

Studies of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays with CMS experiment at LHC

By

CHANDIPRASAD KAR

Enrolment No: PHYS11201504013

National Institute of Science Education and Research, Bhubaneswar

A thesis submitted to the

Board of Studies in Physical Sciences

In partial fulfillment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE



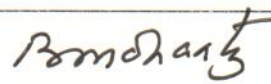
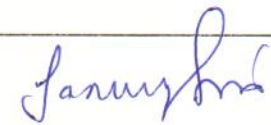
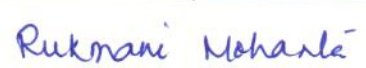
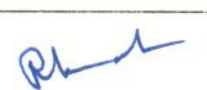
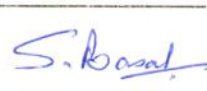
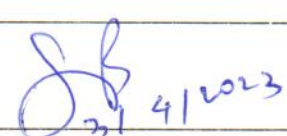
September, 2022



Homi Bhaba National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Chandiprasad Kar** entitled "**Studies of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays with CMS experiment at LHC**" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Prof. Bedangadas Mohanty		Date 3/4/2023
Guide - Prof. Sanjay Kumar Swain		Date 31/4/23
Examiner - Prof. Rukmani Mohanta		Date 3.4.2023
Member 1 - Dr. Prolay Kumar Mal		Date April 3, 2023
Member 2 - Dr. Subhasish Basak		Date 03/04/2023
Member 3 - Dr. Seema Bahinipati		Date 3.4.2023

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date : 31/4/23

Place : NISER

Signature

Co-guide (if any)


Signature

Guide

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Chandiprasad Kar.

Name & Signature of the Student

Chandiprasad Kar

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Chandiprasad Kar.

Name & Signature of the Student

Chandiprasad Kar

List of Publications arising from the thesis

Journal

1. CMS Collaboration, Albert M Sirunyan,...,C. Kar,.. et al., “Measurement of properties of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and search for $B^0 \rightarrow \mu^+ \mu^-$ with the CMS experiment.”, JHEP 04 (2020) 188, [https://doi.org/10.1007/JHEP04\(2020\)188](https://doi.org/10.1007/JHEP04(2020)188) (C. Kar is one of the primary author.)
2. CMS Collaboration, Albert M Sirunyan,...,C. Kar,.. et al., “Measurement of $B_s^0 \rightarrow \mu^+ \mu^-$ decay properties and search for the $B^0 \rightarrow \mu^+ \mu^-$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV”, arXiv:2212.10311v1, <https://doi.org/10.48550/arXiv.2212.10311>, Submitted to Physics Letters B (C. Kar is one of the primary author.)

Proceedings

1. CMS Collaboration, Chandiprasad Kar, “Rare B decays at CMS”, Proceedings of 19th International Conference on B-Physics at Frontier Machines, Published in proceeding of sciences. PoS (BEAUTY2020) 025, DOI: 10.22323/1.391.0025
2. Kar, C. (2022), “Recent Measurement on the $B \rightarrow \mu^+ \mu^-$ properties with CMS Data”, In: Mohanty, B., Swain, S.K., Singh, R., Kashyap, V.K.S. (eds) Proceedings of the XXIV DAE-BRNS High Energy Physics Symposium, Jatni, India. Springer Proceedings in Physics, vol 277. Springer, Singapore. https://doi.org/10.1007/978-981-19-2354-8_17

Symposium and Conferences

1. Participated and presented at Pheno2021: Phenomenology 2021 Symposium Title: “Recent results in Heavy Flavour and Quarkonia Physics from CMS” held at University of Pittsburgh, Video only (Virtual only) from 24-26 May 2021
2. Participated and presented at DAE-HEP 2020: 24th DAE-BRNS High Energy Physics Symposium 2020 Title: “Recent measurement on $B \rightarrow \mu\mu$ properties by using CMS data”, from 14-18 Dec 2020, Bhubaneswar (India)

3. Participated and presented at Beauty2020: 19th International Conference on B-Physics at Frontier Machines Title: "Search for rare decays" from 21-24 Sep 2020, Video only (Virtual World)
4. Participated and presented at CKM2018: International Workshop on the CKM Unitarity Triangle Title: "Dgamma and B Lifetime measurements at CMS" at Physikalisches Institut, Universität Heidelberg, Heidelberg (Germany) from 17-21 Sep 2018

Chandiprasad Kar,

Name & Signature of the Student

Chandiprasad Kar

DEDICATION

I would like to dedicate this to my late father, Hari Prasad Kar, and my mother, Bishnupriya Kar, as well as to my older brother, Ambika Prasad Kar. It is their love that enables me to see a clear picture of promise in every period of despair. I would also like to dedicate this to my wife, Abhipsa Mishra, for all her unwavering love, constant support, and endless encouragement.

ACKNOWLEDGEMENTS

Foremost, I want to offer my endeavor to GOD almighty for the wisdom he bestowed upon me, the strength, peace of mind, and good health in order to finish my thesis.

I express my sincere gratitude to my supervisor, Prof. Sanjay Kumar Swain, for imparting his knowledge and expertise in this study. I want to thank him for allowing me to work in CMS Collaboration, especially in the $B_s^0 \rightarrow \mu^+ \mu^-$ analysis group. Furthermore, I would like to extend my heartfelt thanks to Prof. Urs Langenegger, Prof. Kai-Feng Chen, and Dr. Dmytro Kovalskyi for their invaluable contributions to this analysis. Their guidance and support were crucial in navigating through the complexities of this research project. Their willingness to answer my questions and provide feedback on my work was a constant source of motivation for me, and I am truly grateful for their mentorship.

Completing my doctoral dissertation would not have been possible without the encouragement and support of my doctoral committee members. I am grateful to Prof. Bedangadas Mohanty, Dr. Prolay Kumar Mal, Dr. Subhashis Basak, and Dr. Seema Bahinipati for their valuable feedback and insightful comments on my research topic. Their diverse perspectives and expertise helped me to refine my work and provided me with a deeper understanding of the subject matter.

In addition, I would like to express my gratitude to NISER and DAE for providing financial support during my doctoral studies. Their generous support has enabled me to pursue my academic and research goals and has been instrumental in my success. I am truly grateful for the opportunity to receive financial assistance from such prestigious institutions.

I would like to thank Professors Rukmani Mohanta, Basanta Kumar Nandi, and Thomas Browder for their helpful comments on my thesis.

I am humbled and grateful for the knowledge and expertise that my supervisor and mentors have shared with me throughout this journey. Their continuous encouragement and support have inspired me to push beyond my limits and to strive for excellence in my work. I am honored to have had the opportunity to work with such exceptional mentors, and I look forward to carrying forward the lessons learned from them in my future endeavors.

I would like to thank all the engineers and technicians for their invaluable contributions to the construction of the CMS experiment and the CERN accelerator teams for the

outstanding performance of the LHC complex.

I thank my very helpful seniors at NISER: Dr. Niladri Bihari Sahoo, Dr. Koushik Mandal, and Dr. Protick Mohanta, for the stimulating discussion related to physics and life. I thank my fellow lab mates at NISER: Prafulla, Alope, Amit, Kuldeep, Deepak, Rakesh, Ashsis, Tribeni, Lipsa, and Priyanka, for all the useful discussion in the clean room and for all the fun we had in last few years. Also, I thank my friends: Gour, Bikash, Samir, Liki, Anupa, Purbasha, Joydev, Rani, Abhilash, Rakesh, Amit, Satya, Nirmal, Roma and Bidyadhar, who stood by my side throughout this journey. Their constant encouragement, support, and belief in me have helped me stay focused and determined, even in the face of adversity.

I would like to express my sincere appreciation and gratitude to Rajit Kumar Mishra and Laxmipriya Mishra, my in-laws, for their support and encouragement in last few years. Their belief in me and their constant support have been a source of inspiration and motivation for me, and I am truly grateful for their presence in my life.

Last but not least, I would like to thank my parents and my family (Abhipsa, my sisters Simu and Mamuni, my older brother Ambika, my sister-in-law Madhusmita, and others) for their unconditional love and constant encouragement. They have been my rock, my shelter in times of trouble, and my inspiration to strive for excellence. In my life, my nephew, Shivansh, has been like a ray of sunshine. His innocent and pure heart has given me the strength to pursue my passion, and for that, I am forever grateful.

Contents

Summary	2
List of Figures	4
List of Tables	17
Abbreviations	21
Chapter 1 Introduction	23
1.1 Decay of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in the SM	30
1.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ SM branching fraction	32
1.3 Time evolution of $B_{(s)}^0$ state	36
1.4 SM predictions	40
1.5 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction in BSM	41
1.6 Status of experimental results	44
Chapter 2 The LHC and CMS experiment	46
2.1 The Large Hadron Collider at CERN	46
2.1.1 The LHC design	48
2.1.2 The LHC luminosity	49
2.2 The Compact Muon Solenoid (CMS) experiments	52
2.2.1 The tracking system	55
2.2.2 The electromagnetic calorimeter (ECAL)	58
2.2.3 The hadronic calorimeter (HCAL)	61
2.2.4 Muon chamber	63
2.3 The trigger	65
2.3.1 Level-1 trigger	66
2.3.2 High-level trigger	66
2.4 Particle reconstruction and identification	67
2.4.1 Tracks and vertices reconstruction	69
2.4.2 Charge particle tracking	69
2.4.3 Muon reconstruction	70
2.5 B physics at the LHC	71

Chapter 3	Measurement of properties of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and search for	
	$B^0 \rightarrow \mu^+ \mu^-$	75
3.1	Analysis overview	75
3.2	Choice of normalization channel	77
3.3	Primary dataset and triggers	78
3.4	MC simulation	81
3.5	Muon selection	82
3.5.1	MVA muon identification	83
Chapter 4	Run 1 + Run 2 (2016) data analysis	88
4.1	Offline event selection	89
4.1.1	Primary vertex selection	89
4.1.2	Variables	90
4.2	Multivariate analysis for signal selection	92
4.2.1	Optimization, training, and characterization	94
4.3	Normalization ($B^+ \rightarrow J/\psi K^+$) and	
	control ($B_s^0 \rightarrow J/\psi \phi$) channel fitting	94
4.4	MC corrections	98
4.5	Data-MC simulation comparison	98
4.6	Validation of muon BDT in the data and MC	100
4.7	Estimation of background and signal events	105
4.8	Unbinned maximum likelihood fit for the branching fraction	108
4.8.1	Analysis strategy	108
4.8.2	BDT categories	111
4.8.3	Fragmentation fraction ratio (f_s/f_u)	113
4.8.4	Upper limit estimation for branching fraction	113
4.9	Unbinned maximum likelihood fit for the effective lifetime	114
4.9.1	Fit model	114
4.9.2	Mass distribution model	115
4.9.3	Decay time distribution model	116
4.9.4	Efficiency as a function of decay time	117

4.9.5	Fit procedure	117
4.9.6	Lifetime of B^+ from data	121
4.10	sPlot fit method for effective lifetime	124
4.11	Systematic uncertainties	125
4.11.1	Systematic uncertainties on the branching fraction	125
4.11.2	Systematic uncertainties on the effective lifetime	131
4.12	Results	141
4.12.1	Branching fraction	141
4.12.2	Effective lifetime using sPlot method	144
4.12.3	Effective lifetime using UML method	144
4.13	Summary	148
Chapter 5	Run 2 data analysis	149
5.1	Introduction	149
5.2	Dataset and trigger	150
5.3	Selection requirement	151
5.3.1	Vertexing variables	151
5.3.2	Offline selection requirement	152
5.4	Background suppression with multivariate analysis	153
5.4.1	Validation of MVA using $B^+ \rightarrow J/\psi K^+$ decay	154
5.4.2	MVA validation in data with $B^+ \rightarrow J/\psi K^+$ events	160
5.5	Normalization and control channel yields	162
5.6	Branching fraction measurement with UML fit	165
5.6.1	Estimation of the signal and background components	166
5.6.2	Nuisance parameters	169
5.6.3	Optimization of analysis d_{MVA} and muon BDT selection	170
5.7	Lifetime measurement with UML fit	172
5.7.1	Lifetime fit validation on $B^+ \rightarrow J/\psi K^+$ decay using d_{MVA}	177
5.8	MC corrections	180
5.8.1	Mass scale and resolution	180
5.8.2	MC corrections for analysis d_{MVA}	183

5.8.3	Decay time corrections	187
5.8.4	Muon fake rate correction	191
5.8.5	Trigger efficiency	193
5.9	Systematic uncertainty	195
5.9.1	Systematic uncertainties on the effective lifetime	195
5.9.2	Systematic uncertainties on branching fraction	198
5.9.3	Signal efficiency dependence on the lifetime	200
5.9.4	Systematics summary	202
5.10	Expected performance for branching fraction measurement	202
5.10.1	Expected performance using $B_s^0 \rightarrow J/\psi \phi$ as normalization	204
5.11	Expected performance for effective lifetime measurement	206
5.12	Fragmentation fraction treatment	207
5.13	Results	209
5.13.1	Branching fraction results with $B^+ \rightarrow J/\psi K^+$ as normalization	209
5.13.2	Lifetime measurement	210
5.14	Summary	213
Chapter 6	Conclusion and Outlook	214
Chapter A	Fake rate measurement	219
A.1	Fake rate measurement procedure	220
A.2	Muon selector	224
A.3	Results	226
Chapter B	More results on unblind fit	232
References		238

Summary

The fundamental constituents of matter and the laws of physics that govern the universe are well explained by the theory known as the Standard Model (SM) of particle physics. However, the model leaves several open questions, indicating the presence of new physics. The search for new physics is one of the primary goals for the scientists behind the building of the large high-energy collider experiments. The recent trend in particle physics follows either direct or indirect search methods for the search beyond Standard Model (BSM) physics. The direct search process is very limited because an enormous amount of center-of-mass energy is required to produce a new heavy particle. In contrast, an indirect search can be done by precisely measuring the properties of the SM decay process. Two such SM processes are the decay of the B -meson (B_s^0 and B^0) to the dimuon final state. The leptonic decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ are the effective flavor-changing neutral current processes, which are disallowed at the tree level in the SM. But they only proceed through the higher-level box and penguin diagrams. Additionally, the decays are helicity suppressed, making such decays very rare compared to others. Several BSM theories predict the enhancement of the branching fractions and effective lifetime. So the precise measurement of the properties will constrain the parameter space of several BSM theories.

In this thesis, the effort made by the CMS experiment to measure the $B_s^0 \rightarrow \mu^+ \mu^-$ properties and search for $B^0 \rightarrow \mu^+ \mu^-$ decay using the data collected by the CMS experiment in 2011 (7 TeV), 2012 (8 TeV), and 2016 (13 TeV) is presented. The signal B_s^0 and B^0 candidates are reconstructed using two oppositely charged muons from a displaced decay vertex. In data, along with signal, several background sources from b -hadrons contribute, and they are categorized into peaking, semileptonic, and combinatorial backgrounds. To control the challenging peaking background coming from the misidentified hadrons (like pion, kaon, and proton), a muon identification algorithm is developed using the Boosted Decision Tree (BDT). To suppress the dominant combinatorial background, a separate BDT discriminator is developed. The branching fraction for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays are extracted by performing an unbinned maximum likelihood (UML) fit to mass, mass uncertainty, and a binary configuration to distinguish the decay signature. Similarly, the effective lifetime of the B_s^0 meson in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay is measured by performing a UML fit to mass and decay time, where decay time uncertainty is treated as a conditional observable. The fit models are validated on the pseudo-experiments to check the model-induced bias on the parameter of interest. Several systematic uncertainties are evaluated

using the normalization channel ($B^+ \rightarrow J/\psi K^+$) and control channel ($B_s^0 \rightarrow J/\psi \phi$). The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction measured from the simultaneous fit to multiple BDT categories is $(2.9 \pm 0.7) \times 10^{-9}$, where the uncertainty is the combination of statistical and systematic uncertainties. This is the first observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ with 5.6 standard deviations by the CMS experiment. There is no evidence of the $B^0 \rightarrow \mu^+ \mu^-$ decay. Therefore, an upper limit on the branching fraction is assigned, $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.6 \times 10^{-10}$, at 95% confidence level (CL). The B_s^0 meson effective lifetime in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay mode, for the first time by the CMS experiment, is $\tau(B_s^0 \rightarrow \mu^+ \mu^-) = 1.70_{-0.44}^{+0.61}$ ps, where the uncertainty includes both statistical and systematic contributions. The measured branching fraction and the effective lifetime values are consistent with the SM predictions and the other experimental results.

The second part of the thesis discusses the same measurement performed using the data collected in 2016-2018 at center-of-mass energy of 13 TeV. This analysis aims to improve the precision of the branching fraction and the effective lifetime, using the previous analysis as a baseline, with more data and novel techniques. To improve the sensitivity of the analysis, the previously developed muon BDT identification algorithm is revisited, and a loose working point is selected. Secondly, the preselection cuts on the kinematic and topological variables have been loosened, and a new advanced multivariate algorithm (MVA) using the XGBoost package is trained to suppress the dominant backgrounds. For the optimized MVA working point, the branching fraction and the effective lifetime are then extracted by performing a UML fit to mass, mass uncertainty, decay time, and decay time uncertainty. Several MC corrections are derived using the $B^+ \rightarrow J/\psi K^+$ decay channel and applied in the signal $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ MC samples. The measured $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = [3.83_{-0.42}^{+0.45}] \times 10^{-9}$ and the effective lifetime of the B_s^0 meson is $\tau_{B_s^0} = 1.83_{-0.20}^{+0.23}$ ps, where the uncertainties are a combination of statistical and systematic uncertainties. No signal $B^0 \rightarrow \mu^+ \mu^-$ is observed, and the upper limit on the branching fraction $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ is set to be less than 1.9×10^{-10} at 95 % CL. The measured values are the most precise to date and consistent with the SM predictions and other experimental results. This novel measurement will have a significant impact on our ability to comprehend the flavour anomalies reported by other experiments and to set bounds on the BSM theory.

List of Figures

1.1	Standard Model in a nutshell [24].	25
1.2	Three coupling constants at different energy scale. α_1 , α_2 , and α_3 are the coupling constants for the electromagnetic, weak, and strong forces, respectively [30].	27
1.3	EW “box” (a) and EW “penguin” (b & c) are the main Feynman diagrams contributing to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay in SM. For the $B^0 \rightarrow \mu^+ \mu^-$ decay s quark is replaced with d quark.	31
1.4	(a) Z' model (b) 2HDM with Pseudo scalar boson (c) 2HDM with Scalar boson are few tree level Feynman diagrams contributing to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay in BSM.	42
1.5	The correlation between the ratio of branching fraction ($\bar{R}$) and $\mathcal{A}_{\Delta\Gamma}$ for different scenario where (a) $S = 0$, (b) $P = 1$, (c) $P \pm S = 1$, (d) $\varphi_P, \varphi_S \in \{0, \pi\}$. The plots are taken from the Ref. [43].	42
1.6	(a) The experimental results of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ decay measured so far are shown in this Figure. The blue markers are for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay from the experiments mentioned in the black marker. The red markers are for the upper limits on $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ decay [49]. The measurements described in Chapter 5 are not mentioned in the plot. (b) The SM branching fraction value evolution with time is shown. Note that the first prediction has a large error because, at that time, top quark was not discovered and many input parameters had large errors.	45
2.1	LHC timeline [52].	47
2.2	Illustration of the CERN accelerator complex [57].	49
2.3	Design of the magnetic dipole and the magnetic field direction in the main ring [55].	50
2.4	Total integrated luminosity versus day delivered to CMS during stable beams and for pp collisions in different year [58].	51
2.5	Stacked histogram to show the average pileup profile in different data taking year. $\langle \mu \rangle$ is the average pileup, which is calculated from the σ_{in}^{pp} as indicated on the plot [59].	52
2.6	Sketch of the CMS coordinate system [60].	54

2.7	Demonstration of the relation between the polar angle θ and pseudo-rapidity η [60].	54
2.8	Longitudinal schematic view of the CMS tracker in rz plane. The pixel detector represented above were replaced in the year 2017. Each line corresponds to a detector module. The strip tracker consists of tracker inner barrel (TIB), tracker inner disk (TID), tracker outer barrel (TOB) and tracker endcap (TEC) [61].	56
2.9	Comparison of the layout of current upgraded pixel detector and original pixel detector [62].	57
2.10	Total thickness of tracker material in the unit of radiation length as a function of pseudo-rapidity [55].	59
2.11	Longitudinal cross-sections of a quadrant of ECAL detector [55].	60
2.12	Longitudinal cross-section of a quadrant of the HCAL [55].	62
2.13	Longitudinal cross-section of a quadrant of the CMS detector showing different chambers of the muon system [55].	64
2.14	Dimuon invariant mass distribution collected by different HLT path in the year 2015 [65].	67
2.15	Interaction of different particle with the detector material [66].	68
2.16	Track reconstruction efficiency for muons and pions as a function of pseudo-rapidity for different transverse momentum [55].	70
2.17	Resolution of muon transverse momentum as a function of transverse momentum in barrel region (left) and endcap region (right) [55].	72
2.18	Standard Model cross-section of proton-proton and proton-antiproton [70] (top) and individual bottom quark production cross-sections are shown (bottom) [71].	74
3.1	Left and right plot shows the sketch of signal and combinatorial background events. Here PV stands for primary vertex.	76

3.2	The performance of the BDTs for the current analysis and the previous analysis for 2016 (top) and 2012 (bottom) conditions. The crosses mark the performance of a particular BDT cut mentioned in the legend, whereas the tilted crosses mark the performance of the official Tight Muon selection. Here the efficiencies are calculated by taking the ratio of number of signal or background event passing the BDT selection to total number signal or background events in MC sample.	86
4.1	Visual representation of a B decay; the decay length l_{3D} , pointing angle α_{3D} , impact parameter δ_{3D} , the dimuon momentum $\vec{p}_{\mu\mu}$ and the longitudinal impact parameter l_z are shown (left figure). An isolated cone is shown in the right plot.	91
4.2	Overtraining checks for the three BDTs for 2016BF. The top row shows the BDT response in a linear scale, the bottom row shows the same plots in a logarithmic scale.	95
4.3	Invariant mass distributions of the selected $B^+ \rightarrow J/\psi K^+$ candidates for the year 2016BF (top), and 2016GH (bottom). The left and right column plots are for channel 0 and channel 1, respectively. The continuous curve (black) is for the combined fit and different hatched histograms are for signal and background components. The solid black points are data.	96
4.4	Invariant-mass distributions of the selected $B_s^0 \rightarrow J/\psi \phi$ candidates for the year 2016BF (top), and 2016GH (bottom). The left and right column plots are for channel 0 and channel 1, respectively. The continuous curve (black) is for the combined fit and different hatched histograms are for signal and background components. The solid black points are data.	97

4.5	$B^+ \rightarrow J/\psi K^+$: Illustrations of the sideband subtraction methodology in 2016BF channel 0. The left column shows the fitted mass distribution indicating the low (red arrows) and high (blue arrows) sidebands together with the signal region (black arrows). The middle column shows the shape of the distributions for the three regions, each normalized to unity. The right column shows the signal distribution obtained after subtracting from the signal-region distribution the contributions from the combinatorial and partially-reconstructed background.	99
4.6	Illustrations of the sideband subtracted data (solid markers) and MC simulation (hatched histogram) comparison for few variables in $B^+ \rightarrow J/\psi K^+$ decay mode: the subleading muon p_T , the flight length significance ($l_{3D}/\sigma(l_{3D})$), N_{trk}^{close} , the pointing angle (α_{3D}), the decay time (t) and the analysis BDT discriminator. The comparison is for 2016BF channel 0. The MC simulation is normalized to the same number of events as the data. The shaded region in the ratio plot indicates a $\pm 20\%$ variation.	101
4.7	Illustrations of the sideband subtracted data (solid markers) and MC simulation (hatched histogram) comparison for few variables in $B_s^0 \rightarrow J/\psi \phi$ decay mode: the subleading muon p_T , the flight length significance ($l_{3D}/\sigma(l_{3D})$), N_{trk}^{close} , the pointing angle (α_{3D}), the decay time (t) and the analysis BDT discriminator. The comparison is for 2016BF channel 0. The MC simulation is normalized to the same number of events as the data. The shaded region in the ratio plot indicates a $\pm 20\%$ variation.	102
4.8	2016 hadron misidentification probability in data and MC simulation. The top and bottom row shows the probabilities for channel 0, and channel 1, correspondingly. The systematic error is calculated from the error of the uncertainty-weighted average of data and MC simulation in the two regions.	103
4.9	Invariant mass distribution of rare semileptonic (left column), and peaking (right column) backgrounds for 2016BF channel 0 (top row) and channel 1 (bottom row).	107

