muon objects are shown in Fig. 3.26. At values of p_T below 200 GeV, the momentum resolution for a muon track is driven by measurements in the silicon tracker. As a particle momentum increases and the curvature of its corresponding track decreases, however, momentum resolution in the tracker becomes limited by position measurement resolution. One can then benefit from the large lever arm and the 3.8 T magnetic field in the region between the tracker and the muon system by including hits in the muon chambers.

Tracker muon reconstruction is more efficient than the global muon one, especially for $p \lesssim 5 \text{ GeV}$, because it requests only one single muon segment. Global muon reconstruction requires instead at least two muon stations, thus it has a smaller muon misidentification and, consequently, a higher purity.

Usually, about 99% of muons produced in pp collisions and crossing the muon acceptance are reconstructed either as Global and Tracker muons.

Different selections can be applied to the muon tracks, balancing between

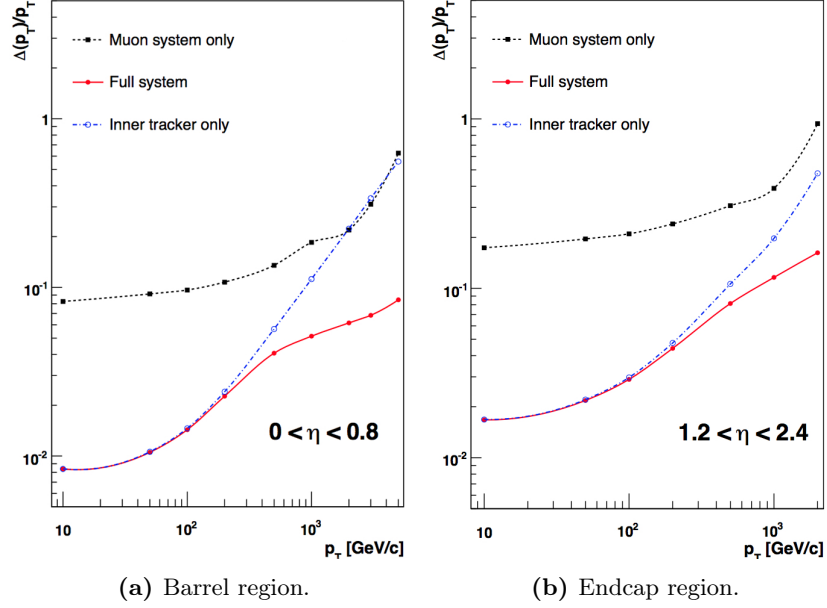


Figure 3.26: Transverse momentum resolution for the three different CMS muon objects, for the barrel (a) and the endcap (a) regions.

efficiency and purity. For this thesis, since the muon misidentification has a non-negligible impact, a high purity selection is chosen: the so-called “tight muon” selection. A tight muon is a muon identified both as tracker and global muon. Furthermore, to suppress hadronic punch-through and muons from decays in flight, the $\chi^2/d.o.f$ of the global fit must be less than ten and the track is required to have at least one muon chamber hit, one pixel hit and six tracker layers. Finally, the track must contain muon segments in at least two muon stations. With these selections, also cosmic muons are reduced at the cost of a slightly lower prompt muon efficiency.

To further reduce the rate of fake muons (kaons, pions, protons), an additional identification algorithm based on a Multivariate Analysis (MVA) is used in this analysis (details will be shown in Sec. 4.3.1).

3.2.7.2.1 Tight muon efficiency with the tag-and-probe method The muon identification and reconstruction efficiencies can be evaluated using a Tag & Probe (TNP) technique [121, 122].

The TNP method relies on muons coming from known resonances decaying into two muons (called dimuons). Strong selections are applied on one of these two muons (the “tag muon”) to strengthen the purity. Very looser selections are applied on the other muon (the “probe muon”). After this preliminary setup, it is possible to apply the interesting selections on the probe muon. The efficiency will

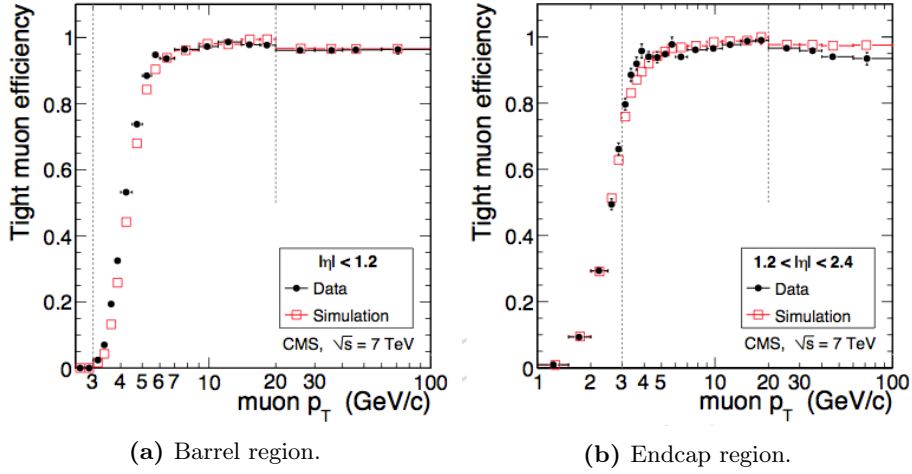


Figure 3.27: TNP tight muon identification and reconstruction efficiencies in data (black points) and MC simulations (red squares), for the barrel (a) and endcap (b) regions. Z -boson events are used for muons with $p_T > 20$ GeV, otherwise J/ψ -meson events are used. The difference on the turn-on curve, where the data points are systematically higher than in simulation, is due to a small difference in the widths of the track-to-segment pulls.

be given by the number of passing probe muons over all probe muons. Typical dimuon resonances are the J/ψ and Υ mesons, and the Z boson. Dimuons are selected through a cut on their invariant mass. Usually there are background events under the dimuon peak, hence, a background subtraction is performed through a fit on the invariant mass shape.

Results on the tight muon selection are shown in Fig. 3.27, for two different $|\eta|$ regions. The efficiency plateau is reached at about $p_T > 10$ GeV. The MC simulation reproduces quite well the efficiency shape.

3.2.7.2.2 Muon identification for non-muon particles Particles other than muons may be wrongly identified as muons. The main sources of muon misidentification are:

- hadron punch-through, in which hadron shower remnants penetrate through the calorimeters and reach the muon system;
- hadron decay-in-flight, as in $\pi^+ \rightarrow \mu^+ \nu_\mu$;
- random matching between an hadron tracker track and the muon stub.

Is is possible to measure these contributions with resonances decaying in kaons, pions and protons, such as $K_S^0 \rightarrow \pi^+ \pi^-$, $\Lambda \rightarrow p \pi^-$, and $\phi \rightarrow K^+ K^-$. With such samples, it is easy to measure the muon misidentification as the fraction of tracks identified as reconstructed muons. Figure 3.28 shows the results on tight

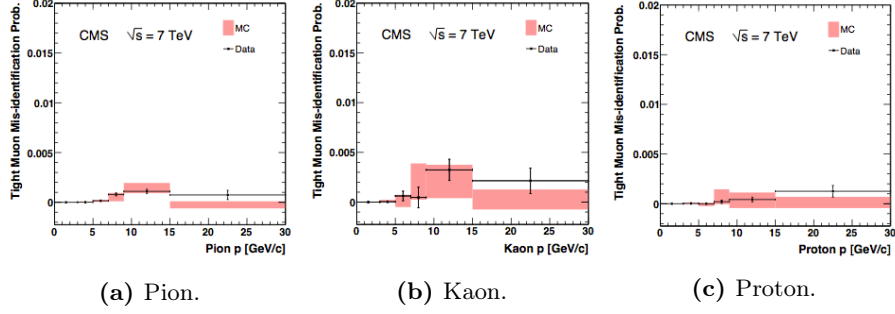


Figure 3.28: Tight muon misidentification for pions (a), kaons (b), protons (c), in data (black lines) and MC simulation (red shaded boxes). Uncertainties are statistical only [119].

muons for pions, kaons and protons. All the misidentification probabilities are well below 1%, reaching even 0.1%, showing the good performance of CMS in the muon identification and in the shielding of the muon system from hadron particles.

3.2.7.3 Primary vertex reconstruction

The aim of the PV reconstruction is the precise determination of the collision points in nucleon-nucleon interactions. For each bunch crossing there can be more than one inelastic collision, and so that many primary vertices. In 2012 the mean number of PVs was 21 (Fig. 3.4), with a tail of the distribution reaching 40, thus an excellent performing PV reconstruction algorithm is essential.

The vertex reconstruction is divided in two steps: vertex finding and vertex fitting. The first step uses only the z coordinate of the tracks: as shown in Sec. 3.2.7.1, the beamspot x and y widths are negligible with respect to the z width. The second step performs three-dimensional vertex fits with the full track information, to find the best estimates of the primary vertex coordinates.

The vertex fitter algorithm is an adaptive vertex fitter [123]. The method is formulated as an iterative re-weighted Kalman filter. The Kalman filter is a least squares estimator, which minimizes the sum $L(\mathbf{v})$ of the squared distances of all tracks from the vertex position:

$$\hat{\mathbf{v}}_{LS} = \arg \min_{\mathbf{v}} L(\mathbf{v}) \quad (3.14)$$

where $\hat{\mathbf{v}}_{LS}$ is the position of the PV. Differentiating Eq. (3.14) we obtain an equation for $\hat{\mathbf{v}}$ which, approximated at the first order, can be solved globally or iteratively.

The least square estimator is not robust, since it is sensitive to mismeasured

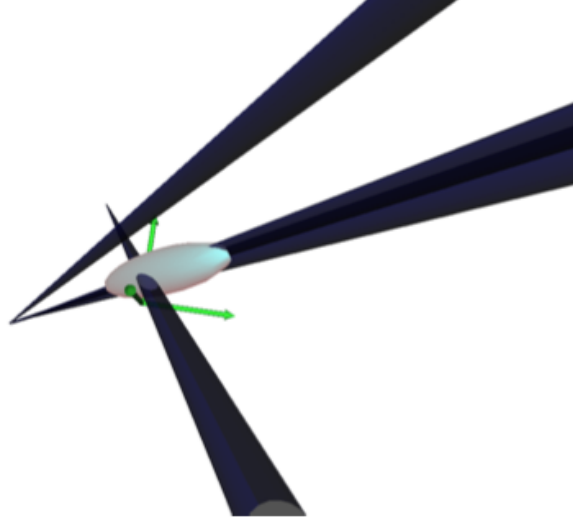


Figure 3.29: Sketch of the adaptive vertex fitter. The event contains four tracks: one incompatible with the other three and two highly collimated, overlapping tracks.

track errors. For this reason, the vertex fitter algorithm uses an adaptive vertex fitter, an iterative procedure in which a weight w_i is associated to every track (down-weighting outliers). The weight can be interpreted as the probability that the track i belongs to a vertex at position \mathbf{v} (see Fig. 3.29 for a sketch of the algorithm).

The clustering algorithm is a deterministic annealing clustering [124]. Given a probability p_{ik} for each track i and vertex k , it finds the most likely distribution of assignments (principle of maximum entropy):

$$\langle \chi^2 \rangle = \sum_{ik} p_{ik} \frac{(z_i - z_k)^2}{\sigma_i^2} \quad (3.15)$$

where z_i and z_k are the z coordinates of the track and the vertex, respectively. The $\langle \chi^2 \rangle$ is gradually decreased, through a “temperature” parameter, keeping the maximum probability state. The algorithm stops when this temperature reaches its minimum.

The PV resolution depends strongly on the number of tracks used in fitting the vertex and their p_T . A data-driven method (split-method) can be used to measure the resolution [125]. It consists in randomly dividing the tracks used in the vertex into two different sets. These two sets are fitted independently to make two different vertices. The distributions of the difference in the fitted vertex positions are fit with a single Gaussian pdf to extract the resolution. Results are in Fig. 3.30 and show good agreement with the simulation. The x

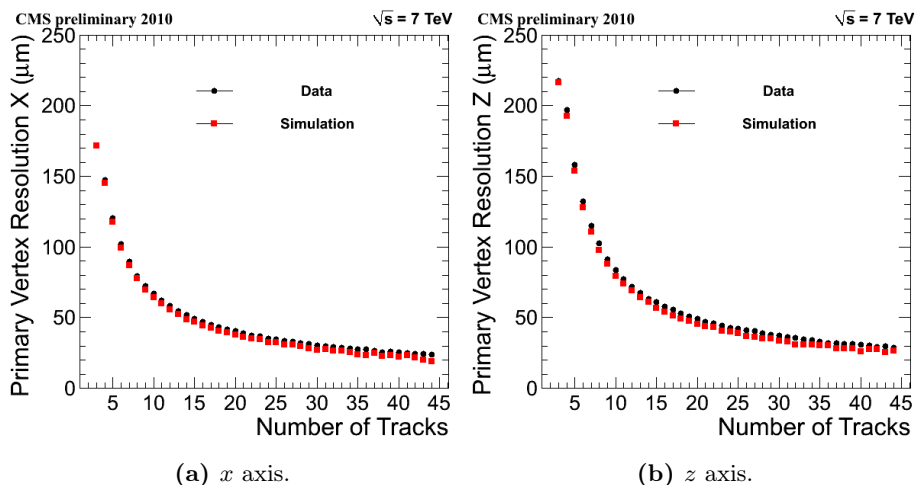


Figure 3.30: Primary vertex resolution along the x axis (a) and along the z axis (b), as a function of the number of tracks used in the fitted vertex [125].

and y resolutions are consistent; when using more than 30 tracks, the transverse resolution is about $20\mu\text{m}$ and the longitudinal resolution is close to $25\mu\text{m}$.

3.2.7.4 Secondary vertex reconstruction

SV reconstruction is done with a kinematic vertex fitting [126]. The mathematical approach used is the least mean squared minimization with Lagrange multipliers. The full decay tree is reconstructed by fitting all tracks and vertices in the current decay applying physics parameter constraints. The constraints depend on the process (the hypothesis) under study. To give an example, for the decay $B_s^0 \rightarrow J/\psi\phi \rightarrow \mu^+\mu^-K^+K^-$ (see Fig. 3.31), the constraint set can be:

- the four final state tracks;
- the two-muon invariant mass must correspond to the J/ψ meson mass;
- the two-kaon invariant mass must correspond to the ϕ meson mass;
- The momentum vector of the B_s^0 meson should point towards the primary vertex.

The total χ^2 of the fit can then be used as a selection criterion. Typical resolutions are around $30\mu\text{m}$ for two-particle decays.

3.2.8 Trigger

The LHC provides billion of events per second. This huge amount of data is too large for being recorded and processed. Therefore an efficient trigger system is needed to select only the most interesting events.

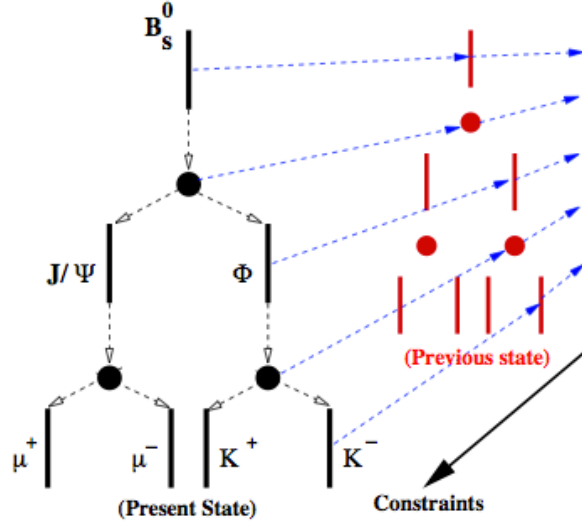


Figure 3.31: Sketch of the decay chain reconstruction for the $B_s^0 \rightarrow J/\psi \phi \rightarrow \mu^+ \mu^- K^+ K^-$ decay in the kinematic vertex fitting.

In CMS the trigger system is basically divided into two steps. The first step is made at hardware level, integrated in the calorimeter and in the muon sub-detectors (Level 1 Trigger (L1)). The second step is software (High Level Trigger (HLT)), executed in a processor farm close to the detector.

3.2.8.1 The Level 1 trigger

The L1 trigger consists of custom-design, easily programmable electronics. The purpose of this step is to reduce the data bandwidth from 40 MHz down to 100 kHz. Only the calorimeters and the muon sub-systems participate in the L1 trigger decisions, because the tracker bandwidth is too large to enable event decisions in time. The L1 trigger has local, regional and global components (Fig. 3.32). The local trigger primitive generators, that are the seeds for the L1 trigger algorithms, are calorimeter energy deposits and muon chambers hit patterns. These trigger primitives are collected by the regional triggers, using pattern logic. The regional triggers sort electron and muon trigger objects, in limited regions of space, ranking them as a function of energy, momentum, and quality. Finally, the global calorimeter and muon triggers collect all these sorted objects and choose the highest ranked ones to be sent to the global trigger. It is the global trigger that has to take the final decision on rejecting or sending the event to the HLT.

The allowed L1 trigger delay between the bunch crossing and the decision is $3.2 \mu\text{s}$. During this time the information is conserved in the electronics buffer of

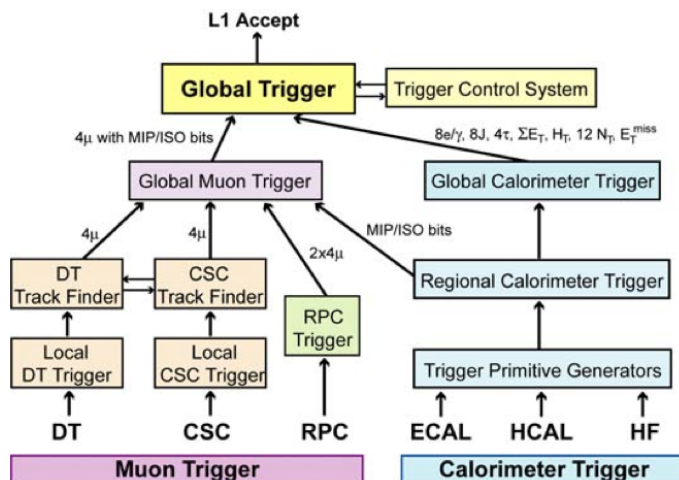


Figure 3.32: Architecture of the L1 trigger.

the sub-detector front-end.

All L1 triggers that acquire data in a given LHC run are grouped together in a so-called L1 trigger menu. As the LHC luminosity was being increased, different L1 trigger menus had to be selected, because the total rate must be maintained in a safe regime below the 1 kHz bandwidth.

There are two main ways to keep the rate low when the luminosity increases: using more tight selections or using a prescale. For the first case, one can request more trigger objects: for example, instead of asking for one single L1 energy deposit in ECAL, one can request two different L1 deposits; or one can raise the energy threshold, moving from 5 GeV to 20 GeV the minimum jet energy. Another possibility is to prescale the trigger, that is a way that allows to skip triggerable events if the trigger has fired before. For example, if a 10 GeV muon trigger has a prescale of 20, it actually fires only one time every twenty events that contain a muon with $p_T > 10$ GeV. This allows to keep triggers with low thresholds, to the detriment of the acquired integrated luminosity that, for the aforementioned trigger, would be $1/20$ of the total luminosity.

Generally, the 2011 and 2012 trigger menus contain about one hundred of different L1 triggers. Some of those can overlap, for example a 36 GeV-jet trigger does not add bandwidth if also a 16 GeV-jet trigger is implemented¹. The total bandwidth is roughly shared between single and double-muon triggers, single and double-jet triggers, and single and double-energy deposit triggers.

This analysis uses two low p_T muons, so it uses the double-muon trigger with the lowest allowed p_T threshold that is present in each menu. Due to the L1

¹if the 16 GeV-jet trigger is not prescaled, otherwise some skipped events from this looser trigger could be collected by the tightest one.

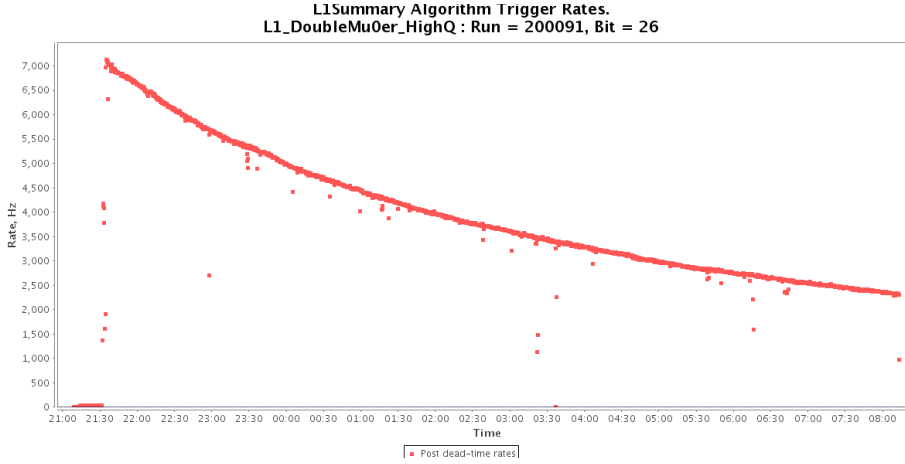


Figure 3.33: L1 double-muon trigger rate as a function of time for a given LHC run. This given trigger does not require a minimum p_T , but the maximum pseudorapidity is set to $|\eta| < 2.2$.

bandwidth limitation, the permitted rate for double-muon triggers is few kHz, since other kinds of triggers are required for other analyses. This means that for the lowest p_T double-muon triggers, the minimum muon p_T was between 0 GeV and 3 GeV and the maximum muon $|\eta|$ between 2.4 and 2.2, depending on the LHC luminosity, that was increasing during the two years.

Figure 3.33 shows the rate of a L1 double-muon trigger used by the $B_s^0 \rightarrow \mu^+\mu^-$ analysis, as a function of time, for a given run. The maximum rate is about 7.1 kHz, measured at the beginning of the run, then, it exponentially decays below 2.5 kHz after 11 hours of data-taking.

3.2.8.2 The High Level Trigger

The final allowed event recording rate is few hundred of events per second and it is selected by the HLT software algorithms that are executed in the processor farm. The information from all the sub-detectors has to be stored in buffers since the time to analyze one event can take up to few seconds (for events with high pileup).

The HLT system is formed by few hundred different software algorithms, running on every reconstructed event. Each of these algorithms, called HLT paths, is divided into different steps, that are producers or filters. Producers make objects relevant to the given HLT path (for example, one can produce electron objects from the ECAL and the tracker). Filters take decisions on these produced objects (in the same example, it can pass events in which there are at least two electrons with $p_T > 20$ GeV). An HLT path gives a final positive response if the event passes its last step. When one of the filters gives a negative

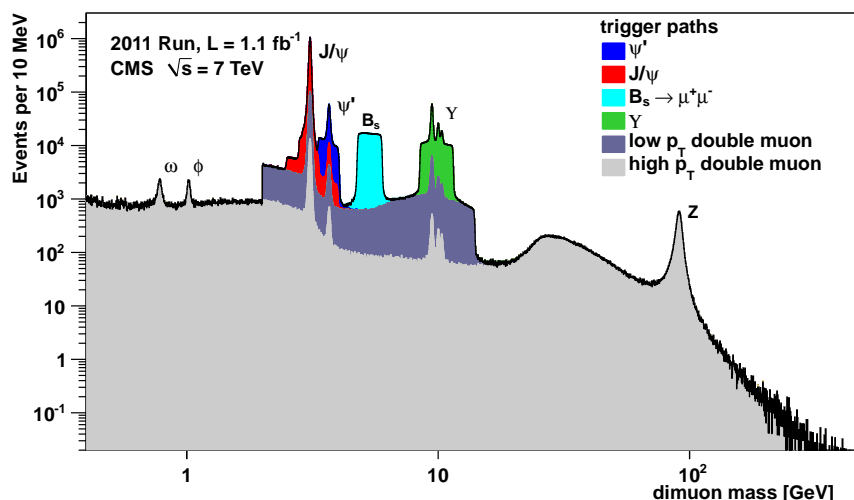


Figure 3.34: Data collected by early July, 2011, superposition of various dimuon trigger paths.

response, the HLT path rejects the event and the subsequent filters and producers of that path are not executed. In this way it is possible to save time and CPU power.

As the instantaneous luminosity increases, the L1 and HLT paths have to tight their selections due to the bandwidth limitation. Figure 3.34 shows the dimuon invariant mass spectrum, as selected by all the different HLT paths, collected with the first 1.1 fb^{-1} 2011 data. It is possible to distinguish many different mass windows selected by different HLT paths.

The first sub-detectors that are used by the HLT methods are the calorimeters and the muon system. The tracker information is added after. For example, muon HLT paths first produce HLT standalone-muon tracks (Sec. 3.2.7.2); if the event passes the standalone-muon filters, the tracker hit information is added to form a global muon track. Then, other filters and producers can be added to better select the event topology of interest.

Since at this stage the processing time is an important aspect, the HLT paths can use just a part of the detector information. For example, muon tracks may be reconstructed not with all the hits but with only the first n hits, or when sufficient precision is achieved. Track resolution can be worse, but much time is saved.

Events are collected in different Primary Datasets (PDs), depending on whichever trigger sets they triggered. For example, all events triggered by paths chosen by B -physics analyses (like this one) belong to the same PD, called in jargon *MuOnia*. Every PD has a maximum data acquisition rate which is

decided by the CMS Physics Management, weighing and gauging the impact of the trigger efficiency for each different physics group.

3.2.9 Event reconstruction and processing

An event that passes at least one HLT path can, then, be saved and reconstructed.

RAW data (RAW), storing the information from all the CMS sub-detectors, are saved in the so-called Tier-0 (T0) at CERN. The T0 does not allow for analysis accesses and the RAW data contain a too low-level information to be used by analyses. There, the RAW data are processed to form RECOⁿstructed data (RECO), containing high level objects ready for the physics analyses.

The T0 also redistributes the PDs to the seven Tier-1 (T1) sites around the world. The T1 sites store data and reprocess them whenever it is needed, for example when new calibration constants are ready.

T1 sites also send RECO data (and Analysis Object Dataset data (AOD), a sub-set of RECO data) to the different Tier-2 (T2) sites, which users can access to run their analyses.

CMS data are saved in the ROOT format [127]. The CMS specific C++ classes form the CMS SoftWare (CMSSW) architecture [128].

MC simulation events contain also:

- the event GENeration data (GEN) from particle guns or softwares like PYTHIA [94], MADGRAPH [129] and EVTGEN [99];
- the simulation (made by GEANT-4 [130, 131]) of the passage of particles through the detector material, and the response of the sub-detectors (SIMulated hits data (SIM));
- the simulation of the electronics digitization (DIGItization data (DIGI)).

Table 3.4 clarifies the main different data types present in CMS, together with their typical size.

Table 3.4: CMSSW main data formats and size per event.

File format	Content	Event size (MB)
DAQ-RAW	Front-end and L1 data	1-1.5
RAW	Raw CMS sub-detectors data	1.5
RECO	Reconstructed physics object (muons, vertices, clusters, etc.)	0.25
AOD	RECO subset for specific physics analysis	0.05
GEN	MC simulation of a given physics process	-
SIM	Simulation of energy deposits (<i>simhits</i>)	-
DIGI	Digitization of the simhits	1.5

Chapter 4

Observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and upper limit on $B^0 \rightarrow \mu^+ \mu^-$ decays

In this Chapter we describe the analysis performed for the observation of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and for the upper limit measurement of the $B^0 \rightarrow \mu^+ \mu^-$ decays. The analysis general criteria are outlined in the first sections. Then, the selections used to discriminate true signal candidate events from the different background events are described in detail, as well as all the background sources. Next, the two methods, the CL_S technique and the unbinned maximum likelihood fit, used to extract the candidate yields, are discussed. Results and interpretations of their implications are reported consequently. Finally, the combination with the LHCb measurements, as well as comments on the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ outlook, are exposed at the end of this Chapter.

4.1 Overview

This analysis searches simultaneously the decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. Signal events are extracted in two ways:

1. with a counting experiment in a signal window in the dimuon invariant mass distribution centered on the $B_{(s)}^0$ meson mass;
2. with an unbinned maximum likelihood fit to the dimuon invariant mass distribution.

To avoid any possible bias due to incorrect discrimination of signal against background events, all the analysis selections are studied and chosen without

using events in a dimuon invariant mass region close to that of the $B_{(s)}^0$ mesons (“blind analysis”). Only after that all selections are fixed, the invariant mass window is revealed (“unblinded”), and the actual number of events measured.

This analysis uses a relative normalization ($B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^+ \mu^- K^\pm$) to avoid the uncertainties of the $b\bar{b}$ production cross-section and of the luminosity measurements. Furthermore, many systematic errors cancel at first order normalizing the signal to a similar decay channel measured in data.

The $B_s^0 \rightarrow \mu^+ \mu^-$ measured branching fraction, then, is:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \frac{N_{\text{obs}}^{B_s^0}}{N_{\text{obs}}^{B^\pm}} \times \frac{\epsilon_{B^\pm}}{\epsilon_{B_s^0}} \times \frac{f_u}{f_s} \times \mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \quad (4.1)$$

where N_{obs}^i is the number of reconstructed signal and normalization candidates.

The total efficiency ϵ_i is factorized into four separate components, that are the acceptance A_i , the analysis selection $\epsilon_{\text{ana},i}$, the muon reconstruction and identification $\epsilon_{\text{reco},i}$, and the trigger request $\epsilon_{\text{trig},i}$:

$$\epsilon_i = A_i \times \epsilon_{\text{ana},i} \times \epsilon_{\text{reco},i} \times \epsilon_{\text{trig},i} \quad (4.2)$$

The factor A_i combines the geometric acceptance and the tracking efficiency: $A_i = N_{\text{acc},i}/N_{\text{gen},i}$, where $N_{\text{gen},i}$ is the total number of generated decays, produced with MC simulations requesting no filters, and $N_{\text{acc},i}$ is the numbers of events requiring minimal selections on all involved tracks, to be in the detector volume. In Eq. (4.2) the ordering is significant, as each efficiency is evaluated on top of the factor on its left.

The determination of the signal efficiency is done through MC simulation samples. It is paramount to validate the MC simulations in data, thus the analysis uses a Control Sample (CS): $B_s^0 \rightarrow J/\psi \phi$, where the J/ψ meson decays into two muons and the ϕ meson decays into two kaons, allowing to compare B_s^0 meson observables directly in data and in MC simulations.

In Eq. (4.1) the factor f_u/f_s is the ratio of probabilities for a b quark to hadronize into a B^\pm or a B_s^0 meson. Assuming f_u to be equal to f_d , we can use the f_s/f_d LHCb measurement, that reports, for $2 < \eta < 5$ [132],

$$\frac{f_s}{f_d} = 0.256 \pm 0.020 \quad (4.3)$$

$\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm)$ and $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$ are the branching fractions for the $B^\pm \rightarrow J/\psi K^\pm$ and $J/\psi \rightarrow \mu^+ \mu^-$ decays, respectively.