- 4.10 Efficiency plots for channel 0 (top row) and channel 1 (bottom row) in 2011, 2012, 2016BF and 2016GH $B^+ \rightarrow J/\psi K^+$ samples (from left to right). For 2011 and 2012, the MC is generated with lifetime 1.67 ps where as for 2016, it is 1.637 ps. The distributions are fitted with efficiency function, as mentioned in Eq. 4.17. 118
- 4.11 Efficiency plots for channel 0 (top row) and channel 1 (bottom row) in 2011, 2012, 2016BF and 2016GH $B_s^0 \rightarrow \mu^+ \mu^-$ samples (from left to right). The efficiency distribution is fitted with the function mentioned in Eq. 4.17. . . 118
- 4.12 Invariant mass and decay time fit projections in $B^+ \rightarrow J/\psi K^+$ decay mode for 2011 channel 0 and 2012 channel 1. Solid blue curve is for the total PDF. The signal and background components are shown as dotted curves. Pull plots are shown in the bottom pad, which indicates the goodness of fit. . . . 122
- 4.13 Invariant mass and decay time fit projections in $B^+ \rightarrow J/\psi K^+$ decay mode for 2016BF channel0 and 2016GH channel 1. Solid blue curve is for the total PDF. The signal and background components are shown as dotted curves. Pull plots are shown in the bottom pad, which indicates the goodness of fit. 123
- 4.14 Dependence of the selection efficiency on the (unknown) effective lifetime of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay for 2016BF channel 0 (left column) and channel 1 (right column). The fit with a polynomial of first degree always passes through the point of the CMS MC simulation because that point has much more statistics than other points. Note that in the plots A is used as an abbreviation for $\mathcal{A}_{\Lambda\Gamma}^{\mu^+\mu^-}$ 127
- 4.15 Data fit using the production process from the MC for 2016 $B^+ \rightarrow J/\psi K^+$ (top row) and $B_s^0 \rightarrow J/\psi \phi$ (bottom row) samples. The left and right columns are for loose and tight BDT selections. The red distribution is for gluon fusion plus flavor excitation, and the green distribution is for gluon splitting. The process fractions are mentioned in the plot legend. 128

4.16	Ratio of trigger efficiency of the normalization sample to the signal sample as a function of the BDT threshold. These plots are used to estimate the systematic uncertainty on the trigger efficiency ratio by varying the BDT selection criteria. The red line corresponds to the default BDT cut, the magenta line is at the BDT value where the BDT-cut related efficiency has dropped by 50%, and the cyan line is displaced by the same distance from the red line. The trigger efficiency ratios corresponding to these lines correspond to the numbers labeled as $\pm 50\%$. The numbers labeled as $\Delta(+)$ or $\Delta(-)$ correspond to the variation of the trigger efficiency ratio over the entire BDT range shown.	129
4.17	Three efficiency functions overlapped on the efficiency histogram for $B^+ \rightarrow J/\psi K^+$ sample in 2011 channel 0 and 2016GH channel 1. The blue, green, and red curves are from Eqs. 4.17, 4.19, and 4.20, respectively.	132
4.18	The top row shows results from 2016BF MC sample and bottom row plots are using 2016GH MC sample. The top-left plot shows the fitted lifetime from different samples (with a generated lifetime in the range of 1.40 ps - 1.80 ps), while using efficiency from 1.47 ps MC sample. Other similar plots are shown with different efficiency functions.	133
4.19	The invariant mass, decay time and decay time uncertainty fit projections from the 8 different channels obtained from the $B_s^0 \rightarrow \mu^+ \mu^-$ signal MC samples (2011, 2012, 2016BF, and 2016GH).	135
4.20	The invariant mass, decay time and decay time uncertainty fit projections from the 8 different channels obtained from the peaking MC samples (2011, 2012, 2016BF, and 2016GH).	136
4.21	The invariant mass, decay time and decay time uncertainty fit projections from the 8 different channels obtained from the semileptonic MC samples (2011, 2012, 2016BF, and 2016GH).	137
4.22	Combinatorial PDF shapes (mass, decay time, decay time uncertainty) for channel 0 and channel 1 from 2011, 2012, 2016BF and 2016GH high side-band data ($m \in [5.45, 5.9]$ GeV) without MVA muon BDT cut, used for toy generation.	138

4.23	The above figure shows B_s^0 lifetime, B_s^0 lifetime positive and negative error and the pull distribution from the 1500 toy sample. The pull is calculated using the generation value of lifetime to be 1.615 ps. The top row is for 2011-2012 era, middle row is for the 2016 era and the bottom row is the combination of all era.	139
4.24	Mass projections from the final UML fits in different categories. The total fit is shown in solid blue curve, and the different background components are shown in dotted curves. The signal $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ components are shown in the red and violet hatched histograms.	143
4.25	Combined mass projections for the high and low BDT-threshold categories. No event weighting is applied.	144
4.26	(Right) Decay time distribution and fit result for the sPlot lifetime fit setup. Note that no requirement on the decay time (> 1 ps) is applied. (Left) Merged mass projection of all the contributing data.	145
4.27	Comparison of the 2D UML fitted results to the expectations for the effective lifetime. From left to right the central value, the upper error, and the lower error are shown.	145
4.28	The data overlayed with fit projections for the dimuon invariant mass (top) and the proper decay time distribution distributions in linear (bottom) scale after merging all the channels. The total fit is shown in solid blue curve, and the different background components are shown in dotted curves and hatched histograms. The signal component is shown in the red hatched histogram.	146
4.29	Fit projection on the dimuon mass and decaytime distributions for 2011 (first column from the left), 2012 (second column from the left), 2016BF (second column from the right), and 2016GH (first column from the right). The top two rows show channel 0, the bottom two rows show channel 1. The slope of the Bernstein polynomial for 2011 chan1 is fixed to zero because of the low statistics.	147

5.1	d_{MVA} distribution for $B_s^0 \rightarrow \mu^+ \mu^-$ MC simulation and data sideband background (left) and ROC curves (right) for three different event splittings.	155
5.2	d_{MVA} and flight length significance distributions for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ events.	156
5.3	Point angle distributions for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ events in MC. All events (left), kaon $p_T > 3.0$ GeV (center), kaon $p_T < 1.5$ GeV (right) are shown. Pointing angle of $B^+ \rightarrow J/\psi K^+$ is measured using B^+ momentum.	157
5.4	Comparison of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ MC distribution for different variables.	158
5.5	Invariant mass distributions along with UML fit projections for $B^+ \rightarrow J/\psi K^+$ events in 2016BF, 2016GH, 2017, and 2018 data. The result of the fit is overlaid (blue curve) and the different components detailed in the legend.	159
5.6	Distributions of the d_{MVA} output (left), the impact parameter with respect to PV in 3D (center) and its significance (right) in 2016BF, 2016GH, 2017, and 2018 data samples (from top to bottom).	161
5.7	The $B^+ \rightarrow J/\psi K^+$ (top row) and $B_s^0 \rightarrow J/\psi \phi$ (bottom row) invariant mass distributions from 2018 channel 0 (left column) and channel 1 (right column) are shown, along with the fit projections.	163
5.8	The simulated distribution of dimuon invariant mass of $B_s^0 \rightarrow \mu^+ \mu^-$ (left), $B \rightarrow h^+ h^-$ (middle) and semileptonic (right) backgrounds and its PDF for 2018 channel 0. All channels and eras share the same semileptonic PDF.	168
5.9	The simulated distribution of the relative mass resolution of $B_s^0 \rightarrow \mu^+ \mu^-$ (left), $B \rightarrow h^+ h^-$ (middle) and semileptonic (right) backgrounds and its PDF for 2018 channel 0.	168

5.10	Significance of B_s^0 (left) and B^0 (right) in terms of different analysis d_{MVA} and muon BDT selections in 2018 channel 0. Other years and channel plots have similar trends which are not shown here. Here X-axis represents the muon BDT requirement and different labels represents the analysis d_{MVA} requirement with benchmark $d_{MVA} > 0.97 + X$ (where X is mentioned in the legend.)	170
5.11	Significance of B_s^0 (left) and B^0 (right) in terms of different analysis d_{MVA} in different years and channels.	171
5.12	The significance of B_s^0 (left) and B^0 (right) signal in terms of different d_{MVA}^{high} with muon BDT > 0.45 , $d_{MVA}^{low} = 0.93$	171
5.13	The relative lifetime error in terms of different requirements of analysis d_{MVA} and muon BDT in 2018 channel 0 (left) and channel 1 (right). Other eras and channels have similar trend. Here x-axis represents the muon BDT requirement and different labels represents the analysis d_{MVA} requirement with benchmark $d_{MVA} > 0.97 + X$ (where X is mentioned in the legend).	172
5.14	Correlation between the B meson point angle and its decay distance (left) and decay time (right) for $B_s^0 \rightarrow \mu^+ \mu^-$ events.	178
5.15	Measured lifetime from the data (red marker) and MC (blue marker) using different d_{MVA} points for 2016BF, 2016GH, 2017, and 2018 data samples.	178
5.16	Invariant mass, decay time and decay time uncertainty fit projections in data after $d_{MVA} > 0.90$ selection for the year 2016BF. The final fit PDF is shown as the solid blue curve. The signal and other background components are shown as the dotted curves. Pull plots are shown in the bottom pads to visualize the goodness of fit.	179
5.17	Fit projection in $\Upsilon(1S) \rightarrow \mu^+ \mu^-$ and $J/\psi \rightarrow \mu^+ \mu^-$ channel from MC and Data for $p_T \in [12.0, 15.0]$ in 2018 channel 0.	181
5.18	Relative mean shift from J/ψ and $\Upsilon(1S)$ for channel 0 (top row) and channel 1 (bottom row) in 2016GH (left column) and 2018 (right column) data samples.	182
5.19	Data and MC comparison before and after reweighting for few important variables in the 2017 + 2018 sample.	186

5.20	Analysis d_{MVA} comparison (in plot it is written as d_{MVA}) in data and MC for both weighted and unweighted distributions for 2016BF (left), 2016GH (center) and 2017 + 2018 (right). The bottom plots represent Data/MC ratios for weighted MC (blue marker) and unweighted MC (red marker). The MC distributions are normalized to the total number of events in data.	186
5.21	Impact of the decay time efficiency correction on MC(left) and data(right) with the $d_{\text{MVA}} > 0.99$ for the year 2016BF. Black markers are the original distributions without correction. Blue markers are the distribution after the MC efficiency correction. Red markers in the right plot are after the data efficiency correction. The distributions are unit normalized.	188
5.22	Ratios of the decay time efficiencies for the tight ($d_{\text{MVA}} > 0.99$) and loose ($d_{\text{MVA}} > 0.90$) selections observed in 2016BF, 2016GH, 2017, and 2018 data and MC simulations for $B^+ \rightarrow J/\psi K^+$ decays. The difference between the two ratios is used to correct the decay time efficiency for $B_s^0 \rightarrow \mu^+ \mu^-$ decays.	192
5.23	HLT trigger efficiency in data and simulated $B_s^0 \rightarrow \mu^+ \mu^-$ events as a function of the muon momentum for channel 0 (left) and channel 1 (right) for 2018.	194
5.24	The results are from 2018 MC sample. The left plot shows the fitted-generated (y-axis) lifetime from different samples (with a generated lifetime in the range of 1.40 ps - 1.80 ps), while using the efficiency from the 1.50 ps MC sample. Similar plots are shown with different efficiency functions.	196
5.25	Combined fit projections for d_{MVA} (0.2, 0.8), d_{MVA} (0.90, 1.0), and d_{MVA} (0.99, 1.0). Single and double exponential functions are shown in red and magenta color. The combined fit is performed after merging data in different channels and eras.	198
5.26	The top row (hypothesis 1) and bottom row (hypothesis 2) of the figure show the B_s^0 lifetime, B_s^0 lifetime positive and negative error and the pull distribution from the 3920 toy sample. The pull is calculated using the generation value of lifetime to be 1.615 ps.	199

5.27	Signal efficiency from 2018 channel 0 (top) and the median of $B_s^0 \rightarrow \mu^+ \mu^-$ BF obtained from the toy MC simulation for different lifetime assumptions (bottom) are shown. Error on each point (in the bottom plot) corresponds to the median of the BF error.	201
5.28	From left to right, the distributions of fitted mean, hi/low uncertainties, pull, and expected significance for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (upper row) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (bottom row) from the pseudo-experiments. The results are obtained from the “2-bin” configuration with fits to 16 categories.	204
5.29	The top row of the figure shows B_s^0 lifetime, B_s^0 lifetime positive and negative error, and the pull distribution from the 5000 toy sample. The pull is calculated using the generation lifetime value of 1.62 ps. The bottom row shows the B_s^0 branching fraction, branching fraction positive and negative error, and the branching fraction pull distributions. The fitted means of the pull distributions are at zero.	207
5.30	The distribution of the B meson p_T from the sPlot fit in data is compared with simulated B meson p_T in both $B^+ \rightarrow J/\psi K^+$, and $B_s^0 \rightarrow \mu^+ \mu^-$ region.	208
5.31	The merged dimuon invariant mass distributions of the unblinded fit in the $d_{\text{MVA}} > 0.99$ category (left) and $0.99 > d_{\text{MVA}} > 0.90$ category (right), with the PDF models used in the generation overlaid.	210
5.32	The left plot is the upper limits on $B^0 \rightarrow \mu^+ \mu^-$ decay branching fraction from the full toy-based CLs technique, and the right plot is the two-dimensional profile likelihood scan for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$. The 2D contours enclose the regions with 1–5 standard deviation coverage, where the 1σ , 2σ , and 3σ regions correspond to 68.3%, 95.4%, and 99.7% the confidence levels, respectively.	211
5.33	The merged dimuon invariant mass, decay time and decay time uncertainty distribution in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit in the data. The channel-wise mass, decay time and decay time uncertainty are projected in Figs. B.3-B.5.	212

5.34	The merged decay time distribution in the signal region [5.28, 5.48] GeV is shown. The combined fit PDF along with signal and different background components are projected.	212
6.1	Summary of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction (top), the B_s^0 meson effective lifetime (middle) in the $B_s^0 \rightarrow \mu^+ \mu^-$ mode and the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction upper limit (bottom) from different experiments.	216
A.1	Muon fake rates from $K_s^0 \rightarrow \pi\pi$ decays as a function of the transverse flight length from MC simulation (left) and data (right) for all the year.	223
A.2	Invariant mass projections from the $K_s^0 \rightarrow \pi\pi$ (top row), $\phi \rightarrow KK$ (middle row) and $\Lambda \rightarrow \mu\pi$ (bottom row) decay. The fit projections for K_s^0 and ϕ are in the μp_T range 8-12.0 GeV, where as projection with full p_T range is shown for Λ . The left and right side plots are before and after soft muon BDT cut, respectively. The green dashed curve is for the signal distribution, the red dotted curve is for combinatorial background and the result of the fit projection is the solid blue curve. The data are shown as the solid black circles with error bars.	225
A.3	Pion fake rates from 2018 data and MC as a function muon p_T for various selections.	227
A.4	Kaon fake rates from 2018 data and MC as a function muon p_T for various selections.	228
B.1	The dimuon invariant mass distributions of the unblinded fit in the $d_{MVA} > 0.99$ categories, with the PDF models used in the generation overlaid. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively.	233

B.2	The dimuon invariant mass distributions of the unblinded fit in the $0.99 > d_{\text{MVA}} > 0.90$ categories, with the PDF models used in the generation overlaid. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively.	233
B.3	The dimuon invariant mass distributions of the unblinded fit in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit. The mass distributions are obtained from the decay time range $[1,14]$ ps. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively. The result of the fit is overlaid (blue line) and the different components detailed in the legend.	234
B.4	The decay time distributions (on a log scale) in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively. The result of the fit is overlaid (blue line) and the different components detailed in the legend.	234
B.5	The decay time uncertainty distributions (on a log scale) in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively. The result of the fit is overlaid (blue line) and the different components detailed in the legend.	236
B.6	Visualization of $B_s^0 \rightarrow \mu^+ \mu^-$ candidate in 2018 data. The two muons produced in the disintegration of the B_s^0 meson are represented by the red lines. Additional curved lines depict charged-particle tracks that also originate from the same collision point. The calorimeter's blue and red towers represent the deposits of hadrons, electrons, and photons, respectively. Tracks originating from other interactions during the event have been eliminated to enhance clarity [105].	237

List of Tables

1.1	A summary of fundamental forces.	23
3.1	Kinematic selections at HLT level for the signal and different used trigger.	80
3.2	Primary analysis triggers and their parameters. Charge, DOCA, VtxProb, $\cos \alpha$, and LxySignif stand for the charge of two muons, distance of closest approach between the two muons, dimuon vertex probability, cosine of the pointing angle, and the flight length significance of the dimuon vertex in xy-plane, respectively. HLT DoubleMu4 3 Bs is used for the signal channel. HLT DoubleMu4 3 Jpsi Displaced and HLT DoubleMu4 3 Jpsi are used for the normalization as well as control channels.	81
4.1	Preselection cuts on variables for BDT training/testing.	93
4.2	Branching fractions of all the decay modes used in this analysis.	106
4.3	The final BDT selection used for the branching fraction determination in different channel, category, and data taking period. In 2011, only one category is selected because of low statistics.	112
4.4	Optimized BDT discriminator boundaries per channel, and running period for the effective lifetime determination.	113
4.5	The 2D UML fit results to the individual channels, using the the signal MC as well as collision data. The events samples are taken after the BDT cut as mentioned in the 2nd column.	121
4.6	A comparison of the effective lifetime from $B^+ \rightarrow J/\psi K^+$ data and the difference with PDG lifetime for different eras and channels.	124
4.7	Summary of systematic uncertainty sources described in the text. The uncertainties quoted for the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ are relative uncertainties. The Era dependence systematic uncertainty is derived from the MC correction method for era-dependent variation on the ratios between $N(\mu\mu)$ from $B \rightarrow \mu^+ \mu^-$ sideband and $B^+ \rightarrow J/\psi K^+$ events.	130
4.8	The 2D UML fit results to the individual channels from $B^+ \rightarrow J/\psi K^+$ channel, using the collision data as well as the variation of BDT efficiency.	133

4.9	Lifetime results in different ϕ region of B meson obtained from the total 2016 $B^+ \rightarrow J/\psi K^+$ sample in different channel. The last column is the difference from the simultaneous fit to channel 0 and channel 1 result to the lifetime obtained from the total ϕ region (0 to 2π).	140
4.10	Summary of the lifetime systematics. The contributions that are included in other items are indicated by (*).	141
4.11	Summary of the fitted yields (for $B_s^0 \rightarrow \mu^+\mu^-$, $B^0 \rightarrow \mu^+\mu^-$, the combinatorial background for $4.9 < m_{\mu\mu} < 5.9$ GeV, and the $B^+ \rightarrow J/\psi K^+$ normalization), the average p_T of the $B_s^0 \rightarrow \mu^+\mu^-$ signal, and the ratio of efficiencies between the normalization and the signal in all 14 categories of the 3D UML branching fraction fit. The size of peaking background is approximately 5–10% of the $B^0 \rightarrow \mu^+\mu^-$ signal. The average p_T (in GeV) is calculated from the MC simulation and has negligible uncertainties. The errors shown include the statistical and systematic components. It should be noted that the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ yield errors (1) are determined from the branching fraction fit and (2) also include the normalization errors.	142
5.1	Selection summary without analysis d_{MVA} .	152
5.2	$B_{(s)}^0 \rightarrow \mu^+\mu^-$ d_{MVA} input variables and their importance scores.	153
5.3	d_{MVA} input variables used to match $B^+ \rightarrow J/\psi K^+$ to $B_{(s)}^0 \rightarrow \mu^+\mu^-$.	157
5.4	$B^+ \rightarrow J/\psi K^+$ event yield weighted by the trigger prescale for the nominal and alternative fits.	164
5.5	$B_s^0 \rightarrow J/\psi \phi$ event yield weighted by the trigger prescale for the nominal and alternative fits.	164
5.6	Models used to describe the signal background mass and mass resolution. Here the CB, G and Keys stands for Crystal-Ball, Gaussian and kernel estimator function. The width parameter (σ_{CB}) of the CB-line is a conditional parameter linearly depending on the dimuon mass resolution. Here κ is the multiplicative factor.	166
5.7	Branching fractions of charmless two body B decays and 3-body B decays.	168

5.8	Detail list of nuisance parameters and their purpose and implementation.	169
5.9	Summary of the mean shift between data and MC for B_s^0 in different years and channels.	183
5.10	Ratio of the mass resolution scales in data and MC for B_s^0 for different years and channels.	183
5.11	Efficiencies obtained from MC samples with (labelled as MC _{corr}) and without (labelled as MC) XGBoost reweighting. The efficiencies obtained from the sWeighted data are labelled as Data.	185
5.12	Normalization corrections for $B_s^0 \rightarrow \mu^+\mu^-$ derived using two different methods: XGBoost reweighting and the efficiency ratio between data and MC in $B^+ \rightarrow J/\psi K^+$ events.	187
5.13	Lifetime obtained in $B_s^0 \rightarrow \mu^+\mu^-$ MC using $B_s^0 \rightarrow \mu^+\mu^-$ only corrections (τ_s) and a mix of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ corrections for $d_{\text{MVA}} > 0.99$ ($\tau_{s'}$). The generated lifetime for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ are 1.609 ps and 1.638 ps.	190
5.14	Lifetime obtained in $B^+ \rightarrow J/\psi K^+$ data using the correction from $B^+ \rightarrow J/\psi K^+$ data for $d_{\text{MVA}} > 0.99$. Here the MC value is 1.638 ps.	191
5.15	Systematics obtained from the correction for $d_{\text{MVA}} > 0.99$. Here the PDG value is 1.638 ps.	191
5.16	Pion and kaon fake rate ratios for data over MC simulation for 2016, 2017, and 2018.	193
5.17	Summary of the systematic uncertainties for the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ branching fraction measurements.	202
5.18	Summary of systematic uncertainties for $B_s^0 \rightarrow \mu^+\mu^-$ effective lifetime measurement (ps).	202
5.19	Expected yields for each component and for each category included in the UML implementation. These yields are used in the toy event generation for the expected performance estimations.	203
5.20	Summary of expected relative uncertainties on $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$, as well as the expected significance from the pseudo-experiments in “1-bin” and “2-bin” configurations.	204

5.2.1	Summary of expected relative uncertainties on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, as well as the expected significance from the pseudo-experiments in “1-bin” and “2-bin” configurations using $B_s^0 \rightarrow J/\psi \phi$ normalization.	205
A.1	Pion fake rate ratio between data and MC simulation for various selections in 2018.	229
A.2	Kaon fake rate ratio between data and MC simulation for various selections in 2018.	229
A.3	Proton fake rates in MC, data and the ratio between data and MC simulation for various thresholds with all available data. For the default and BDT > 0.60 selection, we estimate the fake rate at 95% CL.	230
B.1	A comparison of branching fraction fits with subsets of data.	232
B.2	The χ^2 test for the mass projection for each category after the unblinded fit.	232
B.3	A comparison of branching fraction and lifetime with subsets of data.	235
B.4	The χ^2 test for the mass and decay time projection of each category from the unblinded lifetime fit.	235

LIST OF ABBREVIATIONS

LHC (Large hadron collider)

QCD (Quantum chromodynamics)

SM (Standard model)

BSM (Beyond Standard model)

FCNC (Flavor Changing Neutral Current)

COM (Center-of-mass)

NP (New physics)

CP (Charge-parity)

CMS (Compact Muon Solenoid)

ATLAS (A Toroidal LHC Apparatus)

ALICE (A Large Ion Collider Experiment)

LHCb (LHC Beauty Experiment)

CERN (Conseil Européenne pour la Recherche Nucléaire)

HLT (High-level trigger)

L1 (Level 1)

PF (Particle flow)

PDF (Parton distribution functions)

SV (Secondary vertex)

PV (Primary vertex)

DCA (Distance of closest approach)

BDT (Boosted decision tree)

MVA (Multivariate analysis)

UML (Unbinned maximum likelihood)

$\mu\mathbf{m}$ (Micrometer)

mm (Milimeter)

GeV (Giga electron volt)

TeV (Tera electron volt)

Hz (Hertz)

MHz (Mega hertz)

$\mathbf{fb}^{-1}$ (Femtobarn inverse)

CPU (Central processing unit)

χ^2 (Chi square)

ps (pico second)

η (pseudo-rapidity)

T (Tesla)

MC (Monte Carlo)

PU (Pile up)

Chapter 1

Introduction

The development of particle physics is to better understand the fundamental laws of nature. The Standard Model (SM) of particle physics is the well-established Quantum Field Theory (QFT) developed over sixty years. It describes the constituent of matter and their interactions. The SM predicts that each particle is made up of the combination of the quarks and their antiparticles. The interactions between them are well described by the strong, electromagnetic, and weak forces. As of the SM, there are six quarks (u, d, c, s, t , and b) and six antiquarks. There are also six leptons ($e^-, \nu_e, \mu^-, \nu_\mu, \tau^-, \nu_\tau$) and six antileptons. The quarks can interact through all three forces, whereas leptons do not interact via the strong force. Gravity, one of the fundamental forces of nature, is not included in the SM because its QFT is not developed. The electromagnetic interaction is mediated by mass-zero spin-1 boson (photon), and only the particles which have electric charge interact with the photon. Similarly, the weak interaction is governed by massive spin-1 bosons (Z^0 and $W^\pm$). Every particle interacts weakly except the photon. Finally, the quarks interact strongly via the gluon field. Table [1.1](#) shows different fundamental forces and their range. The gluon

Table 1.1: A summary of fundamental forces.