The $B^0 \rightarrow \mu^+ \mu^-$ branching fraction formula is similar to Eq. (4.1), but without the f_u/f_s ratio.

The chosen Normalization Sample (NS) has a high cross-section, that allows for a high statistic sample. An other interesting normalization decay, for the $B_s^0 \rightarrow \mu^+ \mu^-$ case, could be $B_s^0 \rightarrow J/\psi \phi$, such that the f_s/f_u ratio would not be needed, but its branching fraction is still not well measured (its relative uncertainty is about 25%) [4]:

$$\mathcal{B}(B_s^0 \rightarrow J/\psi \phi) = (1.09_{-0.23}^{+0.28}) \times 10^{-3} \quad (4.4)$$

Instead, for the chosen normalization channel, the overall branching fraction relative uncertainty is around 3.4%:

$$\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) = (1.016 \pm 0.033) \times 10^{-3} \quad (4.5)$$

$$\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.93 \pm 0.06) \times 10^{-2} \quad (4.6)$$

The background level is very different in the forward region of the detector, compared to the barrel region. Furthermore, the mass resolution in the CMS tracker depends strongly on the pseudorapidity of the reconstructed particles (see Sec. 3.2.7.1.5). Therefore, this analysis is performed into two distinct categories: if both muons fulfill $|\eta| < 1.4$, then the $B_{(s)}^0$ candidate belongs to the “barrel category”; otherwise it belongs to the “endcap category”.

4.1.1 Sources of background events

This analysis searches for two reconstructed tracks, identified as muons and with opposite charge. The two tracks must come from a common vertex, well displaced (few millimeters) from the pp primary vertex. Furthermore, the dimuon invariant mass must be compatible with that of the $B_{(s)}^0$ meson.

The reconstructed tracks are identified as muons by the three CMS muon sub-detectors (see Sec. 3.2.6). Thus, background events can be composed by “true” muons, generated by the different physics processes, or “fake” muons (such as kaons, pions and protons), misidentified wrongly as reconstructed muons. Mainly, the misidentification may be due to a punch-through of the particle inside the muon stations or due to a hadron decay-in-flight.

From the analysis point of view, the main background sources can be grouped into three categories:

1. combinatorial background formed by muons coming from b or c weak decays: $q \rightarrow X \mu \bar{\nu}$; in this category are present also events where a true muon is combined with a hadron, misidentified as a muon;
2. rare B^0 , B^\pm , B_s^0 and Λ_b^0 semileptonic decays (like $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$), in which a hadron is misidentified as a muon. Since the neutrino is not

reconstructed, this category forms a continuous background in the invariant mass, that starts at the mass of the decayed hadron and goes down to lower values;

3. rare dihadronic B^0 , B^\pm , B_s^0 and Λ_b^0 decays (like $B_s^0 \rightarrow K^+K^-$), which form a broad peak in the B^0 invariant mass region.

A further background component is the possible cross-feed between the $B_s^0 \rightarrow \mu^+\mu^-$ and the $B^0 \rightarrow \mu^+\mu^-$ decays, in which a B_s^0 event could be identified as a B^0 event due to the detector resolution, or vice versa.

The first background source is without structure in the dimuon mass distribution. No MC simulation is used to study this combinatorial background: the number of generated events would be too consuming both in terms of CPU time and disk storage, to have a several times larger dataset than the 25 fb^{-1} of collected data. To overcome this difficulty, the dimuon invariant mass side-bands are used (defined in Sec. 4.3.2).

The second and third sources present shapes in the dimuon mass distribution that can help to separate them from the signal events (see Sec. 4.9.2). MC simulations are used to study the impact of these rare backgrounds on the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ BF.

Without any strong selection, the first category contains by far the largest fraction of background events, as it is roughly one hundred thousand times bigger than the SM signal expectations.

Regarding the rare decays, despite the very low CMS misidentification probability (see Sec. 4.3.1) and their small BF, they have a non-negligible effect (see Sec. 4.7).

4.2 Primary dataset and triggers

In CMS, collected data must pass a scrutiny that certifies the quality of each run and of each single luminosity section; only those fractions that pass this certification can be used by analyses. There are two types of certification, one which checks the performances of the inner tracker and the muon stations (called “muon certification”); one that includes also the performance of the two calorimeters (called “golden certification”). Since this analysis uses reconstructed tracks and muons, but no information from calorimeters, the first kind of certification is used.

The total integrated luminosity, corresponding to the muon certification, is equal to $(5.0 \pm 0.1)\text{ fb}^{-1}$ for the 2011 7 TeV run and $(20.0 \pm 0.5)\text{ fb}^{-1}$ for the 2012 8 TeV run [133]. These absolute values of the integrated luminosity are,

however, not used anywhere in this analysis, since all inherent variables are normalized through the NS channel.

The total L1 and HLT rate bandwidths are shared between different analysis groups. A careful study is made before and during the data acquisition, to optimize the full rate to each physics group needing.

All L1 triggers involved in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis require the presence of two high quality muons objects in the event. During the 2011 data taking, muons were requested to have $|\eta|^{\text{L1}} < 2.4$. During the 2012, due to the increasing luminosity, to keep the rates under control while retaining the best quality events, the maximum pseudorapidity was decreased to $|\eta|^{\text{L1}} < 2.2$ and the minimum transverse momentum was increased to $p_T^{\text{L1}} > 3 \text{ GeV}$.

Two different sets of HLT trigger paths are used by this analysis: one acquires events in the signal invariant mass region, the other collects normalization and control events. Both trigger paths save the events in the *MuOnia* PD (Sec. 3.2.8.2).

MuOnia is the PD principally used by the *B*-physics group in CMS. Mainly, it collects events that request dimuon candidates with an invariant mass between the J/ψ and the Υ masses, gauging between the needs of the various analyses and the maximum allowed rate, which was around 30 Hz for the *B*-physics during 2011 and 2012.

The HLT trigger selection for both signal and normalization channels is based on two opposite-charge L3 muons, the highest quality of HLT identified and reconstructed muons.

In CMS, the *B*-physics analyses usually rely on HLT paths that apply cuts on the transverse momentum and on the pseudorapidity of muons and dimuons. With increasing instantaneous luminosity (during the two years, the luminosity raised from about 5×10^{32} to about $8 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$), the threshold of the muon p_T was consequently raised and additional requirements on the dimuon mass and momentum and on SV observables, were implemented.

An important element to consider is the presence of the pileup, especially in the 2012 run (Sec. 3.1.3). Two uncorrelated muons coming from different primary vertices could be coupled to form a displaced dimuon. On every *B*-physics dimuon trigger a selection based on the distance of the closest approach between the two muon tracks is applied: the minimum 3D distance d_{ca}^{L3} between the two muon tracks must satisfy $d_{ca}^{L3} < 0.5 \text{ cm}$. For muons coming from the same vertex, this distance is of the same order of the tracker L3 vertex resolution (tens of microns).

To avoid overfilling the rate band with unproductive events, some loose selections related to the secondary vertex are utilized. For this reason, an excellent tracker resolution is imperative to keep rate and background under

Table 4.1: Kinematic selections at HLT level for the signal channel, for each different used trigger.

HLT path	$\min p_{T,\mu_1}^{L3}$	$\min p_{T,\mu_2}^{L3}$	$\max \eta _\mu^{L3}$	$\min p_{T,\mu\mu}^{L3}$
	[GeV]	[GeV]		[GeV]
Barrel 2011	4.0	4.0	1.5	3.9
Endcap 2011	4.0	4.0	2.4	5.9
Barrel 2012	3.0	4.0	1.8	4.9
Endcap 2012	4.0	4.0	2.2	6.9

control.

Next sections are dedicated to the descriptions of the two HLT path sets used by this analysis. Despite the different instantaneous luminosities, signal and normalization triggers changed not much during each year, thus only the strictest versions of each HLT path are described.

4.2.1 HLT paths for the signal sample

In a search for rare decays, the signal trigger efficiency must be the highest possible.

Since in the barrel the resolution is better and the background level is lower than in the endcaps, two different trigger paths are used, with different $|\eta|$ ranges: the one with the strictest muon η range can apply the lowest p_T requirements. The pseudorapidity request, $|\eta| < 1.8$, is chosen to be wider than the analysis barrel category cut (see last paragraph of Sec. 4.1), to take into account the slightly worse trigger resolution.

All signal paths request the dimuon invariant mass to be around the $B_{(s)}^0$ meson masses: $m_{\mu\mu} \in [4.8, 6.0]$ GeV. Additional kinematic requests are shown in Table 4.1. Finally, a very loose cut is requested on the dimuon vertex minimum probability: $p > 0.05\%$. This quantity is the SV fit p -value evaluated from the χ^2 of the vertex, given the two degrees of freedom, corresponding to the two muon tracks used by the vertex builder (Sec. 3.2.7.4).

The total allowed rate for the B_s^0 triggers during the two years of data taking was about 5 Hz. Figure 4.1 shows a typical rate trend for the central B_s^0 trigger used during a 2012 run. The maximum rate, measured at the beginning of the run, is about 3.3 Hz, and it decreases to about 1.3 Hz at the end of the data acquisition.

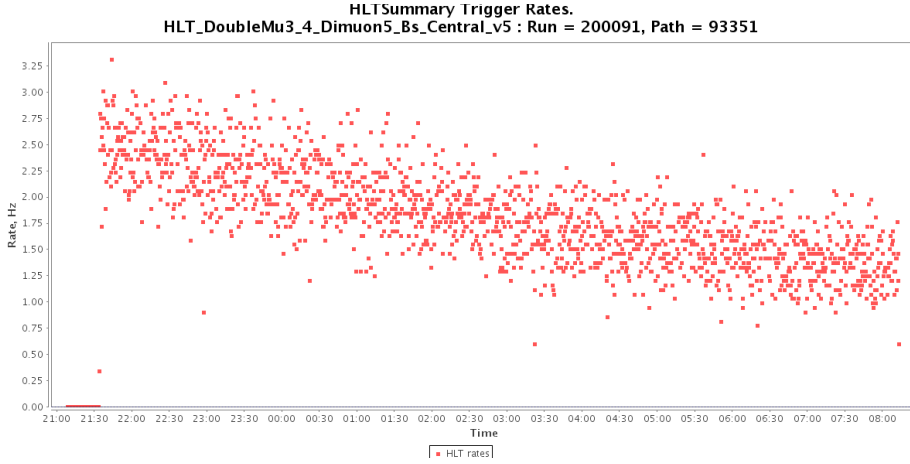


Figure 4.1: HLT double muon trigger rate as a function of time for a given LHC run. For this given trigger, the minimum p_T requests are 4 and 5 GeV for the two muons, respectively, and 4.9 GeV for the dimuon candidate. Furthermore, the dimuon candidate must have an invariant mass between 4.8 and 6 GeV and the two muons must have $|\eta| < 1.8$.

4.2.2 HLT path for the normalization and control samples

Both normalization and control channels rely on the reconstruction of J/ψ mesons at the HLT level. The high CMS performances on muon reconstruction allow for a very clean and high purity J/ψ meson signal-over-background ratio (see, for example, Fig. 3.34 and Fig. 4.2).

The J/ψ meson has by far the highest cross-section with respect to the other resonances involved in the B -physics analyses. To not saturate the B -physics HLT rate, much stricter selections than for the $B_{(s)}^0$ signals are requested, which involve kinematical and SV selections.

The HLT path used by this analysis selects on well-displaced J/ψ mesons coming from B -meson decays. The requirements are:

- two L3 muons with $p_{T,\mu}^{L3} > 4$ GeV and $|\eta|_{\mu}^{L3} < 2.2$. The high-rapidity region $|\eta|_{\mu}^{L3} > 2.2$ is removed by the L1 requests;
- dimuon vertex fit minimum probability greater than 15%;
- dimuon invariant mass centered around the J/ψ meson, in $[2.9, 3.3]$ GeV, and dimuon transverse momentum $p_{T,\mu\mu}^{L3} > 6.9$ GeV;
- transverse life-time significance $L_{xy}^{L3} > 3$, where L_{xy}^{L3} is the ratio of the transverse distance between the dimuon vertex and the PV, over its uncertainty;

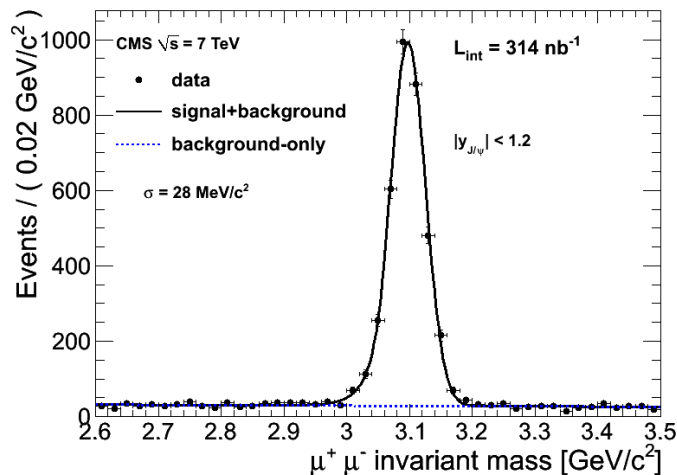


Figure 4.2: Opposite-sign dimuon invariant mass distribution around the J/ψ meson peak, in the $|\eta| < 1.2$ range, fitted with a Crystal Ball [134] plus an exponential pdf [135].

- cosine of the transverse pointing angle α_{xy} higher than 0.9. The transverse pointing angle is the angle between the dimuon momentum and the vector from the PV to the dimuon vertex, in the transverse plane. For a perfectly aligned tracker, this angle must approach zero, compatibly with the tracker resolution.

The total rate for this displaced J/ψ trigger was one of the highest allowed (around 8 Hz), as J/ψ mesons coming from B decays are used also by many other B -physics analyses.

4.3 Off-line signal selection

Variables used to discriminate between signal and background events are related to the single muon reconstruction, the primary vertex, the B candidate and its secondary vertex.

After loose preselections, the most discriminating variables are sent as inputs to MVA methods, that separate as much as possible signal events from background events.

In this section all such variables are described.

4.3.1 Single muon selection

The inner track of the muon candidate is required to be of *highPurity* quality (see Sec. 3.2.7.1.4). The muon momentum is taken from the inner track measurement since at these energies the muon sub-detectors do not improve the tracker

resolution (Fig. 3.26). Furthermore, the two muons in the final state are requested to correspond to the HLT L3 muons used for triggering the event. In this way we have a complete correspondence between the trigger and the off-line analysis objects.

Muon identification is an important aspect in this analysis, especially for the $B^0 \rightarrow \mu^+ \mu^-$ search, as it will be clearer later. Therefore, it is crucial to have the lowest possible “fake muon misidentification” and, at the same time, keep a high muon efficiency. The different behavior of muon, pion, and kaon tracks is studied on data and on MC simulations. Many phenomena can occur such that a non-muon particle is identified as a muon track: for example, protons have a non-null probability to traverse the detector matter and to reach the muon sub-detectors. An other possibility is that a kaon decays into a muon before reaching the muon stations. They all contribute to the muon misidentification.

As a first step, each muon track undergoes the tight muon selection, as described in Sec. 3.2.7.2. This is a standard selection for low p_T muons, widely used by many CMS physics analyses. These selections are especially optimized to suppress muons coming from cosmic rays and from decays-in-flight or punch-through of pions and kaons.

4.3.1.1 MVA muon identification

The tight muon identification selects a high quality and clean sample of well reconstructed muons. However, the misidentification rate, that is, the probability to identify a muon track given a non-muon particle, is a particularly important aspect of this analysis. The peaking rare background (Sec. 4.1.1) sits on the same invariant mass region of the B^0 meson and it is composed by decays that have not negligible branching fractions (Sec. 4.7). Thus, a wrong misidentification value could lead to falsely enhance the B^0 yield. For this reason, care must be taken, and a special treatment is performed to further reduce the muon misidentification expected rate.

After the tight muon selection, an MVA analysis is performed to further discriminate muon tracks from “fake” tracks. This analysis uses a Boosted Decision Tree (BDT) method to separate true muons from fake muons [136].

A decision tree is a binary tree structured classifier (see also Fig. 4.3). Binary decisions are taken on one single variable repeatedly at a time until a stop criterion is reached. Each split uses the variable that at each step (or node) gives the best separation between signal and background when being cut on. The same variable may be used at several nodes, while others might not be used at all. This produces a ranking of the variables: the more used a variable is, the higher rank it gets.

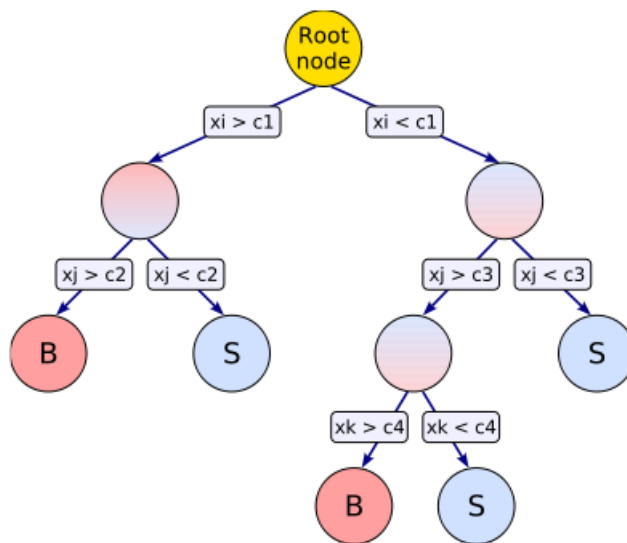


Figure 4.3: Sketch view of a decision tree. Starting from the top, a sequence of left-right splits using the discriminating variables “ x_i ” is applied to the data. The leaf nodes at the bottom end of the tree are labeled “S” for signal and “B” for background depending on the majority of events that end up in the respective nodes.

The overall phase space is split in this way into many regions that are eventually classified as signal or background, depending on the majority of training events that end up in the final leaf node.

To enhance the classifier performance a boosting technique can be used, sequentially applying the BDT algorithm to reweighed (boosted) versions of the training data and then taking a weighed majority vote, thus produced.

The boosting of a decision tree extends the tree to several trees which form the so-called “forest”. The trees are derived from the same training ensemble by reweighing events, and are combined into a single classifier which is given by the average of the individual decision trees.

The boosting technique stabilizes the response of the decision trees with respect to fluctuations in the training sample and it is able to considerably enhance the performance with respect to a single tree.

Many single muon variables can be used as handles in the discrimination. They can be divided into the following classes:

1. basic kinematics distributions;
2. inner track fit informations;
3. global track fit informations;
4. muon chamber informations;

5. calorimeter informations and isolation.

Variables for the first class include the muon transverse momentum, pseudo-rapidity, azimuthal angle, charge, and impact parameters. For kaons and pions the η distributions have spikes at the intersection between the CMS barrel and endcaps, and the impact parameters have larger tails than for real muons.

Inner-track studied variables are the number of tracker layers with measurement, the number of valid tracker and pixel hits, the number of degrees of freedom of the inner track fit. On the average, tracks of fake muons show a lower number of valid hits than those of real muons. In addition to these, two more variables related to the quality of the inner track fit can be taken into account: the reduced χ^2 of the fit and the “kink” discriminator: at each tracker layer, the extrapolated states from the two halves of the tracker tracks are compared to get a χ^2 , the highest value of the χ^2 across all layers is used as the discriminator. This variable is useful for distinguishing decays in flight of particles (as in $K^+ \rightarrow \mu^+ \nu_\mu$) inside the track: the two daughter particles each receive a kink due to the decay and change their directions with respect to the mother particle. Fake muons have higher tails in the kink and χ^2 distributions, and come from those particles that have undergone a decay inside the tracker volume.

For what concerns the global track fit, the most discriminating variables are mainly related to the matching between its inner (tracker) and outer (standalone) parts. One useful variable is the global track tail probability $-\ln P$, where P is the probability of the global track fit, which depends on the χ^2 and on the number of degrees of freedom of the global track. Other two global variables are defined in the following way. One starts with the inner and outer tracks of the muon, and propagates the first one outwards and the second one inwards to a common plane. The momenta and positions of the intersections of each track with the plane are compared, and their compatibility is assessed by computing a χ^2 of the match. These variables are particularly sensitive to tracks with kinks in the volume outside the tracker (for example due to in-flight-decays or to material interaction) and can also discriminate against global tracks made from a wrong inner-outer match.

A powerful variable in the muon system is that related to the matching between the global muon track and the segments reconstructed locally in the muon stations (“segment compatibility”). More specifically, for each muon track the algorithm

- checks how many CSC and DT stations were supposed to be crossed;
- measures the distance between hits in the muon detector and the muon track;

- controls if in each layer the muon track passes close to a boundary;
- depending on the number of crossed stations, weights more the outer stations, as this more likely corresponds to the muon hypothesis.

Based on these informations a single weight variable is built with range $[0, 1]$, such that 0 corresponds to the non-muon hypothesis and 1 corresponds to the muon hypothesis.

Variables related to calorimeter information do not show a strong discriminating power between real and fake muons. Moreover, given that the golden certification is not used, they are not employed.

The MVA study starts from collected reconstructed tight muons. For the signal input, a MC sample of muons coming from B_s^0 dimuon decays are used. For the background input, a MC sample of kaons, identified as tight muons, coming from hadronic B decays, are used.

The BDT method of the ROOT TMVA package [136] is used. Since statistics is limited (especially for fake muons, as the starting misidentification ratio is around the permill level), only the most powerful variables are used, as given by the BDT ranking output (see Fig. 4.4 for the most discriminating used variables).

Results of the BDT training are analyzed looking at the kaon rejection versus the muon efficiency. The improvement over the tight muon identification is such that it is possible to reduce the kaon misidentification by half, merely loosing a 10% efficiency on the true muon identification, as shown in Fig. 4.5. Therefore, given that the rare peaking background is made by two misidentified hadrons, it is possible to reduce this source by a factor 4 around the B^0 mass, loosing only 1/5 of B^0 mesons, improving the final sensibility of the analysis.

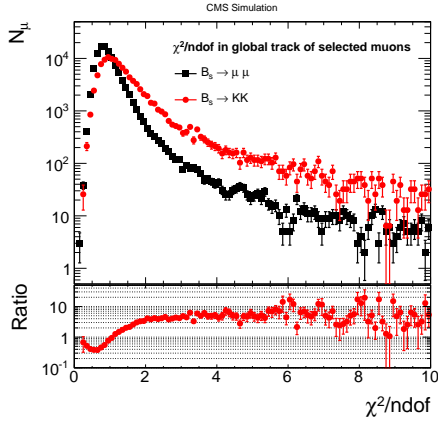
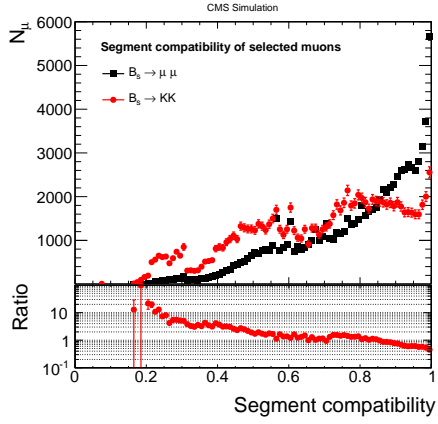
4.3.1.2 Validation of the muon identification

Track and muon efficiencies are validated in data through a TNP technique (see Sec. 3.2.7.2.1), using the $J/\psi \rightarrow \mu^+ \mu^-$ decay.

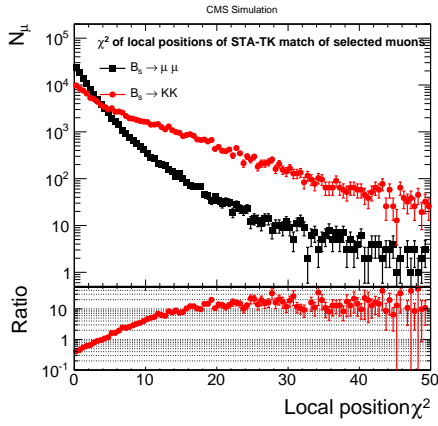
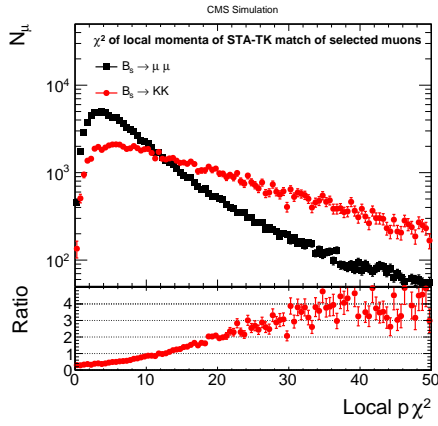
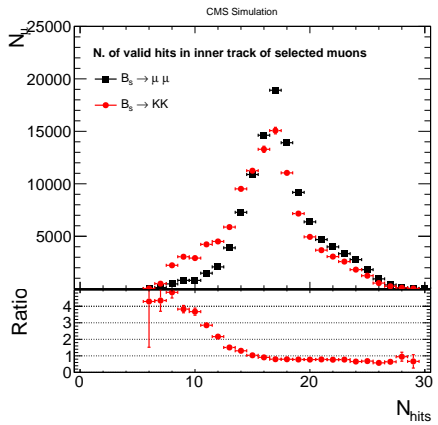
It is possible to measure the muon misidentification in data using primary datasets collected without muon triggers, to obtain an unbiased sample of muon candidates (see also Sec. 3.2.7.2.2). For kaons and pions, the following decay chain is used:

$$D^{*+} \rightarrow D^0 \pi^+ \rightarrow K^- \pi^+ \pi^+ \quad (4.7)$$

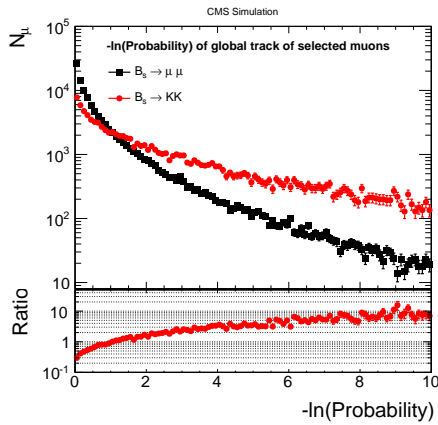
Two opposite-charge hadron tracks are combined to form a D^0 candidate. The D^0 candidate is then combined with a low momentum track (as this slow pion receives a low energy kink) to form the D^{*+} candidate. Additional requirements on the displaced vertex are applied to discriminate the signal against the combinatorial

(a) Global track fit normalized χ^2 .

(b) Segment compatibility.

(c) Local position χ^2 .(d) Local momentum χ^2 .

(e) Tracker valid hits.



(f) Global track tail probability.

Figure 4.4: Muon (black) and kaon (red) distributions of some of the most discriminant variables used by the muon identification MVA.

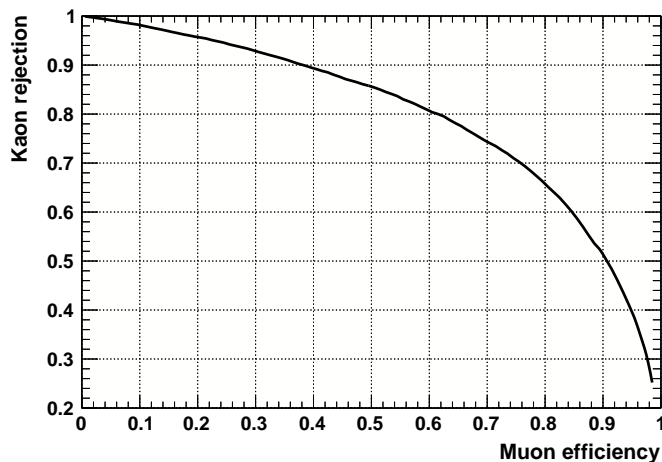


Figure 4.5: Kaon rejection versus muon efficiency for the muon MVA study.

Table 4.2: Muon misidentification probability ranges for pions, kaons and protons.

Hadron	Misidentification (10^{-3})
π	0.5–1.3
K	0.8–2.2
p	0.4–1.5

background. The pion and kaon identification is done by requiring that the charge of the two pions is the same.

Examples of the $D^{*+} - D^0$ mass difference distributions are shown in Fig. 4.6. The peaks in the mass difference distributions are fitted with a double Gaussian pdf on top of a Fermi pdf that parametrizes the background.

In addition to the D^{*+} method, other channels are used to validate the MVA results, with the kaon and pion rates obtained from the normalization and control channels, from the decay $K_s \rightarrow \pi^+ \pi^-$ and from the MC simulations. These results are shown in Fig. 4.7, where the pion and kaon misidentification rates are plotted versus the average pion or kaon transverse momentum in each sample.

For protons, $\Lambda^0 \rightarrow p \pi^-$ decays are used.

In general, the misidentification values show agreement between the MC predictions and data. Therefore the misidentification predicted by the MC simulations is used and the high available statistics makes it possible to use p_T and η dependent measurements. The probabilities for pions, kaons and protons are shown in Table 4.2. The systematic uncertainty on the hadron fake rates is

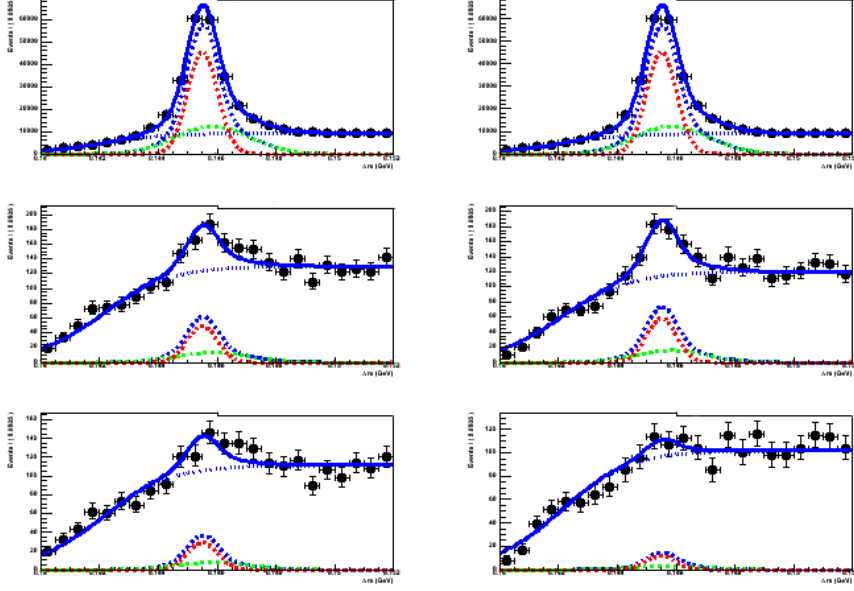


Figure 4.6: Mass difference distributions between the D^{*+} and D^0 mesons, obtained from datasets collected with non-muon triggers. The left-upper plot shows all selected events. The left-middle plot shows events where a pion is misidentified as a muon, while the right-middle plot is the same but for kaons. Here the tight muon identification is applied. The lower plots show the same but using the MVA muon identification on the top of the tight muon. Finally the right-upper plot is for pure events where neither the pion nor the kaon is misidentified. The blue line shows the full pdf, the dashed blue line shows the signal contribution, made by the sum of the two Gaussian pdfs, in red and green dashed lines. The fine-dashed blue line represents the background part.

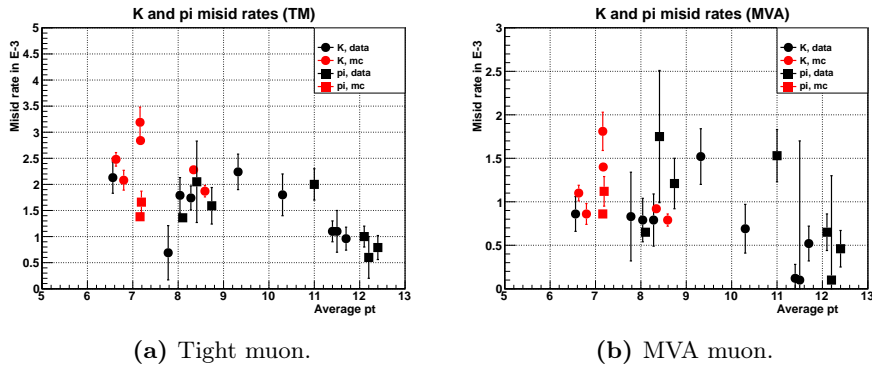


Figure 4.7: Muon misidentification rates for the tight muon identification (a) and for the MVA identification (b), determined from different data (black) and MC (red) samples, for kaons (circles) and pions (squares), plotted versus the average transverse momentum, for particles with $p_T > 4$ GeV. In addition to the D^{*+} data also kaon results from the normalization and control channels are included.

Table 4.3: Invariant mass region definitions for the signal channel.

Definition	Invariant mass range (GeV)
overall mass window	[4.9, 5.9]
$B^0 \rightarrow \mu^+ \mu^-$ signal window	[5.2, 5.3]
$B_s^0 \rightarrow \mu^+ \mu^-$ signal window	[5.3, 5.45]
blind window	[5.2, 5.45]
side-band windows	[4.9, 5.2] \cup [5.45, 5.9]

taken from the spread between the various results obtained with these different methods. The relative uncertainties on the misidentification rates are estimated to be at 50%.

4.3.2 Discriminating variables for $B_{(s)}^0$ -signal decay modes

Once two muon tracks with opposite charge are found in an event, and their invariant mass is within the overall window, they form the “ B candidate” (see Fig. 4.8 for the kinematic distributions of the candidate).

The signal acceptance, evaluated with a MC simulation produced with no generator filters, requests the signal candidate to be in the detector volume:

- both generated muons coming from the B decay must be within the geometric acceptance: $|\eta| < 2.5$ and $p_T > 3.5$ GeV;
- both muons must have a reconstructed *highPurity* track with $|\eta| < 2.4$ and $p_T > 3.5$ GeV.

The invariant mass region under study is divided into four separate sub-ranges, defined in Table 4.3.

The SV is determined with the kinematic vertex fitter (Sec. 3.2.7.4), applying a geometric constraint. The vertex fit reduced χ^2 is one of the input variables for the candidate MVA.

Given the secondary vertex and its momentum, a matching PV is chosen on the distance of the closest approach along the z axis. To avoid any geometrical bias due to displaced tracks, all tracks forming the B candidate are removed from the chosen PV and this is “refitted”, that is, it is reconstructed with the remaining tracks (this is valid also for the NS and CS samples that have three or four tracks, respectively). The PV average track weight $\langle w \rangle = N_{\text{d.o.f}} + 2 / (2N_{\text{trk}})$, where $N_{\text{d.o.f}}$ is the total number of PV degrees of freedom and N_{trk} is the total number of tracks belonging to that PV, is used as a loose cut on the PV reconstruction, as it is higher for well reconstructed PVs.

The main background is made by the combination of two uncorrelated muons (Fig. 4.9), that can come from two different b or c decays. The inner tracker

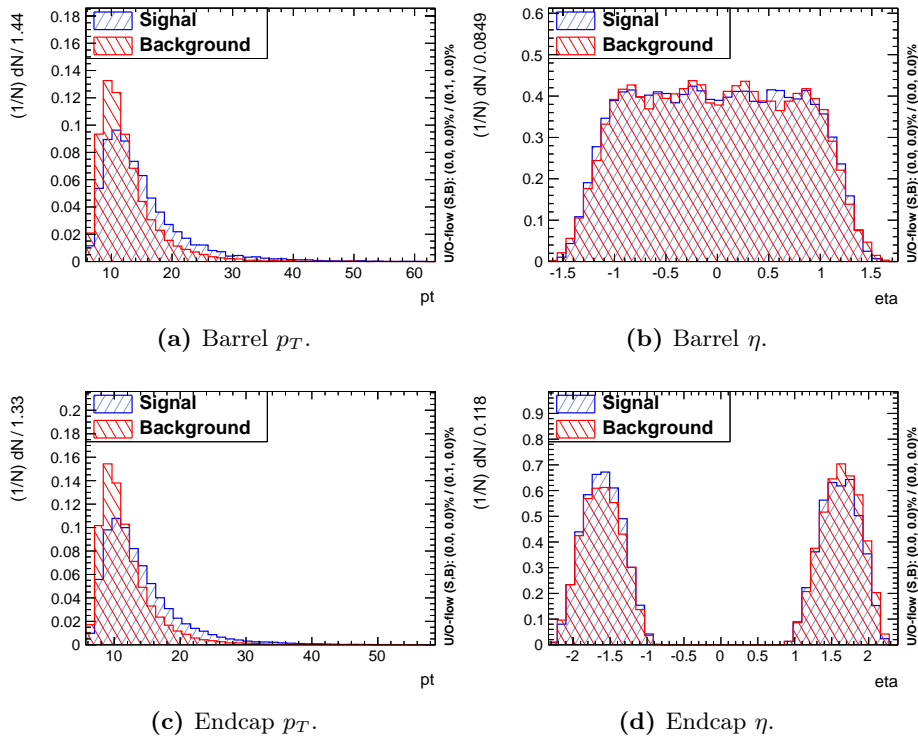


Figure 4.8: Normalized B -candidate kinematic distributions for signal MC events (in blue) and data side-bands (in red).

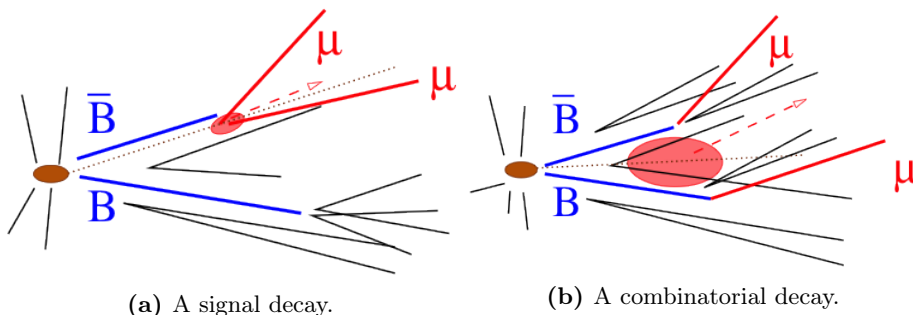


Figure 4.9: Sketch of signal (a) and combinatorial background (b) events.

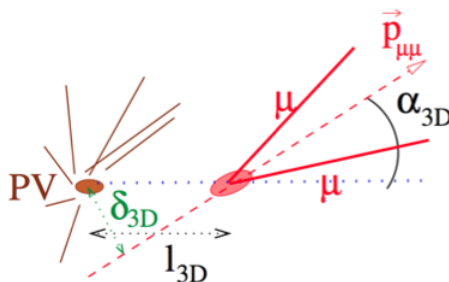


Figure 4.10: Visual sketch of a displaced decay: the impact parameter δ_{3D} , the decay length ℓ_{3D} and the pointing angle α_{3D} variables are shown. The two muon tracks (μ) are combined to form the dimuon momentum $\vec{P}_{\mu\mu}$.

resolution is fundamental to discriminate these fake dimuons from real signals. Their SV is badly reconstructed, their trajectories do not fit well with coming from the same spatial point, as they should if coming from a real decay, and the dimuon reconstructed track does not fit with coming from an existing PV.

All these elements are investigated with the SV variables described in this section.

An other feature is that since many of these muons can come from decay chains, other charged particles are present around them (mainly pions and kaons), and this can be checked looking for other close-by tracks. The so-called “isolation” variables (Sec. 4.3.2.1) explore this effect.

The candidate three-dimensional impact parameter with respect to the PV δ_{3D} and its significance $\delta_{3D}/\sigma(\delta_{3D})$ are used in the event selection. The same is true for the significance of the 3D flight length, $(\ell_{3D}/\sigma(\ell_{3D}))$. Other important variables for displaced vertices are the pointing angle α_{3D} , that is the angle in three dimensions between the B -candidate momentum and the vector from the PV to the SV, and the distance of closest approach between the two muon tracks d_{ca}^{max} . A sketch of the variables related to the SV is shown in Fig. 4.10.

Fig. 4.11 shows the above-mentioned variables for the 2012 barrel category

for the signal MC simulation (in blue) and the data side-bands (in red). It is possible to appreciate how different these distributions look like for signal and background events, thanks to the good CMS tracker resolution.

4.3.2.1 Isolation

Many different isolation variables are used to separate the signal, which is an isolated candidate, from background events containing tracks coming from jets. The first isolation variable I is associated to the PV:

$$I = \frac{p_T(B)}{p_T(B) + \sum_{\text{trk}} p_T} \quad (4.8)$$

where $p_T(B)$ is the candidate transverse momentum. The sum is extended to all other tracks in a cone with $\Delta R = 0.7$, $p_T > 0.9 \text{ GeV}$ and that are not associated to any PV, but have the distance of the closest approach to the SV d_{ca} such that $d_{ca} < 500 \mu\text{m}$, or that are associated to the same PV as the candidate. This last request is made to avoid a pileup dependence of the variable, as, in this way, it does not count tracks coming from other PVs. The ΔR and the minimum p_T constants have been chosen to maximize the separation power of this variable. Fig. 4.12a shows the distribution for I . For B candidates without any other charged tracks above the p_T cut-off, the isolation is equal to $I = 1$. The pronounced dip in the distribution, just below $I = 1$, arises from the minimum transverse momentum requirement that implies a maximum value of I , depending on the p_T of the B candidate.

In addition to this isolation, two other candidate variables check for track activity around the SV:

- the minimum distance of the closest approach d_{ca}^0 of tracks (either associated to no PV or to the same PV as the candidate) to the SV;
- The number of close tracks N_{trk} with $d_{ca} < 300 \mu\text{m}$ and $p_T > 0.5 \text{ GeV}$.

To have further handles on the isolation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay, also the activity around each single muon track is computed. The muon isolation I_μ is defined as:

$$I_\mu = \frac{|\vec{p}_\mu|}{|\vec{p}_\mu| + \sum_{trk} |\vec{p}_{trk}|}, \quad (4.9)$$

where the sum includes all tracks in a cone with radius $\Delta R < 0.5$, with $p_T > 0.5 \text{ GeV}$, that are not part of the B_s^0 candidate and either associated to the same primary vertex as the B_s^0 candidate or are not associated to any primary vertex but have a distance of closest approach less than 1 mm to the muon trajectory.

The parameters for the single muon isolation have been optimized to achieve maximal separation between signal and background while being independent of

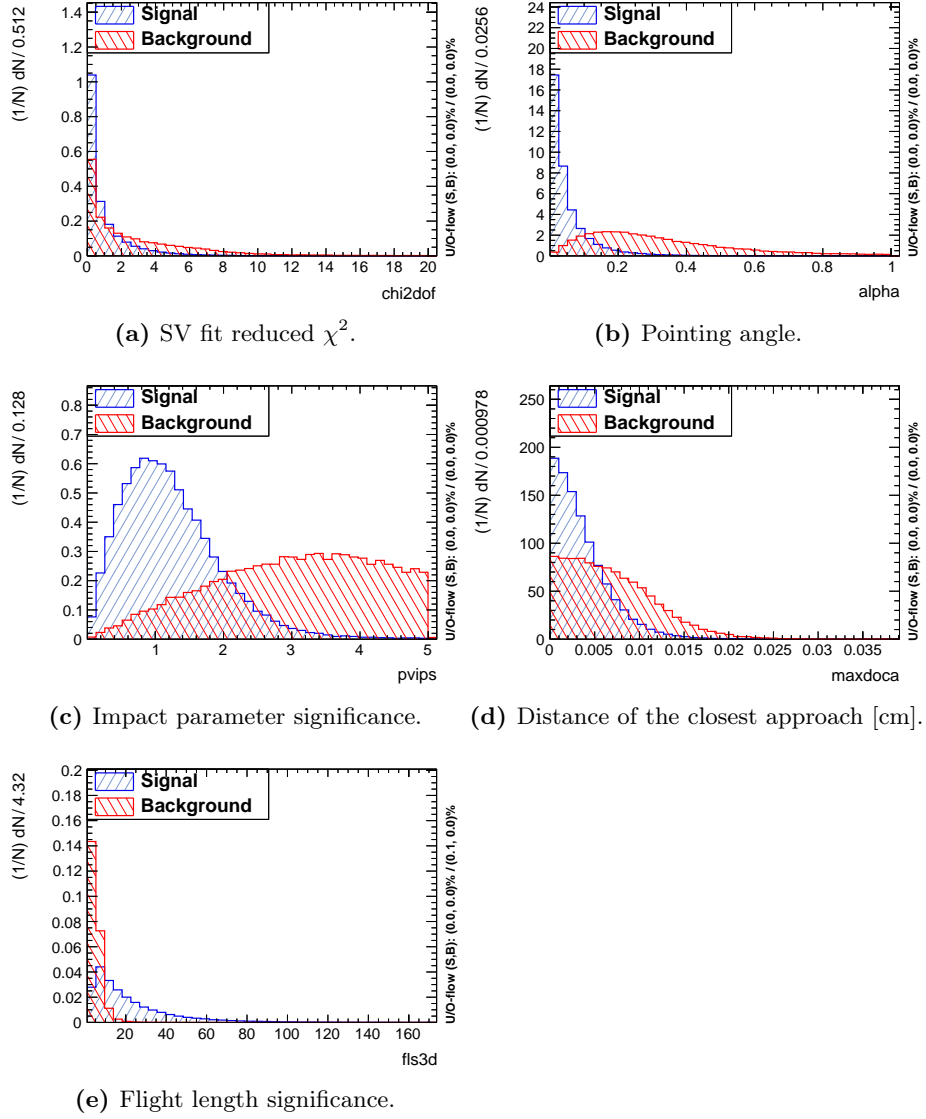


Figure 4.11: Normalized B -candidate distributions of relevant variables for signal MC events (in blue) and data side-bands (in red).

the level of pileup.

All these distributions are shown in Fig. 4.12.

4.4 Normalization channel event selection

The chosen NS channel is the decay $B^\pm \rightarrow J/\psi K^\pm$. In principle, various channels could be used, each with advantages and disadvantages:

- the decay $B^\pm \rightarrow J/\psi K^\pm$ offers the minimal number of additional tracks next to the dimuon. The disadvantage is that the ratio of B^\pm and B_s^0 hadronization is known only to 8%. The $B^\pm \rightarrow J/\psi K^\pm$ and the $J/\psi \rightarrow \mu^+\mu^-$ BF's carry an additional 3.4% (from Eq. (4.5) and Eq. (4.6));
- the decay $B_s^0 \rightarrow J/\psi \phi$ has a branching fraction relative uncertainty of 32%. Furthermore, since this channel has two hadronic tracks, the tracking uncertainty, which will be discussed in Sec. 4.8.2.1, is twice larger than in the channel $B^\pm \rightarrow J/\psi K^\pm$;
- the decay $B_s^0 \rightarrow K^+K^-$ would be a very interesting normalization channel but there is no displaced hadronic trigger and no kaon identification in CMS. Hence, it cannot be used.

For these reasons the chosen normalization channel in CMS is the decay $B^\pm \rightarrow J/\psi K^\pm$.

For the NS acceptance, the requests are the same as for the signal (Sec. 4.3.2), but with further requirements on the kaon track:

- the generated kaon track must fulfill $p_T > 0.4 \text{ GeV}$ and $\eta < 2.5$;
- the reconstructed *highPurity* kaon track must have $p_T > 0.5 \text{ GeV}$ and $\eta < 2.4$.

Only events that pass the triggers on displaced J/ψ mesons, as described in Sec. 4.2.2, are selected. Thus, the off-line reconstruction of $B^\pm \rightarrow J/\psi K^\pm$ candidates starts from two muons with opposite charge, coming from a SV. Each other track in the event is coupled to the muon pair, to form B^\pm candidates, and the kaon mass hypothesis is assigned to this track. The two muon tracks and the kaon track must be of *highPurity* quality and be within the acceptance of the inner tracker ($p_T > 0.5 \text{ GeV}$ and $|\eta| < 2.4$ for the 2011 data or $|\eta| < 2.1$ for the 2012 data). The distance of closest approach between all pairs among the three tracks is required to be less than 1 mm. The dimuon invariant mass $M_{\mu\mu}$ must be compatible with that of a J/ψ meson, taking into account the CMS track resolution: $M_{\mu\mu} \in [3.0, 3.2] \text{ GeV}$. All three tracks are used in the vertex fitting.

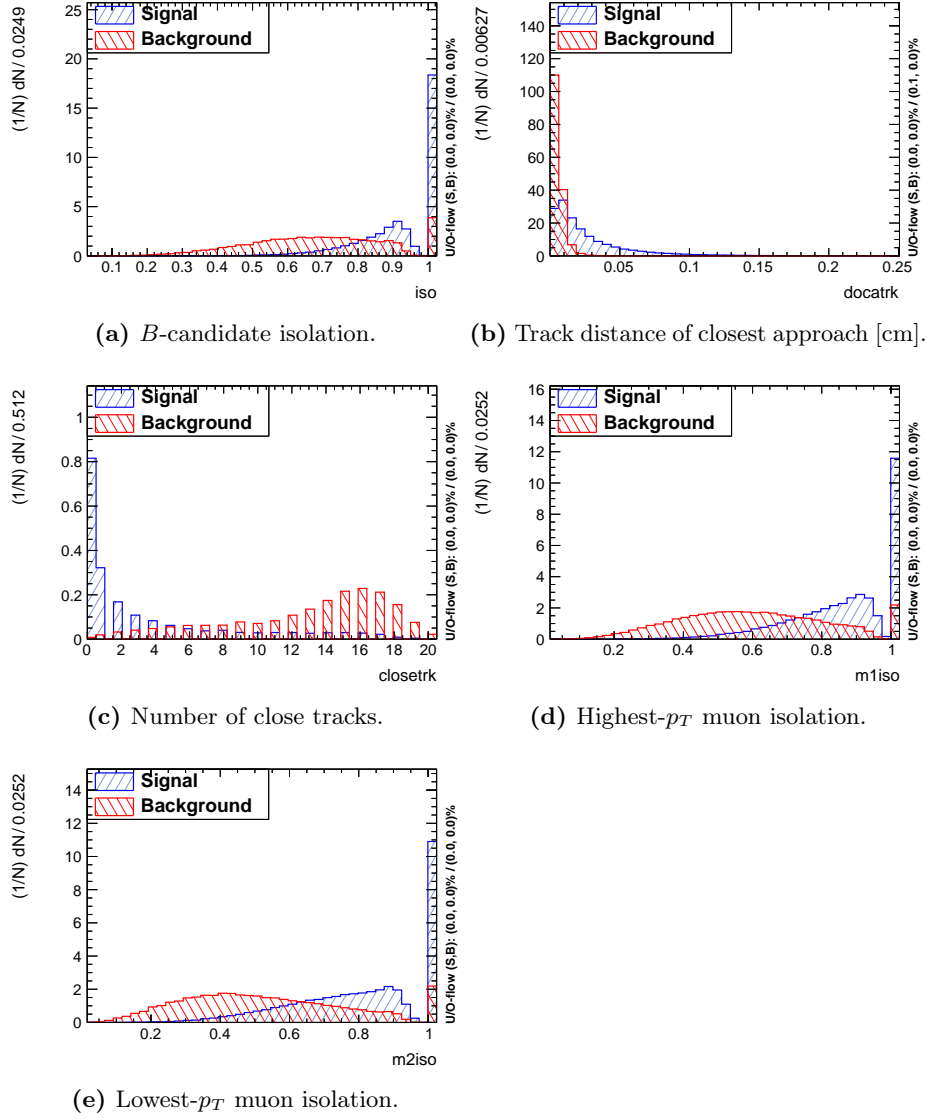


Figure 4.12: Normalized isolation distributions for signal MC events (in blue) and data side-bands (in red).

Table 4.4: Invariant mass region definitions for the normalization channel.

Definition	Invariant mass range (GeV)
overall mass window	[4.90, 5.90]
low side-band	[5.05, 5.15]
signal window	[5.2, 5.35]
high side-band	[5.4, 5.5]

Table 4.4 defines the side-bands and signal regions for the normalization channel. These mass regions are adopted for the side-band subtraction used in the comparison of $B^\pm \rightarrow J/\psi K^\pm$ decays in data and MC simulation.

In Fig. 4.13 the invariant mass distributions of the normalization channel are shown for the barrel and endcap categories, for the two years of data taking.

The B^\pm yield is parametrized with a double Gaussian pdf in the barrel category and a single Gaussian pdf in the endcap category. Since all tracks close to the same vertex of the J/ψ candidate are used to construct the B^\pm candidate, the background has a combinatorial part, consisting of fake candidates where the muon pair is associated to one uncorrelated track. This component has a non resonant shape and an exponential pdf is chosen.

The decays $B^\pm \rightarrow J/\psi \pi^\pm$, which enter the mass window with a wrong mass assignment, are modeled with a Landau function, where the width and peak position are fixed to the expectation from the MC simulation. The $B^\pm \rightarrow J/\psi \pi^\pm$ normalization is constrained relatively to the $B^\pm \rightarrow J/\psi K^\pm$ yield.

The background presents also a step around 5.15 GeV which is due to partially reconstructed decays, like, for example, $B^0 \rightarrow J/\psi K^{*0} \rightarrow \mu^+ \mu^- K^- \pi^+$, where the π^+ is not reconstructed. To parametrize it, simulated $B^0 \rightarrow J/\psi(\mu^+ \mu^-)X$ events are produced, which are then reconstructed in the same way as the $B^\pm \rightarrow J/\psi K^\pm$ events. The invariant mass plots after applying all selection criteria are shown in Fig. 4.14 for the barrel and endcap regions. These yields are fitted with an error function whose parameters are used in an additional background pdf in the final fit.

A possible cause of systematic effect is the SV reconstruction. In the SV algorithm it is possible to constraint the dimuon invariant mass to be consistent with the J/ψ mass. With this approach, the B^\pm mass resolution is much narrower and better displaced from the edge of the partially reconstructed B^0 decays (Fig. 4.15). Since the peak of the correctly reconstructed B^\pm candidates is better separated from the partially reconstructed B mesons, the fit can be limited to a narrower range excluding the region of partially reconstructed B mesons. A 5% systematic uncertainty is assigned checking the difference in the number of B^\pm estimated with and without the J/ψ constraint. This uncertainty is added to

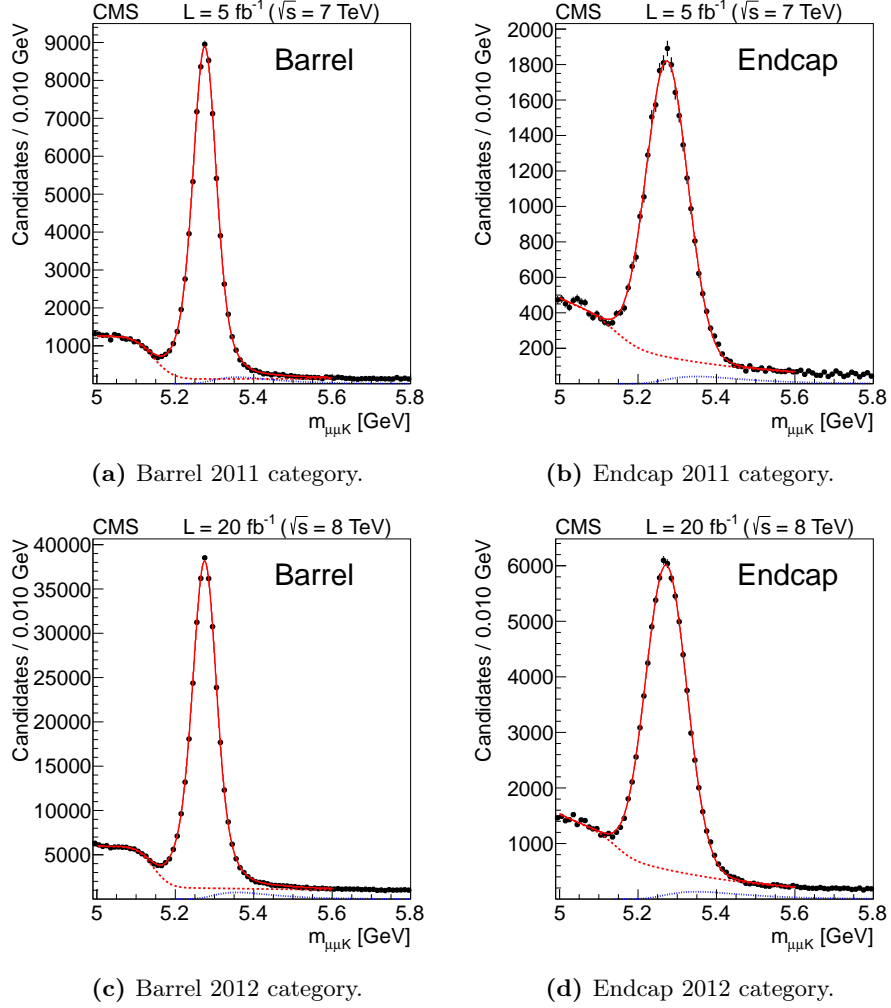


Figure 4.13: $B^\pm \rightarrow J/\psi K^\pm$ invariant mass distributions for the four categories. The red line shows the total pdf. The red dashed line indicates the pdf used to describe the combinatorial background and the contribution from partially reconstructed B decays. The blue dotted line indicates the component from $B^\pm \rightarrow J/\psi \pi^\pm$.

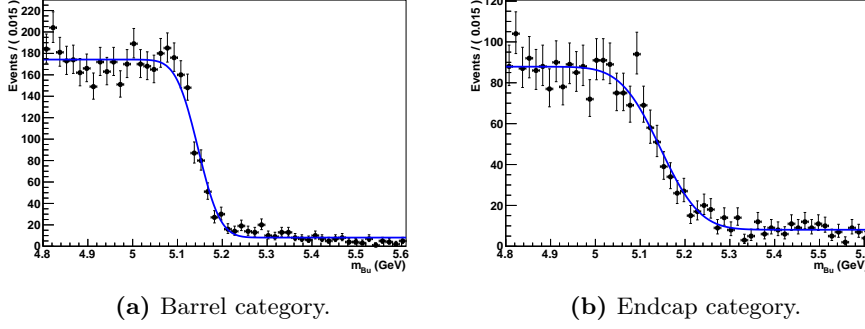


Figure 4.14: Simulated partially reconstructed $B^0 \rightarrow J/\psi(\mu^+\mu^-)X$ decays for the two $|\eta|$ categories. A Fermi pdf (blue line) is superimposed to the MC points (black dots).

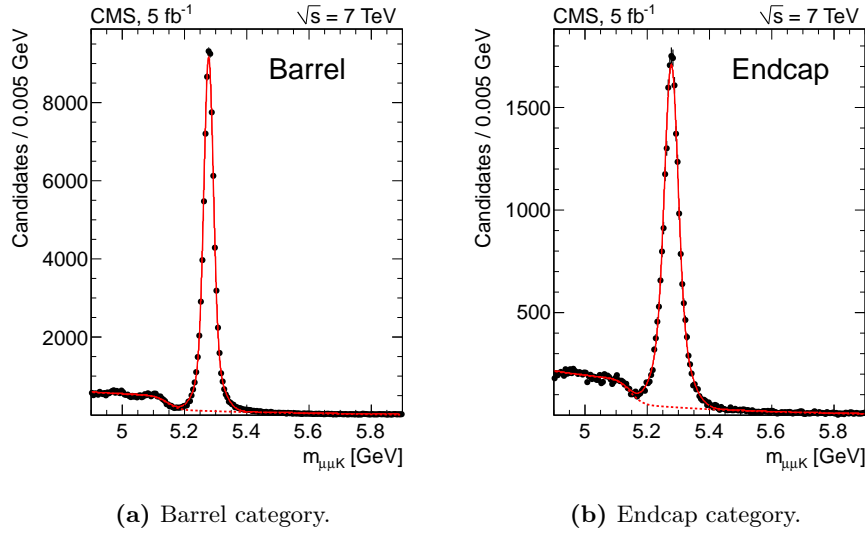


Figure 4.15: Invariant mass distribution for $B^\pm \rightarrow J/\psi K^\pm$ candidates where the invariant mass of the two muons is constrained to the J/ψ world average mass. The legend for the superimposed pdf is the same as in Fig. 4.13.

the $B^\pm \rightarrow J/\psi K^\pm$ statistical uncertainty coming from the fit results, which will be discussed in Sec. 4.10, to form the total NS uncertainty.

4.5 Control sample event selection

It is fundamental to check that the signal MC simulation well reproduces the B_s^0 meson distributions in data. For this purpose, $B_s^0 \rightarrow J/\psi \phi \rightarrow \mu^+ \mu^- K^+ K^-$ decays are used as a validation of the MC simulations and as a way to estimate the systematic uncertainty of the analysis efficiency for $B_s^0 \rightarrow \mu^+ \mu^-$, based on the level of agreement between data and MC simulation. The validation of the MC simulation is especially important for the isolation variables.

The decay $B_s^0 \rightarrow J/\psi \phi$ has four final state particles, thus lowering on average the muon transverse momenta compared to the signal decay. Since the muon p_T requirements stay at the same numerical values for the signal decay and the normalization and control channels, the p_T distribution of the B_s^0 candidates is expected to be harder.