Force	Carriers	Range [m]
Strong	Gluons	10^{-15}
Electromagnetic	Photons	∞
Weak	Z and $W^\pm$ bosons	10^{-18}
Gravity	Gravitons(not discovered yet)	∞

interacts with quarks and gluons through color charge. Three types of color charge exist in nature, i.e., red, green, and blue, with their respective anticolors. In nature, the quarks and antiquarks combine to form mesons (a pair of quark-antiquark) and baryons (three quarks system), which are colorless. There are also four and five quark systems that exist in na-

ture. LHCb experiments have recently observed five quark-bound states in the $J/\psi p$ decay channel and found some evidence for four quark-bound states in the $D^- K^+$ channel [1, 2].

The development of Quantum Electro-Dynamics (QED) theories in the 1940s had put the foundation for the SM. Many theorists like Dirac [3], Schwinger [4], and Feynman [5, 6] have contributed to understanding it better and better. Later in the decade of 1960s and 1970s, the other major part was developed- Higgs [7], Brout and Englert [8], known as the BEH mechanism, and Guralnik, Hagen, and Kibble [9] suggested a model so that vector bosons could acquire mass in a gauge-invariant way. Later Glashow [10], Salam [11], and Weinberg [12] built a unified theory of the electromagnetic and weak interaction called electro-weak interaction. Subsequently, in 1964, Gell-Mann and Zweig [13] introduced a quark model containing u , d , and s quarks to explain the properties of hadrons. Glashow-Iliopoulos-Maiani (GIM) [14] introduced a fourth quark (c) to explain the suppression of Flavor Changing Neutral Current (FCNC) interaction. Four years later, during the “November Revolution” [15, 16], the observation of $(c\bar{c})$ bound state known as J/ψ was made. In 1973, the theory of strong interaction, also known as Quantum Chromo-Dynamics (QCD) was formulated. This theory suggests that the quarks are real particles, carrying color charge, and massless gluons are quanta of strong interaction. In the same year, Gross and Wilczek [17], and Politzer [18] proposed the theory of asymptotic freedom, which states that at a higher energy scale the quarks behave like free particles. The third generation lepton τ was discovered in 1975 by the Martin Perl group [19]. Other SM quarks like bottom [20] and top [21] were discovered in 1977 and 1995, respectively. The last missing piece of the SM was the Higgs boson, observed by CMS [22] and ATLAS [23] experiments at LHC in 2013. The quarks, leptons, and bosonic force carriers of the SM are shown in Fig. 1.1.

The SM has been experimentally verified at different energy scales with extraordinary precision. Nevertheless, the SM could not explain a few fundamental questions and hint

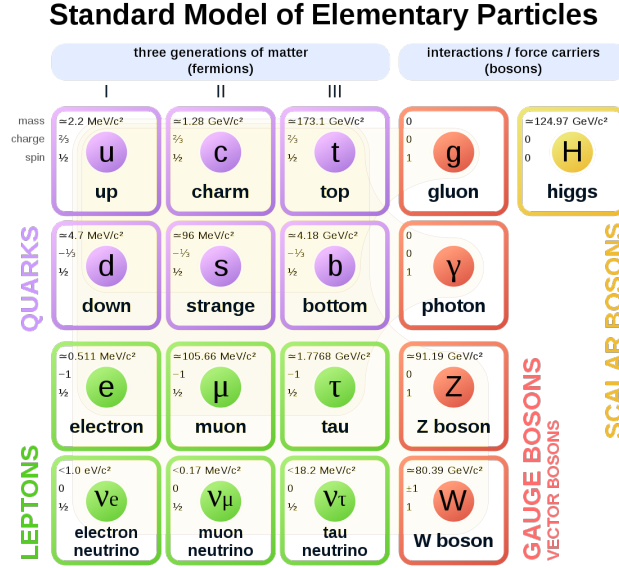


Figure 1.1: Standard Model in a nutshell [24].

for an extension of the SM. The open questions are discussed below,

- The hierarchy problem: It is referred to as a question in particle physics, why the weak force is 10^{24} times stronger than the gravitational force? Technically, the question is that why Higgs boson mass is smaller compared to Planck mass (or the grand unification energy). One would expect that the quantum correction received from all the energy would make the mass of the Higgs huge, which is comparable to Planck mass scale. The loop correction contribution can be from self-interaction, gauge loops, and fermion loops. The correction of mass varies as a function of Λ , up to which SM is valid and can be expressed as

$$\delta m_H^2 \propto \frac{\Lambda^2}{32\pi^2} \quad (1.1)$$

The experimentally measured mass of Higgs is 125 GeV, which is much lesser than the Planck mass. An unnatural fine-tuning is required to make Higgs mass compatible with the radiative corrected mass. Several theories like supersymmetry (SUSY), extra

dimension [25], little Higgs [26] model etc. can solve this issue. In SUSY, each fermionic particle has a boson super-partner and vice versa. The presence of a virtual particle of opposite spin inside the loop cancels the contribution from different loops to the Higgs mass.

- Another open question is, why is the gravitational force not included like the other three forces in the SM?
- The flavor puzzle: As shown in Fig. 1.1, the quarks are observed to be in three flavors. The change in flavor in decay is the result of weak interaction. The flavor changing neutral current (FCNC) process in the SM involves either up-type or down-type flavors (but not both) and either lepton or neutrino flavors (not both). Within the SM, the FCNC process is highly suppressed. Interestingly, the SM doesn't explain why the flavor parameters are small and hierarchical. This is known as the flavor puzzle of the SM. After measuring the neutrino mass and mixing in experiment, the puzzle becomes deeper.
- Neutrino oscillation, i.e., neutrino flavor continuously converting to other flavors, is observed by experiments [27]. The SM expects the neutrino mass to be zero, but the theoretical neutrino oscillation needs to have a non-zero mass which doesn't agree with the SM expectation. There has been a different cosmology study that constrains neutrino mass to be very small, i.e. $\sum m_\nu < 12 \text{ eV}$ [28]. Again, this raises the question of why the top quark's mass is so large compared to neutrino mass [29].
- Most recent cosmological observations estimate the energy content of the universe is constituent of 5% of visible matter, 69% of dark energy, and 26% of dark matter. Many theories like SUSY have given new unknown dark matter candidates to explain the results from the observation; for example, the formation of stars, big bang nucle-

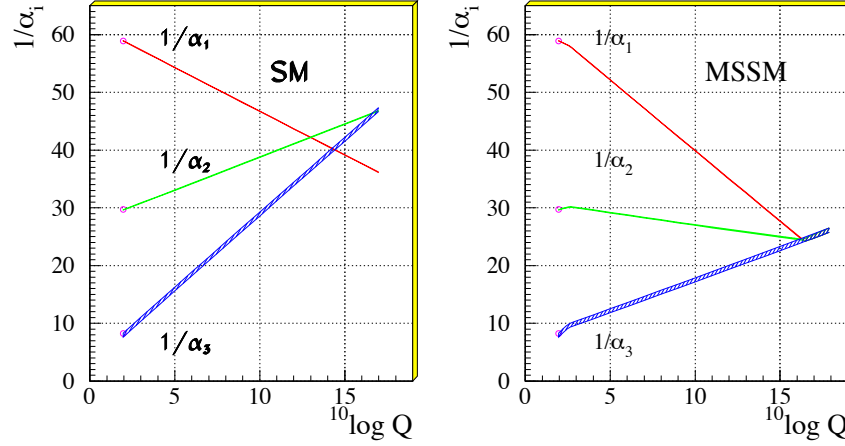


Figure 1.2: Three coupling constants at different energy scale. α_1 , α_2 , and α_3 are the coupling constants for the electromagnetic, weak, and strong forces, respectively [30].

osynthesis etc. Dark energy is required to understand the expansion of the galaxy over time. Another observation that can't be explained by the SM is the matter-antimatter asymmetry in the universe. The SM predicts the presence of matter and antimatter should be equal in-universe, but a large amount of missing antimatter is giving a hint for a new physics model.

- Grand Unification Theory (GUT) is the unification of the strong interaction and electro-weak interaction. The SM doesn't unify these two interactions as shown in the left Fig. 1.2, but BSM theory, like MSSM, combines them at energy scale $\Lambda_{GUT} \approx 10^{16}$ GeV (right plot in Fig. 1.2). Going further with higher energy, all three forces, along with gravity, combine to form a “theory of everything”. As the strength of gravitational interaction is very weak, it is expected to see the impact of gravity on particle physics interactions at an energy scale more than $\Lambda_{Planck} \approx 10^{19}$ GeV.

For these reasons, the SM is believed to be a low-energy spontaneously broken theory. Several new physics (NP) theories predict the existence of new particles or phenomena which can resolve many of the SM issues discussed previously. So far, no experimental

discovery has shed light on the existence of such a theory that may define nature correctly. The *Large Hadron Collider* (LHC), the most gigantic and complicated scientific apparatus ever assembled, has been compelling the experimental energy and luminosity frontier, in the hopes of seeing evidence for the existence of these elusive particles. In LHC, proton bunches collide at a regular interval of 25 ns and produce billions of proton-proton (pp) events per second. Out of these events, only a few interesting events are selected. Later the events are analyzed in the hope of finding the NP particle. The searches for NP at LHC and other experiments are carried out using two approaches: a direct search and an indirect search.

- Direct searches involve scanning of NP particles that are produced on-shell by using the byproducts of high-energy particle collisions. However, the search is limited by the center-of-mass (COM) energy of the collider experiment that is available for the formation of new particles. One such example is the Higgs boson discovery in 2013 by ATLAS and CMS Collaborations [22, 23]. Many NP candidate searches are ongoing at LHC, but no evidence has been found to date. However, the searches still constrain the parameter space of NP theories.
- Indirect searches do not need to be dependent on the colliding particle COM energy. Here the purpose is to precisely measure the SM process properties and compare them with the theoretical predictions. The deviation could hint at the presence of massive off-shell NP particles inside the quantum loops. In the SM, the existence of nearly all particles have been inferred indirectly before direct searches. As an example the Z -boson discovery was first observed indirectly by neutral current neutrino scattering before the direct discovery. Thus the precise measurement of the SM processes at low energy can indirectly probe the massive particle at a high energy scale.

Even though some interesting anomalies are observed in heavy flavor physics results,

the deviations are not significant enough to claim any discovery of new physics. Mainly the anomalies are observed in the measurement of different types of the $b \rightarrow sl^+l^-$ decay processes, where lepton $l \in (e, \mu)$. For example, the branching fraction of the $B_s^0 \rightarrow \phi\mu^+\mu^-$ process measured by the LHCb experiments [31], where the observed differential branching fraction in different $q^2(q = m_{\mu\mu})$ bins show a 3.6 standard deviation (σ) with respect to the SM prediction. The angular observables for the $B^{0/+} \rightarrow K^{*0/+}\mu^+\mu^-$ decays, where the measurements show a $\sim 3\sigma$ tension with the SM predictions. The results were confirmed by the ATLAS, CMS, and LHCb [32, 36] experiments.

The anomalies discussed above are all from the highly suppressed decays in the SM, and this is no coincidence; highly suppressed processes and interactions are of particular interest for indirect NP searches. In this thesis, such highly suppressed decays of the B_s^0 and B^0 to the dimuon final states are discussed in the context of indirectly searching for NP. The two independent analyses of the measurement of the $B_s^0 \rightarrow \mu^+\mu^-$ decays properties and search for the $B^0 \rightarrow \mu^+\mu^-$ decay using pp collision data by the CMS experiment during both the first and second experimental operations of the LHC are presented in Chapters 4 and 5.

This Chapter is organized as follows. The motivation of the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ decay with some generality and the theoretical branching fraction calculation in the SM are discussed in Sections 1.1 and 1.2, respectively. The time evolution of states and the SM predictions of the $B_{(s)}^0$ meson decay properties are discussed in Sections 1.3 and 1.4, respectively. A brief discussion of how NP can influence such decays alongside a description of some well-motivated NP models is given in Section 1.5. Finally, the current status of the experimental results are discussed in Section 1.6.

1.1 Decay of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ in the SM

In the SM, mesons are formed by the combination of quark and antiquark pairs, which are bound together by a strong force. The neutral B mesons, B_s^0 and B^0 , are the bound states of $\bar{b}s$ and $\bar{b}d$, respectively. Due to their neutral nature, they can oscillate to their antiparticle $\bar{B}_s^0$ ($b\bar{s}$) and $\bar{B}^0$ ($b\bar{d}$). These particles exist for $\sim 10^{-12}$ s before decaying to their daughter particles (such as leptons, lighter mesons, hadrons, etc.). Out of many possible decays, decays into a pair of oppositely charged muons are especially interesting and extremely rare. The theoretical branching fractions are of the order of 10^{-9} and 10^{-10} for B_s^0 and B^0 mesons, respectively. The rare occurrence is the result of different sources of suppressions, which are described below.

The first source is due to the compositeness of the B_s^0 and B^0 mesons. The constituent quarks of B_s^0 and B^0 have equal charge magnitudes. In the decay to two muons, only the quark flavor changes. This variety of decay is known as the Flavor Changing Neutral Current (FCNC) process. The weak forces, with the exchange of W bosons, act as a mediator in these decay processes. Because it is the only interaction where the quark flavor is not conserved. However, the GIM mechanism prohibits FCNCs in the SM from occurring at the tree level and allows via more complicated higher-order processes. The dominant one is the Z penguin with a top loop and the W box diagram [37], as indicated in Fig. 1.3. The lack of decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ at tree level results in the suppression compared to other $B_{(s)}^0$ decay modes.

Although the change of flavor between the quarks is allowed in weak force, the coupling strength between the different quark flavors is different in magnitudes. The coupling strength is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [38, 39]. In the SM, there are two types of quark that exist depending upon their charge: up-type quarks of charge $+\frac{2}{3}e$, which include u , c , and t , and down-type quarks of charge $-\frac{1}{3}e$, which in-

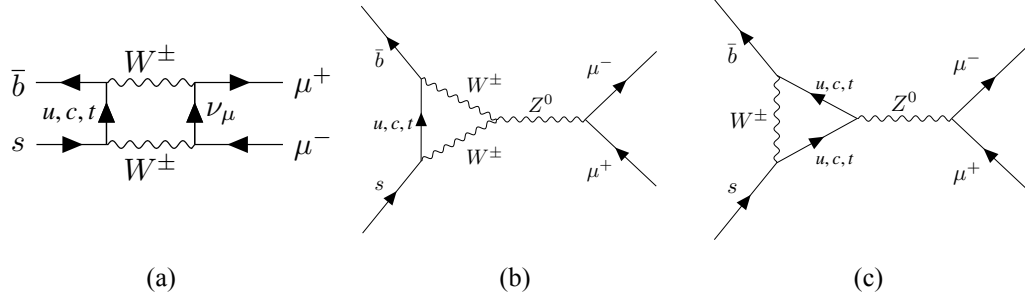


Figure 1.3: EW “box” (a) and EW “penguin” (b & c) are the main Feynman diagrams contributing to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay in SM. For the $B^0 \rightarrow \mu^+ \mu^-$ decay s quark is replaced with d quark.

clude d , s , and b . The weak interaction couples all up-type quarks to the flavor eigenstate of down-type quarks. The relation between the flavor and mass eigenstates is supported via the CKM matrix as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{bd} & V_{bs} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.2)$$

where d' , s' , and b' are the flavor eigenstates, and d , s , and b are the mass eigenstates. The matrix elements V_{ij} represent the quark transition strength between the mass eigenstates of quarks i and j . The CKM matrix has four free parameters, three rotation angles, and one phase. The presence of a non-zero phase makes the CKM matrix complex and the CP symmetry breaks. One of the widely used representations for the CKM matrix proposed by Wolfenstein, suggests the hierarchy in the magnitudes of the matrix elements, i.e., $V_{ud} < V_{cs} < V_{tb} \sim 1$, with smaller contributions outside the diagonal. The representation becomes

$$V_{CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - i\eta) & -\lambda^2 A & 1 \end{pmatrix} + O(\lambda^4) \quad (1.3)$$

The parameter λ has been experimentally measured roughly to be 0.22. The other parameters, like A , ρ , and η , are estimated to be within 0.1-1. The observed structure of the CKM matrix reveals the quark transition is favorable within the same generation, but it becomes

stronger when one goes more away from the diagonal. In the case of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, a single off-diagonal element is required to describe the transition, as shown in Fig. 1.3, which adds suppression to the decay. Figure 1.3 shows that quarks like u , c , and t can contribute inside the loop, and the t quark contribution is large compared to others in the SM. This follows mainly because of two reasons: the top quark has a large mass compared to other quarks, and the large coupling strength of the b -quark to the t -quark compared to others.

The final source of suppression is from the helicities of the muons in the final state. Since both the B^0 and B_s^0 mesons have spin zero when it decays, the angular momentum of both the muons should align in opposite directions. In the SM, the weak forces only couple to left-handed particles and right-handed antiparticles, which lead to the muons having opposite helicities. When a particle is massless, at a high energy scale, the left-handed states have negative helicity, while the right-handed states have positive helicity. However, muons are not massless; therefore, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are suppressed with a factor of $(m_\mu/M_{B_{(s)}^0})^2 \sim 4 \times 10^{-4}$.

Overall, the decays of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ are highly suppressed in the SM compared to other $B_{(s)}^0$ meson decay modes. Various BSM theories offer these decays as an excellent probe for searching NP contributions because the decay rates will be enhanced from the SM in the presence of new heavy particles.

1.2 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ SM branching fraction

The simplest observable of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is the branching fraction ($\mathcal{B}$), which allows us to study the effect of NP contribution by comparing the experimental observation and the SM prediction. By definition, it means the probability of $B_{(s)}^0$ meson events decay into a pair of opposite muons. Using a time-dependent decay rate, $\Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-)$, the

branching fraction can be written as

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \equiv \frac{1}{2} \int_0^\infty \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle dt \quad (1.4)$$

where $\langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle$ is the untagged decay rate at time t , which can be defined as

$$\langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle|_{t=0} = \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) + \Gamma(\bar{B}_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \quad (1.5)$$

and the branching fraction becomes

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{th} \equiv \frac{\tau_{B_{(s)}^0}}{2} \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle|_{t=0} \quad (1.6)$$

where $\tau_{B_{(s)}^0}$ is the mean lifetime of the $B_{(s)}^0$ meson. In the experiment, the number of $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ meson was produced equally, and the untagged decay rate does not discriminate between the particles and antiparticles. The instantaneous decay rate as a function of transition matrix element $|\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|$ evaluated from Fermi's golden rule can be written as

$$\Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-)|_{t=0} = \frac{1}{16\pi} \frac{1}{M_{B_{(s)}^0}} \sqrt{1 - 4 \left(\frac{m_\mu}{M_{B_{(s)}^0}} \right)^2} |\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|^2 \quad (1.7)$$

where $M_{B_{(s)}^0}$ and m_μ are the masses of the relevant $B_{(s)}^0$ meson and muon, respectively. The various transition processes contributing to the decay can be described by the matrix element $|\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|$. The calculation of matrix elements in weak decay such as $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ is complicated by the existence of strong interaction. In contrast to QED, where the contribution from the higher-order diagrams is suppressed by coupling constant ($\frac{1}{137} \ll 1$) and can be ignored, the strong QCD coupling mostly relies on the momentum transfer of the process. At low energy, the strong interaction between the quarks becomes larger, which confines the quarks and requires a non-perturbative approach, for instance, lattice calculation. The decays of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ include interactions on different energy scales, which can be explained by the effective theory of Operator Product Expansion method.

According to this theory, the momentum transfer p through the weak mediator boson is smaller than their actual mass. Hence, the scale with mediator mass as $\mathcal{O}(p^2/M(W)^2)$ can be ignored, resulting in a sum over the charged current vectors to a series of local operators multiplied by the effective coupling constant, known as “Wilson coefficient”. An effective weak Hamiltonian can be expressed as

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\lambda) \mathcal{O}_i(\lambda), \quad (1.8)$$

where $C_i(\lambda)$ are the Wilson coefficients, $\mathcal{O}_i$ are the local operators of suitable decays, V_{CKM}^i are the CKM matrix elements, G_F is the fermi coupling constant and the factor λ is an arbitrary scale separating high- and low- energy contributions. The product of $C_i(\lambda)$ and V_{CKM}^i defines the strength of the operator in Hamiltonian. The transition amplitude can be then written as

$$|\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)| = \langle \mu^+ \mu^- | \mathcal{H}_{eff} | B_s^0 \rangle = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i \langle \mu^+ \mu^- | C_i(\lambda) \mathcal{O}_i(\lambda) | B_s^0 \rangle, \quad (1.9)$$

The local operator $\mathcal{O}_i$ connects the initial and final state of the decay and usually contributes to the largest source of uncertainties in the amplitude estimation. The choice of λ is chosen in such a way that the final transition amplitude must be independent of λ . Usually, the decaying hadron mass is used. The Wilson coefficients describe the short scale physics processes at higher than λ . The Wilson coefficients depends on the $W^\pm$, Z^0 , H^0 , and t quark masses, which arise in the loops of Feynman diagrams. Their values are calculated by matching the expression from the effective theory to the decay amplitude measured from the full theory. This matching is performed at a high factorisation scale, $\lambda = \mathcal{O}(M_W)$, where the contributions from the logarithmic term $\ln(M_W/\lambda)$ are small and the $C_i(\lambda)$ values run down to the $\mathcal{O}(m_b)$ scale.

The leptonic final state of the B_s^0 meson decay does not couple directly to the initial state quarks. Therefore, the four-point matrix element, $\langle \mu^+ \mu^- | \mathcal{O}_i(\lambda) | B_s^0 \rangle$, can be written

as a product of the hadronic and leptonic part, where the hadronic part contains all non-perturbative strong interaction effects. The operator product contains all possible Lorentz invariant combinations. Out of them, only a few of the non-vanishing contributions to the decay amplitudes are

$$\mathcal{O}_{10}^L = (\bar{q}\gamma_\mu P_L b) \otimes (\bar{\mu}\gamma_\mu \gamma^5 \mu), \quad \mathcal{O}_{10}^R = (\bar{q}\gamma_\mu P_R b) \otimes (\bar{\mu}\gamma_\mu \gamma^5 \mu), \quad (1.10)$$

$$\mathcal{O}_S^L = (\bar{q}P_L b) \otimes (\bar{\mu}\mu), \quad \mathcal{O}_S^R = (\bar{q}P_R b) \otimes (\bar{\mu}\mu) \quad (1.11)$$

$$\mathcal{O}_P^L = (\bar{q}P_L b) \otimes (\bar{\mu}\gamma_5 \mu), \quad \mathcal{O}_P^R = (\bar{q}P_R b) \otimes (\bar{\mu}\gamma_5 \mu) \quad (1.12)$$

where $q = s$, and d quarks. The left-handed and right-handed operators are related by replacing $P_L \leftrightarrow P_R$. The operator $\mathcal{O}_{10}^L$ contributes significantly in the SM and comes from electro-weak W and Z^0 penguin diagrams. However, the exchange of scalar and pseudo-scalar particle operators $\mathcal{O}_S^{L(R)}$ and $\mathcal{O}_P^{L(R)}$ contributions are tiny in the SM. The operator $\mathcal{O}_{10}^R$ describes the equivalent interactions as $\mathcal{O}_{10}^L$ but instead for right handed currents that are forbidden in the SM. Using the Eqs. [1.10](#) to [1.12](#), the model independent Effective Hamiltonian can therefore be written as

$$\mathcal{H}_{eff} = \frac{G_F \alpha_{EM}}{\sqrt{2}\pi} \{V_{t(d,s)}^* V_{tb} \sum_i^{10,S,P} (C_i^L(\lambda) \mathcal{O}_i^L + C_i^R(\lambda) \mathcal{O}_i^R)\} \quad (1.13)$$

where $\alpha_{EM} = e^2/4\pi$, is the QED fine structure constant. Since the contribution of up-type quarks in the internal loop diagrams are negligible, only the top quark contribution is considered, and the corresponding CKM matrix elements are V_{td}^* , V_{ts}^* , and V_{tb} .