The CS acceptance requests one more kaon track fulfilling the same acceptance requirements of Sec. 4.4.

The trigger paths are the same as for the normalization sample (Sec. 4.2.2), since they are based on displaced J/ψ mesons. All four involved tracks must fulfill the same selections described in Sec. 4.4 for the NS selection: *highPurity* tracks with $p_T > 0.5$ GeV and $|\eta| < 2.4$ for the 2011 data or $|\eta| < 2.1$ for the 2012 data. The two opposite-sign muons are combined to form a J/ψ candidate, while the other two opposite-sign tracks are assumed to be kaons, that must have an invariant mass $M_{KK} \in [0.995, 1.045]$ GeV, to form the ϕ meson candidate. To further suppress background events, the two kaons must have $\Delta R < 0.25$ in the $\eta\phi$ plane.

Figure 4.16 shows the invariant mass distributions of the control sample in the barrel and endcap channels. As for the NS, a double Gaussian pdf is used for the signal in the barrel category and a single Gaussian pdf in the endcap category. The background is modeled by an exponential pdf, plus a Gaussian component for the $B^0 \rightarrow J/\psi K^{*0}$ decays, where the shape is taken from a fit to the $B^0 \rightarrow J/\psi K^{*0}$ MC simulation and the normalization is let floating in the final fit.

Table 4.5 defines the side-bands and signal regions for the CS channel.

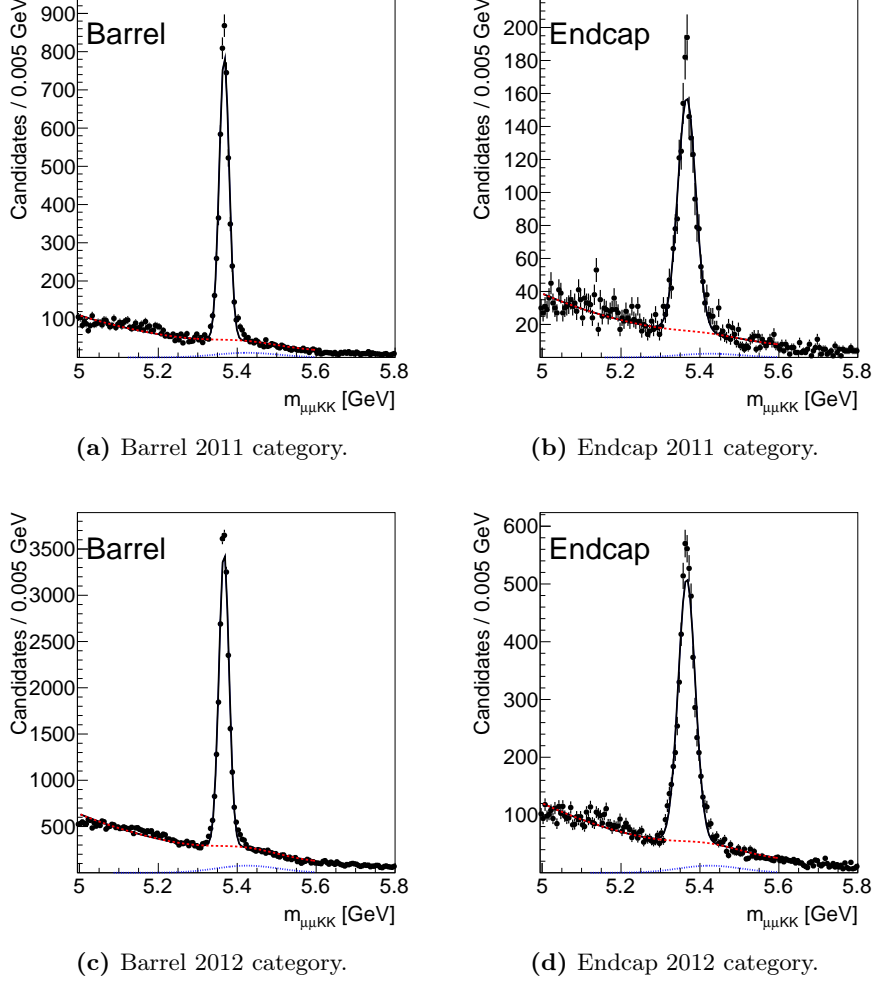


Figure 4.16: $B_s^0 \rightarrow J/\psi\phi$ invariant mass distributions for 2011 (top) and 2012 (bottom) in the barrel (left) and the endcap channel (right). The black line shows the full pdf. The red dashed line indicates the function used to describe the combinatorial background and the contribution from partially reconstructed B decays. The blue dotted line indicates the component from $B^0 \rightarrow J/\psi K^{*0}$.

Table 4.5: Invariant mass region definitions for the control channel.

Definition	Invariant mass range (GeV)
overall mass window	[4.90, 5.90]
low side-band	[5.10, 5.20]
signal window	[5.27, 5.47]
high side-band	[5.5, 5.6]

4.6 B_s^0 -signal selection with multivariate analysis

The signal variables described in Sec. 4.3.2 are used as inputs to a BDT discriminator (see also Sec. 4.3.1.1 for a brief description of the BDT technique). In this way it is possible to obtain the best separation, enhancing the significance of the signals.

4.6.1 Preselection

Before the BDT analysis, a loose preselection is required to remove outliers in the BDT input distributions, that could bias the BDT algorithm and lead to too coarse selection criteria within a specific decision tree. Furthermore, the preselection may have an influence on the MVA performance, so it is optimized to obtain the best separation. In Table 4.6 the applied preselection requirements are summarized.

4.6.2 Event sample splitting

Given the huge amount of collected background data, it is not possible to use a comparable statistics from a MC simulation to train the background sample of the BDT. Thus, the data dimuon mass side-band events are used (see Table 4.3 for the definition of the side-bands).

It is absolutely imperative to avoid any possible bias. To avoid a possible systematic effect when using background events from data, they are splitted into three distinct subsets, and three different BDT are used:

- subset-0: analyzed by BDT-0, trained on subset-1 events, tested on subset-2 events
- subset-1: analyzed by BDT-1, trained on subset-2 events, tested on subset-0 events
- subset-2: analyzed by BDT-2, trained on subset-0 events, tested on subset-1 events

The division of the event sample into different types is based on the event number, that is checked to yield to an unbiased splitting.

4.6.3 BDT results

The BDT pdf distributions for the $B_s^0 \rightarrow \mu^+ \mu^-$ signal and the side-band background are shown in Fig. 4.17. The response distributions for training and control samples are compared with a Kolmogorov-Smirnov test against a possible

Table 4.6: Preselection for the BDT training.

Variable	minimum	maximum	unit
$p_{T,B}$	5.00	9999.00	GeV
p_{T,μ_1}	4.00	999.00	GeV
p_{T,μ_2}	4.00	999.00	GeV
ℓ_{3D}	-	2.00	cm
$\ell_{3D}/\sigma(\ell_{3D})$	0.00	200.00	
$\ell_{xy}/\sigma(\ell_{xy})$	3.00	-	
$\chi^2/\text{d.o.f.}$	-	20.00	
ℓ_z	-	1.00	cm
$\ell_z/\sigma(\ell_z)$	-	5.00	
δ_{3D}	-	0.10	cm
$\delta_{3D}/\sigma(\delta_{3D})$	-	5.00	
d_{ca}^0	-	0.10	cm
α_{3D}	-	1.00	
I	0.00	-	
N_{trk}	-	21.00	
d_{ca}^{max}	-	0.25	cm

overtraining, that could be caused by too small training sub-samples. All tests are passed with high probability.

A table of the background rejection versus the signal efficiency, for each working point, is shown by colored lines in Fig. 4.18 for the three sub-sets. After the BDT training and testing steps, the BDT results from these three sub-sets are merged to obtain a combined efficiency selection, shown by black points.

This analysis is optimized for the observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay, so

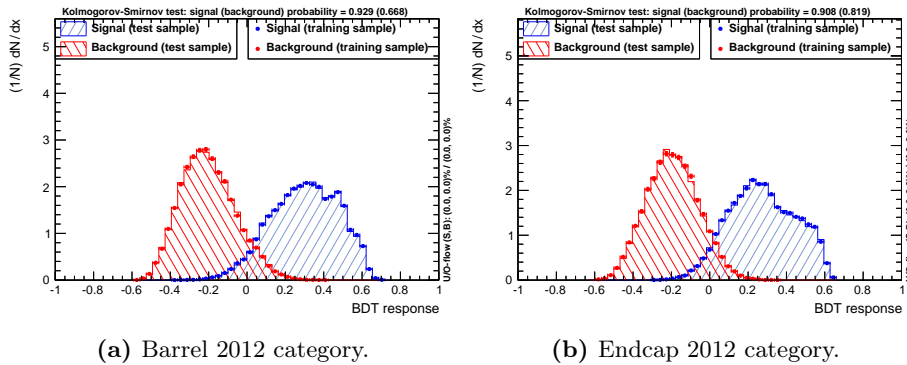
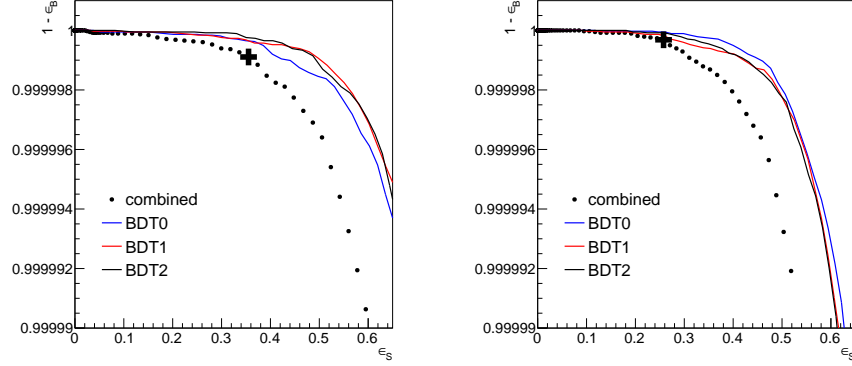


Figure 4.17: MVA over-training control plot for the type-0 2012 data categories. On top the Kolmogorov-Smirnov test results are shown.



(a) Barrel 2012 category.

(b) Endcap 2012 category.

Figure 4.18: Receiver Operating Characteristic (ROC) curves for the 2012 data categories. The solid cross shows the analysis working point.

the figure of merit for the BDT working point is the maximum significance:

$$\text{f.o.m.} = \frac{S}{\sqrt{S+B}} \quad (4.10)$$

where S is the the number of signal events and B is the number of background events. In Eq. (4.10), the expected signal is based on the MC simulation luminosity, while the expected background is determined interpolating form the side-bands. Still in Fig. 4.18, the operating point of the analysis, where the expected significance is maximized, is indicated by the solid cross.

To check against possible further biasing, a similar study is performed on a data background sample, statistically completely independent from the dimuon sample: B candidates where one or both muon candidates fail the MVA muon identification (Anti-Muon Sample (AMS)) have overall characteristics that are very similar to the real dimuon sample. This resemblance of the AMS is also true for the normalization and control samples and is due to the fact that even though the AMS muon candidates fail the very strict muon identification, most of them are nevertheless muons (having passed the default tight muon selection).

Results are shown in Fig. 4.19. The difference on the maximum between the dimuon sample optimization (solid markers) and the optimization on the AMS (histogram) is negligible.

In Fig. 4.20 the BDT distributions for the 2012 run are overlayed for NS and CS in data and MC simulations. The MC simulation samples follow smoothly the distributions in data. The final selection is applied on these two samples and

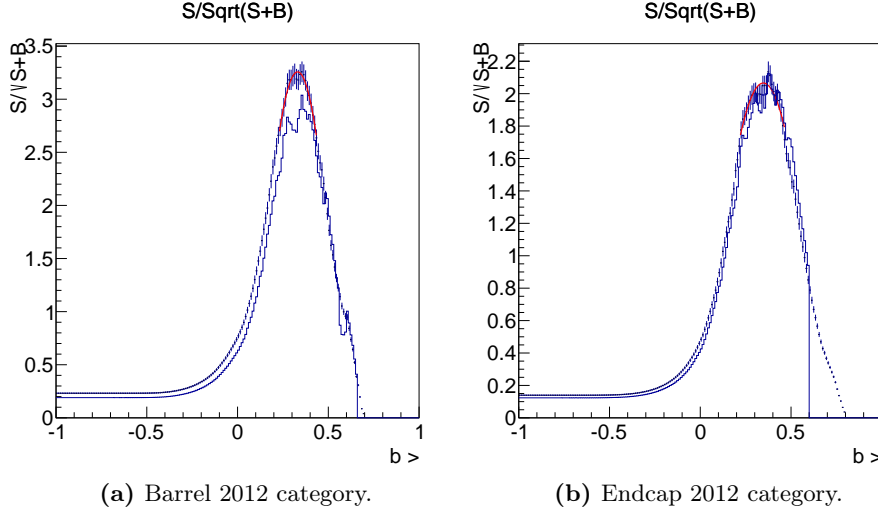


Figure 4.19: Significance versus the BDT selection in 2012 data. The solid markers show the result on the dimuon sample, while the histograms provide the result on the AMS. The red line indicates a fit with a polynomial of second degree around the maximum of the dimuon sample. The AMS is smaller than the dimuon sample and does not contain signal events in the endcap category for a BDT output higher than 0.6.

the difference in the efficiencies between data and MC is taken as systematic uncertainty on the BDT (Table 4.8 in Sec. 4.8.4 shows the numerical values of the BDT systematic uncertainties).

4.6.4 BDT pileup studies

The number of primary vertices N_{PV} is a function of the luminosity and of the total cross-section (see Sec. 3.1.3) and some variables used by this analysis could in principle depend on the pileup.

In Fig. 4.21 some of the used variables are shown, for the 2012 signal MC barrel category, after the preselections of Sec. 4.6.1 and with a BDT selection $b > 0$. Low-pileup events are defined as $N_{PV} < 6$ and high-pileup events are those with $N_{PV} > 15$. For each variable, the two distributions are compared with a binned Kolmogorov-Smirnov test. There is no visible discrepancy for any variable. The same conclusions are reached using the normalization and control MC samples, which agree well with the side-band-subtracted data distributions.

Since there is no visible pileup dependence among the input variables, also the BDT selection does not depend on it (Fig. 4.21f). This is corroborated by the study of the side-band data events: in Fig. 4.22 the mean and the RMS of the data side-band BDT distributions are shown versus the number of PVs. For the signal-like region, the same study is done after requiring a high BDT

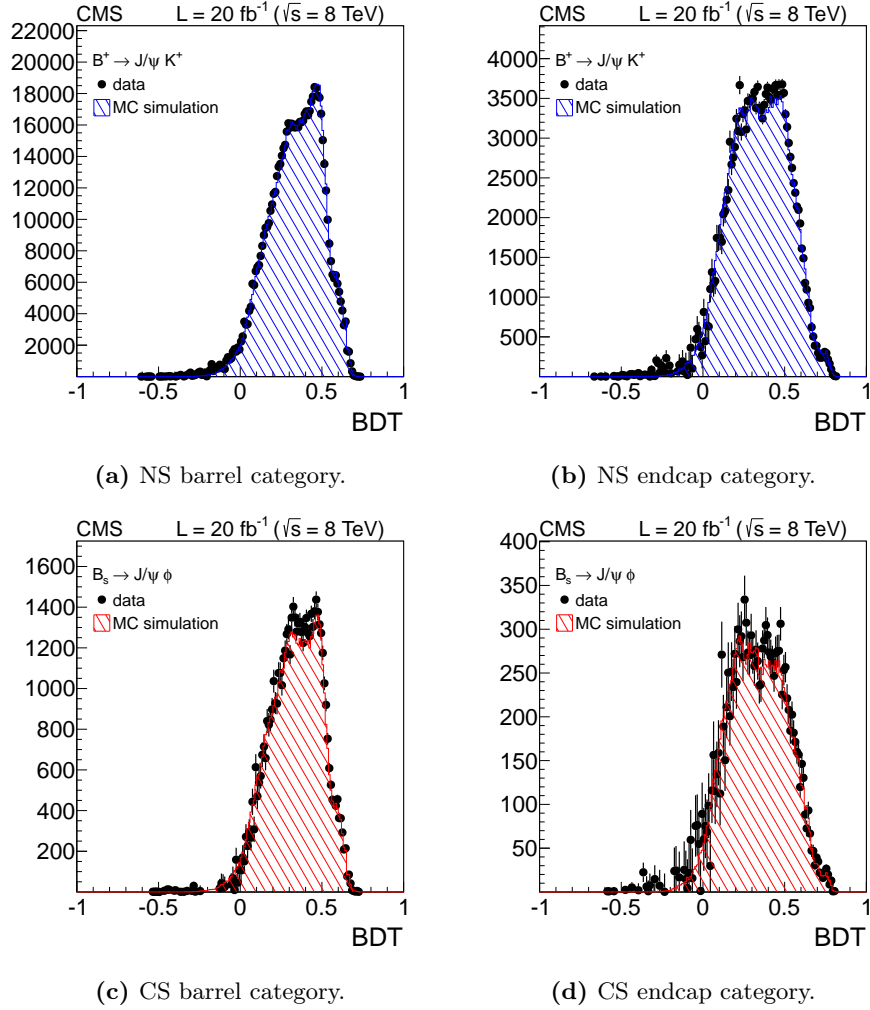


Figure 4.20: Overlay of the distributions of the BDT output b for 2012 for the normalization and control samples in the barrel and endcap categories.

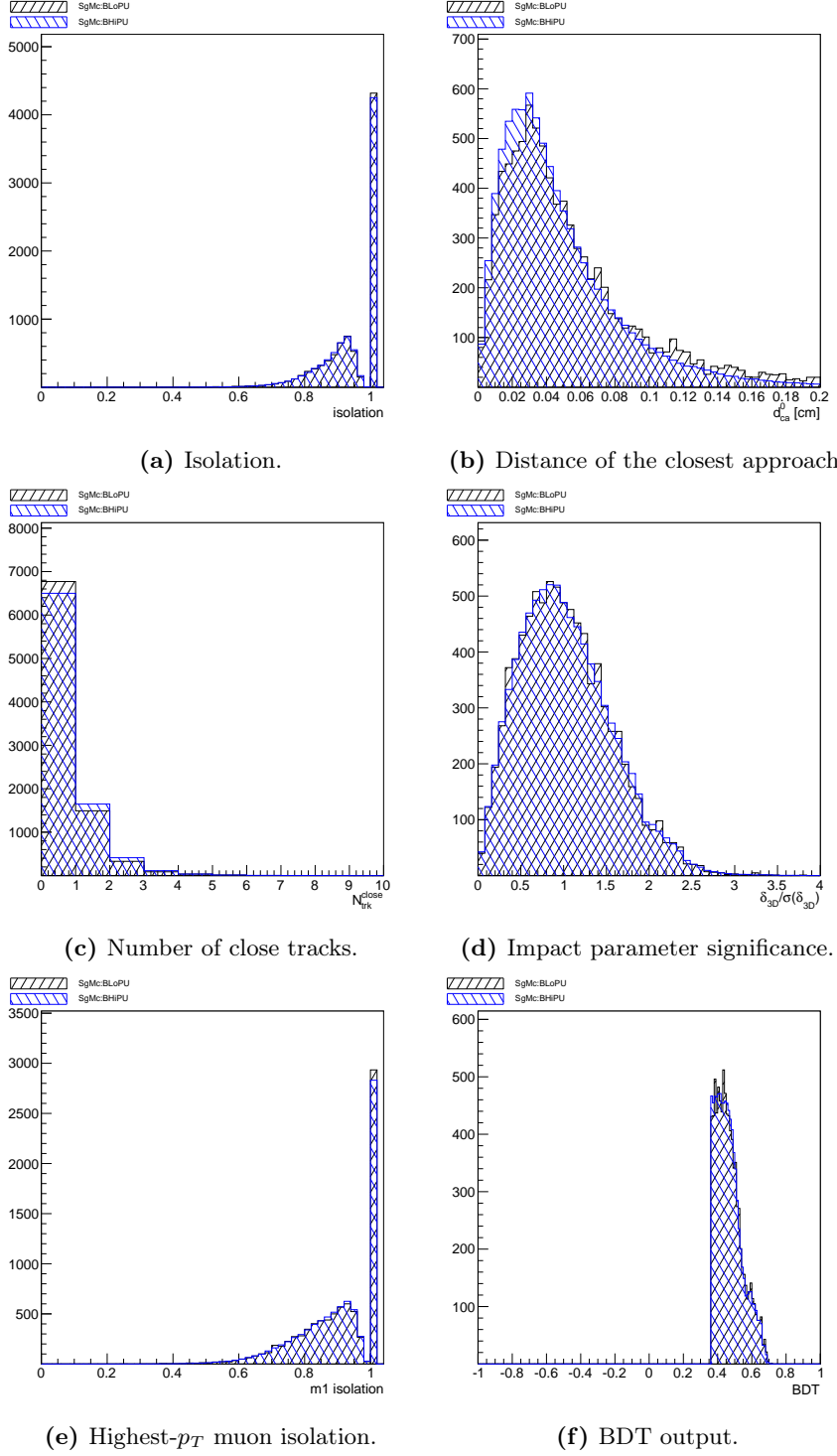


Figure 4.21: Overlay of 2012 signal MC barrel distributions in low (black histograms) and high (blue histograms) pileup events. The BDT output **(f)** is shown after the final BDT cut.

output. Additionally, the BDT efficiency on the side-band data does not present any pileup effect. Therefore, the conclusion is that this analysis does not depend on the pileup.

4.6.5 BDT mass dependence studies

An unbiased analysis requires that the BDT output does not depend on the dimuon invariant mass. This is verified in data checking whether there is a difference between the BDT distributions for the low and the high side-bands (Fig. 4.23). The BDT efficiency is very similar in the two sidebands.

Furthermore, special MC simulation samples generating $B_s^0 \rightarrow \mu^+ \mu^-$ events, with the B_s^0 mass shifted to lower and higher values, are used to control the BDT response for the signal. No mass dependence is found, therefore it is possible to conclude that the BDT analysis does not present a pileup dependence, nor a mass bias.

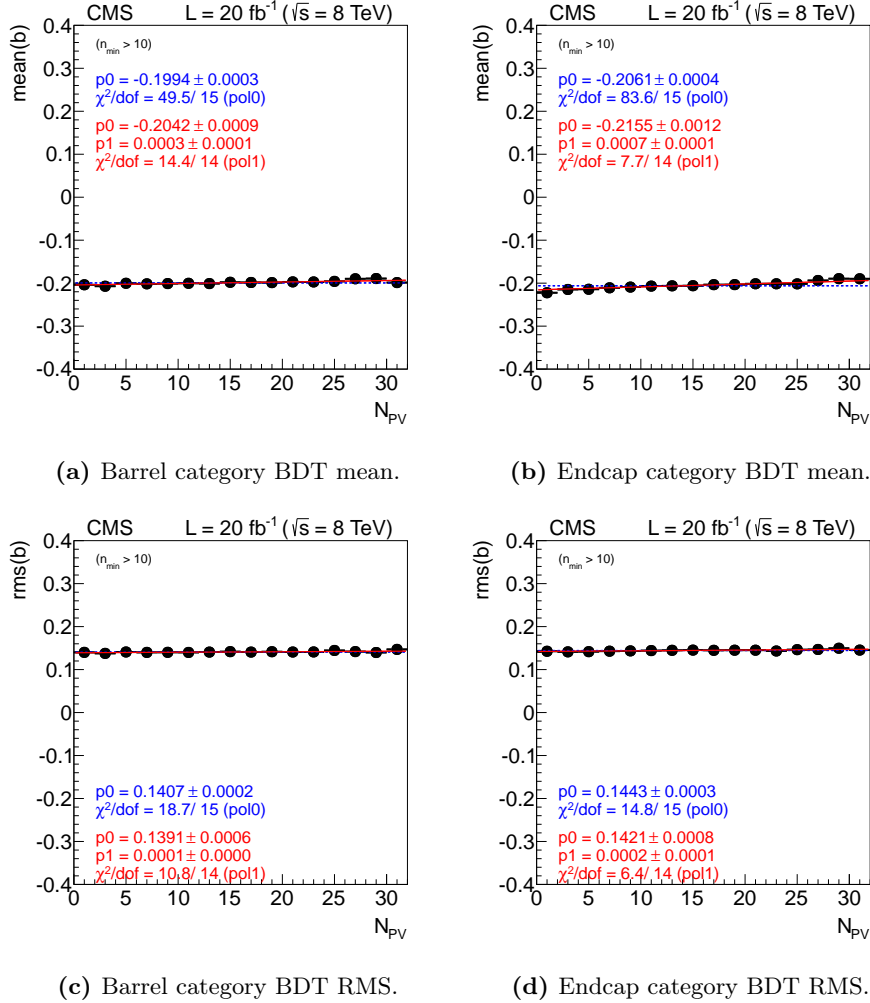


Figure 4.22: Profile histograms showing, for side-band data, the mean of the BDT output b in the barrel (a) and endcap (b) categories, and their RMSs, as a function of the number of primary vertices. The lines indicate fits with a constant and a straight line and the fit parameters are provided in the plot. No requirement on b is applied.

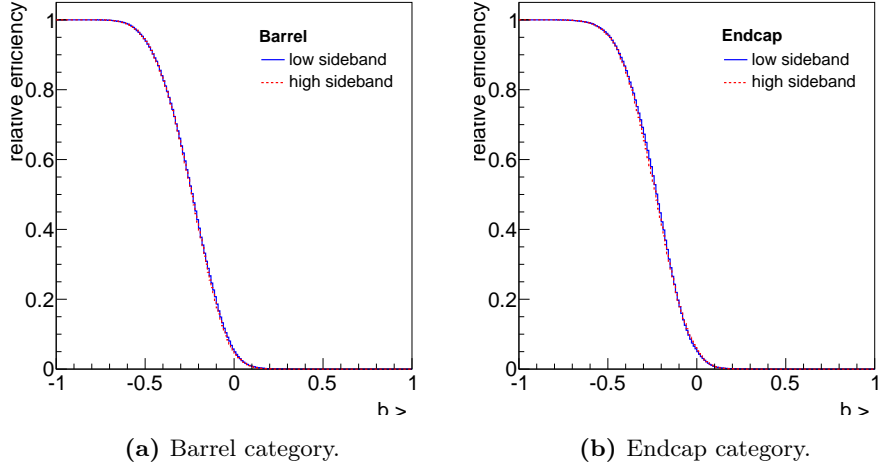


Figure 4.23: BDT efficiencies in data side-bands as a function of BDT output b_{\min} in the barrel (a) and endcap (b) for the low (blue line) and high (red line) dimuon mass side-bands.

4.7 Background determination

The combinatorial background arises from two reconstructed muons (real or misidentified hadrons), originating from separate particle decays. It is possible to estimate the combinatorial contribution under the signal peaks from the high side-band, where no resonant decays are expected. The uncertainty in the interpolation is estimated by comparing the interpolated value with the measured candidate yield in the sample failing the muon identification requirements (the AMS sample). In the Unbinned Maximum Likelihood (UML) fit, this background is parametrized with a linear function (as will be shown in Sec. 4.9.2).

The total contributions of the rare backgrounds are normalized to the reconstructed B^\pm yield in data using the following formula:

$$N(X) = \frac{\mathcal{B}(X)}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)} \frac{f_X}{f_u} \frac{\epsilon_X}{\epsilon_{B^\pm}} N_{\text{obs}}^{B^\pm} \quad (4.11)$$

where X is the given rare decay, f_X is the corresponding hadronization probability and ϵ_X is the analysis acceptance and efficiency. The resulting event yields $N(X)$ are weighted with the corresponding muon misidentification rate for kaons, pions and protons, or the muon identification efficiency, and trigger efficiency.

Table 4.7 shows the BF of the rare background decays studied in this analysis. Figure. 4.24 and Fig. 4.25 show the expected distributions of rare backgrounds in the barrel and endcap categories.

The systematic uncertainties on amount of rare (peaking and non-peaking)

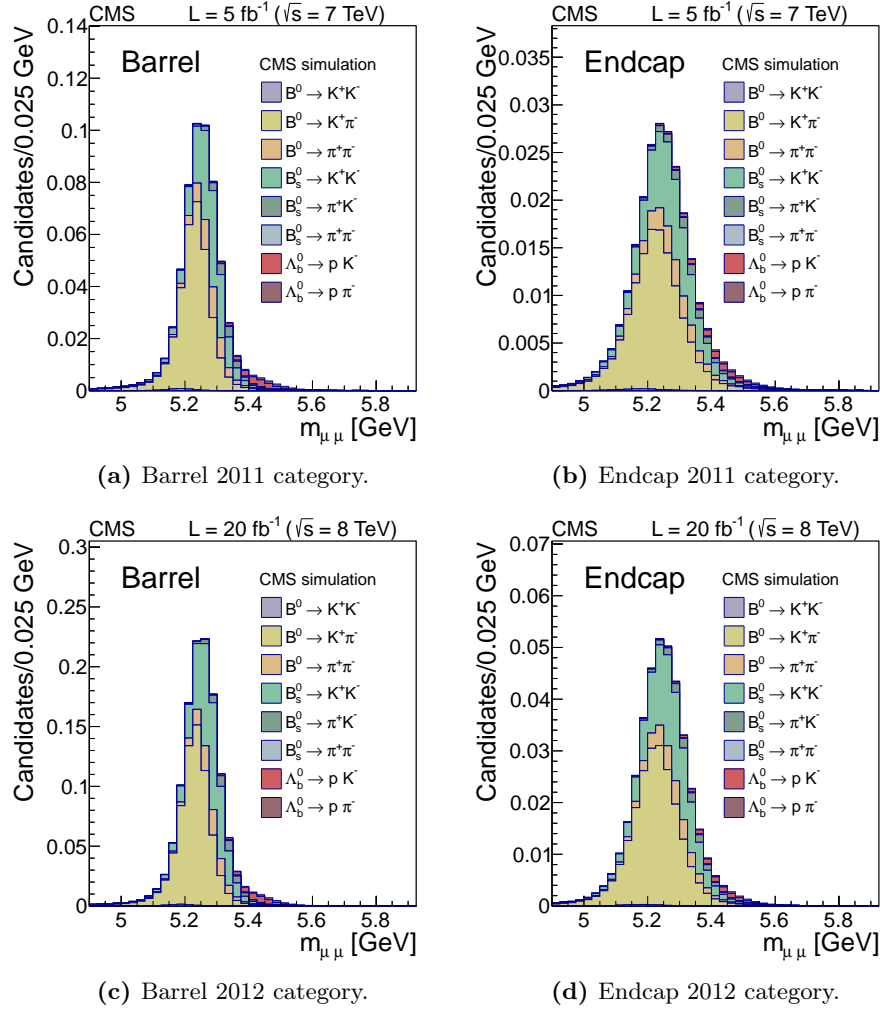


Figure 4.24: Invariant mass distributions of rare peaking backgrounds for each category.

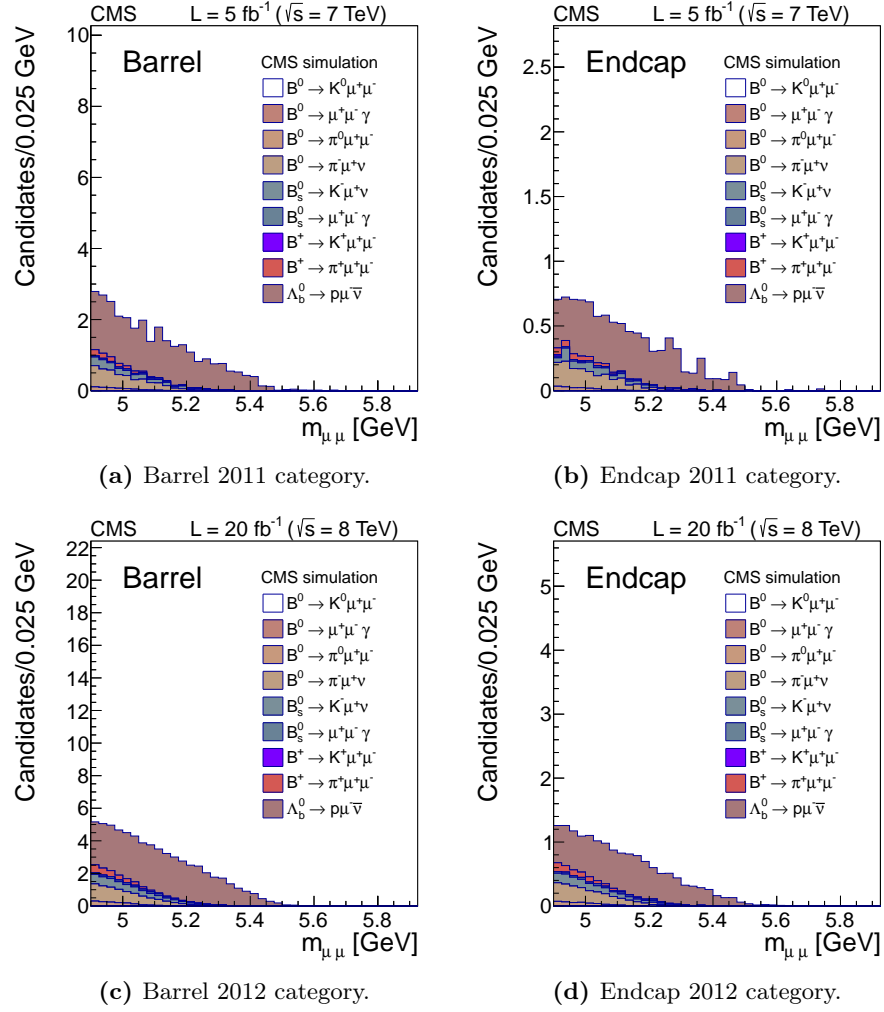


Figure 4.25: Invariant mass distributions of rare semileptonic B decays for each category.

Table 4.7: Rare branching fractions (\mathcal{B}) included in the analysis.

Decay	\mathcal{B}	Reference
$B_s^0 \rightarrow K^+ K^-$	$(2.54 \pm 0.38) \times 10^{-5}$	[51]
$B_s^0 \rightarrow K^+ \pi^-$	$(5.00 \pm 1.10) \times 10^{-6}$	[51]
$B_s^0 \rightarrow \pi^+ \pi^-$	$(7.30 \pm 1.40) \times 10^{-7}$	[51]
$B^0 \rightarrow K^+ K^-$	$(1.30 \pm 1.00) \times 10^{-7}$	[51]
$B^0 \rightarrow K^+ \pi^-$	$(1.95 \pm 0.06) \times 10^{-5}$	[51]
$B^0 \rightarrow \pi^+ \pi^-$	$(5.11 \pm 0.22) \times 10^{-6}$	[51]
$\Lambda_b^0 \rightarrow p \pi^-$	$(3.50 \pm 1.00) \times 10^{-6}$	[4]
$\Lambda_b^0 \rightarrow p K^-$	$(5.50 \pm 1.40) \times 10^{-6}$	[4]
$B_s^0 \rightarrow K^- \mu^+ \nu_\mu$	$(1.42 \pm 0.07) \times 10^{-4}$	[51]
$B^0 \rightarrow \pi^- \mu^+ \nu_\mu$	$(1.42 \pm 0.07) \times 10^{-4}$	[51]
$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$	$(6.50 \pm 6.50) \times 10^{-4}$	[137]
$B^\pm \rightarrow \mu^+ \mu^- \pi^\pm$	$(2.30 \pm 0.60) \times 10^{-8}$	[138]
$B^\pm \rightarrow \mu^+ \mu^- K^\pm$	$(4.80 \pm 0.40) \times 10^{-7}$	[4]
$B^0 \rightarrow \mu^+ \mu^- \pi^0$	$(1.20 \pm 0.60) \times 10^{-8}$	[4]
$B^0 \rightarrow \mu^+ \mu^- K^0$	$(3.80 \pm 0.80) \times 10^{-7}$	[4]
$B^0 \rightarrow \mu^+ \mu^- \gamma$	$(1.34 \pm 0.27) \times 10^{-10}$	[139]
$B_s^0 \rightarrow \mu^+ \mu^- \gamma$	$(1.89 \pm 0.38) \times 10^{-8}$	[139]

backgrounds arise from the uncertainty on the muon misidentification probability (Sec. 4.3.1) and from the branching fractions. A 50% relative uncertainty is assigned on the pion, kaon, and proton fake rates. The fake rate uncertainties of the different hadrons are treated as uncorrelated while for the same hadrons correlations are included.

The knowledge about branching fractions varies considerably: some of them are known at the few percent level, some other have not been measured, but the theory allows an informed guess, while for others ($\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}_\mu$, in particular) not even the model-based expectations agree and the resulting uncertainties are basically 100%.

4.8 Further systematic studies and cross-checks

Many possible sources of systematic or biasing effects are checked and estimated to corroborate the final results of this analysis.

4.8.1 Yield ratio for different run ranges

Various conditions can affect the data taking during the runs, like the LHC luminosity and the trigger selections. As this analysis relies on a normalization channel, these effects are mitigated. It is, nevertheless, important to check the data acquisition as two different triggers are used for the signal and the NS and

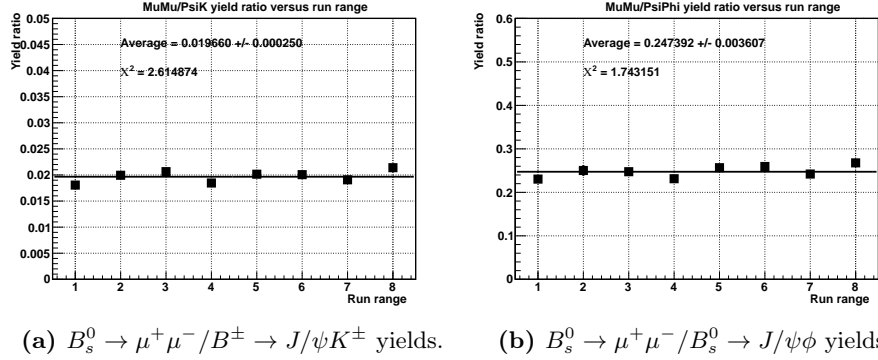


Figure 4.26: Ratios of signal trigger yields divided by $B^\pm \rightarrow J/\psi K^\pm$ yields (a), and ratio of signal trigger yields divided by $B_s^0 \rightarrow J/\psi \phi$ yields (b), for different runs. A constant line is fit and a χ^2 test result is superimposed in each plot.

CS samples. In order to check this, data are split into six periods of roughly equal luminosity.

The signal trigger yield divided by the normalization and control sample yields are shown in Fig. 4.26, where a 0-order polynomial fit included. As the fit χ^2 test is low and the average has a low error, these ratios can be assumed as constant.

4.8.2 Acceptance and efficiency systematics

The uncertainty on the acceptance and efficiency of the signal has contributions from the detector response (track reconstruction, muon identification, trigger efficiency) and from the uncertainty on the production mechanisms that influence the phase space distributions of the muons, and thus the overall geometrical acceptance. To first order these effects cancel out as the signal is normalized to the NS sample. The residual effects are studied in this section.

4.8.2.1 Track reconstruction, muon identification and trigger efficiency uncertainties

The analysis relies on MC simulation for the evaluation of the efficiencies. The single muon identification and trigger efficiencies can be determined with the TNP method directly on data with a J/ψ sample. To determine a possible systematic effect in the MC samples, the efficiency ratio $\epsilon_{B^\pm} / \epsilon_{B_s^0}$ is determined with the TNP approach in data and in MC simulation. The difference between the two samples is taken as systematic error on the simulation of the detector identification and trigger efficiencies. Results confirm the good behavior of the MC simulation, as the efficiency ratio difference for the muon identification is

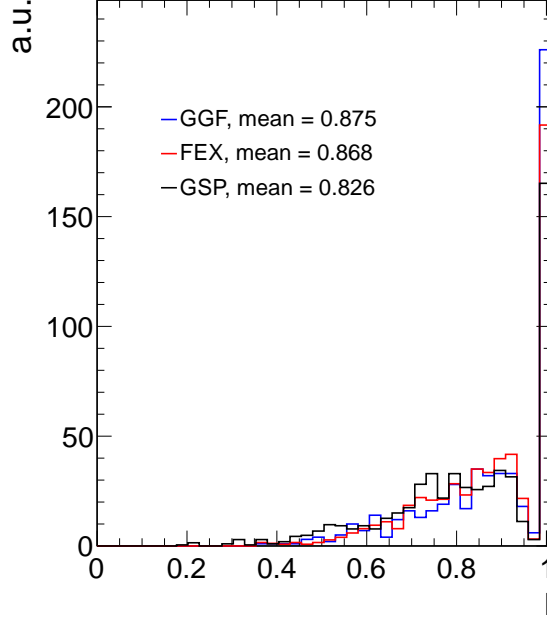


Figure 4.27: Comparison of the isolation distribution divided into the three different QCD $b\bar{b}$ production processes.

below 2%, while the trigger ratios are at most 3% apart.

The tracking efficiency uncertainty affects the signal track reconstruction and isolation efficiencies, and it is evaluated for isolated muons and hadrons in [122], using MC simulation, the TNP method and the decay ratio $D^0 \rightarrow K^-\pi^+\pi^-\pi^+/D^0 \rightarrow K^-\pi^+$. Following the cited paper, the total uncertainty on hadron tracking efficiency is estimated to be 4%, dominated by the statistical uncertainty of the method.

4.8.2.2 Production mechanisms

The three main QCD $b\bar{b}$ production mechanisms (PC, FE and GS) present different distributions in p_T and $\Delta\phi$ (see also Sec. 2.7.4). It is important to validate the MC simulation, as the geometrical acceptance and isolation variables depend on them (Fig. 4.27).

Measurements of the $b\bar{b}$ pair angular correlations have been done in CMS, for instance in [140], where data exhibit a sizable fraction of the $b\bar{b}$ pairs produced at small opening angles. Furthermore, it is shown that difference of the cross-section ratios for the regions $\Delta R < 0.8$ and $\Delta R > 2.4$ increases to 60% in data with respect to PYTHIA simulation.

An estimation on the impact of the different production mechanisms is

performed studying the acceptance as a function of the PYTHIA $b\bar{b}$ production η range, for each of the three production mechanisms. The maximum variation of the acceptance ratio between signal and normalization is taken for an estimation of the systematic uncertainty, 3.5% for the barrel, and 5.0% for the endcap categories.

To investigate whether the data indicate any significant discrepancy of the MC simulation production process mixture, variables that could have sensitivity to the mixture can be studied: gluon splitting events might have more activity around the reconstructed B candidate and less activity outside a cone around the B candidate; instead, gluon-gluon fusion events should have a larger activity back-to-back in the transverse plane.

This is measured by summing over track transverse momenta and by searching for other muons from semileptonic B -decays. In particular the studied “activity” observables are:

- the radius between the B candidate and a third muon, $\Delta R(C, \mu^{\text{OS}})$;
- the transverse momentum of a third muon, $p_{T,\mu}^{\text{OS}}$;
- the relative outside isolation $I_{\text{OS}}^{\text{rel}} = F_{\text{OS}}/p_T$, where F_{OS} is the sum of the scalar transverse momentum of all tracks with $p_T > 0.9 \text{ GeV}$ outside a cone of $R = 1.0$ around the B candidate, where the tracks must be associated either to the B candidate PV or not associated to any PV.

Their distributions on data and on MC simulations are shown in Fig. 4.28 for the normalization and control samples. It is possible also to try to measure the single contributions of the three production mechanisms, fitting data using the sum of the production pdfs, as taken from the MC simulations, and extracting their relative contributions. A reasonable agreement between data and MC simulations is found in all these cases, thus there are no strong reasons to question the production process mixture in the MC simulation.

4.8.2.3 Mass scale and resolution uncertainties

The imperfect knowledge of the mass scale determines a bias that can be estimated comparing dimuon resonances in data and MC simulation. The $B_{(s)}^0$ candidates are in an invariant mass region between the charmonia and bottomia mesons, so it is natural to control a possible systematic effect in the mass scale looking at the J/ψ , $\psi(2S)$ as well as the $\Upsilon(nS)$ ($n = 1, 2, 3$) mesons.

Figure 4.29a shows the shifts on the peak positions and the linearly interpolated peak shift for the $B_{(s)}^0$ meson. At the $B_{(s)}^0$ mass scale, the observed mass shift in data is less than 8 MeV in the barrel category and less than 12 MeV in

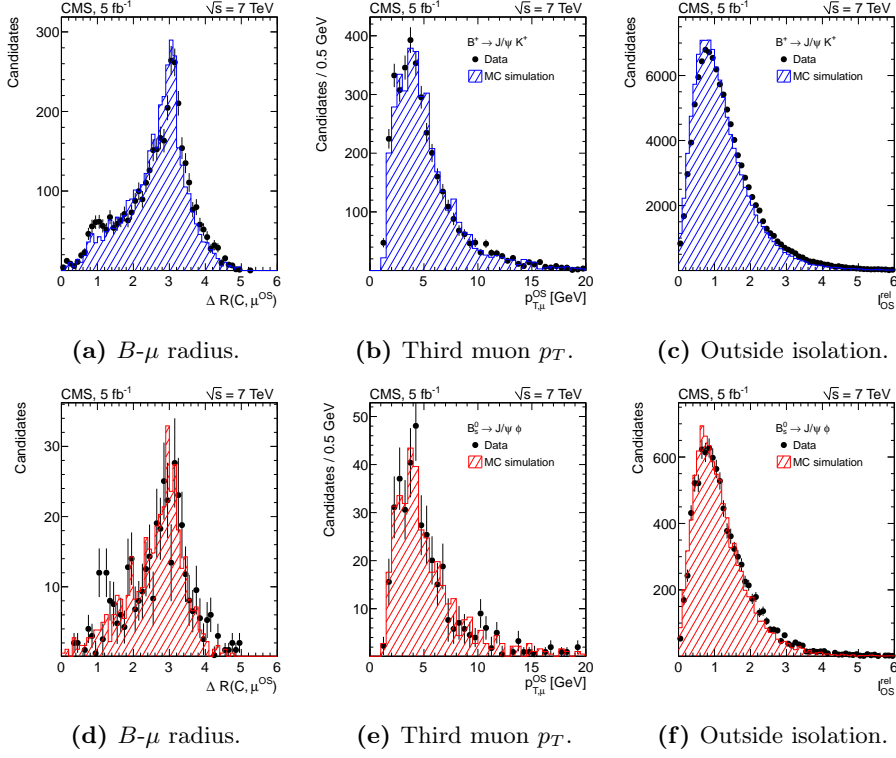


Figure 4.28: Activity distributions in data and MC simulations for the normalization (top) and control (bottom) samples. For their definition, see Sec. 4.8.2.2.

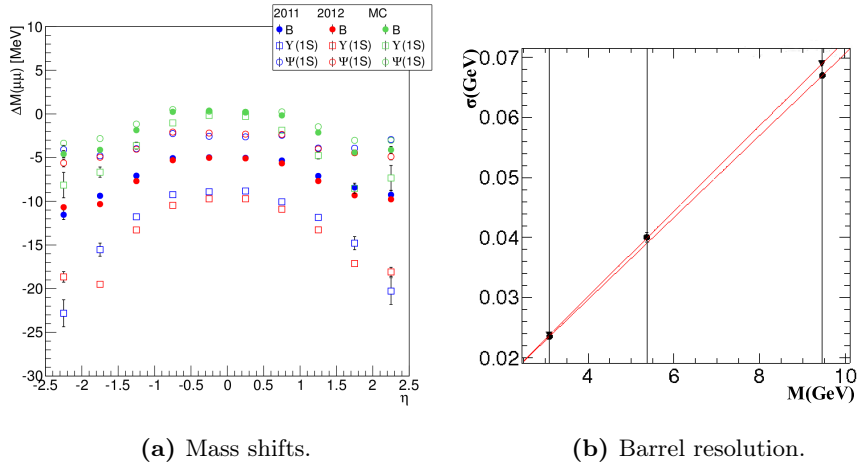


Figure 4.29: Shifts on the peak positions versus η (a). The empty points show the measured shifts on the J/ψ and $\Upsilon(1S)$ mesons, while the full circles show the interpolated (blue or red) or simulated (green) shifts at the $B_{(s)}^0$ mass value. Mass resolution in data (triangle) and MC simulations (circle) for different dimuon invariant masses, in the barrel category (b).

Table 4.8: Summary of the relative systematic uncertainties. All uncertainties are provided as percentages. For rows where two numbers are give the first one corresponds to the 2011 dataset and the second to 2012. When the two errors are assumed to be the same, only one number is given.

Category	Barrel (%)	Endcap (%)
f_s/f_u : production ratio of u and s quarks	8.0	8.0
acceptance: production processes	3.5	5.0
mass scale and resolution	5.0	5.0
efficiency (signal): data/MC simulation	9.5 - 3.3	7.9 - 2.3
efficiency (normalization): data/MC simulation	0.5 - 2.3	0.5 - 1.1
efficiency (normalization): kaon track efficiency	4.0	4.0
efficiency trigger	3.0	3.0
efficiency muon identification	2.0	2.0
normalization: fit on the pdf	5.0	5.0

the endcap category. In the MC simulation the mass shift is less than 2 MeV in the barrel category and less than 5 MeV in the endcap one.

The mass resolution depends strongly on the track pseudorapidity, due to several factors like the amount of material traversed by the track, the pixel detector hit resolution, the number and location of the pixels, the lever arm. The resolution in the MC simulations is studied against data for the charmonia and bottomia resonances and also on the NS and CS samples. MC simulations reproduce all these features reasonably well (Fig. 4.29b).

4.8.3 Production ratio f_s/f_u

The hadronization ratio f_s/f_u is used as measured by the LHCb collaboration [132], where a small p_T dependence is found. To test for possible pseudorapidity or transverse momentum dependence in the CMS acceptance, a differential measurement of the ratio $B^\pm \rightarrow J/\psi K^\pm$ over $B_s^0 \rightarrow J/\psi \phi$ is performed in p_T and η bins. In each bin, the yields of normalization and control sample events are fit and corrected by the bin efficiencies. From the normalization and control sample branching fractions, a value for f_s/f_u is computed for each bin. Results are found to be consistent with flat across the p_T bins for both barrel and endcap categories, and a conservative systematic uncertainty of 5% is assigned.

4.8.4 Summary of systematic uncertainties

In Table 4.8 the systematic uncertainties of the analysis are summarized.

4.9 Statistical models

As this is a "blind" analysis, the signal box is uncovered only after freezing all selections. The following section describes the statistical models used to extract the final results on the "unblinded" data.

4.9.1 Binned analysis

In the binned analysis the invariant mass range $[4.9, 5.9]$ GeV (see also Table 4.3) is divided into the $B_s^0 \rightarrow \mu^+ \mu^-$ signal window $[5.3, 5.45]$ GeV, the $B^0 \rightarrow \mu^+ \mu^-$ signal window $[5.2, 5.3]$ GeV and the background region, consisting of the complement of the above.

The number of entries in each region is considered a random variable satisfying Poissonian statistics: let us call, for the barrel category, N_s^B the number of candidates in the B_s^0 signal window, N_d^B the number of candidates in the B^0 signal window, and N_b^B the number of candidates in the background window. Analogously, variables N_s^E, N_d^E, N_b^E are defined for the events in the endcap category. For simplicity, all considerations below are written for the barrel category, although they are also equally valid for the endcap one.

The background observable is modeled as

$$N_b^B \sim \text{Pois}(\nu_b^B + \nu_{b,\text{rare}}^B) \quad (4.12)$$

where the expected number of combinatorial background events ν_b^B enters the model as a nuisance parameter and is determined from the side-bands; $\nu_{b,\text{rare}}^B$ is the expected number of rare (peaking and semileptonic) background events.

The ratio of combinatorial background in the signal windows with respect to the side-band window is given by the scale factors τ_s^B and τ_d^B . In total, the Poissonian for the signal windows are of the following form

$$N_s^B \sim \text{Pois}(\tau_s^B \nu_b^B + \nu_{s,\text{rare}}^B + P_{ss}^B \mu_s \nu_s^B + P_{sd}^B \mu_d \nu_d^B) \quad (4.13)$$

$$N_d^B \sim \text{Pois}(\tau_d^B \nu_b^B + \nu_{d,\text{rare}}^B + P_{ds}^B \mu_s \nu_s^B + P_{dd}^B \mu_d \nu_d^B) \quad (4.14)$$

The following additional notation is used for $i, j \in \{s, d\}$: P_{ij}^B is the probability for a reconstructed $B_j^0 \rightarrow \mu\mu$ decay to be in the B_i^0 signal window; μ_i is the signal strength of $B_i^0 \rightarrow \mu\mu$, that is the ratio of true branching fraction to SM branching fraction.

The expected number of reconstructed decays ν_i , assuming the SM hypothesis

$\mathcal{B}_{\text{SM}}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, is

$$\nu_i = \frac{\mathcal{B}_{\text{SM}}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)} \frac{f_s}{f_u} \frac{\epsilon_{B_s^0}}{\epsilon_{B^\pm}} N_{\text{obs}}^{B^\pm} \quad (4.15)$$

where all variables are the same as for Eq. (4.1).

The model, thus, consists of six Poissonian observables (N_s^E , N_s^B , N_d^E , N_d^B , N_b^E , N_b^B).

The constraints for the combinatorial background and cross-feed is given by the Poissonians. All other constraints of the nuisance parameters are implemented as bifurcated Gaussians. The total likelihood function for each year is given by

$$L(N_{s,d,b}^{B,E}; \mu_s, \mu_d, \nu) = \left(\prod_{\substack{k \in \{s,d,b\} \\ \ell \in \{B,E\}}} \text{Pois}(N_k^\ell) \right) \times \left(\prod_k \text{BifurGauss}(\nu_k) \right) \quad (4.16)$$

where ν stands for all nuisance parameters entering the model.

The total likelihood function combining the 2011 and 2012 measurements is simply given by the product of twice the function above:

$$L_{\text{comb}} = L_{2011} L_{2012} \quad (4.17)$$

4.9.1.1 Upper limit

The upper limit is computed using the CL_S method [141]. When computing the upper limit for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay, the branching fraction of $B^0 \rightarrow \mu^+ \mu^-$ is assumed unknown and enters the computation as a nuisance parameter as well. When computing the upper limit for the $B^0 \rightarrow \mu^+ \mu^-$ decay, it is the $B_s^0 \rightarrow \mu^+ \mu^-$ decay that is treated as an additional nuisance parameter.

To test whether a fixed signal strength μ_s is in the confidence interval, the ratio of profiled likelihoods test statistic is used:

$$q_{\text{ul}} = \frac{L(N_{s,d,b}^{B,E}; \mu_s, \hat{\hat{\nu}}(\mu_s))}{L(N_{s,d,b}^{B,E}; 0, \hat{\hat{\nu}}(0))}, \quad (4.18)$$

where the double hat means the conditional maximum likelihood estimate for a given fixed parameter of interest. When evaluating CL_b , the signal strength is set to zero.

Toys are generated randomizing all nuisance parameters according to their constraint function (Frequentist-Bayesian Hybrid approach). For each toy experiment, the observed value in the side-band is kept fix at the actual observed value.

Table 4.9: Predicted number of SM signal events $N_{\text{signal}}^{\text{exp}}$ and expected number of signal and background events $N_{\text{total}}^{\text{exp}}$ for the four categories. The event numbers refer to the B^0 and B_s^0 signal regions, respectively.

Category		$N_{\text{signal}}^{\text{exp}}$	$N_{\text{total}}^{\text{exp}}$
2011	B^0 Barrel	0.27 ± 0.03	1.3 ± 0.8
	B_s^0 Barrel	2.97 ± 0.44	3.6 ± 0.6
	B^0 Endcap	0.11 ± 0.01	1.5 ± 0.6
	B_s^0 Endcap	1.28 ± 0.19	2.6 ± 0.5
2012	B^0 Barrel	1.00 ± 0.10	7.9 ± 3.0
	B_s^0 Barrel	11.46 ± 1.72	17.9 ± 2.8
	B^0 Endcap	0.30 ± 0.03	2.2 ± 0.8
	B_s^0 Endcap	3.56 ± 0.53	5.1 ± 0.7

In Table 4.9 the expected yields are shown for all categories. The expected upper limits for the $B_s^0 \rightarrow \mu^+ \mu^-$ BF are:

$$5.9_{-1.6}^{+1.7} \times 10^{-9} \quad \text{SM} \quad (4.19)$$

$$1.8_{-0.5}^{+0.8} \times 10^{-9} \quad \text{Background only} \quad (4.20)$$

Instead, the expected upper limits for $B^0 \rightarrow \mu^+ \mu^-$ are:

$$6.3_{-2.0}^{+2.7} \times 10^{-10} \quad \text{SM} \quad (4.21)$$

$$5.4_{-1.6}^{+2.3} \times 10^{-10} \quad \text{Background only} \quad (4.22)$$

4.9.2 Unbinned maximum likelihood fit analysis

There are four contributions to the total mass yield: the two signals yields, $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, the combinatorial background and the rare background, subdivided into the semileptonic and the peaking contributions.

The shape parameters of every contribution, but the combinatorial one, are fixed to MC simulation events. The number of events of each non-rare contribution is let floating, to be fitted to the unblinded event yield, while the rare contribution yields are constrained, as nuisance parameters, to the expectations.

Thus, the mass likelihood of each independent category is:

$$L = N_{B_s^0} F_{B_s^0} + N_{B^0} F_{B^0} + N_{\text{comb}} F_{\text{comb}} + N_{\text{peak}} F_{\text{peak}} + N_{\text{semi}} F_{\text{semi}} \quad (4.23)$$

where N_i is the number of events and F_i is the pdf, for each contribution i .

A Per-Event Error (PEE) pdf is implemented: the width parameters of the B_s^0 and B^0 mesons are not constant, but rather they are taken from each event,

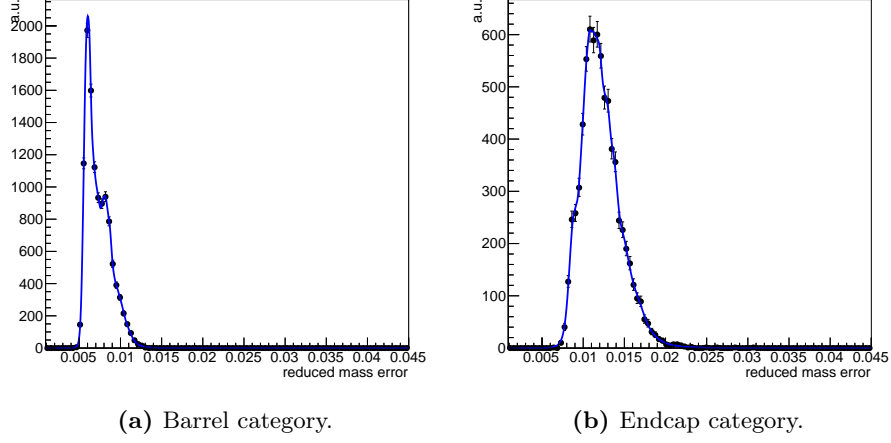


Figure 4.30: Examples of reduced mass error distributions for the barrel (a) and endcap (b) categories. The pdf (blue line) is plotted over the B_s^0 distribution (black points).

from the candidate mass error $\sigma(\eta)$. To avoid any dependence between $\sigma(\eta)$ and the mass of the candidate, since the detector resolution has a linear dependence with respect to the dimuon invariant mass, we divide the mass error by the event mass value, obtaining a “reduced” mass error: $\sigma_r(\eta) \equiv \sigma(\eta)/m$.

The total pdf for each signal is a single Crystal Ball pdf [134], multiplied by the probability distribution of its reduced mass resolution M^R [142]:

$$F_{B_s^0} = CB(m|\mu_{B_s^0}, k_{B_s^0} \times \sigma(\eta), \alpha_{B_s^0}, n_{B_s^0}) \times M_{B_s^0}^R(\sigma_r(\eta)) \quad (4.24)$$

for the B_s^0 meson, and similarly for the B^0 meson:

$$F_{B^0} = CB(m|\mu_{B^0}, k_{B^0} \times \sigma(\eta), \alpha_{B^0}, n_{B^0}) \times M_{B^0}^R(\sigma_r(\eta)) \quad (4.25)$$

The parameter k_i is a global scale factor that takes into account a possible improper error estimate. The reduced mass error distribution is fitted with a kernel estimation pdf [143] that is a superposition of Gaussian kernels, one for each data point (see Fig. 4.30 for examples of reduced mass error distributions). Figure 4.31 shows examples of the fitted invariant mass distribution for the B_s^0 and B^0 mesons.