The branching fraction of the decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ can be written using Eq. [1.13](#) as,

$$\begin{aligned} \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) &= \frac{\tau_{B_{(s)}^0} G_F^4 M_W^4 \sin^4(\theta_W)}{8\pi^5} |C_{10}^{SM} V_{tq}^* V_{tb}|^2 f_{B_{(s)}^0}^2 M_{B_{(s)}^0} m_\mu^2 \\ &\quad \times \sqrt{1 - \frac{4m_\mu^2}{M_{B_s^0}^2} (|P|^2 + |S|^2)} \end{aligned} \quad (1.14)$$

where $\tau_{B(s)^0}$ is the average lifetime of the $B(s)^0$ meson, θ_W is the weak mixing angle and $\mathcal{C}_{10}^{SM}$ is the SM parameter of $\mathcal{C}_{10}$. The factor $f_{B(s)^0}$ is derived from hadronic matrix element. The pseudo-scalar and scalar operator can be written as

$$P = \frac{\mathcal{C}_{10}^R - \mathcal{C}_{10}^L}{\mathcal{C}_{10}^{SM}} + \frac{M_{B(s)^0}^2}{2\mu} \left(\frac{m_b}{m_b + m_{(d,s)}} \right) \left(\frac{\mathcal{C}_P^R - \mathcal{C}_P^L}{\mathcal{C}_{10}^{SM}} \right) \quad (1.15)$$

$$S = \sqrt{1 - \frac{4m_\mu^2}{M_{B(s)^0}^2} \frac{M_{B(s)^0}^2}{2m_\mu}} \left(\frac{m_b}{m_b + m_{(d,s)}} \right) \left(\frac{\mathcal{C}_S^R - \mathcal{C}_S^L}{\mathcal{C}_{10}^{SM}} \right), \quad (1.16)$$

In the SM, the coefficients P and S are real, $P = 1$ and $S = 0$. However, in BSM theories, the values are different, resulting a change in the branching fraction.

Another interesting observable along with branching fractions is the ratio of branching fractions of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays. By taking the ratio, the dependence of the Wilson coefficient and $|V_{tb}|$ cancels and results in the dependence on the CKM matrix elements, $|V_{td}|$ and $|V_{ts}|$. The ratio of branching fractions can be expressed as,

$$\mathcal{R} \equiv \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = \frac{\tau_{B^0}}{1/\Gamma_H^s} \frac{|V_{td}|^2}{|V_{ts}|^2} \frac{f_{B^0}^2}{f_{B_s^0}^2} \frac{M_{B^0}}{M_{B_s^0}} \sqrt{\frac{1 - \frac{4m_\mu^2}{M_{B^0}^2}}{1 - \frac{4m_\mu^2}{M_{B_s^0}^2}}} \quad (1.17)$$

In other words, the measurement of the ratio of branching fractions tests the coupling strength described by the CKM matrix in the SM and verifies whether the BSM theory follows the same or not. The benefit of using this observable is that the theoretical uncertainty is small because the dominant contribution from the Wilson coefficient and $|V_{tb}|$ cancel.

1.3 Time evolution of $B(s)^0$ state

As discussed in the previous section, the $B(s)^0$ oscillates to its antiparticle $\bar{B}(s)^0$ with the evolution of time. Oscillation occurs as a result of the constituent quark transition to different quark flavors with the interchange of W bosons. Therefore, the state which propagates

through time is the superposition of the $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ states. The time-integrated branching fraction is evaluated from the time evolution of the states. The time evolution between the particle and antiparticle is given by Time-dependent Schrödinger equation, which is written as

$$i \frac{d}{dt} \begin{pmatrix} |B_{(s)}^0(t)\rangle \\ |\bar{B}_{(s)}^0(t)\rangle \end{pmatrix} = \begin{pmatrix} M - i\frac{\Gamma}{2} \end{pmatrix} \begin{pmatrix} |B_{(s)}^0(t)\rangle \\ |\bar{B}_{(s)}^0(t)\rangle \end{pmatrix} \quad (1.18)$$

where M and Γ are the mass and decay time Hermitian matrices. A Hermitian matrix has the following properties, $M_{12}^* = M_{21}$ and $\Gamma_{12}^* = \Gamma_{21}$. Furthermore, the charge, parity, and time inversion invariance add additional constraints such as $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. According to quark mixing, for any $t > 0$, the particle state is the superposition of the $B_{(s)}^0$ and $\bar{B}_{(s)}^0$ states. The eigenstates can be defined as

$$|B_H\rangle = p|B_{(s)}^0\rangle - q|\bar{B}_{(s)}^0\rangle, \quad |B_L\rangle = p|B_{(s)}^0\rangle + q|\bar{B}_{(s)}^0\rangle, \quad (1.19)$$

where the coefficients p and q should satisfy the condition $|p|^2 + |q|^2 = 1$. $|B_H\rangle$ and $|B_L\rangle$ are known as the heavy and light mass eigenstates with eigenvalues of $(m_{H,L} - i\Gamma_{H,L}/2)$. The off-diagonal elements of the $B_{(s)}^0 - \bar{B}_{(s)}^0$ mixing matrix represents the mass difference Δm and decay width difference $\Delta\Gamma$ between the two mass eigenstates. The relationships can be written as,

$$m \equiv \frac{m_L + m_H}{2}, \quad \Delta m \equiv m_H - m_L \quad (1.20)$$

$$\Gamma \equiv \frac{\Gamma_L + \Gamma_H}{2}, \quad \Delta\Gamma \equiv \Gamma_L - \Gamma_H \quad (1.21)$$

The difference Δm and $\Delta\Gamma$ are the measurable quantity, and it is defined in such a way that the Δm value results always positive, whereas $\Delta\Gamma$ can yield either sign. A new parameter is introduced and defined as

$$\lambda_{\mu\mu} = \frac{q}{p} \frac{\bar{A}_{\mu\mu}}{A_{\mu\mu}} \quad (1.22)$$

where $\bar{A}_{\mu\mu} = \mathcal{M}(\bar{B}_{(s)}^0 \rightarrow \mu^+ \mu^-)$ and $A_{\mu\mu} = \mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ to simplify the decay rate equation. The final time integrated total decay rate can be written as

$$\langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle = \mathcal{N} |\mathcal{A}_{\mu\mu}|^2 (1 + |\lambda_{\mu\mu}|^2) e^{\frac{-\Gamma_s t}{2}} \left(\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \mathcal{A}_{\Delta\Gamma} \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right) \quad (1.23)$$

where $\mathcal{N}$ is the additional factor in Eq. 1.7. The parameter, $\mathcal{A}_{\Delta\Gamma}$, which depends on final state can be defined as

$$\mathcal{A}_{\Delta\Gamma} = \frac{2Re(\lambda_{\mu\mu})}{1 + |\lambda_{\mu\mu}|^2} \quad (1.24)$$

Again the decay rate equation can be simplified in terms of heavy and light mass eigenstates as [40]

$$\begin{aligned} \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle &= \mathcal{N} |\mathcal{A}_{\mu\mu}|^2 (1 + |\lambda_{\mu\mu}|^2) ((1 - \mathcal{A}_{\Delta\Gamma})e^{-\Gamma_L t} + (1 + \mathcal{A}_{\Delta\Gamma})e^{-\Gamma_H t}) \\ &= R_H e^{-\Gamma_H t} + R_L e^{-\Gamma_L t} \end{aligned} \quad (1.25)$$

The above equation shows the total decay rate as a sum of heavy and light mass eigenstates.

The parameter $\mathcal{A}_{\Delta\Gamma}$ is therefore can be rephrased as a function of R_H and R_L

$$\mathcal{A}_{\Delta\Gamma} = \frac{R_H - R_L}{R_H + R_L} \quad (1.26)$$

The value of $\mathcal{A}_{\Delta\Gamma}$ can lie in the range -1 to +1. If the heavy mass eigenstate decays to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, $\mathcal{A}_{\Delta\Gamma} = +1$ and -1 for the light mass eigenstate decay.

The time-dependent decay rates are used in the evaluation of the branching fraction. The theoretical branching fraction as in Eq. 1.6, can be rewritten as,

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{th}} = \frac{\tau_{B_{(s)}^0}}{2} (R_H + R_L) \quad (1.27)$$

Similarly, the measured time integrated branching fraction is

$$\begin{aligned} \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} &\equiv \frac{1}{2} \int_0^\infty \langle \Gamma(B_{(s)}^0(t) \rightarrow \mu^+ \mu^-) \rangle dt \\ &= \frac{1}{2} \left(\frac{R_H}{\Gamma_H} + \frac{R_L}{\Gamma_L} \right) \\ &= \frac{\tau_{B_{(s)}^0}}{2} (R_H + R_L) \left[\frac{1 + \mathcal{A}_{\Delta\Gamma} y_{(s)}}{1 - y_{(s)}^2} \right] \end{aligned} \quad (1.28)$$

where $y_{(s)}$ is given by $y_{(s)} = \Delta\Gamma_{(s)}/2\Gamma_{(s)}$. Therefore, the relation between the measured and theoretical branching fraction is given below

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} = \left[\frac{1 + A_{\Delta\Gamma} y_{(s)}}{1 - y_s^2} \right] \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{th}} \quad (1.29)$$

For $B_s^0 - \bar{B}_s^0$ oscillation, the difference in lifetime between heavy and light states is sizeable. But in the case of $B^0 - \bar{B}^0$ oscillation, the difference is extremely small, and the value of y is ignorable, which results in the theoretical branching fraction being equivalent to the experimental one. As of Eq. 1.14, the branching fraction is dependent on the complex parameters P and S . Then the Eq. 1.24 can be rewritten as

$$\mathcal{A}_{\Delta\Gamma} = \frac{|P|\cos 2\varphi_P - |S|\cos 2\varphi_S}{|P|^2 + |S|^2} \quad (1.30)$$

where φ_P and φ_S are the CP-violating phases. In the SM, the maximum value $\mathcal{A}_{\Delta\Gamma}$ can take is +1 due to $P=1$ and $S=0$. The final state of $B_s^0 \rightarrow \mu^+ \mu^-$ decay is a CP odd state, and only the heavy mass eigenstate has a CP odd final state. This leads to the conclusion that only the heavy mass eigenstate can decay as $B_s^0 \rightarrow \mu^+ \mu^-$ in SM. The value of $|P|$ and $|S|$ parameters can be altered by the NP model, which results in the change in branching fraction and $\mathcal{A}_{\Delta\Gamma}$. So a measurement of $\mathcal{A}_{\Delta\Gamma}$ will allow us to understand the NP contribution by comparing the branching fraction to the SM expectation. The value of $\mathcal{A}_{\Delta\Gamma}$ can be measured from the time-integrated decay rate equation. This is achieved by separating the $B_s^0 \rightarrow \mu^+ \mu^-$ decay into $|B_s^0\rangle$ and $|\bar{B}_s^0\rangle$ initial states. However, this approach is not feasible because the decay is rare. Another method to measure the $\mathcal{A}_{\Delta\Gamma}$ is from the effective lifetime. The effective lifetime can be measured from the untagged decay rate as

$$\tau_{\mu^+ \mu^-} \equiv \frac{\int_0^\infty t \langle \Gamma(B_s(t) \rightarrow \mu\mu) \rangle dt}{\int_0^\infty \langle \Gamma(B_s(t) \rightarrow \mu\mu) \rangle dt} \quad (1.31)$$

and it can be expressed as a function of $\mathcal{A}_{\Delta\Gamma}$, then the equation becomes

$$\tau_{\mu^+ \mu^-} \equiv \frac{\tau_{B_s}}{1 - y_s^2} \left(\frac{1 + 2\mathcal{A}_{\Delta\Gamma} y_s + y_s^2}{1 + y_s^2} \right) \quad (1.32)$$

The measurement of the effective lifetime not only allows us to measure the branching fraction but also the $B_s^0 - \bar{B}_s^0$ mixing parameters [41]. According to SM, only the CP odd state, heavy mass eigenstate, contributes to $B_s^0 \rightarrow \mu^+ \mu^-$ decay, which leads to an effective lifetime of $\tau_{\mu^+ \mu^-} \equiv \tau_H \equiv \frac{1}{\Gamma_H}$. That is why measuring an effective lifetime is complementary to the branching fraction for studying SM and NP effects.

1.4 SM predictions

The SM branching fraction can be written from Eq. 1.14 by assuming $P = 1$ and $S = 0$ as

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{\tau_{B_{(s)}} G_F^4 M_W^4 \sin^4(\theta_W)}{8\pi^5} |C_{10}^{SM} V_{tq}^* V_{tb}|^2 f_{B_{(s)}}^2 M_{B_{(s)}} m_\mu^2 \times \sqrt{1 - \frac{4m_\mu^2}{M_{B_{(s)}}^2}} \quad (1.33)$$

By putting all the latest results in Eq. 1.33, the branching fractions are [42],

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (3.66 \pm 0.14) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) &= (1.03 \pm 0.05) \times 10^{-10} \end{aligned} \quad (1.34)$$

The relative uncertainty on the branching fractions are 4-5%. The theoretical uncertainty is calculated from the Eq. 1.33, including next-to-leading order QED effects and also the virtual photon exchange contributions, and next-to-next-to-leading order QCD effects. The recent progress in the lattice QCD to determine $f_{B_{(s)}}^0$ helps to achieve the smallest uncertainty in the branching fraction. The dominant sources of theoretical uncertainty come from CKM parameters, followed by $f_{B_{(s)}}^0$ and then top-quark mass (m_t). Figure 1.6 (b) shows the evolution of theoretical $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction.

The ratio of branching fraction can be written easily from Eq. 1.17 as,

$$\mathcal{R} \equiv \frac{\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)}{\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)} = \frac{\tau_{B^0}}{1/\Gamma_H^s} \frac{|V_{td}|^2 f_{B^0}^2 M_{B^0}}{|V_{ts}|^2 f_{B_s^0}^2 M_{B_s^0}} \sqrt{\frac{1 - \frac{4m_\mu^2}{M_{B^0}^2}}{1 - \frac{4m_\mu^2}{M_{B_s^0}^2}}} = 0.0281 \pm 0.0016 \quad (1.35)$$

From theoretical point of view, it has a small theoretical uncertainty compared to the individual branching fractions. That is why it is an excellent probe to test BSM theories and flavor coupling strength.

Another, interesting variable effective lifetime of B_s^0 meson is determined from Eq. [1.32](#), as

$$\tau_{\mu^+\mu^-} = 1.620 \pm 0.007 \text{ ps} \quad (1.36)$$

where the input values to the Eq. [1.32](#) are, $\mathcal{A}_{\Delta\Gamma} = +1$, mean lifetime $\tau_{B_s^0} = 1.515 \pm 0.004$ ps, and $y_s = 0.065 \pm 0.003$. The above lifetime is the lifetime for heavy mass eigenstates, as SM predicts only the heavy mass state decay to dimuon states. A similar calculation for $\mathcal{A}_{\Delta\Gamma} = -1$ can be done and the lifetime for the light mass eigenstate is $\tau_L = 1.423 \pm 0.005$ ps. The difference in lifetime between the two mass eigenstates is 0.197 ps. Hence a precise measurement of the lifetime will be able to distinguish the $\mathcal{A}_{\Delta\Gamma} = +1$ and $\mathcal{A}_{\Delta\Gamma} = -1$.

1.5 $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fraction in BSM

Many BSM theories predict the branching fractions of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ either to be small or large compared to SM predictions. Furthermore, a measurement of $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions and the effective lifetime of the B_s^0 meson in the $B_s^0 \rightarrow \mu^+\mu^-$ decay can put a constraint on the parameter space available for NP. A few of the NP models are discussed in this section. As shown in Eqs. [1.14](#) and [1.24](#), the branching fraction is dependent on $|S|$, $|P|$, and $\mathcal{A}_{\Delta\Gamma}$, and these parameters are affected independently by the new physics model. SM predicts the value of $\mathcal{A}_{\Delta\Gamma} = +1$, $|S| = 0$, and $|P| = 1$. Here, four different scenarios will be discussed from Ref [\[43\]](#): (A) $S = 0$, (B) $P = 1$, (C) $P \pm S = 1$ and (D) $\varphi_P, \varphi_s \in \{0, \pi\}$.

1.5.0.0.1 Scenario A: Pseudo-scalar dominated New Physics ($S = 0$) In this scenario, there is no contribution from the scalar particle. The term $(\mathcal{C}_S^R - \mathcal{C}_S^L)$ in Eq. [1.16](#) becomes zero, which leaves the Wilson coefficients $\mathcal{C}_{10}^{L,R}$ and $\mathcal{C}_P^{L,R}$ to choose any non-SM value.

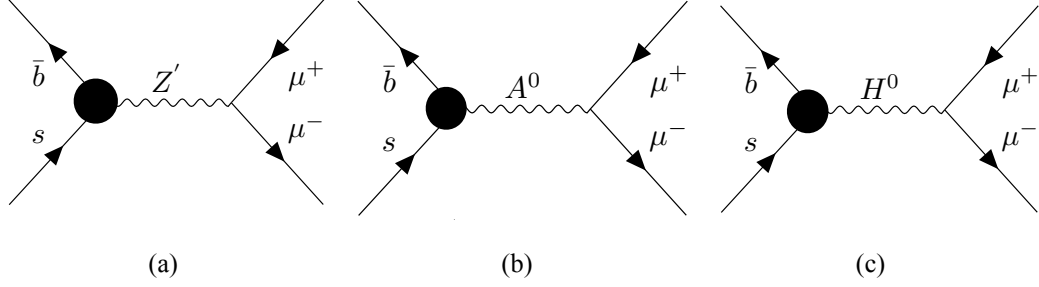


Figure 1.4: (a) Z' model (b) 2HDM with Pseudo scalar boson (c) 2HDM with Scalar boson are few tree level Feynman diagrams contributing to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay in BSM.

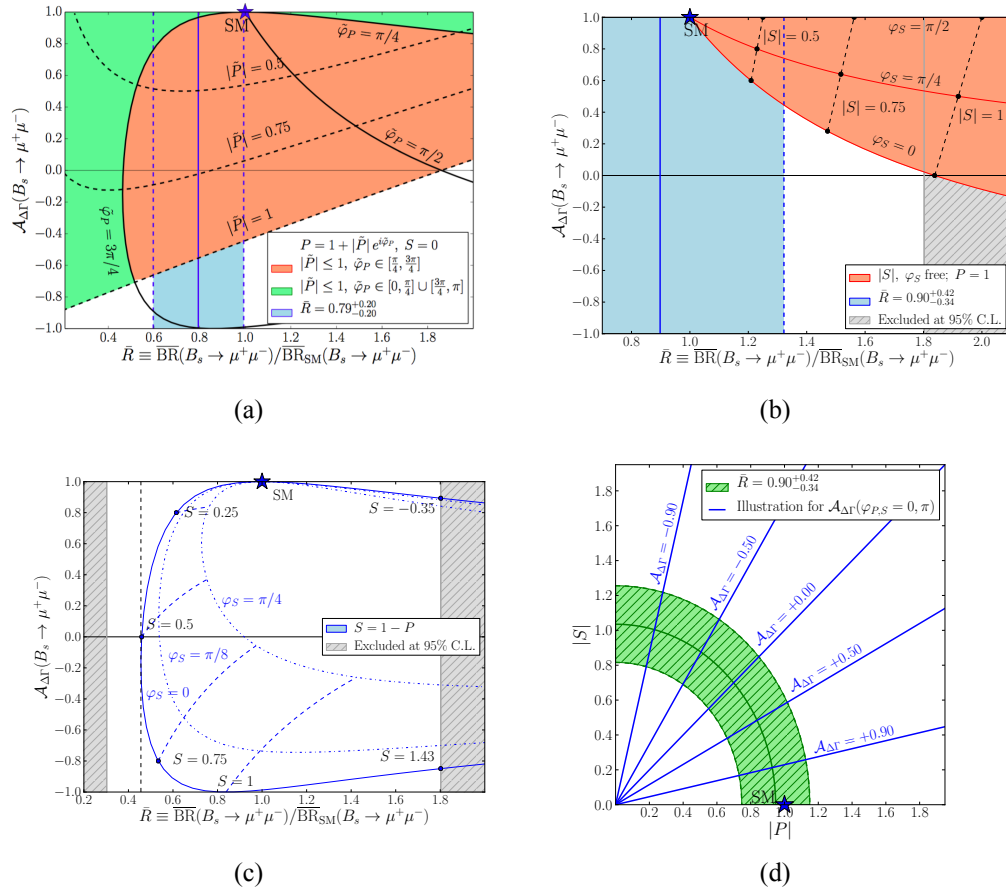


Figure 1.5: The correlation between the ratio of branching fraction ($\bar{R}$) and $\mathcal{A}_{\Delta\Gamma}$ for different scenario where (a) $S = 0$, (b) $P = 1$, (c) $P \pm S = 1$, (d) $\varphi_P, \varphi_S \in \{0, \pi\}$. The plots are taken from the Ref. [43].

Models with pseudo-scalars and scalars, only scalars which couple both left-right symmetrically with quarks so that $C_S^R = C_S^L$, fall in this category. One example of this scenario is Constrained Minimal Flavor Violation (CMFV) [44]. In CMFV, the CP-violating interactions and the coupling of quark flavors have the same Yukawa structure as the SM. This hypothesis is independently describing the flavor structures of the NP models. All the Wilson coefficients vanish except for C_{10} , which is real. However, the NP only affects the ratio of branching fractions. Any deviation of the branching fractions ratio from the SM or MFV will hint at a new flavor structure in theoretical models. In other models like the Z' model, new operators arise at the tree level to allow the FCNC process, as shown in Fig. 1.4. In Fig. 1.5, the correlation between the ratio of branching fractions and the $\mathcal{A}_{\Delta\Gamma}$ for the $S = 0$ scenario, is shown.

1.5.0.0.2 Scenario B: Scalar dominated New Physics ($P = 1$) In this scenario, Eq. 1.15 simplifies to $C_{10}^R = C_{10}^{SM}$ and $C_{10}^L = C_P^{R,L} = 0$ and the pseudo-scalar meson that couples symmetrically (left-right) to the quarks is accounted for $C_P^R = C_P^L$. Here the effect of NP is carried by the scalar particle. In this case, the $\mathcal{A}_{\Delta\Gamma}$ and the branching ratio both depend on the $2\varphi_S$ and $|S|$. As shown in Fig. 1.5, for no CP-violating phase ($\varphi = 0$) the value of $|S|$ pushes $\mathcal{A}_{\Delta\Gamma}$ to zero. Moreover, after considering the experimental bounds, $|S| < 0.5$ is allowed. Two Higgs Doublet Models (2HDMs) with a scalar H^0 particle is one of the examples, and the Feynman diagram is shown in Fig. 1.4.

1.5.0.0.3 Scenario C: Mixed (pseudo-)scalar dominated New Physics ($P \pm S = 1$) In this case, the NP effects are realized after giving equal status to P and S . If we consider $4m_\mu^2/M_{B_s}^2 \simeq 1$ (in Eq. 1.16) and $C_{10}^R - C_{10}^L \simeq C_{10}^{SM}$ (in Eq. 1.15), the coefficients in this scenario are related as,

$$C_S^{R(L)} = \pm C_P^{R(L)} \quad (1.37)$$

For example, in case of 2HDMs, the condition $P \pm S = 1$ will be satisfied if the coupling of the H_0 , A_0 are left-right asymmetric and either $\mathcal{C}_{S,P}^L$ and $\mathcal{C}_{S,P}^R$ are large. The MSSM model can also contribute to the decay $B_s^0 \rightarrow \mu^+ \mu^-$, if the contribution of vector-axial vector operators (as shown in Fig. 1.4), $\mathcal{C}_{10}^L$ are tiny. In the MSSM, the scalar operator contribution is enlarged by large $\tan(\beta)$ effects, which helps to understand the decoupling regime [45].

1.5.0.0.4 Scenario D: $\varphi_P, \varphi_S \in \{0, \pi\}$ In this scenario, the assumption is that there is no CP violating phase in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay mode. In other words, all the Wilson coefficients will take real values. In Fig. 1.5, it can be seen that how the ratio of branching fractions and $\mathcal{A}_{\Delta\Gamma}$ are able to determine the value of the parameters $|S|$ and $|P|$.