For the combinatorial background a polynomial with one free parameter is used:

$$F_{\text{comb}} = P(1)(m|C) \times M_{\text{comb}}^R(\sigma_r(\eta)) \quad (4.26)$$

The parameter C is let floating in the final fit. This functional form is validated on independent data samples. The side-bands in the near-low BDT region,

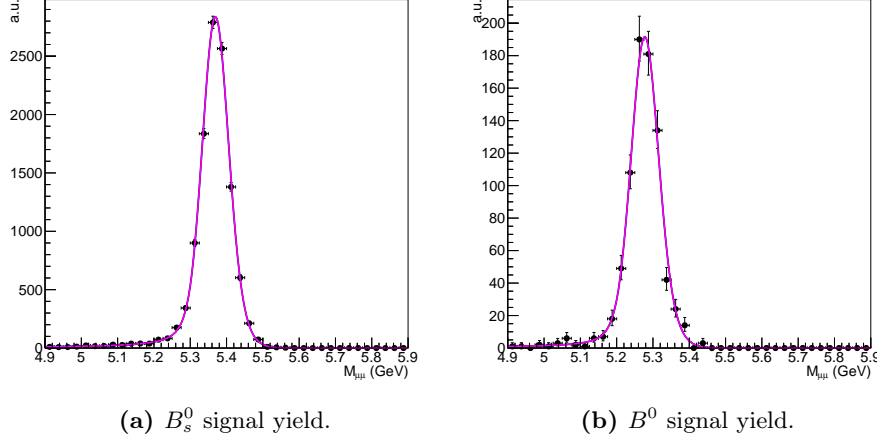


Figure 4.31: Examples of invariant mass distributions for the B_s^0 (a) and B^0 (b) signals. The pdf (purple line) is plotted over the B_s^0 and B^0 distributions (black points).

which are not used by the fit, and the AMS sample, that contains tight dimuon candidates, that are very similar to the signal candidates, confirm the form of this background.

The peaking background is fitted with the sum of a Gaussian and a Crystal Ball pdfs sharing the same mean. The addition of the Gaussian pdf is due to the fact that the peaking background is made by different contributions ($B_s^0 \rightarrow K^+K^-$, $B^0 \rightarrow \pi^+\pi^-$, $B^0 \rightarrow K^+\pi^-$, ...), whose peaks have different values because of the different meson masses and also because of the wrong muon mass assignment to the charged tracks.

The semileptonic background has a decreasing shape, that is parametrized with the Gaussian kernels method.

Therefore, the rare background pdfs are

$$F_{\text{peak}} = (f_{\text{peak}}^G G(m|\mu_{\text{peak}}, \sigma_{\text{peak}}^G) + (1 - f_{\text{peak}}^G) CB(m|\mu_{\text{peak}}, \sigma_{\text{peak}}^{CB}, \alpha_{\text{peak}}, n_{\text{peak}})) \times M_{\text{peak}}^R(\sigma_r(\eta)) \quad (4.27)$$

$$F_{\text{semi}} = G_k(m) \times M_{\text{semi}}^R(\sigma_r(\eta)) \quad (4.28)$$

where f_{peak}^G is the fraction of the Gaussian contribution to the peaking background.

The rare background distributions and the pdfs are shown in Fig. 4.32. The numbers N_{peak} and N_{semi} of peaking and semileptonic events are constrained to the normalization decay $B^\pm \rightarrow J/\psi K^\pm$ (see Sec. 4.9.2.2).

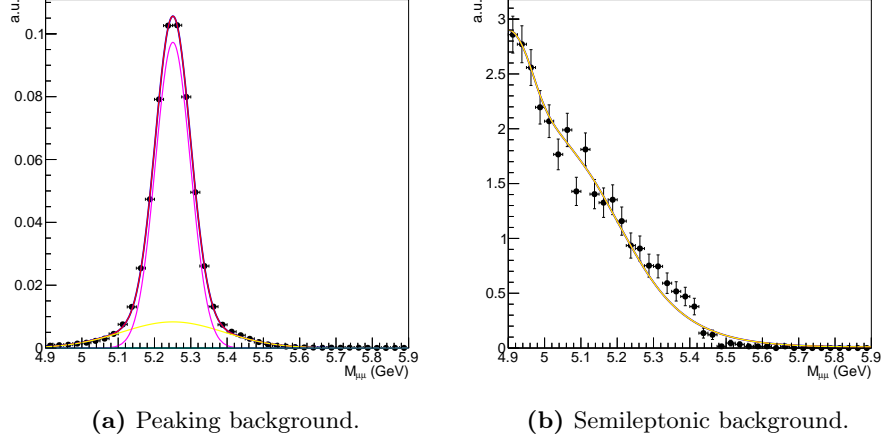


Figure 4.32: Examples of invariant mass distributions for the peaking (a) and semileptonic (b) backgrounds. Black points are from MC simulations. For the peaking background the total pdf (red line) is shown, together with the Crystal Ball (purple line) and the Gaussian (yellow line) contributions. For the semileptonic background the total pdf (yellow line) is shown.

4.9.2.1 The mass likelihood

There are four statistically independent categories (the two $|\eta|$ regions for the two data-taking years), therefore, the mass pdf is the product of four likelihoods of Eq. (4.23) type:

$$L_{\text{tot}} = \prod_{i=0}^3 L_i \quad (4.29)$$

The signal yields $N_{B_s^0,i}$ are linear functions of the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$: $N_{B_s^0,i} = K_{B_s^0,i} \times \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$, where the constant $K_{B_s^0,i}$ is

$$K_{B_s^0,i} = (N_{B^\pm,i}) \times \left(\frac{f_s}{f_u} \right) \times \left(\frac{\epsilon_{B_s^0,i}}{\epsilon_{B^\pm,i}} \right) \times \left(\frac{1}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)} \right) \quad (4.30)$$

The same is true for the B^0 signal, but without the f_s/f_u ratio:

$$K_{B^0,i} = (N_{B^\pm,i}) \times \left(\frac{\epsilon_{B^0,i}}{\epsilon_{B^\pm,i}} \right) \times \left(\frac{1}{\mathcal{B}(B^\pm \rightarrow J/\psi K^\pm) \times \mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)} \right) \quad (4.31)$$

Thus, Eq. 4.23, for each category i , can be put into the following form:

$$L_i = \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) K_{B_s^0,i} F_{B_s^0,i} + \mathcal{B}(B^0 \rightarrow \mu^+\mu^-) K_{B^0,i} F_{B^0,i} + N_{\text{comb},i} F_{\text{comb},i} + N_{\text{semi},i} F_{\text{semi},i} + N_{\text{peak},i} F_{\text{peak},i} \quad (4.32)$$

Table 4.10: Categorized BDT boundaries for each category.

Category	Boundaries
barrel 2011	0.1, 0.31, 1
endcap 2011	0.1, 0.26, 1
barrel 2012	0.1, 0.23, 0.33, 0.44, 1
endcap 2012	0.1, 0.22, 0.33, 0.45, 1

4.9.2.2 Constraints

The constants $K_{j,i}$, shown in the previous section, are products of parameters P_l (one for each parenthesis) that have total uncertainties σ_l . We add these systematic errors as Gaussian nuisance parameters to the likelihood. The Gaussian parameters are constant: the mean $P_{0,l}$ is equal to the estimator and the sigma parameter is set to its uncertainty. Thus for each parameter P_l of Eq. (4.30) and Eq. (4.31):

$$P \sim G(P_l | P_{0,l}, \sigma_l) \quad (4.33)$$

Eq. (4.33) is valid also for the estimations of the rare background yields: N_{peak} is constraint with a Gaussian and N_{peak} is constrained with a Log-Normal pdf. The total mass likelihood becomes:

$$L_{\text{tot}} = \prod_{i=0}^3 L_i L_i^{\text{constr}} \quad (4.34)$$

4.9.2.3 The categorized BDT method

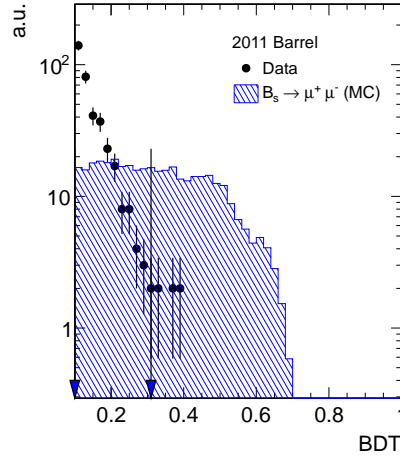
To take full information from the BDT analysis, the invariant mass distributions are sub-divided in bins of the BDT distribution (obtaining the so-called “categorized BDT” method). Eq. (4.34) is still valid, but the number of terms grows as a function of the number of BDT bins. The binning, illustrated in Fig. 4.33, is chosen to have the same event yield in each bin. Taking into account the different integrated luminosities between the two years of data-taking, twelve bins are obtained (Table 4.10). The region below the BDT discriminator $b < 0.1$ is completely dominated by the combinatorial background.

The final pdf is studied with MC toy experiments, and shows no significant bias on the yield evaluations.

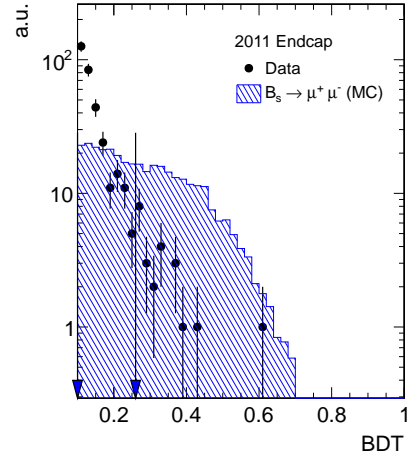
4.9.2.4 Significance calculation

For the B_s^0 significance, the null and the alternative hypotheses are the following:

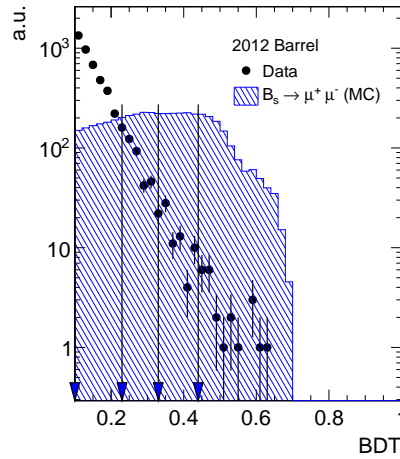
- **Null Hypothesis** \mathcal{L}_0 : $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$.



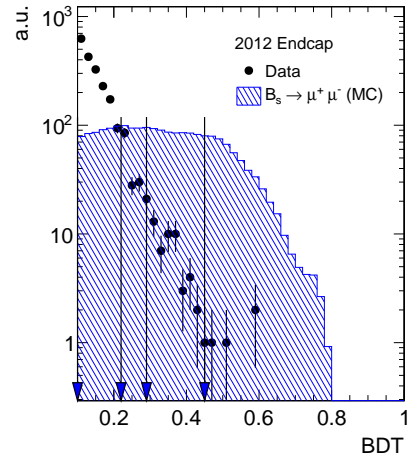
(a) Barrel 2011 category.



(b) Endcap 2011 category.



(c) Barrel 2012 category.



(d) Endcap 2012 category.

Figure 4.33: Illustration of the BDT categories in 2011 and 2012 with data events with $4.9 < m < 5.9$ GeV and signal MC events, normalized to the data.

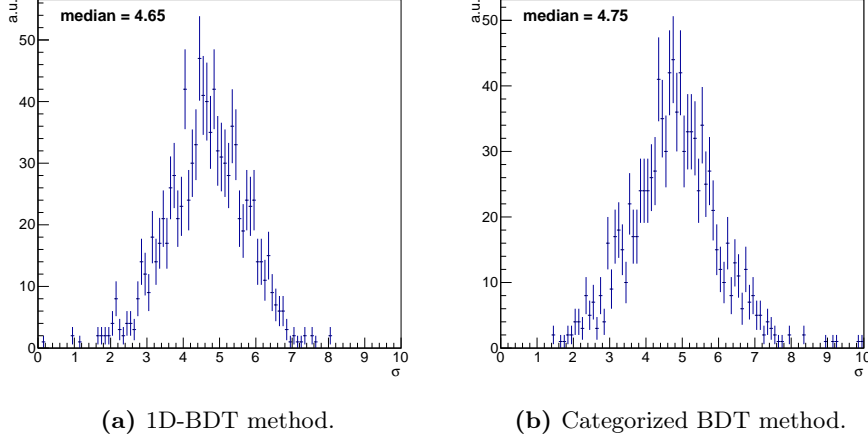


Figure 4.34: MC significance distributions expressed in Gaussian sigmas, evaluated with a likelihood ratio test, for the B_s^0 hypothesis, for the 1D-BDT method (a) and the categorized BDT method (b).

- **Alternative Hypothesis \mathcal{L}_1 :** $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is let floating.

The same is true also for the B^0 significance evaluation, replacing $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ with $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$.

The significance is evaluated through the use of the Wilks' theorem [144], that states that the logarithm of the likelihood ratio $\mathcal{L}_1/\mathcal{L}_0$ is asymptotically distributed as a χ^2 distribution with N degrees of freedom:

$$-2 \ln \frac{\mathcal{L}_1}{\mathcal{L}_0} \sim \chi^2(N) \quad (4.35)$$

Using MC toy experiments it is possible to evaluate the SM expected distributions of the significance (Fig. 4.34). The B_s^0 median of the significance distribution, taking into account all systematic effects, is 4.7 standard deviations (σ) for the 1D-BDT method and 4.8σ for the categorized BDT method. Given that the expected significance is higher, the chosen method for the measurement of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is the categorized BDT method.

Table 4.11: Final efficiencies for the signal and normalization channels. The $B^0 \rightarrow \mu^+ \mu^-$ efficiencies are very close to the $B_s^0 \rightarrow \mu^+ \mu^-$ efficiencies and are, therefore, not shown.

Efficiency	Barrel 2011	Endcap 2011	Barrel 2012	Endcap 2012
$B_s^0 \rightarrow \mu^+ \mu^-$				
A	0.034 ± 0.001	0.024 ± 0.001	0.033 ± 0.001	0.024 ± 0.001
ϵ_{ana}	0.229 ± 0.007	0.170 ± 0.005	0.196 ± 0.006	0.103 ± 0.003
ϵ_{reco}	0.47 ± 0.02	0.66 ± 0.05	0.57 ± 0.02	0.65 ± 0.05
ϵ_{trig}	0.83 ± 0.03	0.73 ± 0.04	0.63 ± 0.02	0.61 ± 0.04
$\epsilon[10^{-2}]$	0.30 ± 0.04	0.20 ± 0.02	0.23 ± 0.03	0.09 ± 0.01
$B^\pm \rightarrow J/\psi K^\pm$				
A	0.010 ± 0.001	0.005 ± 0.001	0.011 ± 0.001	0.006 ± 0.001
ϵ_{ana}	0.25 ± 0.01	0.19 ± 0.01	0.23 ± 0.01	0.13 ± 0.01
ϵ_{reco}	0.50 ± 0.02	0.60 ± 0.05	0.60 ± 0.02	0.56 ± 0.05
ϵ_{trig}	0.77 ± 0.02	0.60 ± 0.04	0.55 ± 0.02	0.49 ± 0.03
$\epsilon[10^{-2}]$	0.098 ± 0.008	0.036 ± 0.004	0.082 ± 0.007	0.021 ± 0.003

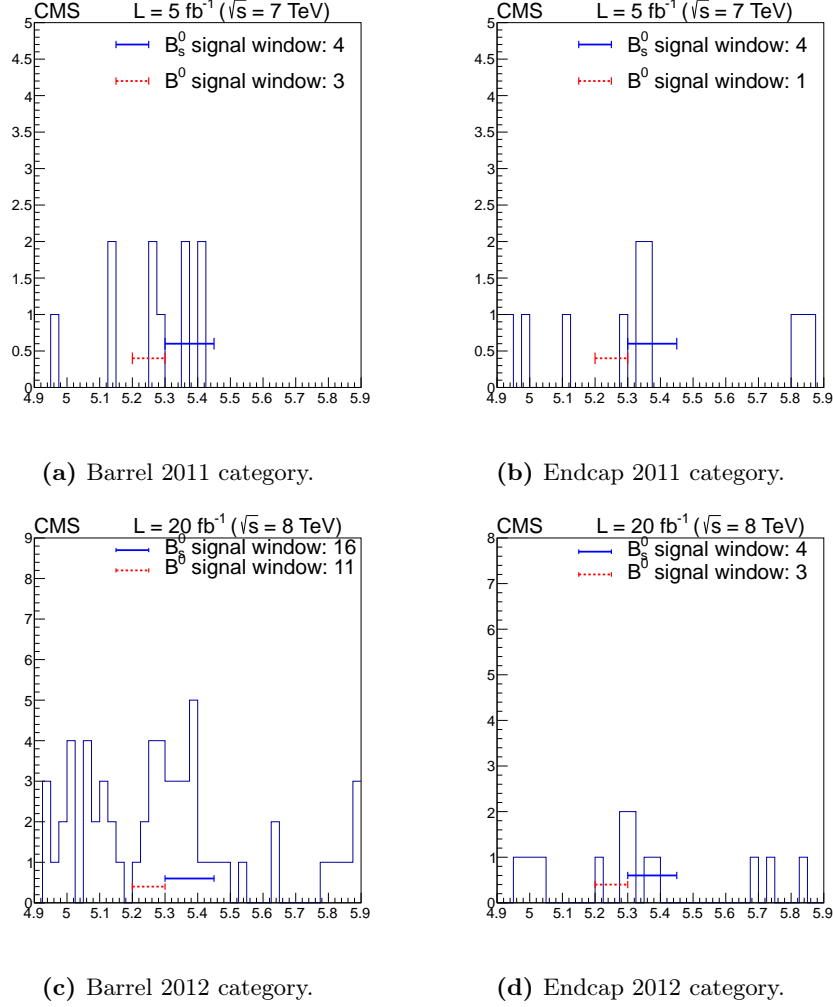
4.10 Results

After the final choice of all selections, the analysis is “unblinded”: the blind window (Table 4.3) is open and the results on the searches for the B_s^0 and B^0 mesons can be extracted.

Table 4.11 shows the final efficiencies ϵ_i for each category.

In Fig. 4.35 the unblinded dimuon invariant mass distribution in the barrel and endcap categories are shown. In Fig. 4.36 the same final events are shown in a 2D plot of BDT output versus invariant mass. The expected MC distributions for the $B_s^0 \rightarrow \mu^+ \mu^-$ candidates are superimposed to the data distributions. These invariant mass distributions are given to the binned analysis (Sec. 4.10.1) and to the unbinned likelihood fit (Sec. 4.10.2) to extract the final $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ results, respectively.

The expected and observed yields in the two blind windows are shown, together with the NS yields, in Table 4.12. The observed numbers of candidates in each category are within one or two standard deviations close to the expected values.

**Figure 4.35:** Dimuon invariant mass distributions.**Table 4.12:** Expected and observed number of signal and background events in the B^0 and B_s^0 mass windows, in the four 1D-BDT categories. Last row shows the fitted NS yield, $N_{B^\pm}^{\text{obs}}$, together with its total uncertainty.

Category	$N_{B^0}^{\text{exp}}$	$N_{B^0}^{\text{obs}}$	$N_{B_s^0}^{\text{exp}}$	$N_{B_s^0}^{\text{obs}}$	$N_{B^\pm}^{\text{obs}}$
Barrel 2011	1.3 ± 0.8	3	3.6 ± 0.6	4	$(71.2 \pm 4.1) \times 10^3$
Endcap 2011	1.5 ± 0.6	1	2.6 ± 0.5	4	$(21.4 \pm 1.1) \times 10^3$
Barrel 2012	7.9 ± 3.0	11	17.9 ± 2.8	16	$(309 \pm 16) \times 10^3$
Endcap 2012	2.2 ± 0.8	3	5.1 ± 0.7	4	$(69.3 \pm 3.5) \times 10^3$

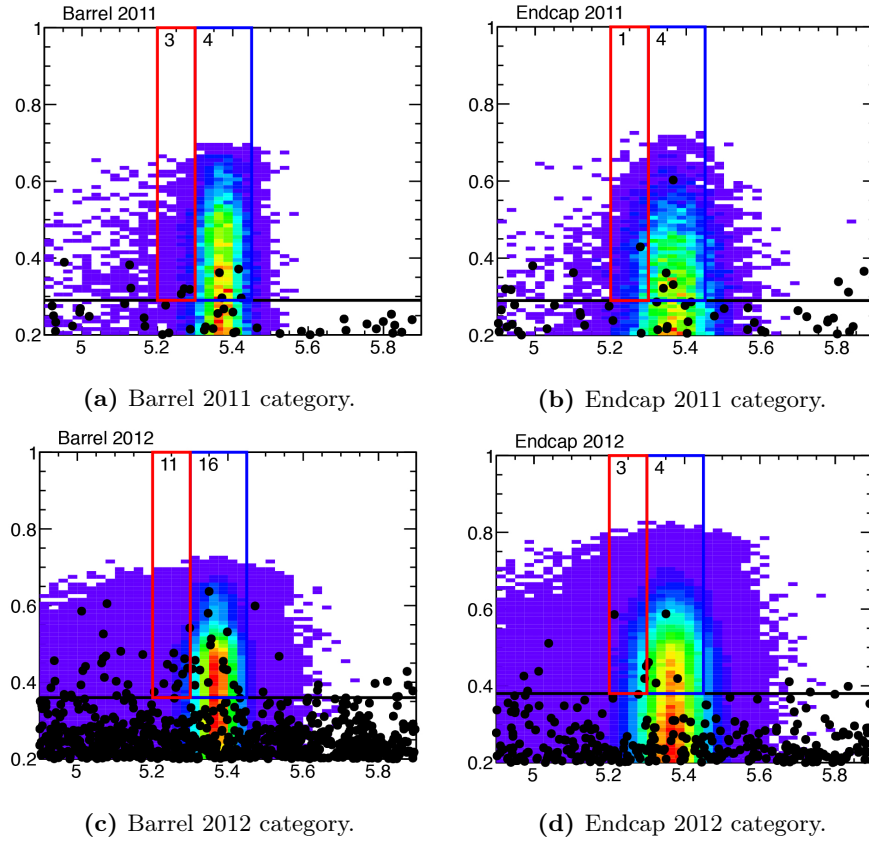


Figure 4.36: Analysis BDT output versus the invariant mass distributions. The horizontal black line corresponds to the 1D-BDT selection. The red and blue lines delimit the B^0 and B_s^0 blind windows, respectively. On top, the numbers of observed events in these ranges are shown. The colored histograms show the distributions for the $B_s^0 \rightarrow \mu^+ \mu^-$ signal MC.

4.10.1 Results with the binned analysis

These results are evaluated using the binned CL_s method, that gives the following upper limits:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \leq 4.9 \times 10^{-9} \quad \text{at 95\% CL} \quad (4.36)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) \leq 1.1 \times 10^{-9} \quad \text{at 95\% CL} \quad (4.37)$$

The p -value of the $B_s^0 \rightarrow \mu^+ \mu^-$ signal, assuming unknown $B^0 \rightarrow \mu^+ \mu^-$ signal strength, is 2.2×10^{-3} , corresponding to 2.9σ . Instead, the p -value of the $B^0 \rightarrow \mu^+ \mu^-$ signal assuming unknown $B_s^0 \rightarrow \mu^+ \mu^-$ signal is 2.8×10^{-2} , corresponding to 1.9σ .

The p -value of neither $B_s^0 \rightarrow \mu^+ \mu^-$ nor $B^0 \rightarrow \mu^+ \mu^-$ signal is

$$p(\mu_s = 0, \mu_d = 0) = 3.0 \times 10^{-6} = 4.5 \sigma \quad (4.38)$$

and the p -value for both to be consistent to the SM hypothesis is

$$p(\mu_s = 1, \mu_d = 1) = 4.3 \times 10^{-1} = 0.2 \sigma \quad (4.39)$$

If the observed excess is assumed from a signal, the branching fractions can be measured to be

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.6_{-1.0}^{+1.2}) \times 10^{-9} \quad (4.40)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (4.9_{-2.7}^{+3.1}) \times 10^{-10} \quad (4.41)$$

where the errors are related to the standard 1σ uncertainty.

The corresponding confidence level plots are shown in Fig. 4.37 for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, and Fig. 4.38 for $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. Finally, the two-sided interval confidence levels are shown in Fig. 4.39.

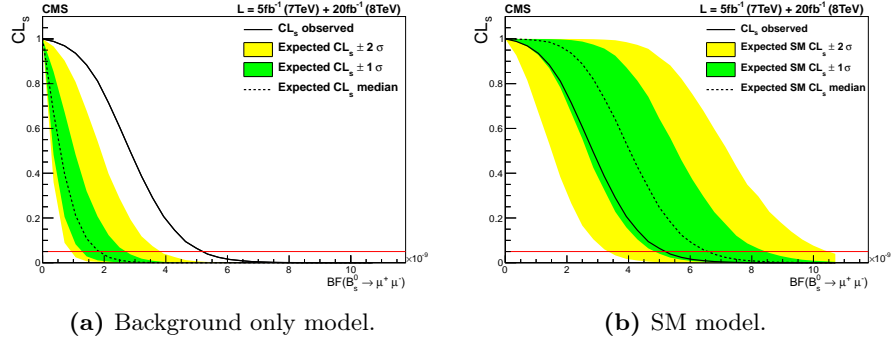


Figure 4.37: $B_s^0 \rightarrow \mu^+ \mu^-$ confidence level plots: observed CL_s curve with expected curve from background only model (a) and SM model (b).

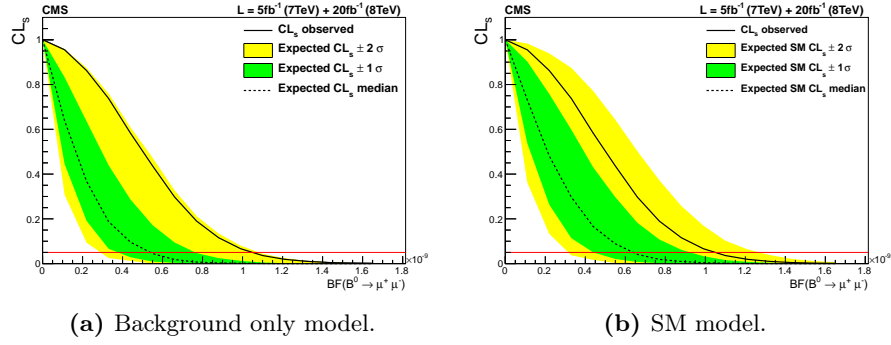


Figure 4.38: $B^0 \rightarrow \mu^+ \mu^-$ confidence level plots. Top plots show the observed CL_s curve with expected curve from background only model (a) and SM model (b).

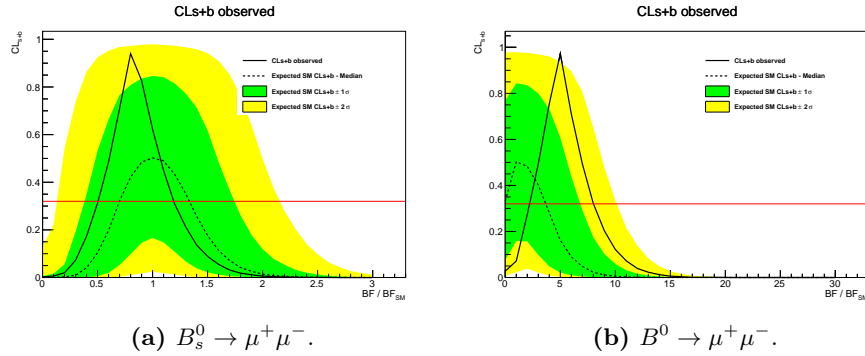


Figure 4.39: CL_s two-sided interval confidence levels for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (a) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (b).

4.10.2 Results with the UML fit

The UML simultaneous fit technique is performed on the unblinded data for both methods. Results of the fitting are shown in Fig. 4.40, for the 1D-BDT method, and in Fig. 4.41 for the categorized BDT (Sec. 4.9.2.3) method.

For the categorized BDT fit the measured signal branching fractions are:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0_{-0.9}^{+1.0}) \times 10^{-9} \quad (4.42)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (3.5_{-1.8}^{+2.1}) \times 10^{-10} \quad (4.43)$$

and the significances for the alternative hypotheses described in Sec. 4.9.2.4 are:

$$\text{sign}(\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)) = 4.3 \sigma \quad (4.44)$$

$$\text{sign}(\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)) = 2.0 \sigma \quad (4.45)$$

The significance, given no $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ nor $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ decays, is:

$$\text{sign}(\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \cap \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)) = 4.7 \sigma \quad (4.46)$$

The profile likelihoods, from which the significances of the alternative hypothesis are extracted, are shown in Fig. 4.42 for both branching fractions.

The statistical uncertainty can be estimated redoing the fit after fixing all nuisance parameters to their post-fit expected values. The systematic total uncertainty is obtained by subtracting in quadrature the statistical uncertainty from the total uncertainty:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0_{-0.8}^{+0.9}(\text{stat.})_{-0.4}^{+0.6}(\text{syst.})) \times 10^{-9} \quad (4.47)$$

The $B_s^0 \rightarrow \mu^+ \mu^-$ BF uncertainty is statistically dominated.

For the 1D-BDT method, the $B_s^0 \rightarrow \mu^+ \mu^-$ BF significance is $\text{sign}(\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)) = 4.8 \sigma$ and the double null hypothesis significance is $\text{sign}(\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \cap \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)) = 5.4 \sigma$.

As illustration, Fig. 4.43 shows all data from the different categories, combined together. For these plots, the individual categories are weighed with $w = S/(S+B)$, where S is the signal yield and B is the background yield, determined at the B_s^0 peak position of each category. The overall normalization is set such that the fitted B_s^0 signal corresponds to the total yield of the individual contributions.

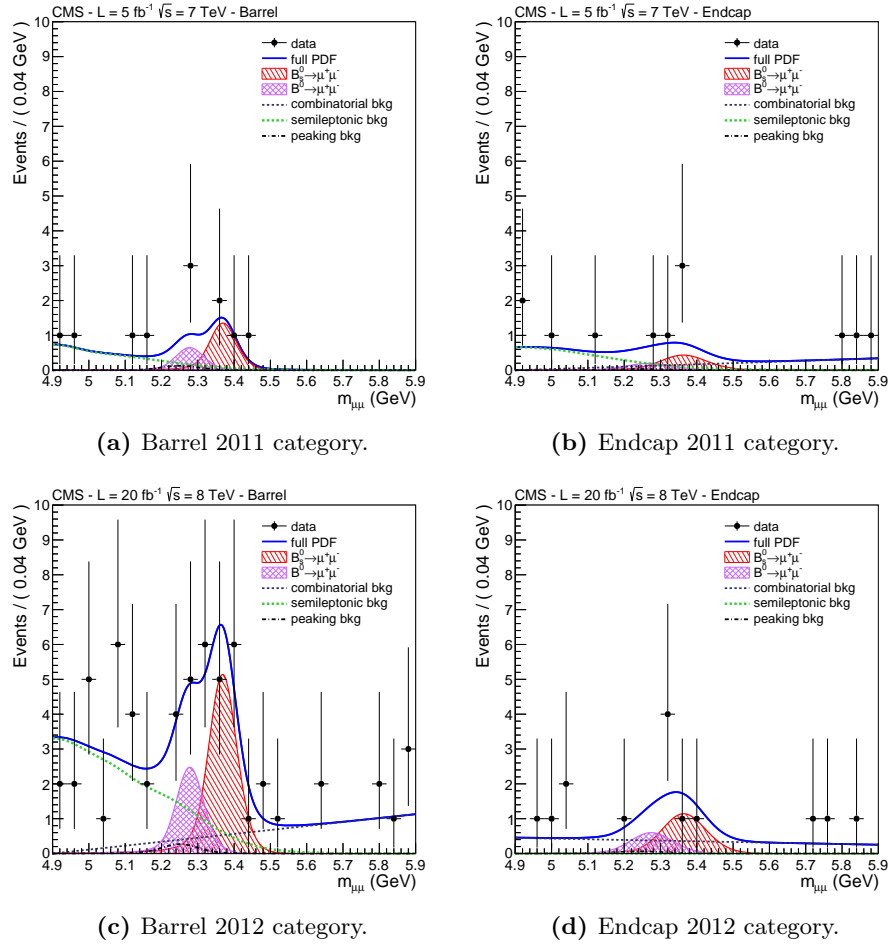


Figure 4.40: 1D fit result on the dimuon invariant mass distribution for 2011 (top) and 2012 (bottom) in the barrel (left) and the endcap channel (right).

4.10. RESULTS

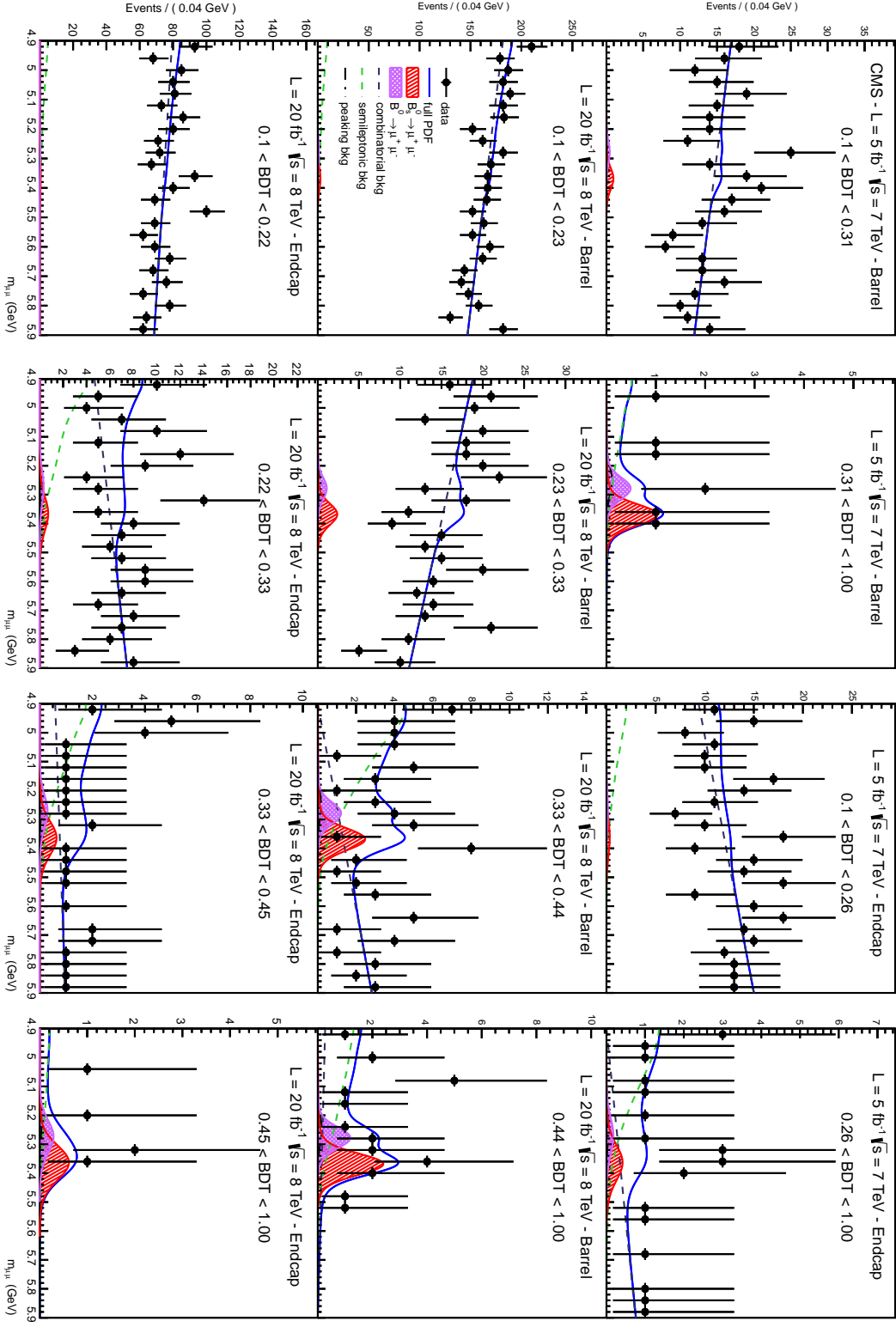


Figure 4.41: Categorized BDT fit result. The top row shows the 2012 samples, divided in two barrel and two endcap categories. The middle row displays the four 2012 barrel categories and the bottom row the four 2012 endcap categories.

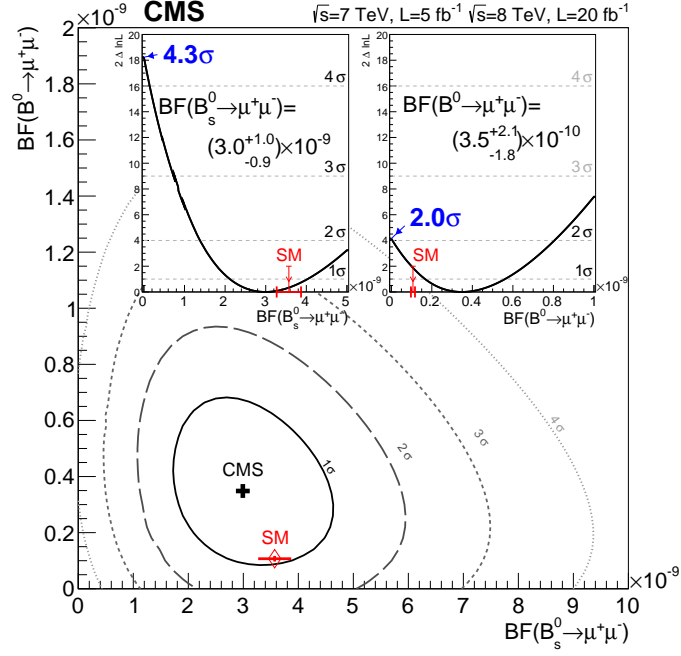


Figure 4.42: Categorized BDT method: scan of the ratio of the joint likelihood for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. As insets, the likelihood ratio scan for each of the branching fractions when the other is profiled together with other nuisance parameters; the significance at which the background-only hypothesis is rejected is also shown.

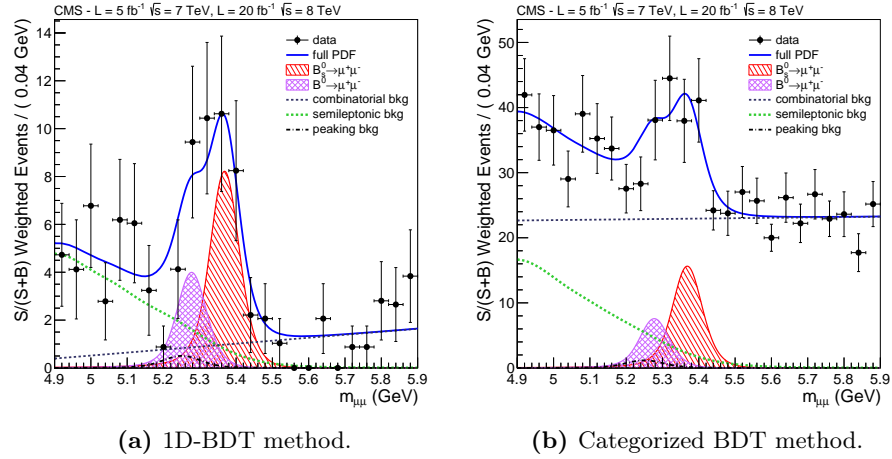


Figure 4.43: Weighted invariant mass distributions, based on $S/(S+B)$, for the 1D-BDT (a) and categorized BDT (b) methods.

4.10.3 Interpretations of the results

With the LHC run 1 data it has been finally possible to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ decay after decades of experimental efforts. This striking result has been obtained thanks to the large amount of events produced by the LHC in just two years of excellent run operations. The high performances of the CMS detector, especially the silicon tracker and muon stations, allowed for a very high muon reconstruction efficiency and high geometrical coverage. This is especially true for the low p_T muons that are used in this analysis. The strong fake muon rejection and the high separation between signals and background sources have been achieved thanks to the BDT techniques applied on the muon and candidate variables. This sample of almost pure muons allowed the maximum likelihood fit to extract a clear signal with a remarkable significance.

The $B_s^0 \rightarrow \mu^+ \mu^-$ BF measurement is fully compatible with the SM expectation. A lower than expected significance is obtained (4.3σ instead of 4.8σ , Sec. 4.9.2.4), since the measured BF is slightly lower than the SM value, but within uncertainty. The $B^0 \rightarrow \mu^+ \mu^-$ BF measurement is three times the SM expectations, but with a not compelling significance of 2σ . The $B^0 \rightarrow \mu^+ \mu^-$ upper limit evaluated with the CL_S method is set to $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) \leq 1.1 \times 10^{-9}$ at 95% CL, ten times higher than the SM expectation value.

An other interesting way to check the results is to parametrize the B_s^0 and B^0 yields with a signal strength parametrization normalized to the SM expectations, instead of using $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ as free parameters in the likelihood, as described in Sec. 4.9.2.1. Let us define an overall signal strength parameter μ , applicable both to $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, and a parameter λ_{ds} that decouples the BF of the two mesons:

$$\mu = \frac{\mathcal{B}(B \rightarrow \mu^+ \mu^-)}{\mathcal{B}_{\text{SM}}(B \rightarrow \mu^+ \mu^-)} \quad (4.48)$$

$$\lambda_{ds} = \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}_{\text{SM}}(B^0 \rightarrow \mu^+ \mu^-)} \bigg/ \frac{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}_{\text{SM}}(B_s^0 \rightarrow \mu^+ \mu^-)} \quad (4.49)$$

The UML fit with these two parameters represents a compatibility test of the results against the SM expectations on the single BFs and also on their ratio. The simultaneous fit of these two parameters gives the following results, with the BDT categorized method:

$$\mu = 0.84^{+0.31}_{-0.25} \quad (4.50)$$

$$\lambda_{ds} = 3.9^{+3.7}_{-2.2} \quad (4.51)$$

If we fix the signal ratio between the two states to be as in the SM prediction,

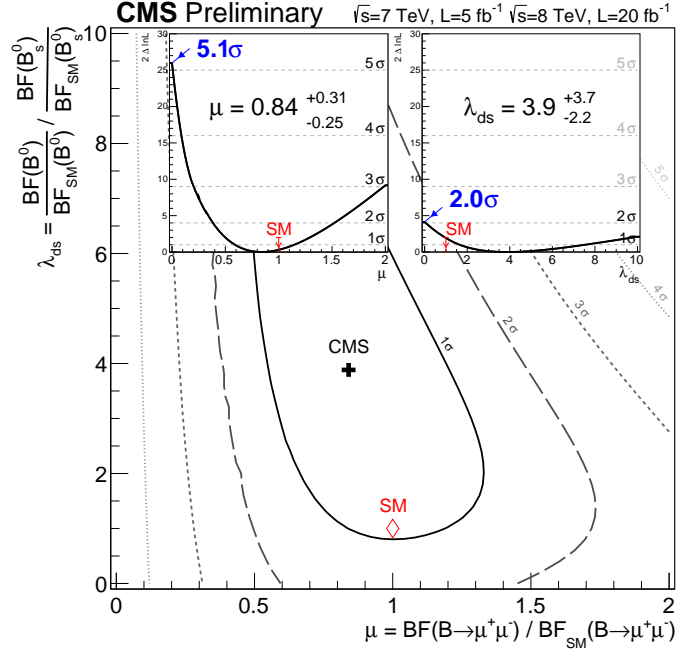


Figure 4.44: Categorized BDT method: scan of the ratio of the joint likelihood for μ and λ_{ds} . As insets, the likelihood ratio scan for each of the parameters when the other is profiled together with other nuisance parameters; the significance at which the hypothesis wherein the parameter is zero is rejected is also shown. The 5.1σ significance for the parameter μ is due to the assumed alternative hypothesis that considers the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields as one unique signal parameter.

($\lambda_{ds} = 1$), the fitted signal strength is

$$\mu = 1.01^{+0.31}_{-0.26} \quad (4.52)$$

Instead, setting the signal strength to the SM one ($\mu = 1$), the BF ratio parameter λ_{ds} is

$$\lambda_{ds} = 3.1^{+2.0}_{-1.7} \quad (4.53)$$

The 1D and 2D likelihood ratio scans are shown in Fig. 4.44. The 5.1σ significance for the parameter μ is due to the assumed alternative hypothesis that considers the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ yields as one unique signal parameter. With this parameterization it is still evident that the $B_s^0 \rightarrow \mu^+ \mu^-$ BF is SM fully compatible. Furthermore, it is clear that the $B^0 \rightarrow \mu^+ \mu^-$ signal, even if it is about three times higher than the SM expectation, does not have statistical significance to claim a substantial deviation from the SM predictions.

4.10.4 CMS-LHCb combination

After the publication of the shown CMS results, a new measurement of f_s/f_d was published by the LHCb collaboration [145]:

$$\frac{f_s}{f_d} = 0.259 \pm 0.015 \quad (4.54)$$

which changes very slightly the $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ result of Eq. (4.42), that uses $f_s/f_d = 0.256 \pm 0.020$ from Eq. (4.3).

The LHCb collaboration reports the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ decay, with a 4σ significance, and the $B^0 \rightarrow \mu^+\mu^-$ decay, with a 2σ significance [86]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.9^{+1.1}_{-1.0}(\text{stat})^{+0.3}_{-0.1}(\text{syst})) \times 10^{-9} \quad (4.55)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (3.7^{+2.4}_{-2.1}(\text{stat})^{+0.6}_{-0.4}(\text{syst})) \times 10^{-10} \quad (4.56)$$

Therefore, a combination of the measurements can be implemented.

A preliminary study has been published using only the reported final results. The combination is performed using ensembles of simplified pseudo-experiments, where the distributions are modeled with variable-width Gaussian functions [2]. The main correlation between the two analyses is given by f_s/f_d , as it is used by both experiments to extract the $B_s^0 \rightarrow \mu^+\mu^-$ BF normalized to the $B^\pm \rightarrow J/\psi K^\pm$ decay¹. Other correlations, like the $B^\pm \rightarrow J/\psi K^\pm$ BF and the rare peaking and semileptonic assumed branching fractions, are negligible.

Assigning a 100% correlation to f_s/f_d , the combined results are:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.9 \pm 0.7) \times 10^{-9} \quad (4.57)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (3.6^{+1.6}_{-1.4}) \times 10^{-10} \quad (4.58)$$

They are shown, together with the SM expectations, also in Fig. 4.45.

The significance of the B_s^0 signal reaches more than 5σ , while the B^0 significance still remains below 3σ . Thus, the empirical observation of the $B_s^0 \rightarrow \mu^+\mu^-$ decay can be claimed.

While results shown in this section are performed with an approximated method, an exact combination will be published when the analysis teams finish to combine the two complete likelihood functions.

4.10.5 Impact on new physics

The $B_s^0 \rightarrow \mu^+\mu^-$ BF empirical measurement has a strong impact on high energy physics. The accuracy of the achieved measurement sets stringent boundaries to

¹The LHCb analysis uses also the $B^0 \rightarrow K^+\pi^-$ decay as a normalization channel.

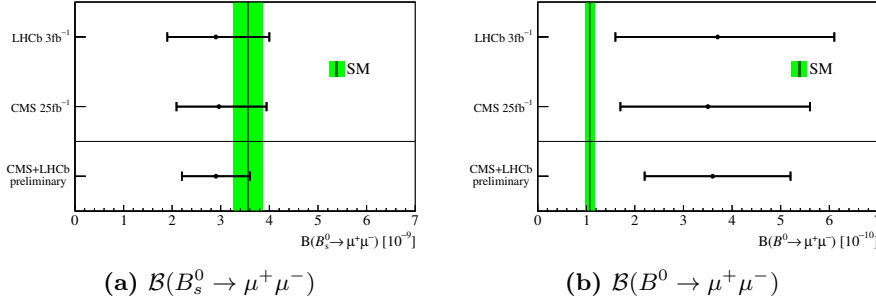


Figure 4.45: Comparison of the latest CMS and LHCb results, the combined value, and the SM prediction (vertical line) for the time-integrated branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (a) and for $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (b). The width of the vertical band represents the uncertainty in the SM prediction. The error bars represent the combined statistical and systematic uncertainties.

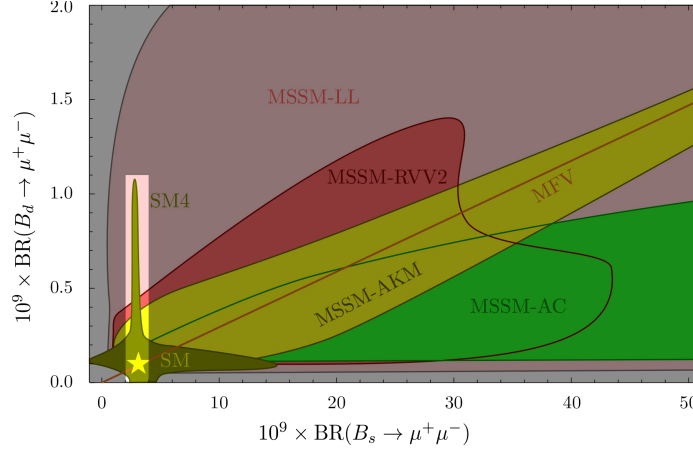


Figure 4.46: Correlation between the branching fractions, as shown in Fig. 2.6, but with the excluded area (in gray) due to the CMS measurement of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and the upper limit on $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, by the author.

the NP models which now have to confront with a precise measurement of this important decay.

The first consequence is that since the $B_s^0 \rightarrow \mu^+ \mu^-$ BF is compatible with the SM and it has a small uncertainty, the phase space of possible BSM contributing parameters to this BF is notably reduced. An exhaustive review on the impact on NP models is not practical here as many new papers are being published; a non comprehensive list of the most recent publications using the combined CMS-LHCb results is [146–148]. Thus, here we will say a few general considerations.

Figure 4.46 show the same B^0 - B_s^0 correlation plot of Fig. 2.6, but with the forbidden region due to the CMS result shaded in gray. None of the shown NP models are ruled out because everyone allows for results close to the SM, but the

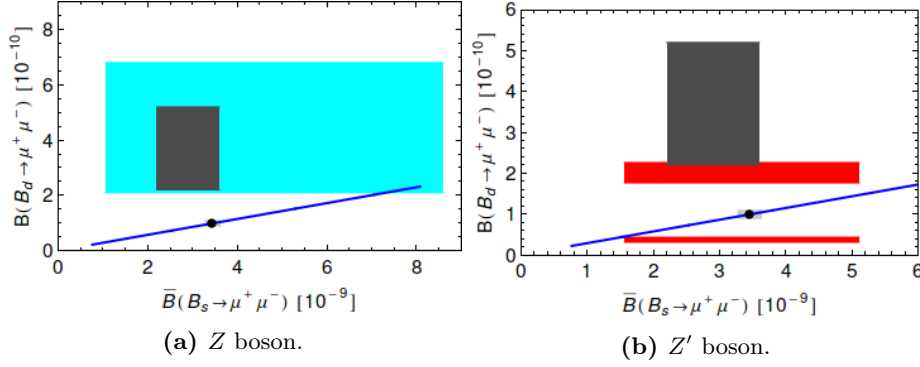


Figure 4.47: $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ versus $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ with NP contributions coming from a heavy neutral boson. The SM prediction is represented by the light gray area with black dot and the MFV prediction by the blue line. The dark gray rectangle shows the combined CMS-LHCb experimental results. The cyan rectangle shows the allowed region for a Z scenario (a), while the red regions are due to a Z' contribution (b) [149].

large enhancing effects, predicted by many of them, are not allowed anymore, especially for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. For the $B^0 \rightarrow \mu^+ \mu^-$ decay, there is still space of large enhancements with respect the SM prediction.

Even if the two rare decay results are compatible with the SM, there is some tension for the $B^0 \rightarrow \mu^+ \mu^-$ BF, that could be explained within NP scenarios. For instance, in addition to the SUSY models already discussed, new contributions could come from a new neutral heavy gauge boson Z' , with purely left-handed complex couplings and real coupling to muons, or from the SM Z boson with left-handed complex couplings (Fig. 4.47).

In any case, the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ experimental value does not present a large significance, for any of the two collaboration measurements.

Improvements on the CMS measurement can occur from a better understanding of the rare backgrounds, and especially of the hadronic two-body B decays, which populate the invariant mass region around the B^0 mass value. In this aspect, better estimations of the expected rare backgrounds can come from a better evaluation of the muon misidentification, that now has assigned a 50% relative uncertainty, and from better measurements of the BFs of these hadronic B decays. In fact, for some of them, the relative uncertainty can go up to 100% (Table 4.7).

However, the $B^0 \rightarrow \mu^+ \mu^-$ BF measurement is dominated by statistical uncertainty as it is for the $B_s^0 \rightarrow \mu^+ \mu^-$ BF measurement, therefore, more data is needed. Next upgrades on the LHC collider and on the CMS sub-detectors point to this aim, as described in the next section.

4.10.6 Future prospects

The LHC run 1 ended on December 2012. After about two years of technical stop (the Long Shutdown 1 (LS1)), the center-of-mass collision energy will increase to 14 TeV. During this Run 2 period (2015-2017), it is expected that the LHC will deliver an integrated luminosity of the order of 100 fb^{-1} to the CMS and ATLAS detectors. During Run 3 (2019-2021) the instantaneous luminosity of the machine will increase to $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, adding additional 300 fb^{-1} .

The CMS detector, on the same time, thanks to the upgrades due to the new technologies, will have an improved muon trigger capability and, starting from 2017, an improved pixel detector [150]. The muon system will increase the momentum resolution and have the capability to cope with the increased pileup conditions. The pixel detector will improve the resolution in the vertexing by roughly 50%.

Starting from 2023, the LHC will undergo a major luminosity upgrade (HL-LHC): the luminosity for CMS will reach $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, with the goal to deliver a total integrated luminosity of 3 ab^{-1} , still at the same center-of-mass energy of 14 TeV.

The CMS experiment is expected to get improvements to the muons system and a new inner tracker, with higher granularity, to cope with a mean expected pileup of up to 140 interactions per bunch crossing. In particular, the new inner tracker system will have L1 tracking capabilities, in order to reconstruct efficiently all tracks with $p_T > 2 \text{ GeV}$. The muon system will also have improved coverage in the forward direction and extended trigger capabilities.

The outer tracker material budget will diminish by roughly a factor of 2 in the central region ($|\eta| < 1$) and about a factor of 3 in the intermediate region around $1.2 < |\eta| < 1.5$. This, combined with a smaller silicon sensors pitch, will improve the momentum resolution by about a factor of 1.5 in the barrel region ($|\eta| < 1.4$) and 1.2 elsewhere. This will help separate the B^0 signal from the left tail of the B_s^0 signal, which now becomes a background to the B^0 measurement.

It is possible to do some rough estimations of the improvements on the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ measurements after each new LHC run, rescaling the collected luminosity, and taking into account the increased center-of-mass energy, 14 TeV, which should increase the $b\bar{b}$ cross-section by about 75% as compared to the 8 TeV center-of-mass cross-section [151].

For simplicity, the effects of any possible improvements on the analysis strategy (for instance better selections and fits, as well as the usage of additional normalization channels) are not considered, and the current trigger efficiency is supposed to be maintained equal to the one used in this analysis.

The most sensible aspect is the pileup. The increased pileup will impact the

Table 4.13: Expected uncertainties in the BF measurements for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ at different integrated luminosity values \mathcal{L} . We also report the 1σ range of significance for $B^0 \rightarrow \mu^+ \mu^-$ and, finally, the relative uncertainty on the B^0 to B_s^0 branching fraction ratio.

$\mathcal{L}(\text{fb}^{-1})$	$\delta\mathcal{B}/\mathcal{B}(B_s^0)$	$\delta\mathcal{B}/\mathcal{B}(B^0)$	B^0 sign.	$\delta \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)}$
100	15%	66%	0.5–2.4 σ	71%
300	12%	45%	1.3–3.3 σ	47%
3000	12%	18%	5.4–7.6 σ	21%

isolation variables, due to a non negligible chance of another vertex being close to the B -meson production one. Also, the resulting larger number of tracks could potentially affect pattern recognition, requiring tightening of reconstruction criteria. Moreover, the typical resolution in the track longitudinal impact parameter in the endcaps (0.5 to 1 mm) is expected to be comparable to the average pileup vertex separation. The significance for $B^0 \rightarrow \mu^+ \mu^-$ may not be limited by the signal yield, but by the separation between B^0 and B_s^0 signal distributions. For this reason, only the barrel part of the detector will be used for the HL-LHC analysis.

On the other hand, the systematic uncertainty from f_s/f_u could reasonably be reduced to the level of 5%, and also the systematics due to the rare backgrounds could go down, thanks to a better determination of the misidentification rate and a better measurement of the $B \rightarrow hh'$ decays.

Table 4.13 shows the expected results, assuming the SM BFs.

With the increased data sample sizes it will be possible to reduce both systematic and statistical uncertainties leading to high precision measurements of $B^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$, which would allow more stringent tests of the SM.

Chapter 5

Conclusions

The standard model (SM) gives a very accurate prediction for the Branching Fraction (BF) of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay and makes its measurement very sensitive to new physics. Any possible small deviation from the SM prediction allows to observe new contributions coming from New Physics (NP), thanks to the small SM theoretical uncertainty. On the other hand, the same very small SM expectation value of the BF and the huge background expected at the LHC are major challenges for the observation and the extraction of the BF measurement for this decay.

In this thesis we presented the first measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ BF and also the most recent upper limit of the $B^0 \rightarrow \mu^+ \mu^-$ BF, performed with the CMS experiment at the CERN LHC. The 25 fb^{-1} sample of collected data corresponds to the total 2011 and 2012 LHC runs for pp collisions at the center-of-mass energies of $\sqrt{s} = 7 \text{ TeV}$ and 8 TeV , respectively.

Signal selections and efficiencies are evaluated with Monte Carlo (MC) simulations, validated with many data-driven-methods. The $B_s^0 \rightarrow J/\psi \phi$ decay is used to validate the B_s^0 MC simulation. The $B^\pm \rightarrow J/\psi K^\pm$ decay is used as the normalization channel, to avoid a dependence on the uncertainties of the $b\bar{b}$ production cross section and luminosity measurements and to cancel out to first order many systematic uncertainties related to the analysis efficiencies. Systematic effects due to the analysis and detector efficiencies are evaluated on data to be below 10%.

Different background sources can mimic the signal topology. Events in which two uncorrelated muons come from two different B decays represent the largest background, combinatorial in nature, five orders of magnitude larger than the SM expected signal yield. A Boosted Decision Tree (BDT) method is used to separate the $B_{(s)}^0$ candidates from this combinatorial background. Variables related to the displaced secondary vertex topology and to the isolation of the

muon and dimuon candidates are used to train the classifier.

Rare two and three-body B decays, like $\Lambda_b^0 \rightarrow pK^-$ and $B^0 \rightarrow \pi^- \mu^+ \nu_\mu$, can be wrongly selected as signals due to misidentified muons. A BDT multivariate technique is used to optimize the muon identification against kaons, pions and protons, coming from these B decays, forming the rare peaking and semileptonic backgrounds. We obtained a further 50% misidentification rejection at the cost of 10% of muon efficiency, on top of the standard CMS tight muon selection.

The B^0 upper limit is extracted with a counting experiment with the CL_S method in the B^0 invariant mass window.

The $B_s^0 \rightarrow \mu^+ \mu^-$ BF is measured with a simultaneous unbinned maximum likelihood fit on twelve independent data categories. This method allowed a 4.3σ discovery of this decay.

The final $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ BF measurements and upper limits are:

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.0_{-0.9}^{+1.0}) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &\leq 1.1 \times 10^{-9} \text{ at } 95\% \text{ CL}\end{aligned}$$

The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is fully compatible with the SM predictions. Its relative uncertainty, about 30%, is the most precise up-to-date and it is at the same level of the LHCb results, showing the ability of CMS to perform high precision measurements in B -physics with muons.

Since the $B_s^0 \rightarrow \mu^+ \mu^-$ BF measurement does not present deviations from the SM expectation, it sets stringent limits on new physics parameters. For instance, this result gives important constraints on the supersymmetry phase space, as relevant as those coming from the direct searches.

New physics effects on the $B_s^0 \rightarrow \mu^+ \mu^-$ BF, if any, are small. Precise measurements of the BFs for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$ and of their ratio, have become essential to distinguish possible deviations of these decays from the SM expectations.

The CMS and LHCb measurements are statistically limited, therefore new data will improve the experimental precision for both decays, for both experiments. In this prospect, the next few years look encouraging as the CMS detector will have the possibility to collect hundred of inverse femtobarn of integrated luminosity. The main challenge will be the pileup. Better measurements can be obtained if the detectors will be able to live with an average of up to 140 primary vertices per bunch crossing. Under this aspect, the expected CMS upgrades, especially those for the inner tracker sub-detector, seem promising.

Given these results, the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ BF high precision measurements can be considered critical for the indirect searches of new physics

in the next future.