To conclude, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay is recognized as a “golden channel” for NP models. Although this decay has yet to reveal the NP contribution, recent measurements still provide ample space to look for the effect of NP. Similarly, the measurement of the effective lifetime of the B_s^0 meson in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay draws attention to investigate the sensitive parameter $\mathcal{A}_{\Delta\Gamma}$ in the search for NP.

1.6 Status of experimental results

The search for the $B^0 \rightarrow \mu^+ \mu^-$ decay started at e^+e^- colliders, then the saga continued along with the $B_s^0 \rightarrow \mu^+ \mu^-$ decay in the pp collider (LHC) by different experiments. Over a number of decades, the sensitivity of the decay increased significantly both on the experimental and theoretical sides and can be seen in Fig. 1.6. In 2013, CMS and LHCb experiments both individually reported evidence of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay with a statistical significance of 4 standard deviations (σ) [46, 47]. Subsequently, a combined analysis by CMS and LHCb [48] claimed the observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay at 5σ significance. Later the ATLAS, CMS, and LHCb experiments performed the analysis with more data to

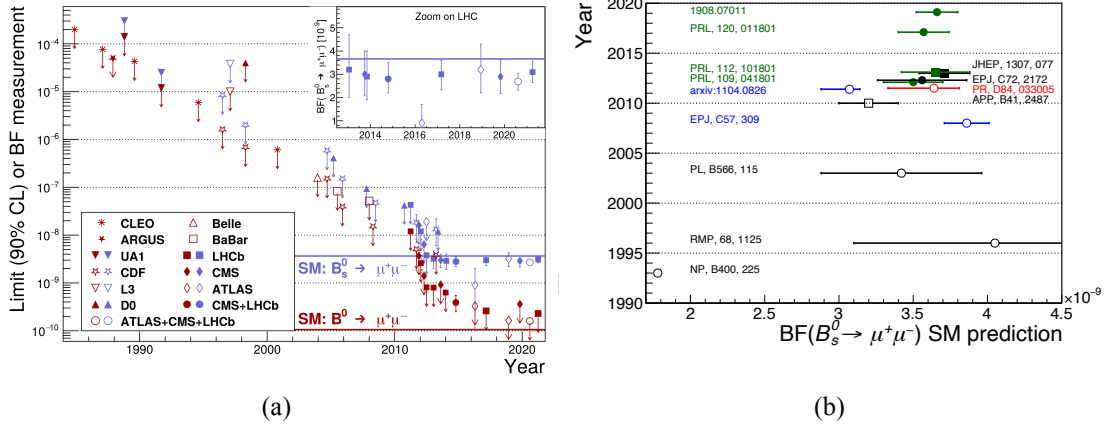


Figure 1.6: (a) The experimental results of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ decay measured so far are shown in this Figure. The blue markers are for the $B_s^0 \rightarrow \mu^+\mu^-$ decay from the experiments mentioned in the black marker. The red markers are for the upper limits on $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ decay [49]. The measurements described in Chapter 5 are not mentioned in the plot. (b) The SM branching fraction value evolution with time is shown. Note that the first prediction has a large error because, at that time, top quark was not discovered and many input parameters had large errors.

confirm the excess observed in the combined analysis and improved the uncertainties of the properties measurements. Unfortunately, no significant excess of the $B^0 \rightarrow \mu^+\mu^-$ decay has been observed so far.

The most precise results on the branching fractions and the effective lifetime are obtained from the CMS experiment, using the data collected between 2016-2018. The measured values are

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)_{\text{CMS}} &= (3.83^{+0.44}_{-0.41}) \times 10^{-9} \\ \mathcal{B}(B^0 \rightarrow \mu^+\mu^-)_{\text{CMS}} &= (0.37^{+0.75}_{-0.68}) \times 10^{-10} \\ \tau_{\mu^+\mu^-} &= 1.83^{+0.23}_{-0.20} \text{ ps}\end{aligned}\tag{1.38}$$

The observed $B^0 \rightarrow \mu^+\mu^-$ signal significance is less than 1σ , so an upper limit on the branching fraction was estimated using CLs [50] method. The values are $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 1.5(1.9) \times 10^{-10}$ at 90 (95) % confidence level (CL). More details will be discussed in this thesis.

Chapter 2

The LHC and CMS experiment

The measurements of the different properties of the B meson described in this thesis are based on the data collected in between 2011-2018 by the Compact Muon Solenoid (CMS) detector situated at the Large Hadron Collider (LHC) experiment [51]. This chapter mainly focuses on the LHC's detail and various aspects of the CMS detector. Section 2.1 briefly introduces the LHC and its design along with the operational experiments. Section 2.2 discusses the CMS experiment, providing the detailed information about its sub-detector and performance. A brief overview of the trigger system that is used to select the interesting events is discussed in Section 2.3. The reconstruction and identification algorithm used for the particles are discussed in Section 2.4. Finally, in Section 2.5 the B physics program at LHC is discussed.

2.1 The Large Hadron Collider at CERN

The LHC is a circular accelerator and collider installed at CERN (Conseil Européenne pour la Recherche Nucléaire) near the city of Geneva on the border of France and Switzerland. The LHC is built in the same old tunnel as the Large Electron-Positron (LEP) Collider. The LEP was colliding electron-positron during the year 1984-1989.

The circumference of the LHC is 27 km long. It is designed to accelerate and collide protons as well as heavy ions. To date, LHC is the biggest particle physics experiment ever built and is operating at the highest energy compared to any other collider in the world. LHC started functioning in 2008, the first colliding proton-proton center-of-mass (COM) energy was 7 TeV. But because of some technical fault in the LHC ring magnet, the LHC operation was stopped. The faulty magnet was replaced, and in 2011 the data taking was resumed. Then in 2012, the energy was pushed to 8 TeV. Later, LHC underwent upgrades

to replace the electronics, and in 2015, it achieved 13 TeV COM energy and successfully operated until 2018.

After a long shutdown, LHC became operational again in mid 2022 with slightly higher COM energy and expected to continue until 2026. Then, a major upgrade is scheduled to cope with the high luminosity, after which phase-2 data taking will begin. The LHC timeline is shown in Fig. 2.1. At LHC, two proton beams of particles move in opposite directions

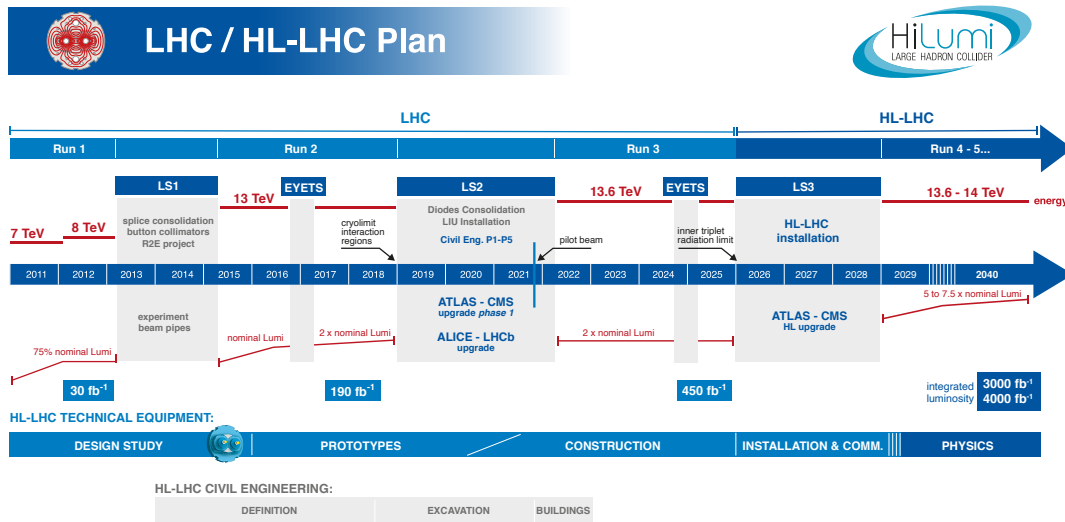


Figure 2.1: LHC timeline [52].

and collide in a few interaction points, which are arranged around the LHC. To study the product of collision, huge massive detectors are placed surrounding the interaction point. There are four different main experiments which are;

- ATLAS (A Toroidal LHC Apparatus) [53]
- ALICE (A Large Ion Collider Experiment) [54]
- CMS (Compact Muon Solenoid) [55]
- LHCb (LHC Beauty Experiment) [56]

The ALICE experiment was specially designed to study physics from the heavy ion (Pb-Pb) collisions. The main focus of this experiment is to study the quark-gluon plasma created in extreme temperature and density conditions during the collision of heavy ions. ATLAS and CMS are the two multipurpose detectors designed to study the new physics effect and precision measurements. They can analyze the data from both proton-proton and heavy ion collisions. LHCb was designed to study the properties of the charm and beauty mesons which are produced in the forward region in the LHC. One of the main goals of the LHC was the discovery of the Higgs boson, the last missing element of the SM. In addition to the Higgs boson, the LHC was designed to verify and explore various beyond Standard Model theories, such as supersymmetry, extra dimensions, and dark matter.

2.1.1 The LHC design

The accelerated proton and heavy ion particles achieve the maximum energy through multiple stages. The different stages are as follows: Protons are stripped from the hydrogen atoms by removing the electrons. Then the protons are accelerated up to 50 MeV by going through the LINear ACcelerator (LINAC2). Again the accelerated protons from LINAC2 are injected into the circular Proton Synchrotron Booster (PSB), which is 157 m long. The booster accelerates the protons to 1.4 GeV and delivers them to the Proton Synchrotron (PS). The PS further accelerates them to 25 GeV and creates 72 bunches with a spacing of 25 ns. Each bunch contains around 10^{11} protons. In the next stage, the bunches are fed to the Super Proton Synchrotron (SPS) to achieve energy up to 450 GeV. Finally, the bunches are transferred to the LHC ring and accelerated up to 7 TeV. The LHC main ring is composed of two rings which accelerate protons or heavy ions in opposite directions. A schematic view of the LHC ring is shown in Fig. [2.2](#). The rings cross at four different places and split into eight octets. The proton bunches are injected into two LHC main rings in octants 2 and 8. In each octant, the dipole magnets are able to control the bunches. Figure [2.3](#) shows

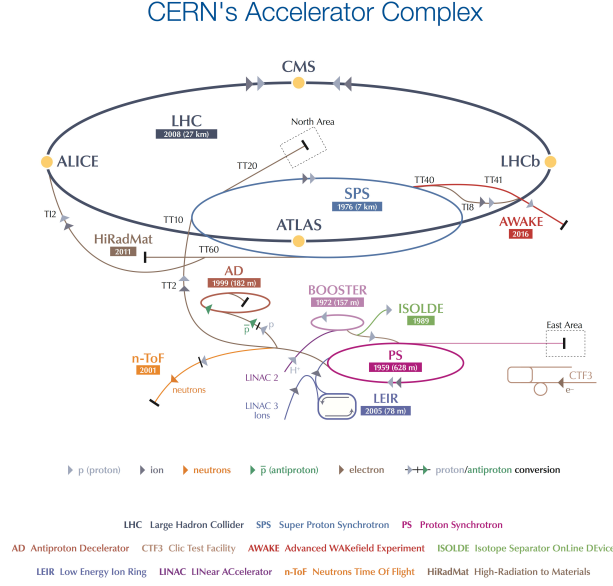


Figure 2.2: Illustration of the CERN accelerator complex [57].

the schematic diagram for complex dipole magnets, which produce a powerful magnetic field of strength upto 8.33 Tesla. The magnetic field is so strong that it can bend the 8 TeV proton beams and keep the beams circulating inside the ring. The operating temperature of all the 1232 superconducting dipole magnets is 1.9 K, which is achieved by liquid helium. The dipole magnets also operate under an ultra-high vacuum, which acts as an insulating medium from the surrounding. In addition to the dipole magnet, there are other magnets (such as quadrupoles, octupoles, sextupoles, decapoles) that are utilized to focus and correct several features of the beam. In total, approximately 9600 magnets are used [51].

2.1.2 The LHC luminosity

In each physics process, the event rate n_{event} can be written as the product of its production cross-section (σ_{event}) at a particular energy, and the luminosity $\mathcal{L}$. The mathematical formula is

$$n_{event} = \mathcal{L} \sigma_{event} \quad (2.1)$$

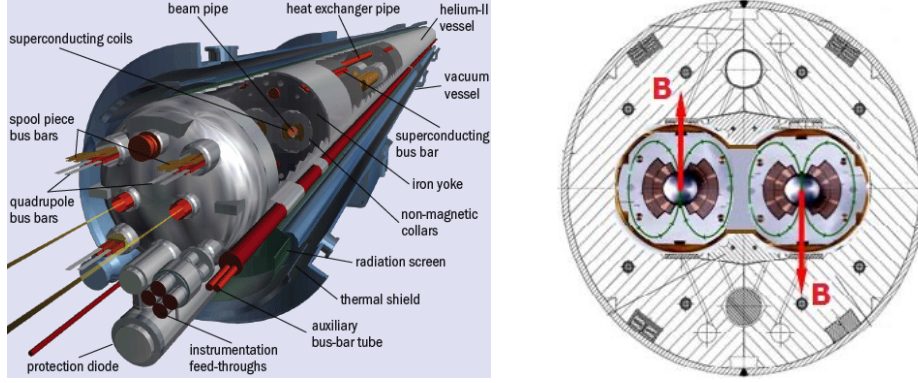


Figure 2.3: Design of the magnetic dipole and the magnetic field direction in the main ring [55].

where the $\mathcal{L}$ depends on the beam parameters, which can be expressed as,

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.2)$$

where N_b is the number of protons per bunch, f_{rev} is the revolution frequency. It also depends on the number of bunches per beam n_b , the relativistic gamma factor γ_r , the normalized rms transverse beam emittance ϵ_n , the betatron function at the Interaction Point (IP) β^* and the geometric luminosity reduction factor F . The design peak luminosity of the LHC is $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ at the beginning of each run, but luminosity goes down with time due to beam collisions. As shown in Fig. 2.1, LHC has completed physics operation and collected the data from the two eras, which lasted from 2011 to 2013 and 2015 to 2018. The evolved integrated luminosity over the period of time is shown in Fig. 2.4. The achievement of high instantaneous luminosity brings a challenge, i.e., pileup (PU), defined as the number of collisions happen simultaneously in a bunch crossing. The average PU is directly varies with the instantaneous luminosity, as

$$\langle PU \rangle = \frac{\mathcal{L} \sigma_{pp}^{inel}}{n_b f} \quad (2.3)$$

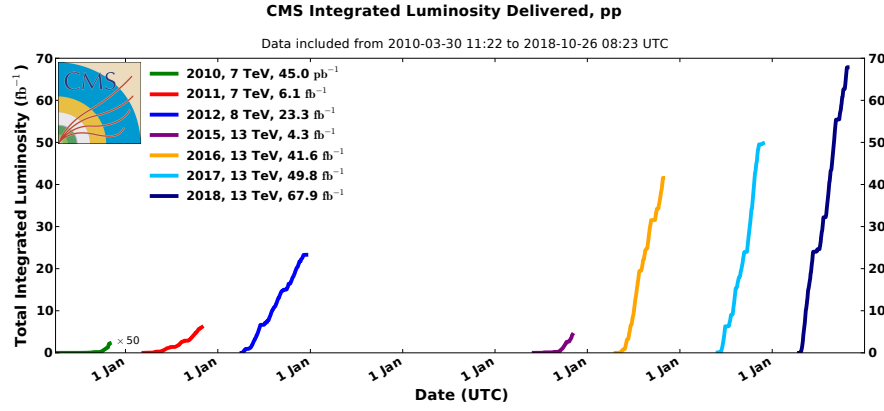


Figure 2.4: Total integrated luminosity versus day delivered to CMS during stable beams and for pp collisions in different year [58].

where σ_{pp}^{inel} is the pp cross-section for inelastic collision. n_b and f are the number of bunch per beam and the revolution frequency. The average pileup per year is shown in Fig. 2.5. With the increase in the pileup values, the resolution of particles and the efficiency of the detector will degrade.

From 2010 to 2013 (referred as Run 1), LHC delivered approximately 29 fb^{-1} of pp collision of data which led to the discovery of the Higgs boson in July 2012. It then underwent a major upgrade to cope with the harsher data-taking conditions for Run 2 operations. Run 2 operation began in 2015 and ended in 2018. Throughout this period, the instantaneous luminosity of LHC pushed beyond the design luminosity. The maximum luminosity reached to $\mathcal{L} = 2.14 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ by the end of 2018. In 2016, 2017, and 2018 the delivered integrated luminosities by the LHC were around 41, 50, and 68 fb^{-1} , respectively. In mid-2022, LHC resumed operation after a prolonged shutdown, and the pp COM energy increased to 13.6 TeV. By the end of 2025, the expected integrated luminosity by LHC will be 450 fb^{-1} , which is two times more than in Run 2. The above data-taking duration is referred to as the Run 3 data-taking period. From the future projection studies, it is found that a major discovery of new physics will not be achieved without a significant increase in integrated luminosity. So the goal of the LHC phase-2 upgrade is to achieve a peak lumi-

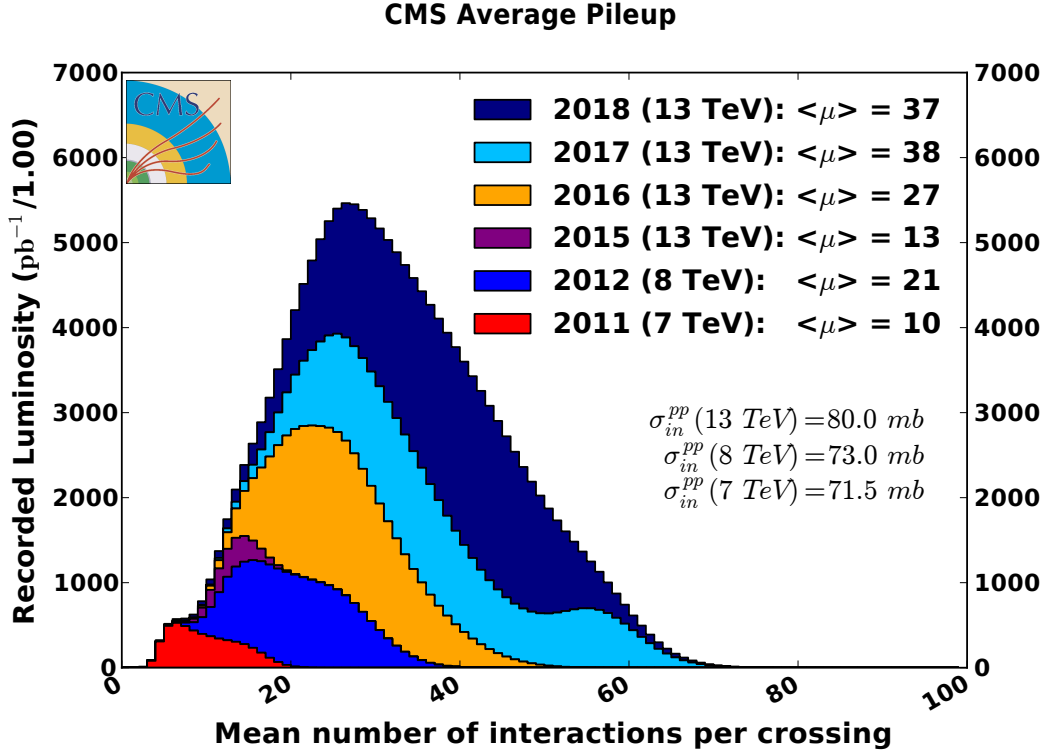


Figure 2.5: Stacked histogram to show the average pileup profile in different data taking year. $\langle\mu\rangle$ is the average pileup, which is calculated from the σ_{in}^{pp} as indicated on the plot [59].

osity of $5 \times 10^{34} \text{ fb}^{-1}$. The total integrated luminosity of 3000 fb^{-1} is expected at the end of phase-2 operation.

2.2 The Compact Muon Solenoid (CMS) experiments

The Compact Muon Solenoid experiment is one of the general purpose detectors situated at the interaction point 5 at the LHC at CERN. It has a cylindrical geometry with a central “barrel” part and two “endcap” parts on both sides and covers a full solid angle to measure all the particles originating from collisions. CMS detector is 28.7 m long and has a diameter of 15 m, for a total weight of 14,000 tons, known as the heaviest LHC experiment. Its subsystems are lined up concentrically around the beam line. The innermost part of the de-

tector close to the beam pipe is tracker. This is followed by two layers of calorimeter. Both the tracker and calorimeters are inside the solenoid magnet. The magnetic field produced by the solenoid is 3.8 Tesla. Surrounding the solenoid, the muon system is present. As the name suggests, muon detection in CMS is very precise. The main motivation behind building the CMS detectors are the following,

- search for SM and BSM Higgs boson decaying into photons, τ leptons, b quarks, W and Z bosons,
- search for new physics processes that decay into multi-leptons or multi-jets or missing transverse energy or any combination of the three,
- study CP violation in the decay of B mesons (like $B_s^0 \rightarrow J/\psi\phi \rightarrow \mu^+\mu^-K^+K^-$),
- search for rare decays like, $B^0 \rightarrow \mu^+\mu^-$, $\tau \rightarrow 3\mu$, and $B^0 \rightarrow \mu^+\mu^-\mu^+\mu^-$,
- study top quark properties, EW physics, QCD, and jet physics at the TeV scale.

Therefore, CMS is known as a multipurpose experiment. The main features of the CMS detector are discussed below,

- as the name suggests, the muon identification and the transverse momentum (p_T) resolution of CMS is excellent compared to other detectors at LHC. CMS can reconstruct muon from $p_T \approx 1$ GeV to $p_T \approx 1$ TeV with the help of inner tracker and muon systems,
- good electromagnetic energy resolution,
- good primary and secondary vertex localization,
- measurement of missing transverse energy with high resolution.

The CMS detector follows a right-handed coordinate system where the origin is set to the nominal interaction point. The x-axis of the Cartesian coordinate system points radially inward with respect to the LHC ring, the y-axis points vertically upward, and the z-axis points along the counter-clockwise beam direction. Since the CMS has a cylindrical geometry, polar coordinates are also used. The conventional Cartesian and polar coordinates are illustrated in Fig. 2.6. The azimuthal angle θ is measured from the x-axis in the xy plane. The polar angle is measured from the z-axis. The pseudo-rapidity of a particle is defined as $\eta = -\ln[\tan(\theta/2)]$. The pseudo-rapidity varies from 0 to infinity, as shown in Fig. 2.7.

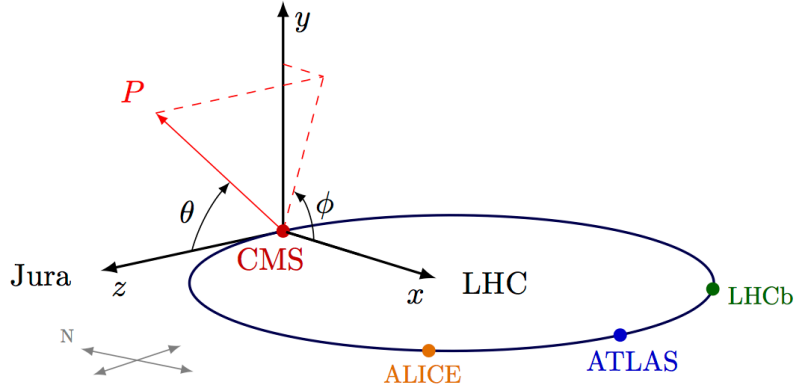


Figure 2.6: Sketch of the CMS coordinate system [60].

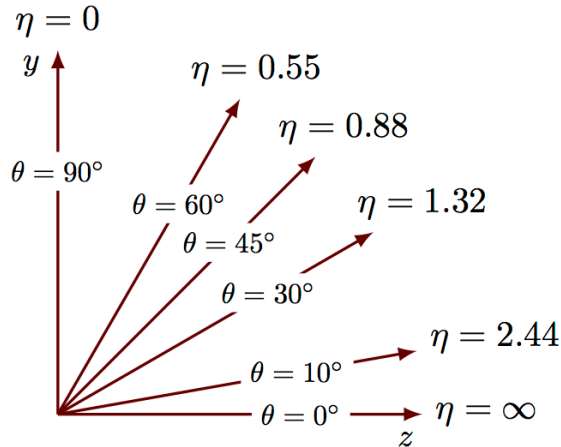


Figure 2.7: Demonstration of the relation between the polar angle θ and pseudo-rapidity η [60].