Bibliography

- [1] Serguei Chatrchyan et al. “Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ with the CMS Experiment”. *Phys.Rev.Lett.* 111 (2013), p. 101804. DOI: 10.1103/PhysRevLett.111.101804. arXiv: 1307.5025 [hep-ex].
- [2] CMS collaboration and LHCb collaboration. *Combination of results on the rare decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ from the CMS and LHCb experiments*. Tech. rep. CMS-PAS-BPH-13-007. CERN-LHCb-CONF-2013-012. Geneva: CERN, July 2013.
- [3] Ce Meng, Hao Han, and Kuang-Ta Chao. “X(3872) and its production at hadron colliders” (2013). arXiv: 1304.6710 [hep-ph].
- [4] J. Beringer et al. “Review of Particle Physics (RPP)”. *Phys.Rev.* D86 (2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001.
- [5] Michael E. Peskin and Dan V. Schroeder. *An Introduction To Quantum Field Theory (Frontiers in Physics)*. Westview Press, 1995. ISBN: 0201503972.
- [6] Chen-Ning Yang and Robert L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. *Phys.Rev.* 96 (1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191.
- [7] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. *Phys.Rev.Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [8] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. *Phys.Rev.Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.
- [9] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble. “Global Conservation Laws and Massless Particles”. *Phys.Rev.Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

- [10] S.L. Glashow. “Partial Symmetries of Weak Interactions”. *Nucl.Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.
- [11] Abdus Salam and John Clive Ward. “Electromagnetic and weak interactions”. *Phys.Lett.* 13 (1964), pp. 168–171. DOI: 10.1016/0031-9163(64)90711-5.
- [12] Steven Weinberg. “A Model of Leptons”. *Phys.Rev.Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.
- [13] G. Arnison et al. “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV”. *Phys.Lett.* B122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.
- [14] G. Arnison et al. “Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/ c^2 at the CERN SPS Collider”. *Phys.Lett.* B126 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.
- [15] M. Banner et al. “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider”. *Phys.Lett.* B122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2.
- [16] P. Bagnaia et al. “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN anti-p p Collider”. *Phys.Lett.* B129 (1983), pp. 130–140. DOI: 10.1016/0370-2693(83)90744-X.
- [17] O.W. Greenberg. “Quarks: Resource Letter Q-1”. *Am.J.Phys.* 50 (1982), pp. 1074–1089. DOI: 10.1119/1.12922.
- [18] S.L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. *Phys.Rev.* D2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.
- [19] J.J. Aubert et al. “Experimental Observation of a Heavy Particle J”. *Phys.Rev.Lett.* 33 (1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.
- [20] J.E. Augustin et al. “Discovery of a Narrow Resonance in e^+e^- Annihilation”. *Phys.Rev.Lett.* 33 (1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.
- [21] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. “Advantages of the Color Octet Gluon Picture”. *Phys.Lett.* B47 (1973), pp. 365–368. DOI: 10.1016/0370-2693(73)90625-4.

- [22] H. David Politzer. “Reliable Perturbative Results for Strong Interactions?”. *Phys.Rev.Lett.* 30 (1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346.
- [23] D.J. Gross and Frank Wilczek. “Ultraviolet Behavior of Nonabelian Gauge Theories”. *Phys.Rev.Lett.* 30 (1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.
- [24] Martin L. Perl et al. “Evidence for Anomalous Lepton Production in e^+e^- Annihilation”. *Phys.Rev.Lett.* 35 (1975), pp. 1489–1492. DOI: 10.1103/PhysRevLett.35.1489.
- [25] S.W. Herb et al. “Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions”. *Phys.Rev.Lett.* 39 (1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252.
- [26] F. Abe et al. “Observation of top quark production in $\bar{p}p$ collisions”. *Phys.Rev.Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 [hep-ex].
- [27] S. Abachi et al. “Search for high mass top quark production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”. *Phys.Rev.Lett.* 74 (1995), pp. 2422–2426. DOI: 10.1103/PhysRevLett.74.2422. arXiv: hep-ex/9411001 [hep-ex].
- [28] K. Kodama et al. “Observation of tau neutrino interactions”. *Phys.Lett.* B504 (2001), pp. 218–224. DOI: 10.1016/S0370-2693(01)00307-0. arXiv: hep-ex/0012035 [hep-ex].
- [29] Serguei Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. *Phys.Lett.B* (2012). DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [30] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. *Phys.Lett.B* (2012). arXiv: 1207.7214 [hep-ex].
- [31] Chris Quigg. “Unanswered Questions in the Electroweak Theory”. *Ann.Rev.Nucl.Part.Sci.* 59 (2009), pp. 505–555. DOI: 10.1146/annurev.nucl.010909.083126. arXiv: 0905.3187 [hep-ph].
- [32] T. Nakano and K. Nishijima. “Charge Independence for V-particles”. *Prog.Theor.Phys.* 10 (1953), pp. 581–582. DOI: 10.1143/PTP.10.581.
- [33] M. Gell-Mann. “The interpretation of the new particles as displaced charge multiplets”. English. *Il Nuovo Cimento* 4 (2 1956), pp. 848–866. ISSN: 0029-6341. DOI: 10.1007/BF02748000. URL: <http://dx.doi.org/10.1007/BF02748000>.

- [34] CMS collaboration. *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*. Tech. rep. CMS-PAS-HIG-13-005. Geneva: CERN, 2013.
- [35] ATLAS collaboration. *Combined coupling measurements of the Higgs-like boson with the ATLAS detector using up to 25 fb⁻¹ of proton-proton collision data*. Tech. rep. ATLAS-CONF-2013-034. Geneva: CERN, Mar. 2013.
- [36] ATLAS collaboration. *Study of the spin of the new boson with up to 25 fb⁻¹ of ATLAS data*. Tech. rep. ATLAS-CONF-2013-040. Geneva: CERN, Apr. 2013.
- [37] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. *Phys.Rev.Lett.* 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.
- [38] Makoto Kobayashi and Toshihide Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. *Prog.Theor.Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.
- [39] J. Charles et al. “CP violation and the CKM matrix: Assessing the impact of the asymmetric *B* factories”. *Eur.Phys.J.* C41 (2005), pp. 1–131. DOI: 10.1140/epjc/s2005-02169-1. arXiv: hep-ph/0406184 [hep-ph].
- [40] Lincoln Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. *Phys.Rev.Lett.* 51 (1983), p. 1945. DOI: 10.1103/PhysRevLett.51.1945.
- [41] Gerhard Buchalla, Andrzej J. Buras, and Markus E. Lautenbacher. “Weak decays beyond leading logarithms”. *Rev.Mod.Phys.* 68 (1996), pp. 1125–1144. DOI: 10.1103/RevModPhys.68.1125. arXiv: hep-ph/9512380 [hep-ph].
- [42] D.W. Duke and R.G. Roberts. “Determinations of the QCD Strong Coupling α_s and the Scale Λ_{QCD} ”. *Phys.Rept.* 120 (1985), p. 275. DOI: 10.1016/0370-1573(85)90112-7.
- [43] Siegfried Bethke. “World Summary of α_s (2012)” (2012). arXiv: 1210.0325 [hep-ex].
- [44] Stephen P. Martin. “A Supersymmetry primer” (1997). arXiv: hep-ph/9709356 [hep-ph].
- [45] CMS collaboration. *CMS Supersymmetry Physics Results*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>.
- [46] ATLAS collaboration. *ATLAS Supersymmetry (SUSY) searches*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResults>.

- [47] T. Inami and C.S. Lim. “Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $k_L \rightarrow \mu\bar{\mu}$, $K^+ \rightarrow \pi^+\nu\bar{\nu}$ and $K^0 \leftrightarrow \bar{K}^0$ ”. *Prog.Theor.Phys.* 65 (1981), p. 297. DOI: 10.1143/PTP.65.297.
- [48] Gerhard Buchalla and Andrzej J. Buras. “QCD corrections to rare K and B decays for arbitrary top quark mass”. *Nucl.Phys.* B400 (1993), pp. 225–239. DOI: 10.1016/0550-3213(93)90405-E.
- [49] Andrzej J. Buras et al. “On the Standard Model prediction for $\mathcal{B}(B_{s,d} \rightarrow \mu^+\mu^-)$ ” (2012). arXiv: 1208.0934 [hep-ph].
- [50] Andrzej J. Buras et al. “Probing New Physics with the $B_s \rightarrow \mu^+\mu^-$ Time-Dependent Rate”. *JHEP* 1307 (2013), p. 77. DOI: 10.1007/JHEP07(2013)077. arXiv: 1303.3820 [hep-ph].
- [51] Y. Amhis et al. “Averages of B-Hadron, C-Hadron, and tau-lepton properties as of early 2012” (2012). arXiv: 1207.1158 [hep-ex].
- [52] Denis Derkach. “Unitarity Triangle Fitter Results for CKM Angles” (2013). arXiv: 1301.3300 [hep-ph].
- [53] R.J. Dowdall et al. “B-meson decay constants from improved lattice NRQCD and physical u, d, s and c sea quarks”. *Phys.Rev.Lett.* 110 (2013), p. 222003. DOI: 10.1103/PhysRevLett.110.222003. arXiv: 1302.2644 [hep-lat].
- [54] CMS collaboration. *Top mass combination*. Tech. rep. CMS-PAS-TOP-11-018. Geneva: CERN, 2012.
- [55] Kristof De Bruyn et al. “Probing New Physics via the $B_s^0 \rightarrow \mu^+\mu^-$ Effective Lifetime”. *Phys.Rev.Lett.* 109 (2012), p. 041801. DOI: 10.1103/PhysRevLett.109.041801. arXiv: 1204.1737 [hep-ph].
- [56] David M. Straub. “Constraints on new physics from rare (semi-)leptonic B decays” (2013). arXiv: 1305.5704 [hep-ph].
- [57] C. Bobeth et al. “Analysis of neutral Higgs boson contributions to the decays $\bar{B}_s \rightarrow \ell^+\ell^-$ and $\bar{B} \rightarrow K\ell^+\ell^-$ ”. *Phys.Rev.* D64 (2001), p. 074014. DOI: 10.1103/PhysRevD.64.074014. arXiv: hep-ph/0104284 [hep-ph].
- [58] C. Bobeth et al. “Enhancement of $\mathcal{B}(B_d \rightarrow \mu^+\mu^-)/\mathcal{B}(B_s \rightarrow \mu^+\mu^-)$ in the MSSM with modified minimal flavor violation and large $\tan\beta$ ”. *Phys.Rev.* D66 (2002), p. 074021. DOI: 10.1103/PhysRevD.66.074021. arXiv: hep-ph/0204225 [hep-ph].
- [59] M. Artuso et al. “B, D and K decays”. *Eur.Phys.J.* C57 (2008), pp. 309–492. DOI: 10.1140/epjc/s10052-008-0716-1. arXiv: 0801.1833 [hep-ph].

- [60] John F. Gunion et al. “Errata for the Higgs hunter’s guide” (1992). arXiv: [hep-ph/9302272](#) [[hep-ph](#)].
- [61] Heather E. Logan and Ulrich Nierste. “ $B_{s,d} \rightarrow \ell^+ \ell^-$ in a two-Higgs doublet model”. *Nucl.Phys.* B586 (2000), pp. 39–55. DOI: [10.1016/S0550-3213\(00\)00417-X](#). arXiv: [hep-ph/0004139](#) [[hep-ph](#)].
- [62] G. D’Ambrosio et al. “Minimal flavor violation: An Effective field theory approach”. *Nucl.Phys.* B645 (2002), pp. 155–187. DOI: [10.1016/S0550-3213\(02\)00836-2](#). arXiv: [hep-ph/0207036](#) [[hep-ph](#)].
- [63] C. Hamzaoui, M. Pospelov, and M. Toharia. “Higgs mediated FCNC in supersymmetric models with large $\tan \beta$ ”. *Phys.Rev.* D59 (1999), p. 095005. DOI: [10.1103/PhysRevD.59.095005](#). arXiv: [hep-ph/9807350](#) [[hep-ph](#)].
- [64] F. Mahmoudi. “Indirect search for New Physics: complementarity to direct searches” (2013). arXiv: [1310.2556](#) [[hep-ph](#)].
- [65] R. Aaij et al. “First Evidence for the Decay $B_s^0 \rightarrow \mu^+ \mu^-$ ”. *Phys.Rev.Lett.* 110 (2013), p. 021801. DOI: [10.1103/PhysRevLett.110.021801](#). arXiv: [1211.2674](#) [[hep-ex](#)].
- [66] David M. Straub. “New physics correlations in rare decays” (2010). arXiv: [1012.3893](#) [[hep-ph](#)].
- [67] Monika Blanke et al. “Rare K and B Decays in a Warped Extra Dimension with Custodial Protection”. *JHEP* 0903 (2009), p. 108. DOI: [10.1088/1126-6708/2009/03/108](#). arXiv: [0812.3803](#) [[hep-ph](#)].
- [68] Monika Blanke et al. “FCNC Processes in the Littlest Higgs Model with T-Parity: a 2009 Look”. *Acta Phys.Polon.* B41 (2010), pp. 657–683. arXiv: [0906.5454](#) [[hep-ph](#)].
- [69] Tobias Hurth et al. “Constraints on New Physics in MFV models: A Model-independent analysis of $\Delta F = 1$ processes”. *Nucl.Phys.* B808 (2009), pp. 326–346. DOI: [10.1016/j.nuclphysb.2008.09.040](#). arXiv: [0807.5039](#) [[hep-ph](#)].
- [70] Andrzej J. Buras et al. “Patterns of Flavour Violation in the Presence of a Fourth Generation of Quarks and Leptons”. *JHEP* 1009 (2010), p. 106. DOI: [10.1007/JHEP09\(2010\)106](#). arXiv: [1002.2126](#) [[hep-ph](#)].
- [71] Kaustubh Agashe and Christopher D. Carone. “Supersymmetric flavor models and the $B \rightarrow \phi K_S$ anomaly”. *Phys.Rev.* D68 (2003), p. 035017. DOI: [10.1103/PhysRevD.68.035017](#). arXiv: [hep-ph/0304229](#) [[hep-ph](#)].

- [72] Graham G. Ross, Liliana Velasco-Sevilla, and Oscar Vives. “Spontaneous CP violation and non-Abelian family symmetry in SUSY”. *Nucl.Phys.* B692 (2004), pp. 50–82. DOI: 10.1016/j.nuclphysb.2004.05.020. arXiv: hep-ph/0401064 [hep-ph].
- [73] Stefan Antusch, Stephen F. King, and Michal Malinsky. “Solving the SUSY Flavour and CP Problems with $SU(3)$ Family Symmetry”. *JHEP* 0806 (2008), p. 068. DOI: 10.1088/1126-6708/2008/06/068. arXiv: 0708.1282 [hep-ph].
- [74] Lawrence J. Hall and Hitoshi Murayama. “A Geometry of the generations”. *Phys.Rev.Lett.* 75 (1995), pp. 3985–3988. DOI: 10.1103/PhysRevLett.75.3985. arXiv: hep-ph/9508296 [hep-ph].
- [75] David M. Straub. “Anatomy of flavour-changing Z couplings in models with partial compositeness”. *JHEP* 1308 (2013), p. 108. DOI: 10.1007/JHEP08(2013)108. arXiv: 1302.4651 [hep-ph].
- [76] Bernard Aubert et al. “Search for decays of $B^0 \rightarrow$ mesons into e^+e^- , $\mu^+\mu^-$, and $e^\pm\mu^\mp$ final states”. *Phys.Rev.* D77 (2008), p. 032007. DOI: 10.1103/PhysRevD.77.032007. arXiv: 0712.1516 [hep-ex].
- [77] H. Albrecht et al. “B Meson Decays Into Charmonium States”. *Phys.Lett.* B199 (1987), p. 451. DOI: 10.1016/0370-2693(87)90952-X.
- [78] C. Albajar et al. “A search for rare B meson decays at the CERN Sp \bar{p} S collider”. *Phys.Lett.* B262 (1991), pp. 163–170. DOI: 10.1016/0370-2693(91)90660-I.
- [79] T. Bergfeld et al. “Search for decays of B^0 mesons into pairs of leptons: $B^0 \rightarrow e^+e^-$, $B^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow e^\pm\mu^\mp$ ”. *Phys.Rev.* D62 (2000), p. 091102. DOI: 10.1103/PhysRevD.62.091102. arXiv: hep-ex/0007042 [hep-ex].
- [80] M.C. Chang et al. “Search for $B^0 \rightarrow \ell^+\ell^-$ at BELLE”. *Phys.Rev.* D68 (2003), p. 111101. DOI: 10.1103/PhysRevD.68.111101. arXiv: hep-ex/0309069 [hep-ex].
- [81] Georges Aad et al. “Search for the decay $B_s^0 \rightarrow \mu^+\mu^-$ with the ATLAS detector”. *Phys.Lett.* B713 (2012), pp. 387–407. DOI: 10.1016/j.physletb.2012.06.013. arXiv: 1204.0735 [hep-ex].
- [82] Serguei Chatrchyan et al. “Search for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ decays”. *JHEP* 1204 (2012), p. 033. DOI: 10.1007/JHEP04(2012)033. arXiv: 1203.3976 [hep-ex].
- [83] R. Aaij et al. “Strong constraints on the rare decays $B_s \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ ”. *Phys.Rev.Lett.* 108 (2012), p. 231801. DOI: 10.1103/PhysRevLett.108.231801. arXiv: 1203.4493 [hep-ex].

- [84] T. Aaltonen et al. “Search for $B_s \rightarrow \mu^+ \mu^-$ and $B_d \rightarrow \mu^+ \mu^-$ decays with the full CDF Run II data set” (2013). arXiv: 1301.7048 [hep-ex].
- [85] Victor Mukhamedovich Abazov et al. “Search for the rare decay $B_s \rightarrow \mu \mu$ ”. *Phys.Rev.* D87 (2013), p. 072006. DOI: 10.1103/PhysRevD.87.072006. arXiv: 1301.4507 [hep-ex].
- [86] R. Aaij et al. “Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment”. *Phys.Rev.Lett.* 111 (2013), p. 101805. DOI: 10.1103/PhysRevLett.111.101805. arXiv: 1307.5024 [hep-ex].
- [87] CMS collaboration. *CMS B-Physics and Quarkonia Results*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsBPH>.
- [88] John C. Collins and Davison E. Soper. “The Theorems of Perturbative QCD”. *Ann.Rev.Nucl.Part.Sci.* 37 (1987), pp. 383–409.
- [89] A.D. Martin et al. “Parton distributions for the LHC”. *Eur.Phys.J.* C63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].
- [90] Elliott D. Bloom et al. “High-Energy Inelastic e - p Scattering at 6° and 10° ”. *Phys.Rev.Lett.* 23 (1969), pp. 930–934. DOI: 10.1103/PhysRevLett.23.930.
- [91] Martin Breidenbach et al. “Observed Behavior of Highly Inelastic electron-Proton Scattering”. *Phys.Rev.Lett.* 23 (1969), pp. 935–939. DOI: 10.1103/PhysRevLett.23.935.
- [92] E. Norrbin and T. Sjostrand. “Production and hadronization of heavy quarks”. *Eur.Phys.J.* C17 (2000), pp. 137–161. DOI: 10.1007/s100520000460. arXiv: hep-ph/0005110 [hep-ph].
- [93] Bo Andersson et al. “Parton Fragmentation and String Dynamics”. *Phys.Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.
- [94] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. *Comput.Phys.Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- [95] SeBong Chun and Charles Buchanan. “A Simple plausible path from QCD to successful prediction of $e^+ e^- \rightarrow$ hadronization data”. *Phys.Rept.* 292 (1998), pp. 239–317. DOI: 10.1016/S0370-1573(97)00047-1.
- [96] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “PYTHIA 6.4 Physics and Manual”. *JHEP* 0605 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

- [97] Rick Field. “Early LHC Underlying Event Data - Findings and Surprises” (2010). arXiv: 1010.3558 [hep-ph].
- [98] J. Pumplin et al. “New generation of parton distributions with uncertainties from global QCD analysis”. *JHEP* 0207 (2002), p. 012. arXiv: hep-ph/0201195 [hep-ph].
- [99] D.J. Lange. “The EvtGen particle decay simulation package”. *Nucl.Instrum.Meth.* A462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.
- [100] Oliver S. Bruning et al. “LHC Design Report. 1. The LHC Main Ring” (2004). Ed. by Oliver S. Bruning.
- [101] ATLAS collaboration. *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*. LHC Tech. Proposal. Geneva: CERN, 1994.
- [102] CMS collaboration. *CMS, the Compact Muon Solenoid : technical proposal*. LHC Tech. Proposal. Geneva: CERN, 1994.
- [103] LHCb collaboration. *LHCb: Technical Proposal*. Tech. Proposal. Geneva: CERN, 1998.
- [104] W Kienzle et al. *TOTEM, Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC: Technical Proposal*. Tech. rep. CERN-LHCC-99-007. LHCC-P-5. Geneva: CERN, Mar. 1999.
- [105] ALICE collaboration. *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC Tech. Proposal. Geneva: CERN, 1995.
- [106] O. Adriani et al. *LHCf experiment: Technical Design Report*. Technical Design Report LHCf. Geneva: CERN, 2006.
- [107] W. Bertsche et al. “The ALPHA experiment: A cold antihydrogen trap”. *AIP Conf.Proc.* 796 (2005), pp. 301–308. DOI: 10.1063/1.2130184.
- [108] W.J. Stirling. private communication. 2013.
- [109] S.Y. Lee. *Accelerator Physics*. World Scientific, 2004. ISBN: 9789812562005. URL: <http://books.google.it/books?id=VTc8Sdld5S8C>.
- [110] Frank Zimmermann and M. P. Zorzano-Mier. *Touschek Scattering in HERA and LHC*. Tech. rep. LHC-PROJECT-NOTE-244. Geneva: CERN, Dec. 2000.
- [111] S. Chatrchyan et al. “The CMS experiment at the CERN LHC”. *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.
- [112] CMS collaboration. *CMS Tracker Detector Performance Results*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.

- [113] Bo Li, Keisuke Fujii, and Yuanning Gao. “Kalman-filter-based track fitting in non-uniform magnetic field with segment-wise helical track model” (2013). arXiv: 1305.7300 [physics.ins-det].
- [114] Vardan Khachatryan et al. “CMS Tracking Performance Results from early LHC Operation”. *Eur.Phys.J.* C70 (2010), pp. 1165–1192. DOI: 10.1140/epjc/s10052-010-1491-3. arXiv: 1007.1988 [physics.ins-det].
- [115] Susanna Cucciarelli et al. *Track reconstruction, primary vertex finding and seed generation with the Pixel Detector*. Tech. rep. CMS-NOTE-2006-026. Geneva: CERN, Jan. 2006.
- [116] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. CMS-NOTE-2006-041. Geneva: CERN, Dec. 2006.
- [117] CMS collaboration. *Tracking and Vertexing Results from First Collisions*. Tech. rep. CMS-PAS-TRK-10-001. 2010. Geneva: CERN, 2010.
- [118] CMS collaboration. *CMS Tracking Results*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTRK>.
- [119] Serguei Chatrchyan et al. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV” (2012). arXiv: 1206.4071 [physics.ins-det].
- [120] G. L. Bayatian et al. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. Geneva: CERN, 2006.
- [121] CMS collaboration. *Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV*. Tech. rep. CMS-PAS-MUO-10-002. 2010. Geneva: CERN, 2010.
- [122] CMS collaboration. *Measurement of Tracking Efficiency*. Tech. rep. CMS-PAS-TRK-10-002. 2010. Geneva: CERN, 2010.
- [123] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. “Adaptive vertex fitting”. *J.Phys.G* G34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.
- [124] K. Rose. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. *Proceedings of the IEEE*. Vol. 86. Nov. 1998, pp. 2210 –2239.
- [125] CMS collaboration. *Tracking and Primary Vertex Results in First 7 TeV Collisions*. Tech. rep. CMS-PAS-TRK-10-005. 2010. Geneva: CERN, 2010.
- [126] K. Prokofiev and T. Speer. “A kinematic and a decay chain reconstruction library”. *Computing in High Energy Physics and Nuclear Physics 2004*. (Sept. 27–Oct. 1, 2004). Interlaken, Switzerland: CERN, 2005.

- [127] R. Brun and F. Rademakers. “ROOT: An object oriented data analysis framework”. *Nucl.Instrum.Meth.* A389 (1997), pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.
- [128] CMS collaboration. *CMSSW Application Framework*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework>.
- [129] Johan Alwall et al. “MadGraph 5 : Going Beyond”. *JHEP* 1106 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].
- [130] John Allison et al. “Geant4 developments and applications”. *IEEE Trans.Nucl.Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.
- [131] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. *Nucl.Instrum.Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [132] R. Aaij et al. “Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics”. *JHEP* 1304 (2013), p. 001. DOI: 10.1007/JHEP04(2013)001. arXiv: 1301.5286 [hep-ex].
- [133] CMS collaboration. *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*. Tech. rep. CMS-PAS-LUM-13-001. Geneva: CERN, 2013.
- [134] M. Oreglia. “A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$ ”. PhD thesis. 1980.
- [135] Vardan Khachatryan et al. “Prompt and non-prompt J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV”. *Eur.Phys.J.* C71 (2011), p. 1575. DOI: 10.1140/epjc/s10052-011-1575-8. arXiv: 1011.4193 [hep-ex].
- [136] Jan Therhaag. “TMVA: Toolkit for multivariate data analysis”. *AIP Conf.Proc.* 1504 (2009), pp. 1013–1016. DOI: 10.1063/1.4771869.
- [137] Z. Ligeti. private communication. 2012.
- [138] R. Aaij et al. “First observation of the decay $B^+ \rightarrow \pi^+\mu^+\mu^-$ ”. *JHEP* 1212 (2012), p. 125. DOI: 10.1007/JHEP12(2012)125. arXiv: 1210.2645 [hep-ex].
- [139] Dmitri Melikhov and Nikolai Nikitin. “Rare radiative leptonic decays $B_{d,s} \rightarrow l^+l^-\gamma$ ”. *Phys.Rev.* D70 (2004), p. 114028. DOI: 10.1103/PhysRevD.70.114028. arXiv: hep-ph/0410146 [hep-ph].
- [140] Vardan Khachatryan et al. “Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV”. *JHEP* 1103 (2011), p. 136. DOI: 10.1007/JHEP03(2011)136. arXiv: 1102.3194 [hep-ex].
- [141] Alexander L. Read. “Modified frequentist analysis of search results (The CL_s method)”. CERN-OPEN-2000-205 (2000).

- [142] Giovanni Punzi. “Comments on likelihood fits with variable resolution”. *eConf* C030908 (2003), WELT002. arXiv: physics/0401045 [physics].
- [143] Kyle S. Cranmer. “Kernel estimation in high-energy physics”. *Comput.Phys.Commun.* 136 (2001), pp. 198–207. DOI: 10.1016/S0010-4655(00)00243-5. arXiv: hep-ex/0011057 [hep-ex].
- [144] S.S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. *Annals Math.Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.
- [145] LHCb Collaboration. *Updated average f_s/f_d b -hadron production fraction ratio for 7 TeV pp collisions*. Tech. rep. LHCb-CONF-2013-011. July 2013.
- [146] Kaustubh Agashe et al. “Constraining RS Models by Future Flavor and Collider Measurements: A Snowmass Whitepaper” (2013). arXiv: 1310.1070 [hep-ph].
- [147] R. Mohanta. “Effect of scalar leptoquarks on the rare decays of B_s meson” (2013). arXiv: 1310.0713 [hep-ph].
- [148] Frederik Beaujean, Christoph Bobeth, and Danny van Dyk. “Comprehensive Bayesian Analysis of Rare (Semi)leptonic and Radiative B Decays” (2013). arXiv: 1310.2478 [hep-ph].
- [149] Andrzej J. Buras. “Towards the Identification of New Physics through Correlations between Flavour Observables” (2013). arXiv: 1309.7791 [hep-ph].
- [150] CMS Collaboration. *Technical proposal for the upgrade of the CMS detector through 2020*. Tech. rep. CERN-LHCC-2011-006. LHCC-P-004. Geneva: CERN, June 2011.
- [151] CMS Collaboration. *CMS reach in $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions for the new LHC runs*. Tech. rep. CMS-PAS-FTR-13-022. Geneva: CERN, 2013.

Acronyms

ALICE A Large Ion Collider Experiment.

ALPHA Antihydrogen Laser PHysics Apparatus.

AMS Anti-Muon Sample.

AOD Analysis Object Dataset data.

ATLAS A Toroidal LHC ApparatuS.

BDT Boosted Decision Tree.

BF Branching Fraction.

BPix Pixel Barrel.

BSM Beyond Standard Model.

CDF Collider Detector at Fermilab.

CERN Conseil Européenne pour la Recherche Nucléaire.

CKM Cabibbo-Kobayashi-Maskawa.

CL Confidence Level.

CMS Compact Muon Solenoid.

CMSSM Constrained Minimal Supersymmetric Standard Model.

CMSSW CMS SoftWare.

CS Control Sample.

CSC Cathode Strip Chamber.

CTF Combinatorial Track Finder.

DIGI DIGItization data.

DT Drift Tube.

EB Barrel calorimeter.

ECAL Electromagnetic CALorimeter.

EE Endcap calorimeter.

EW Electro-Weak.

FCNC Flavor Changing Neutral Current.

FE Flavor Excitation.

FPix Pixel Forward.

GEN event GENeration data.

GIM Glashow-Iliopoulos-Maiani.

GS Gluon Splitting.

GUT Grand Unified Theory.

HB Hadron Barrel.

HCAL Hadronic CALorimeter.

HE Hadron Endcaps.

HEP High Energy Physics.

HF Hadron Forward.

HLT High Level Trigger.

HO Hadron Outer.

IP Interaction Point.

L1 Level 1 Trigger.

LEP Large Electron-Positron Collider.

LHC Large Hadron Collider.

LHCb Large Hadron Collider beauty.

LHCf Large Hadron Collider forward.

LINAC LINear particle ACcelerator.

LO Leading Order.

LS1 Long Shutdown 1.

MC Monte Carlo.

MFV Minimal Flavour Violation.

MSSM Minimal Supersymmetric Standard Model.

mSUGRA minimal SUperGRAvity.

MVA Multivariate Analysis.

NLL Next to Leading Log.

NLO Next to Leading Order.

NP New Physics.

NS Normalization Sample.

OPE Operator Product Expansion.

PC Pair Creation.

PD Primary Dataset.

PDF Parton Distribution Function.

pdf probability density function.

PEE Per-Event Error.

pp proton-proton.

pQCD perturbative QCD.

PS Proton Synchrotron.

PSB Proton Synchrotron Booster.

PV Primary Vertex.

QCD Quantum Chromo-Dynamics.

QED Quantum Electro-Dynamics.

RAW RAW data.

RECO RECOnstructed data.

RF Radio Frequency.

RG Renormalization Group.

RMS Root Mean Square.

ROC Receiver Operating Characteristic.

RPC Resistive Plate Chamber.

SIM SIMulated hits data.

SM standard model.

SPS Super Proton Synchrotron.

SUSY SUpersYmmetry.

SV Secondary Vertex.

T0 Tier-0.

T1 Tier-1.

T2 Tier-2.

TEC Tracker EndCap.

TIB Tracker Inner Barrel.

TID Tracker Inner Disk.

TNP Tag & Probe.

TOB Tracker Outer Barrel.

TOTEM TOTal Elastic and diffractive cross section Measurement.

UML Unbinned Maximum Likelihood.