The radial distance can be expressed as $r = \sqrt{x^2 + y^2}$. Similarly, the momentum of a particle can be expressed as the Lorentz boost invariant form, i.e. $p_T = \sqrt{p_x^2 + p_y^2}$ in the transverse plane.

The momentum of charged particles that pass through the detector is measured with the help of a magnetic field. The magnetic field produced by the superconducting solenoid should be able to bend the particle trajectories, and the momentum can be measured from the bending. The momentum resolution of the charged particle depends on the magnetic field and the solenoid radius,

$$\frac{dp}{p} = \frac{p}{BL^2} \quad (2.4)$$

where L is the length of the trajectory that the particle travelled inside the magnetic field. In other words, it is equal to the radius of the solenoid. The CMS solenoidal magnet generates a strong magnetic field of 3.8 Tesla. The solenoid is 13 m long and 6 m in diameter and is made up of 4 layers of windings of NbTi cable. To achieve the superconducting state, the cables are cooled to 4.5 K by a helium refrigeration system. An accuracy of $20 \mu\text{m}$ is achieved in the measurement of vertex position with the help of a solenoidal magnetic field because it manages the bending of track trajectories in the transverse plane. The size of the solenoid allows efficient track reconstruction up to $|\eta| = 2.5$.

2.2.1 The tracking system

The innermost sub-detector of CMS is a silicon tracking system with a 2.5 m diameter and 5.8 m length. The motivation behind the design of the tracker is to provide a precise and efficient measurement of the trajectories of charged particles appearing from the interaction point. It helps to measure the p_x , p_y , p_z , and charge of the charged track and also allows for reconstructing the primary, secondary, and sometimes higher decay vertices (like K_s^0 and Λ^0). The reason for using silicon semiconductors over other semiconductors like Germanium is that they are abundantly accessible, easy to synthesize and withstand high

temperature.

The CMS tracking system is built with two different technologies to fulfill the challenging requirements: silicon pixel and silicon strips, as shown in Fig. 2.8.

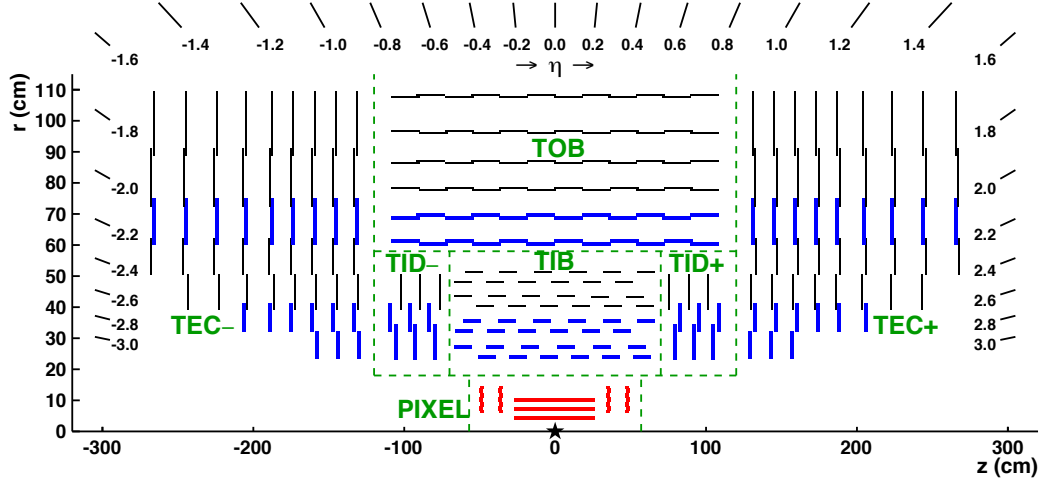


Figure 2.8: Longitudinal schematic view of the CMS tracker in rz plane. The pixel detector represented above were replaced in the year 2017. Each line corresponds to a detector module. The strip tracker consists of tracker inner barrel (TIB), tracker inner disk (TID), tracker outer barrel (TOB) and tracker endcap (TEC) [61].

There are 1440 pixel modules collectively known as the pixel detector, which are arranged in three barrel layers and four disks in the endcap. The active surface of the pixel detector is one square meter. It is present near to the beam pipe and finely segmented to give excellent spatial and time resolution. As I discussed, many simultaneous interactions occur during the time of bunch crossing. In addition, the secondary vertex exists close to the primary vertex, displaced up to a few mm, where the heavy particles such as the τ lepton and hadrons constituting b and c quarks decay to daughter particles. The pixel detector plays a major role in reconstructing the vertex and also delivers the seed to the tracking algorithm. To handle the increase of luminosity, an upgraded pixel detector was installed in March 2017. The layout of the original and upgraded pixel detector is compared in Fig. 2.9.

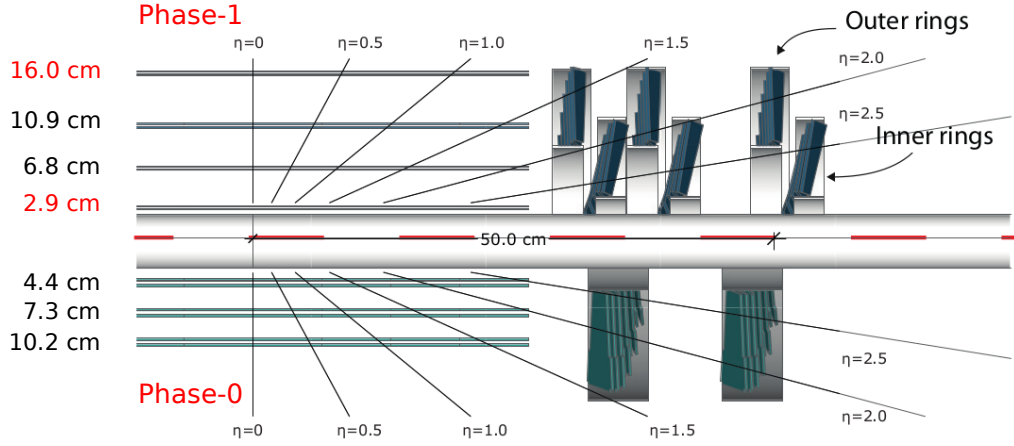


Figure 2.9: Comparison of the layout of current upgraded pixel detector and original pixel detector [62].

The upgraded pixel detector consists of four layers in the barrel and three discs in each endcap. The size of the pixel cell is $100 \times 150 \mu\text{m}^2$, and it provides a vertex spatial resolution of 15-20 μm .

Surrounding the pixel detector, the outer silicon strip tracker is present. The strip tracker consists of tracker inner barrel (TIB), tracker inner disk (TID), tracker outer barrel (TOB), tracker endcaps (TEC). Approximately 9.3 million silicon strips are present, and the sizes differ for inner and outer strips. The TIB covers the region radially from $20 < r < 55 \text{ cm}$, and the TOB covers up to $r = 116 \text{ cm}$. Similarly, the TID and TEC cover the region from $58 < |z| < 124 \text{ cm}$ and $124 < |z| < 282 \text{ cm}$. The detector occupancy decreases with increasing radial distance; therefore, a very good spatial precision is needed close to the beam pipe. The silicon tracker is made up of silicon semiconductors. The working principle of the semiconductor is that it creates an electron-hole pair when a charged particle traverse through it and induces a electronic signal. As it can be seen within a given layer, each module is slightly shifted in r and z with respect to its neighbour to minimize the holes in accep-

tance. The blue line, as in Fig. 2.8, specifies that one module is rotated with respect to the other so that one can accurately determine the hit position in three dimensions. The whole system is placed in -20°C temperature to cope with the high-level radiation from the bunch crossing. The tracker is designed in such a way that inside the tracker volume, there will be less inactive material so that the energy loss due to bremsstrahlung, photon conversion, multiple scattering, and nuclear interactions are small, resulting in small distortion of the energy measurement in the calorimeter. The tracking material and the services like cooling pipes and cables represent the material of radiation length 1.6. With the upgrade of the pixel detector, the tracker material was subsequently reduced by 40% in endcaps and 10% in the barrel, and the IP was improved up to a factor of 1.5 in a longitudinal direction. Fig. 2.10 illustrates the material budget of the current CMS tracker in units of radiation length (X_0) as a function of the pseudo-rapidity,

2.2.2 The electromagnetic calorimeter (ECAL)

The CMS ECAL is made up of homogenous lead tungstate crystals (PbWO_4) to measure precisely the energy of the photons and electrons. The motivation behind the design is that the incident electrons or photons leads to an electromagnetic shower and loose most of its energy within the ECAL volume itself. The energy measurement is performed based on the scintillating light produced by the particles of the shower. Here lead tungstate is used as both absorber and active material. The crystal has certain excellent properties compared to other crystals, that are high density ($\rho = 8.28\text{ g/cm}^3$), small radiation length ($X_0 = 88\text{ mm}$), and small Molière radius (22 mm). These crystals produce fast light yield; for a given crystal, 80% of the light is emitted within 25 ns, which is the time gap between the bunch crossing in LHC. The ECAL layout is shown in Fig. 2.11. It has been separated into two parts; ECAL barrel (EB) and ECAL endcap (EE). ECAL barrel extends up to $|\eta| < 1.497$ and the two endcaps covers $1.497 < |\eta| < 3.0$. The barrel is instrumented with about

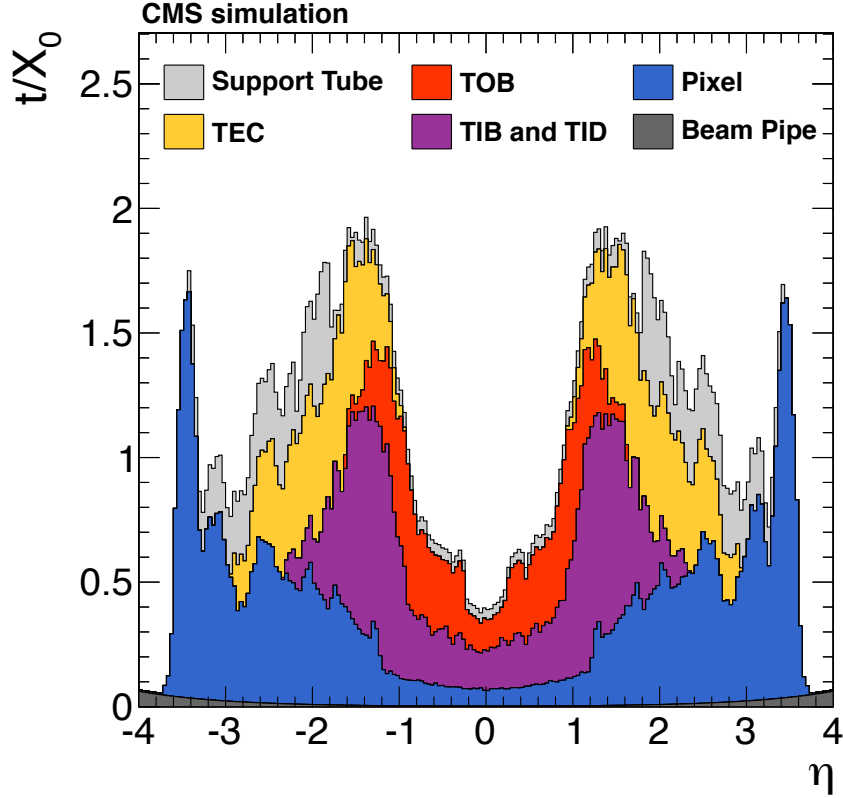


Figure 2.10: Total thickness of tracker material in the unit of radiation length as a function of pseudo-rapidity [55].

61,200 trapezoidal crystals, each with surface area $22 \times 22 \text{ mm}^2$ and 23 cm length. The endcap consists of 7324 crystals of each surface area $28.6 \times 28.6 \text{ mm}^2$ and length 22 cm. The average length of the crystal in the barrel and endcap corresponds to 25.8 and 24.7 radiation lengths, which are sufficient to contain more than 98% of the shower produced by photons and electrons at an energy of 1 TeV. The crystals are tilted up to 3° with respect to the position of the nominal interaction point to avoid the acceptance gaps between the crystals. The silicon avalanche photodiodes (APDs) are used in the barrel region to collect and amplify the scintillation light from the crystal. The APDs are designed for the high particle flux and intense magnetic field. In endcaps, vacuum photo triodes (VPTs) are used. During the period of data taking, the crystals are monitored regularly for transparency loss

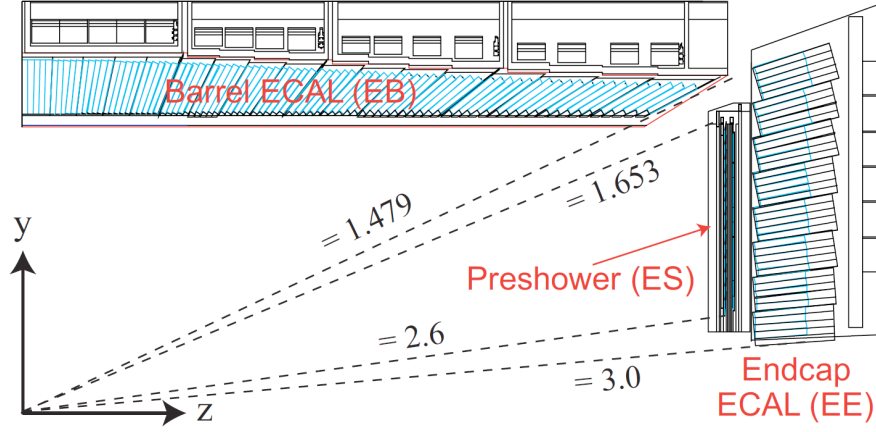


Figure 2.11: Longitudinal cross-sections of a quadrant of ECAL detector [55].

due to radiation damage. Later on, the corrections are applied to compensate for the change in crystal response.

In front of each endcap, a much finer electromagnetic pre-shower detector (ES) is installed to provide an improved spatial resolution in the region $1.65 < |\eta| < 2.6$. The ES is a sampling calorimeter with two layers of lead absorber followed by 22 mm-long silicon strips to calculate the energy sediment and the transverse profile of the shower. In particular, it helps to differentiate between a single high-energy photons and pairs of collimated low-energy photons resulting from $\pi^0 \rightarrow \gamma\gamma$ decays. Over the time, the ES suffers a loss in identification potential because of the large number of neutral pions originating from the tracker material.

The energy measurement in an electromagnetic scintillator is based on the principle of the energy released in the crystals is proportional to the energy of the incident particle. However, the real energy resolution is degraded by various factors. The intrinsic energy

resolution ($\frac{\sigma_E}{E}$), measured with electron test beam studies, is parameterized as,

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2 \quad (2.5)$$

The first term on the right-hand side of Eq. 2.5 represents the stochastic fluctuations in the event-by-event measurements; it is small as a result of the homogeneous nature of the calorimeter. The second term corresponds to the noise contributions from the electronics, the digitization chain, and the pileup. Finally, the third term is due to non-uniformities in the detector response, miscalibration, and energy leakage. Over time, the energy resolution of the CMS detector slightly degrades in the forward region because of the loss of transparency in the crystal due to large doses of radiation. The loss is monitored by injecting laser light into each ECAL crystal and measuring its response during each re-fill of the LHC.

2.2.3 The hadronic calorimeter (HCAL)

The energy of hadrons is measured in HCAL. It is situated outside of the ECAL. When proton-protons collide at the interaction point, the quarks are produced and immediately undergo hadronization. The hadrons interact with the HCAL material and produce a hadron shower, which is initiated by gluons. As a result, narrow jets made mostly of hadrons and photons are produced in the same direction as the quark or gluon that initiated the shower. The HCAL layout is shown in Fig. 2.12. The HCAL sub-detectors has a wide η coverage, up to $|\eta| = 5.2$. The barrel hadronic calorimeter (HB), covering up to $|\eta| < 1.4$, consists of alternating layers of brass absorbers and plastic scintillator tiles. The relative size of the alternate materials is optimized to maximize the hadron interaction length (λ_i) within the volume constrained by the solenoid. Wavelength shifter fibers are embedded in the tiles to transmit the collected light to hybrid photodiodes. The endcap hadronic calorimeter (HE) has a similar design and covers the endcaps in the $1.3 < |\eta| < 3.0$ region. The thickness of HB and HE are between $7\lambda_i$ and $10\lambda_i$. To improve the longitudinal confinement of the

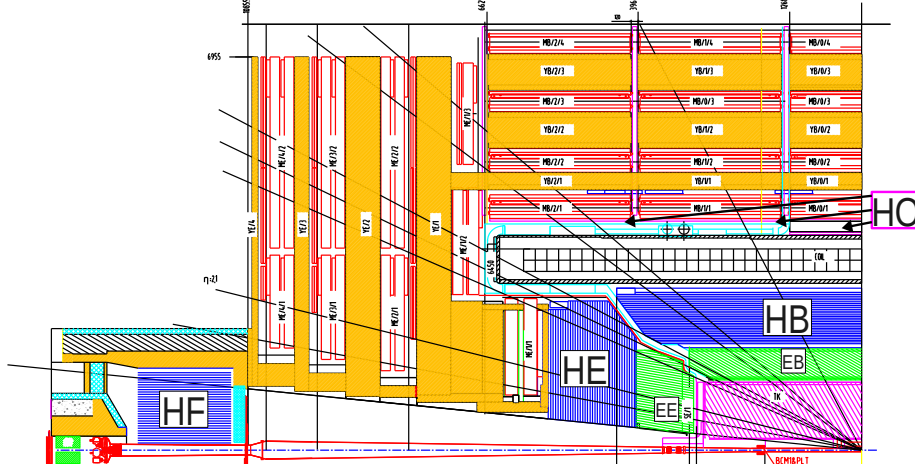


Figure 2.12: Longitudinal cross-section of a quadrant of the HCAL [55].

hadronic showers, an outer hadronic calorimeter (HO) is placed outside the solenoid volume, covering the $|\eta| < 1.4$ regions; it consists of scintillating material rings embedded in the yoke structure, read out by silicon photomultipliers (SiPM). Finally, the $3 < |\eta| < 5.2$ region is covered by the forward hadronic calorimeter (HF) calorimeter, deploying steel absorbers and quartz fibers, which are more resistant to the intense radiation acting on the forward detectors; the Cherenkov light produced by the quartz medium is collected by photomultiplier tubes (PMT). From beam test analyses, the combined resolution of the ECAL + HCAL barrel calorimeters has been measured to,

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84\%}{\sqrt{E}}\right)^2 + (7.4\%)^2 \quad (2.6)$$

The first term accounts for stochastic effects such as statistical fluctuations in the shower development; the second constant term is due to detector effects independent from the energy, such as imperfect calorimeter calibration. The modest energy resolution degrades the calorimeter-based reconstruction of jets and hadronic tau leptons; therefore, the reconstruction of such objects rely on the Particle Flow algorithm, which optimally exploits the whole detector to achieve improved energy and angular resolution. The wide coverage of

the HCAL makes it sensitive to most of the collision products. In order to maintain its performance, certain components of the HCAL were upgraded at the end of 2017. The photodetectors in the HB, HE, and HF were replaced, and the functionality of the readout electronics was expanded. With the introduction of precision timing measurements and longitudinal depth segmentation, the pattern recognition capabilities in the HCAL were improved, providing an additional handle for background rejection.

2.2.4 Muon chamber

As the name and logo of the CMS project suggest, the core specification of the detector design is the precise measurement of muon's momenta. Muons are produced in hard scattering, and their signatures make them appropriate candidates for trigger decisions. Muons are produced with energy ranging from a few GeV to a few TeV. Muons lose less energy when interacting in the tracker material than electrons because they are less subject to radiative effects due to their higher mass. This is one reason why the muon detector, the outermost part of the CMS detector, sits outside the solenoidal volume. The magnetic field of strength 2 T is produced inside the iron structure by the return field of the solenoid magnetic field, which helps to measure the charge and momentum of the muons complementary to the silicon tracker. The muon system is made up of a gas ionization chamber, which is reliable and cost-effective. Three different technologies based on gas ionization chambers were deployed: drift tube (DT) chamber, cathode strip chamber (CSC), and resistive plate chamber (RPC) [55]. To achieve the maximal coverage and maximal reconstruction efficiency, the different chambers are situated in different places, as shown in Fig. 2.13.

The DT chambers are located in the central region ($|\eta| < 1.2$), where there is a low particle rate (muon occupancy and neutron background are low), and the magnetic field is uniform and as weak as 0.4 T. Each chamber consists of twelve levels of drift tubes and each tube is filled with a mix of argon (85%) and CO₂ (15%). Their spatial resolution ranges

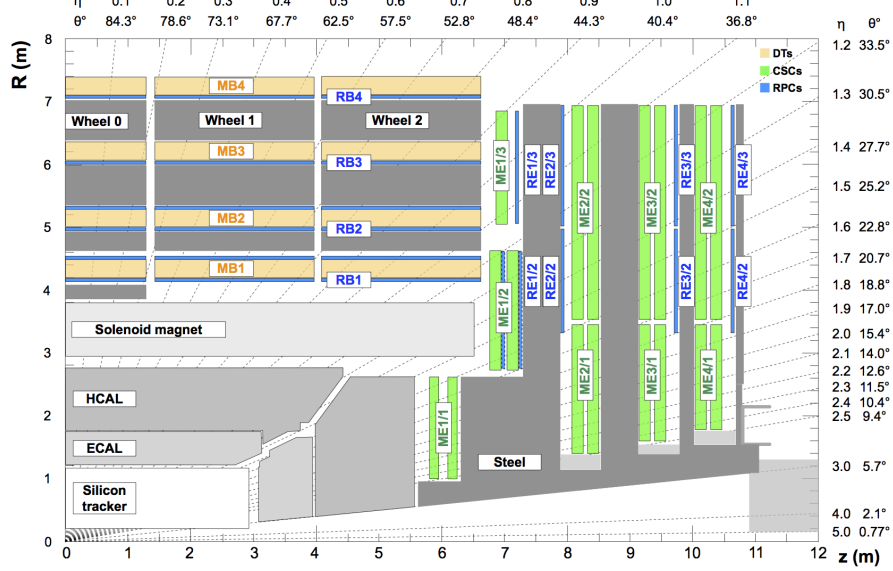


Figure 2.13: Longitudinal cross-section of a quadrant of the CMS detector showing different chambers of the muon system [55].

from about 80 to 120 μm in the (r, ϕ) plane and from 200 to 300 μm in the z -directions. When a muon traverses the gas, its position and angle are inferred from the time needed for the knock-off electrons to drift towards the anode. The CSC detectors are designed to tolerate the strong and non-uniform magnetic field of the endcap regions ($0.9 < |\eta| < 2.4$) and have a very fast response. They have a trapezoidal shape and contain alternate layers of anode wires and cathode strips filled with a mix of argon (45%), CO_2 (50%), and CF_4 (10%). When a muon passes through the chamber, it ionizes the gas, and the position of the muon is obtained from a combination of the induced signals. It has a very fast response, and provides a spatial resolution of 40-150 μm . Finally, the RPC chambers are located in the barrel and endcap regions, covering up to $|\eta| < 1.9$. The main advantage of using a RPC is that it has an excellent time resolution (< 3 ns). They are formed by two gaps of resistive 2 mm-thick layers interlaid with a mixture of gas (95.2% $\text{C}_2\text{H}_2\text{F}_4$, 45% $\text{i-C}_4\text{H}_{10}$ and

0.3% SF₆), operated in avalanche mode. When traversed by a muon, an electron cascade is triggered by the high electric field inside the volume and read out with strips located on the outer surface. They are particularly useful at identifying the bunch crossing associated with a muon track, even in the presence of a high pileup, which is an essential feature of the muon trigger system. At the end of 2017, an additional detector chamber was installed in the region of $1.6 < |\eta| < 2.2$ to increase the redundancy of the muon system in the endcaps. The technology used is the gas electron multiplier (GEM) [63], which is well suited for this purpose as it has a thin profile and excellent rate capability while it is able to withstand the high particle fluxes of the forward region. The upgrade consisted of the pre-installation of 10 prototype chambers; the complete installation of the 144 chambers on the first layers of two endcaps is completed before the start of the Run 3 data-taking period.

2.3 The trigger

The CMS experiment has two levels of the trigger to select exciting events from the billions of events produced per second in a bunch crossing. The current average rate of bunch crossing is 40 MHz at CMS. No technologies are available nowadays to process and store such a large amount of data. Most of the SM physics processes cross-sections are of the order 10^5 pb (pico barn), while the total pp interaction cross section is six order of magnitude higher. So most of the produced events are not interesting for new physics study. The CMS trigger aims to select rapidly events with a possible physics of interest from the billions of events. The trigger systems need to be more efficient in different data-taking conditions and robust against the high instantaneous luminosity and pileup delivered by the LHC. The trigger selection is performed using the produced particle's kinematic properties as input. It has been divided into two steps to reduce the rate. The first step is performed at the hardware level known as the level-1 (L1) trigger, which selects up to 100 kHz of the most

interesting events [64]. The processing time of the L1 trigger is $3.8 \mu\text{s}$. Followed by this, a high-level trigger (HLT), which is software-based, performs a detailed check of events in a computing processor farm close to the detector. The rate after the HLT is reduced to 1 kHz with a processing time $\approx 200 \text{ ms}$.

2.3.1 Level-1 trigger

The L1 trigger is the starting point of the data acquisition chain, which demands flexibility in collecting the events in the harshest scenario. Starting from 2015, LHC has achieved its designed instantaneous luminosity and beyond, which brings more challenges during the data-taking because of the pileup. There are many soft collision that occur during the bunch crossing, which are not interesting for physics searches. With the increase in the pileup, the decision-making of the L1 trigger becomes strong. A major upgrade of the L1 trigger was installed and commissioned between 2015 and 2016. The electronic boards were replaced by Advanced Mezzanine Cards arranged with powerful field-programmable gate arrays. The L1 trigger uses the information from the calorimeters and the muon chamber separately to execute a decision.

2.3.2 High-level trigger

This is the second step of the trigger, which reduces the rate of the output of the L1 trigger from 100 kHz to 1 kHz. HLT is implemented in a computing farm of 32000 cores. The object reconstructions are performed using full detector information, including tracking information. The average processing time of an event is 320 ms. The L1 objects are used as a starting point to reconstruct the HLT candidates. The set of steps performed for HLT reconstruction is commonly known as the “HLT path”. The HLT path consists of binary decisions for object quality referred to as “filters”. The HLT paths are optimized in such a way that the most discriminate filter is applied first and then subsequent filters. The HLT

reconstruction used the information from the calorimeter and muon chamber first to reduce the rate, and then the track reconstruction is performed, which is more CPU extensive. The full HLT selection uses more than 600 paths for various physics and monitoring needs.

The dimuon invariant mass spectrum selected by all the HLT paths with the 3 fb^{-1} of 2015 data is shown in Fig. 2.14. The different mass windows of the HLT are clearly distinguished in the plot.

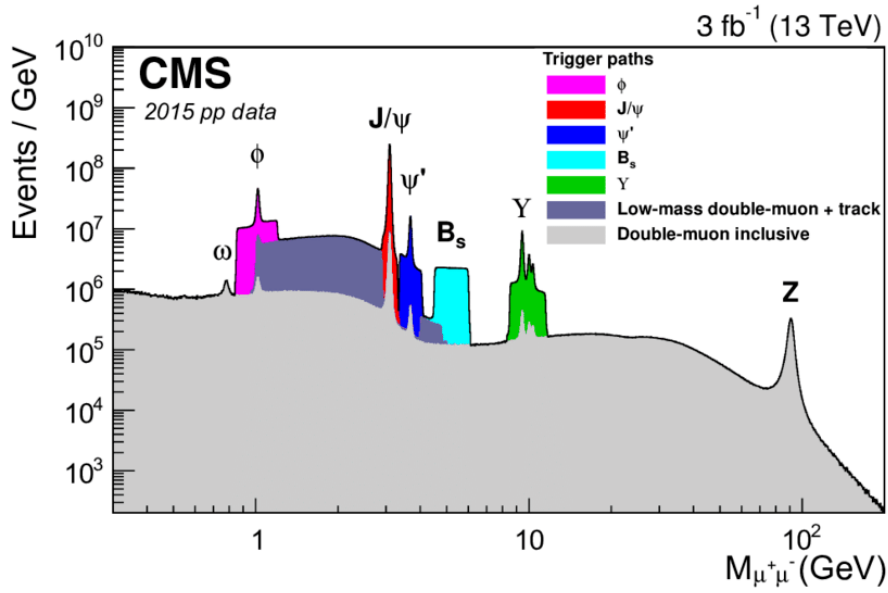


Figure 2.14: Dimuon invariant mass distribution collected by different HLT path in the year 2015 [65].

2.4 Particle reconstruction and identification

During the course of pp collision, every type of particle produced out of collision interacts differently with the sub-detectors portrayed in Fig. 2.15; however, the reality is much more subtle.

Due to the high-density material in the innermost part of the detector, the muon loses a small part of the energy before reaching the muon chamber. The energy and charge of the muons are reconstructed in the inner tracking system and muon chamber. Electrons and

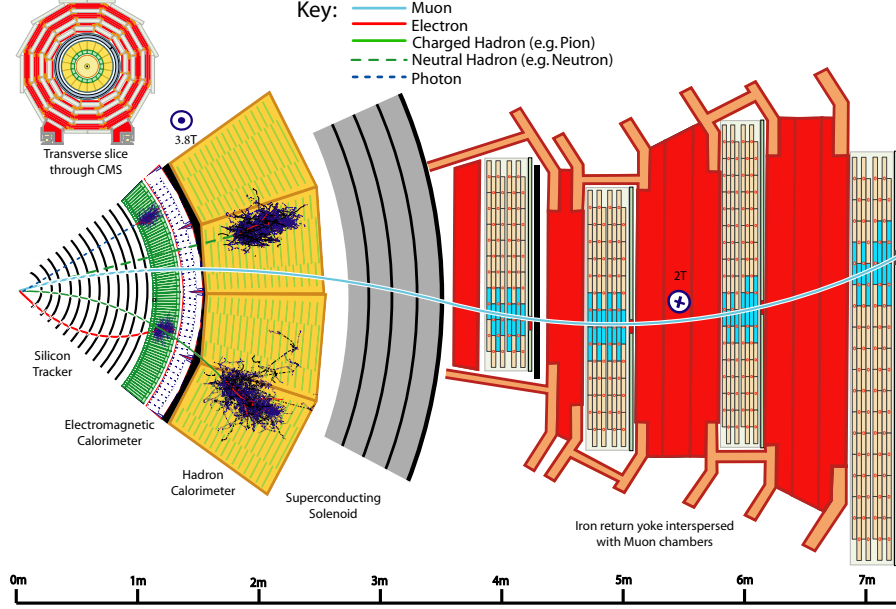


Figure 2.15: Interaction of different particle with the detector material [66].

photons dissipate their energy in the ECAL, and the charge of the electrons is measured in the tracking system. After passing through the ECAL by losing a small part of the energy, hadrons deposit most of the energy in the HCAL. The tracker also provides information about the trajectory of their charged components.

Finally, neutrinos interact negligibly with the detector material and go undetected; however, the presence of the neutrino is accounted for by taking the energy imbalance of the event in the transverse plane. The particle flow (PF) algorithm [67] is the most general reconstruction algorithm in CMS. The algorithm begins with the hit information of the silicon tracker and energy deposition of the calorimeters to construct tracks and clusters. These are the basic inputs to the physics object reconstructions. Before the LHC period, the idea behind the physics object reconstruction was built on the signals collected separately by each sub-detector. Instead of treating the information from each sub-detector separately, a combined way of tracking and clusters of energy for more accurate energy, momentum, and position measurement is the founding principle of the CMS reconstruction algorithm.

The primary goal of the PF algorithm is to reconstruct the relevant properties such as the nature, momentum, and direction of each particle, for electrons, muons, photons, taus, and jets, along with precise particle identification.

2.4.1 Tracks and vertices reconstruction

The CMS tracking system aims to identify and measure particle tracks efficiently. The high energetic particle tracks are well-measured, which means high purity and a low misidentification rate. But for low energy tracks like hadrons, tracking is not so efficient. As a result, a degraded energy resolution of the hadronic track is observed. In a typical LHC event, charged hadrons carry two-thirds of jet energy. Therefore, any inefficiency in the tracking will damage a proper reconstruction of the jet energy and angular resolution.

2.4.2 Charge particle tracking

The trajectory of the charged particle inside the CMS detector is reconstructed by merging the trailing hits left on the tracker layers. The algorithm followed to reconstruct the tracks is known as the combinatorial track finder (CTF) [68]. The CTF algorithm can be split into three steps: seed generation, track finding, and track fitting. The track seeds start by using the hit information from the pixel detector because of the high efficiency, the low occupancy, and a better resolution for the two-dimensional position. The track finding step then uses a Kalman filter algorithm [69] to sequentially extrapolate the track from the beam interaction point to outward. The trajectory is extrapolated to each tracker layer, and hits are assigned to the tracks based on the χ^2 between the predicted and measured positions. Each track is assigned with a quality criteria on the basis of χ^2 and the number of missing hits. Only the tracks that satisfy the best quality criteria are kept for further propagation. In the track fitting step, all valid tracks are refitted with a Kalman filter and a second filter running backward, towards the beam spot, with the result of the first one. Fig. 2.16 shows

the expected performance of the track reconstruction for muons and pions with different transverse momenta. The muon track reconstruction efficiency is about 99% up to $|\eta| < 2.1$, and then drops due to reduced coverage of the forward pixel detector. On the other hand, the lower pion efficiency results from the interaction with the tracker material.

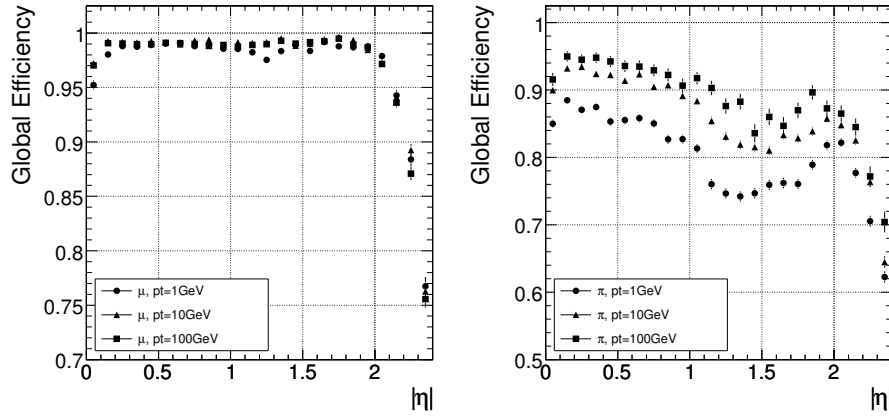


Figure 2.16: Track reconstruction efficiency for muons and pions as a function of pseudorapidity for different transverse momentum [55].

2.4.3 Muon reconstruction

CMS is robust enough to identify and reconstruct muons over the full CMS detector acceptance region with high precision and efficiency. Muon plays a vital role in the discovery of the New Physics processes and precision measurement of the SM process. Out of many produced particles during each beam collision, only a muon survives to the muon system. CMS provides an excellent muon momentum resolution with the help of a solenoidal magnetic field. The muon reconstruction is performed in different stages to exploit the subdetectors information.

- Standalone reconstruction: In the offline reconstruction, the seed is being reproduced from the hits found in the detector. The seed finding algorithm is based on DT and

CSC segments, where only one segment is already sufficient. Eventually, the full muon track is built using the Kalman filter [69] technique. By default, the seed is propagated to the innermost layer of the muon system from which the track reconstruction is started using an inside-out filter to refine the seed. The outside-in filter is then applied and the track is built.

- global reconstruction : The standalone muon tracks are matched with the tracks from inner tracker, tracks reconstructed within the tracker system, to improve the quality of the muon track. The compatible tracks are fitted globally and merged into a global muon track.

Muons are identified as tracker muons if extrapolated tracks from the inner tracker match with one of the track segments from the muon chamber. About 99% of the muons produced are reconstructed either as global muons or tracker muons, or both. Muons of transverse momentum below 10 GeV undergo multiple scattering in the iron yoke material and are often reconstructed as only tracker muons. Sometimes, the high p_T charged hadrons and muons from decay in flight punch through the HCAL volume and are mis-reconstructed as muons, resulting in a contaminated low p_T tracker muons collection. The resolution of the transverse muon momentum as a function of transverse momentum is shown in Fig. 2.17.

2.5 B physics at the LHC

At LHC, when proton-proton moving in the opposite direction collides, they can either go through an elastic or inelastic scattering. The kinetic energy is conserved in the case of elastic scattering but not in the case of inelastic scattering. In inelastic scattering, the internal degrees of freedom of the protons increase, which results in fragmentation. The internal constituent of protons, quarks, and gluons, follow the *parton distribution functions* (*pdfs*), which represents the fraction of proton momentum they carry, x , at a squared energy

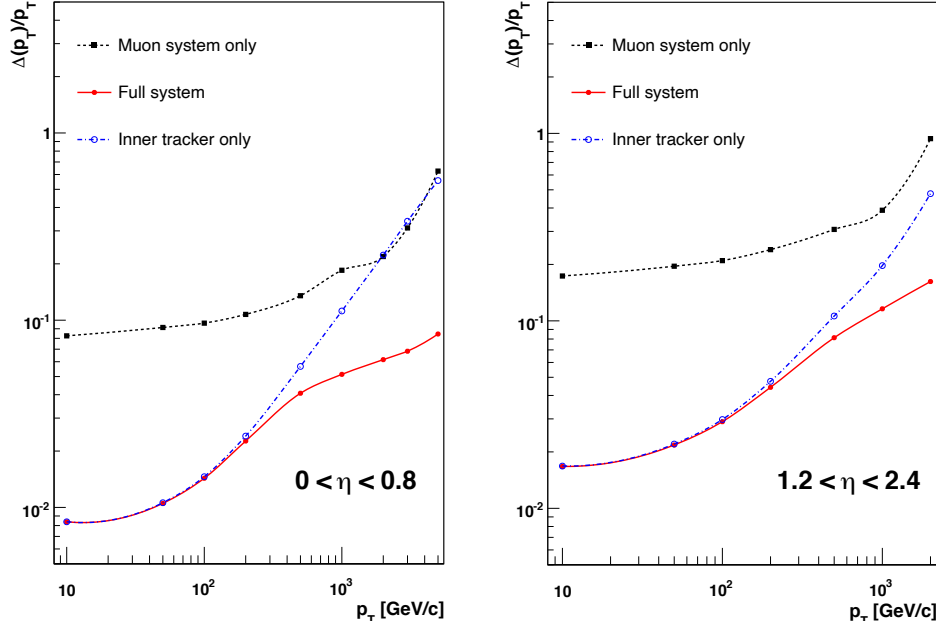


Figure 2.17: Resolution of muon transverse momentum as a function of transverse momentum in barrel region (left) and endcap region (right) [55].

scale Q^2 . Here the partons can take any fraction of momentum; that is why the hadron collider experiments are suitable for studying a wide mass range of hadrons.

***b*-quark production mechanisms**

b-quark production occurs in many ways for a given energy scale Q^2 , and can be divided in three categories,

- Pair Creation: which includes hard QCD scattering ($q\bar{q}$ annihilation and gluon gluon fusion) at leading order and also next to leading order.
- Flavor excitation: it occurs when one *b* quark of a $b\bar{b}$ pair from the proton sea scattering against a gluon from the other proton.

- Gluon splitting: where one gluon splits into a $b\bar{b}$ pair, in the initial or final state.

The $b\bar{b}$ production also can happen from the heavy resonance decay, such as $t \rightarrow bW^+$, $Z \rightarrow b\bar{b}$, and $H \rightarrow b\bar{b}$. Their contribution is negligible compared to QCD production at proton-proton or proton-antiproton collider. The expected b -quark production cross-section as a function of $\sqrt{s}$ is shown in Fig. 2.18. The plot on the right shows the contribution of the different production processes on the different energy scales. For small energies, the dominant contribution is from the pair creation mechanism, followed by flavor excitation and gluon splitting. This is well understood because of the fact that no strong kinematic requirements are required for the flavor excitation mechanism. The situation changes as we increase the energy, the flavor excitation overtakes the pair production, and the contribution from gluon splitting starts to approach the other two. The reason behind this behaviour is not so difficult to understand. Any partonic process contains a hardest $2 \rightarrow 2$ scattering process irrespective of energy, whereas the number of branchings in the shower both in the initial and final state increases with energy due to the higher gluon virtuality.

The $b\bar{b}$ cross-section at $\sqrt{s} = 13$ TeV is estimated to be 500 microbarn. The predominant co-linear production of the $b\bar{b}$ pair in either beam direction allows the LHCb experiment to benefit more compared to ATLAS and CMS. But ATLAS and CMS can compete with LHCb at the same level because of their higher luminosity, which is a factor of 100 higher than that of LHCb. The very large cross-section of the $b\bar{b}$ production, yielding $\mathcal{O}(10^7)$ $b\bar{b}$ pair production per second ($\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$), allows unique access to the rare decays and their precise properties measurement.

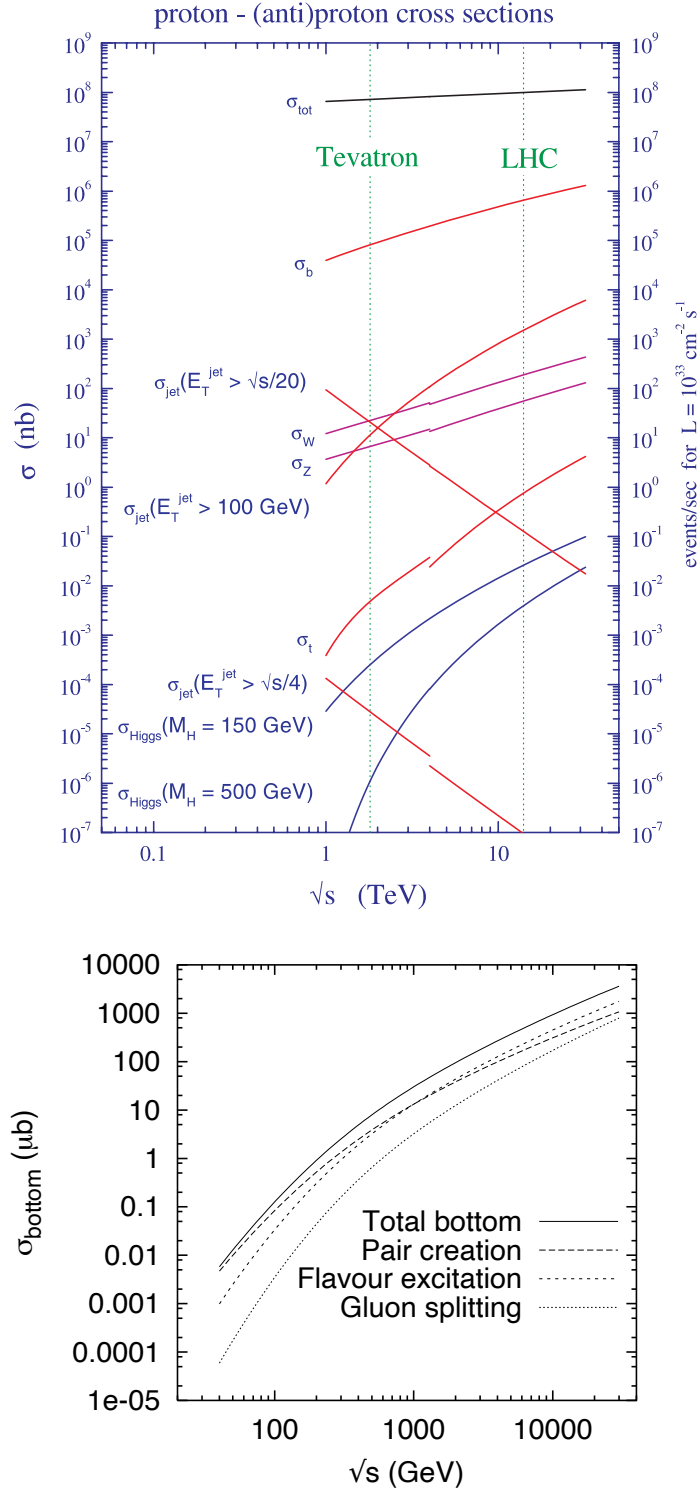


Figure 2.18: Standard Model cross-section of proton-proton and proton-antiproton [70] (top) and individual bottom quark production cross-sections are shown (bottom) [71].

Chapter 3

Measurement of properties of $B_s^0 \rightarrow \mu^+ \mu^-$ decays and search for $B^0 \rightarrow \mu^+ \mu^-$

This short chapter describes event reconstruction and preparation of data for physics analysis. It defines how full event reconstruction is performed for different decay modes and how the normalization channel is chosen. Section 3.1 gives an overview of the event reconstruction. The choice of normalization channel for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yields extractions is discussed in Section 3.2. The dataset and triggers used for both analyses are discussed in Sections 3.3 and 3.4. A dedicated technique for choosing a good muon to suppress the background in data is defined. The detail of the muon selection is discussed in Section 3.5. These discussions are performed with a view to simplifying and unifying discussions for the physics analyses described in Chapters 4 and 5, respectively.

3.1 Analysis overview

We reconstruct the rare leptonic neutral B decays, in particular, $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$, by combining two unlike-sign muons that come from a common vertex. The invariant mass of two good muon tracks is required to be within 4.9 – 5.9 GeV. In data, along with the signal, different sources of background, which mimic the signal topology, are sorted into three categories. The first category, known as the combinatorial background, which is the main background as shown in Fig. 3.1, is formed by the combination of two uncorrelated muons, which can come from two different b or c decays. This contributes to the flat shape in the full invariant mass distribution. The second type of background is the semileptonic background, which results from the three-body decays of B^0 , B^+ , B_s^0 , and Λ_b particles. For example, $B_s^0 \rightarrow K^\pm \mu^\mp \nu / \pi^\pm \mu^\mp \nu$, here the real muon combined with the hadron misidentified as muon to mimic the signal like event. Since the third particle, neutrino (ν), is not

reconstructed, this makes a continual background in the lower side of the signal invariant mass. The final category includes the decay of B^0 , B^+ , B_s^0 , and Λ_b decays into two hadron tracks, such as $B_s^0 \rightarrow \pi^+ \pi^-$, and both hadrons are misidentified as muons. These decays can form a broad peak under the B^0 invariant mass region.

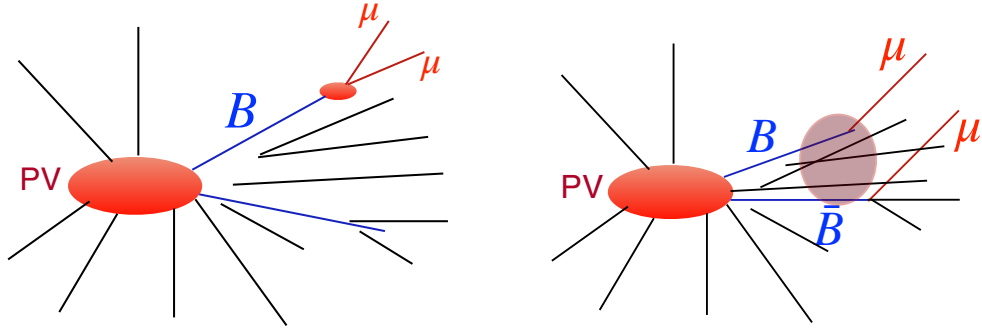


Figure 3.1: Left and right plot shows the sketch of signal and combinatorial background events. Here PV stands for primary vertex.

The event selection for $B^+ \rightarrow J/\psi K^+$ starts from the two opposite charged muons, matched with the trigger of displaced J/ψ mesons, coming from a secondary vertex (SV). Each high purity track in the event is then coupled to the dimuon to form an offline B^+ candidate. CMS has no dedicated particle identification detector, so the kaon mass hypothesis is assigned to each track along with kinematic restrictions ($p_T > 0.6$ GeV and $|\eta| < 2.4$). The two muons of the candidate must have an invariant mass in the range $3.05 < m_{\mu^+ \mu^-} < 3.15$ GeV. The dimuon transverse momentum is required to be larger than 7 GeV (because at the trigger level there is a selection requirement on the dimuon p_T). In addition, the distance of closest approach (DCA) between all three track pairs needs to be less than 1 mm. All three tracks are used in the vertexing to define the B candidate vertex and all subsequently derived quantities.

It is also crucial to reconstruct the $B_s^0 \rightarrow J/\psi \phi$ decay to study various systematic uncertainty related to the analysis efficiency for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. The decay of

$B_s^0 \rightarrow J/\psi \phi (\rightarrow \mu^+ \mu^- K^+ K^-)$ is a four-body decay. The reconstruction starts from two unlike-sign muons, which are later combined and vertexed with two unlike-sign high purity tracks fulfilling $p_T > 0.8$ GeV and $|\eta| < 2.4$. The DCA between all pairs among the four tracks is required to be less than 1 mm, and the two hadron tracks are assigned the kaon mass hypothesis. The two kaons must have an invariant mass of $1.01 < m_{K^+ K^-} < 1.03$ GeV, and have $\Delta R < 0.3$ in the $\eta\phi$ plane. The dimuon mass restriction is similar to the $B^+ \rightarrow J/\psi K^+$ channel. All four tracks are used in the vertexing to define the B candidate vertex and all derived quantities.

3.2 Choice of normalization channel

Given the current precision in the knowledge of the $b\bar{b}$ cross-section at LHC, direct extraction of the branching fraction of the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays would be affected by significant uncertainty. As commonly done in many B physics analyses, the signal branching fraction is calculated by normalizing to another decay channel of a B meson, whose branching fraction is well known and whose characteristics allow a clean and precise reconstruction. The $B^+ \rightarrow J/\psi K^+$ decay mode is the best candidate for our case. Now, the signal branching fractions $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ can be then extracted as

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \mathcal{B}(B^+ \rightarrow J/\psi K^+) \frac{N_{B_s^0 \rightarrow \mu^+ \mu^-}}{N_{B^+ \rightarrow J/\psi K^+}} \frac{\varepsilon_{B^+ \rightarrow J/\psi K^+}}{\varepsilon_{B_s^0 \rightarrow \mu^+ \mu^-}} \frac{f_u}{f_s} \quad (3.1)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = \mathcal{B}(B^+ \rightarrow J/\psi K^+) \frac{N_{B^0 \rightarrow \mu^+ \mu^-}}{N_{B^+ \rightarrow J/\psi K^+}} \frac{\varepsilon_{B^+ \rightarrow J/\psi K^+}}{\varepsilon_{B^0 \rightarrow \mu^+ \mu^-}} \frac{f_u}{f_d} \quad (3.2)$$

where N_x is the number of fitted candidates for the decay x and ε_x is the corresponding full selection efficiency. f_q is the hadronization fractions of a b quark with a q quark to form the respective B meson. The ratio $\frac{f_d}{f_u}$ is expected to be 1 in the SM due to isospin symmetry. Eq. [3.1](#) shows clearly that a precise knowledge of $\frac{f_u}{f_s}$ is necessary for a precise measurement

of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and the current measurement has error at 5% level by LHCb experiment. In addition, the LHCb experiment observed a p_T dependence of the hadronization fraction and is not confirmed by other experiments to date. This can be taken as the disadvantage of using this decay channel as normalization.

Equations 3.1 and 3.2 have the additional benefit of allowing a partial cancellation of most systematic uncertainties due to selection and reconstruction, which affect the numerator and denominator in the same proportion. In this analysis, we do not use the decay $B_s^0 \rightarrow J/\psi \phi$ as a relative normalization since the measurement of the branching fraction of $B_s^0 \rightarrow J/\psi \phi$ in the LHCb experiment (the most precise one to date) is based on the relative normalization to the $B^+ \rightarrow J/\psi K^+$ process. Using the $B_s^0 \rightarrow J/\psi \phi$ will additionally bring a scale factor on the $B_s^0 \rightarrow J/\psi \phi$ branching fraction to account for the difference between CMS and LHCb phase space. Moreover, the measurement of $B_s^0 \rightarrow J/\psi \phi$ has one additional kaon that gives additional systematic uncertainty. The current most precise measurement of the $\mathcal{B}(B_s^0 \rightarrow J/\psi \phi)$ has a 4.75% systematic uncertainty [72], which is much larger than the 2% systematic uncertainty of $\mathcal{B}(B^+ \rightarrow J/\psi K^+)$. Therefore, $B^+ \rightarrow J/\psi K^+$ is used for the normalization, and $B_s^0 \rightarrow J/\psi \phi$ will be mainly used for cross-checking.

In principle, one can also use $B_s^0 \rightarrow K^+ K^-$ decay as normalization, but in CMS, there is no displaced hadronic trigger as well as no hadron identification algorithm. So this option was discarded in CMS.

3.3 Primary dataset and triggers

Data collected by the CMS undergoes scrutiny that certifies the quality of each run and each single luminosity section, only the high-quality data that passes this certification is used for various analyses. Two valid certifications are granted; one is called “muon certification”, which check the performances of the inner tracker and muon chamber, and the

second one, which includes all the sub-detector performance known as “golden certification.” This analysis uses the first one because events are reconstructed using tracks and muons (since no informations from the calorimeters is used). The total integrated luminosity, corresponding to the muon certification, for the years 2011 and 2012 were equal to 5.0 fb^{-1} (7 TeV) and 20.0 fb^{-1} (8 TeV), respectively. During the Run 2 data taking, the total integrated luminosity collected by CMS was 36.0 fb^{-1} in 2016, 42 fb^{-1} in 2017, and 60 fb^{-1} in 2018. In this thesis, I collectively refer to 2016, 2017, and 2018 data as Run 2. Similarly, 2011 and 2012 data is referred to as Run 1. Moreover, despite the natural differences between the Run 1 and Run 2 beam energy conditions, which introduce separate treatments of the datasets, the offline selections of events are designed as close as possible to the Run 1 publication.

In CMS, the signal sample with $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is triggered by two muons with an invariant mass $4.8 < m_{\mu^+ \mu^-} < 6.0 \text{ GeV}$, while the normalization sample is triggered by two muons from a J/ψ decay, forming a secondary vertex with a flight length (distance between PV and SV) significance of 3σ from the beam spot in the transverse plane. Additional requirements are imposed on the DCA between the two muons (0.5 cm), the cosine of the pointing angle of the dimuon momentum with respect to the direction from the beam spot to the secondary vertex ($\cos(\alpha) > 0.9$), and the probability of the vertex fit (changed substantially over time). The HLT trigger selection for both signal and normalization channels are based on L1 informations and using the sophisticated algorithm to decide whether to keep the event or not.

In 2011 and 2012, two signal triggers were used to collect the events from two regions: barrel and endcap, because the resolution is better and the level of background is lower in barrel region compared to endcap region. The details of the kinematic selections at HLT level are described in Table [3.1](#).

In 2016, a major change to the trigger setup was introduced, which only stored the

HLT path	min p_{T,μ_1} [GeV]	min p_{T,μ_2} [GeV]	max $ \eta_\mu $	min $p_{T,\mu\mu}$ [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

Table 3.1: Kinematic selections at HLT level for the signal and different used trigger.

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ dimuon events within $|\eta| < 1.4$. This was done to keep the rate under control to accommodate the increasing LHC luminosity. However, the L1 seeds for the signal and displaced- J/ψ samples were always identical. The change on the trigger introduced in 2016 remained in place for the entire Run 2. Table 3.2 provides all the relevant details about the triggers. The normalization triggers have a variable prescale, whereas the signal trigger is unprescaled (prescale value 1). Prescaling is generally needed if the production cross-section of the process is high. For our case, $B^+ \rightarrow J/\psi K^+$ is abundantly produced and overall $\mathcal{B}(B^+ \rightarrow J/\psi K^+)$ is 10^6 higher than $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. Storing all the events will result in a lack of storage space. A better solution is to design a prescale to control the data collection rate. For instance, if the trigger has a prescale factor of 10, only 1 out of 10 collisions that satisfies the trigger selections will be recorded. At LHC, the instantaneous luminosity of a particular fill (the number assigned every time when the beam injection to the LHC ring happens) decreases with an increase in time. So to maintain the HLT rate under control and maximize the signal efficiency, a variable prescaling scheme is adopted [73]. The triggered events are stored in the dataset named Charmonium for Run 2 and Muonia for Run 1. There are a few other datasets like DoubleEG, EGamma, SinglePhoton, SingleElectron and SingleMuon, which collect events using other triggers such as an electron, photon, jet, etc., are particularly helpful for different validation purposes.

During the initial data-taking period of 2016, the silicon micro-strip detector experienced operational instabilities. This leads to a significant impact on the trigger efficiency

	HLT_DoubleMu4_3_Bs	HLT_DoubleMu4_3_Jpsi	HLT_DoubleMu4_3_Jpsi _Displaced
Periods	2016-2018	2018	2016-2017
Mass	[4.5, 6.0]	[2.9, 3.3]	[2.9, 3.3]
$p_{T_{\mu^+ \mu^-}}^{min}$	4.9	6.9	6.9
$p_{T_{\mu}}$	4.0	4.0	4.0
Charge	opposite	opposite	opposite
DOCA($\mu^+ \mu^-$)	> 0.5	> 0.5	> 0.5
VtxProb	> 0.005	> 0.1	> 0.1
$\cos \alpha$		> 0.9	> 0.9
LxySignif			> 3.0
Prescale			
2016	1		1-8
2017	1		1-14
2018	1	10	

Table 3.2: Primary analysis triggers and their parameters. Charge, DOCA, VtxProb, $\cos \alpha$, and LxySignif stand for the charge of two muons, distance of closest approach between the two muons, dimuon vertex probability, cosine of the pointing angle, and the flight length significance of the dimuon vertex in xy-plane, respectively. HLT_DoubleMu4_3_Bs is used for the signal channel. HLT_DoubleMu4_3_Jpsi_Displaced and HLT_DoubleMu4_3_Jpsi are used for the normalization as well as control channels.

with an increase in the pileup. Therefore, the 2016 data period is divided into two separate running periods, marked as 2016BF, and 2016GH, of roughly equal integrated luminosity.

The signal-to-background ratio (S/B) is very different in the central part of the detector compared to the forward region. That's why the analysis is performed in different “channels”. The channels are defined using the η of the most-forward muon and have different boundaries because of the trigger acceptance (for 2011 and 2012, η ranges up to 2.1, and for Run 2, it is up to 1.4).

3.4 MC simulation

The analysis uses Monte Carlo (MC) simulation samples to determine the signal and normalization efficiency and is used to optimize the analysis selection requirements. The shape

of the backgrounds, where one or two hadrons are misidentified as muons, is also studied using MC samples. The produced MC samples include decay modes like $B \rightarrow h\mu\nu$, $B \rightarrow h\mu\mu$, and $B \rightarrow hh'$, where $h(^{'})$ stands for hadrons like K, π & p . Simulation of MC event performed in the following step:

- first, the event GENERation data (GEN) is done using particle guns or softwares like Pythia [74], and EvtGen [75],
- the generated events are then passed through detector material response based on the Geant4 [76] software, and the response of the sub-detectors (SIMulated hits data (SIM)) are processed,
- the simulation is followed by the electronics digitization (DIGItization data (DIGI)),
- the events are then processed by the CMS event reconstruction algorithms as the data,
- finally, the reconstruction output is reduced to a minimal set of variables, and sufficient to carry out the majority of the physic analyses,

It is impossible to generate sufficient combinatorial background MC events because of limited CPU time and disk storage. To overcome this difficulty, the dimuon mass sidebands are used.

3.5 Muon selection

Muon identification plays an important role in the search for the $B^0 \rightarrow \mu^+ \mu^-$ decay because backgrounds originating from the charmless two body decays, where both hadrons are reconstructed as muons, can create a peaking background that could falsely enhance the B^0 yield. Therefore, it is crucial to maintain a high muon efficiency at the cost of the lowest possible “fake muon misidentification”. There are a few chances of getting a non-muon particle which is identified as a muon track. For example, protons have a non-null

probability of traversing the detector material and reaching the muon sub-detectors. Another possibility is that a kaon/pion decays into a muon before reaching the muon stations. They all contribute to the muon misidentification. The signal candidate reconstruction starts with a selection of two muon candidates satisfying stringent selection requirements. Each muon candidate is required to be reconstructed by both TrackerMuon and GlobalMuon algorithms. It has to pass the loose muon identification (i.e., it has to satisfy particle flow event reconstruction) and the inner track of the muon candidate is required to be of highest purity quality. This set of requirements rejects the majority of poorly reconstructed muon candidates. Furthermore, a requirement is applied to the two muons in the final state to match with the HLT muons. In this way, we have a complete correspondence between the trigger and the offline analysis objects.

3.5.1 MVA muon identification

There are several muon identification algorithm available in CMS for different analysis use. For example, tight muon identification chooses a clean and high-quality sample of well-reconstructed muons. The algorithm employs the particle flow muons, and on top of that few more selections related to the primary vertex are applied [77]. But this algorithm has a high misidentification rate. The misidentification rate is defined as the probability of distinguishing a muon track given a non-muon particle. Therefore, to improve the $B^0 \rightarrow \mu^+ \mu^-$ analysis sensitivity, a Boosted Decision Tree (BDT) method, which is implemented in the TMVA package [78], is developed to separate true muons from misidentified muons. A decision tree is a binary tree structure where 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. Many single muon variables can be used as handles in the BDT. The input variables are divided into the following classes:

- Variables calculated in the muon track reconstruction,
 1. Segment Compatibility: a measure of the standalone muon track's compatibility with the muon hypothesis.
 2. Muon Time Uncertainty: a variable describing the time-of-flight error in the RPC muon subsystem.
- Variables related to general track properties. It is observed that the variables determined with the silicon tracker are sensitive to the quality of the muon track measurement. This reveals that the tracks from charged hadrons and those from particles with a decay-in-flight often have lower quality.
 1. Inner Track χ^2/dof : The χ^2 of the inner track divided by the number of degrees of freedom of the fit.
 2. Outer Track χ^2/dof : The χ^2 of the muon outer track divided by the number of degrees of freedom of the fit.
 3. Inner Valid Fraction: The number of valid hits of the inner track divided by the total number of expected hits of the inner track.
 4. Layers With Hits: The number of tracker layers with hits.
 5. Charge Product: The product of the tracker track's charge and the stand alone muon track's charge.
 6. Tracker Kink: This is also known as change in track curvature. The kink algorithm applied to the global muon's inner track. The kink algorithm takes the difference in ϕ of the predicted track position and the actual recHit, squares it

and divides it by the error in ϕ of the recHit position if it is greater than zero.

The final value corresponds to the sum of these calculated values for all recHits.

- Variables obtained during the global muon reconstruction,
 1. Similarity measure of Local Momentum: It is the χ^2/dof of the momentum matching between the extrapolated silicon tracker and standalone muon at the innermost muon layer.
 2. Similarity measure of Local Position: It is the χ^2/dof of the position matching between the extrapolated silicon tracker and standalone muon at the innermost muon layer.
 3. Tracker Track Relative χ^2 : The sum of χ^2/dof estimates of the hits in the silicon tracker with respect to the global muon track.
 4. Global Kink: The same algorithm calculating the Tracker Kink applied on the global track.
 5. Global Track Tail Probability: the probability of the global muon track χ^2/dof .

The BDT was trained using the above variables, excluding the kinematic variables, to keep the BDT independent of the kinematic properties of the particles. The same set of variables is used in the muon BDT training for Run 1 and Run 2. The training uses muons from the signal samples and misidentified charged hadrons from the background samples. Samples constituting the two body hadronic b decays are called as background samples. The most important discriminating variables are the Similarity measure of Local Position and Segment Compatibility. The performance of a BDT was measured using a Receiver Operating Characteristics (ROC) curve, in this context, is defined as the muon efficiency as a function of the misidentification probability. The BDT response can range from $[-1, 1]$ per event, where the -1 response corresponds to the non-muon event, and +1 corresponds

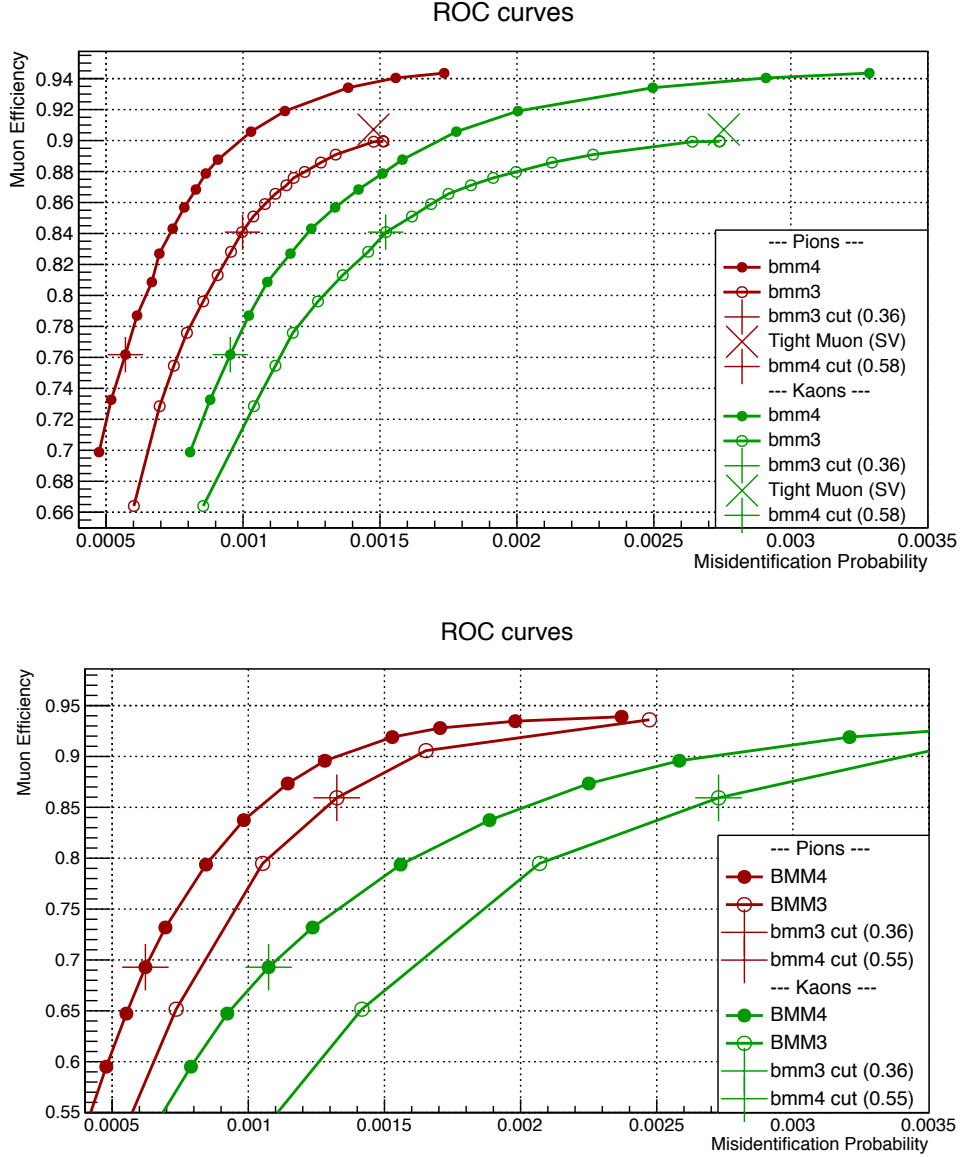


Figure 3.2: The performance of the BDTs for the current analysis and the previous analysis for 2016 (top) and 2012 (bottom) conditions. The crosses mark the performance of a particular BDT cut mentioned in the legend, whereas the tilted crosses mark the performance of the official Tight Muon selection. Here the efficiencies are calculated by taking the ratio of number of signal or background event passing the BDT selection to total number signal or background events in MC sample.

to the muon event. The BDT working point for muon is chosen after scanning muon efficiency and misidentification probability for different BDT points. Figure 3.2 shows the muon efficiency vs misidentification probability for different BDT point. Compared to the previous BDT used in Ref. [79], referred to as bmm3, the muon efficiency for the new muon BDT is about 5% (absolute) higher for the same misidentification rate. The average misidentification rate of the new muon BDT is achieved to be 6×10^{-4} for pions and 10^{-3} for kaons, for both Run 1 and Run 2. However, the tight muon selection gives a higher muon efficiency but has a larger average misidentification rate.

Chapter 4

Run 1 + Run 2 (2016) data analysis

In this Chapter, I describe the analysis performed for the measurement of the properties of $B_s^0 \rightarrow \mu^+ \mu^-$ decay and the search for the $B^0 \rightarrow \mu^+ \mu^-$ decay using the data collected by the CMS experiment in 2011, 2012, and 2016 at center-of-mass energies of 7 TeV, 8 TeV, and 13 TeV, respectively.

The reconstructed event variables and general selection criteria to select the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates are discussed in Section 4.1. The dominant background for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ is the combinatorial background. To suppress this background, a multivariate classifier is developed, and a more detailed study on the optimization is discussed in Section 4.2. Section 4.3 describes the fit model and the corresponding yield obtained in the normalization and control channels. Section 4.4 discusses some corrections developed to fix the yield instability. The classifier developed for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ is validated in the normalization channel, and a comparison between the data and MC distributions is performed to make sure everything works as expected. The comparison method is described in Section 4.5. The muon BDT (discussed in the previous chapter) developed for the selection of a good muon is validated in the data, and the comparison is discussed in Section 4.6. Section 4.7 describes the method to estimate the signal and background yields, which will be used for the validation of the fit method. The fitting strategies for the two branching fraction extractions and validations are described in Section 4.8. For the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime measurement, two independent methods have been developed. One is the extended two-dimensional UML fit, and the second one is the sPlot fit. Both methods are discussed in Sections 4.9 and 4.10. Section 4.11 discusses the possible systematic uncertainties for branching fraction and effective lifetime. Finally, the fit results are summarized in Section 4.12.

4.1 Offline event selection

Several variables are used to differentiate between the signal and background events. Those are related to muon reconstruction, primary vertex selection, and the reconstruction of the secondary vertex. After a loose preselection, the most discriminating variables are fed as input to the BDT classifier, which separates the signal events from the background events. We will be discussing all such variables in this section.

4.1.1 Primary vertex selection

A typical pp collision involves multiple head-on interactions, resulting in a huge number of tracks originating from a variety of vertices. The vertex reconstruction starts with the tracks which are suitable with the beam spot and a low χ^2 . On average, there were about 35 primary collision points per beam crossing in 2016. By default, the primary collision points are ranked in a list according to the sum of the transverse momenta of the tracks coming out of each collision point. Out of many pp interaction points, the b hadron candidates must originate from one of the vertices (known as the PV or primary vertex). There are a few choices available to find the PV

- PV with the highest p_T : The primary vertex for which the sum of the p_T of the tracks in the vertex is the highest.
- PV with the best pointing angle in three-dimension (3D): The primary vertex for which angle between the b hadron momentum and the vector joining the primary vertex to the secondary vertex is smallest.
- PV closest in z : The primary vertex to which the distance of closest approach (in the z -direction) to each PV of the extrapolated trajectory of the B candidate is smallest.

In this analysis, the PV is chosen based on the PV closest in z . In the following, this PV is

referred to as the $b\bar{b}$ PV. In comparison to the other two options, this choice has a better position resolution. The position resolution of the primary vertex is of the order of $\mathcal{O}(10)\mu\text{m}$ in all directions. It is possible that a common track can be used for both B candidate reconstruction and $b\bar{b}$ PV reconstruction, which may create bias in the $b\bar{b}$ PV position. To avoid any geometrical bias due to displaced tracks, the tracks are removed, and the PV is refitted with the remaining tracks with a beam spot constraint. In this fit, based on an adaptive fitting method [80], a weight from 0 to 1 is assigned to each track. The B candidate is rejected if the average track weight of the $b\bar{b}$ PV (excluding the B candidate tracks) is smaller than 0.6.

4.1.2 Variables

After selecting the two good muon tracks of opposite charge, their invariant mass is reconstructed, and a preselection cut on the invariant mass is applied ($4.9 < m_{\mu^+\mu^-} < 6.0$ GeV). In this analysis, the full range of the invariant mass spectrum is divided into four regions; a lower sideband region ($4.9 < m_{\mu^+\mu^-} < 5.2$ GeV), a $B^0 \rightarrow \mu^+\mu^-$ signal window ($5.2 < m_{\mu^+\mu^-} < 5.3$ GeV), a $B_s^0 \rightarrow \mu^+\mu^-$ signal window ($5.3 < m_{\mu^+\mu^-} < 5.45$ GeV), and a higher sideband region ($5.45 < m_{\mu^+\mu^-} < 5.9$ GeV). The $B^0 \rightarrow \mu^+\mu^-$ and $B_s^0 \rightarrow \mu^+\mu^-$ signal windows are blinded in data throughout the validation.

Many variables with the highest discriminating power in the offline analysis are calculated in the 3D space. The 3D impact parameter (δ_{3D}), its uncertainty ($\sigma(\delta_{3D})$), and its significance ($\delta_{3D}/\sigma(\delta_{3D})$) are determined with respect to the $b\bar{b}$ PV. The flight length (l_{3D}), distance between $b\bar{b}$ PV to SV, and its significance ($l_{3D}/\sigma(l_{3D})$) are the important variables to distinguish between background and signal. Another important variable is the pointing angle, α_{3D} . It is the 3D angle between the B momentum direction and the $b\bar{b}$ PV to SV vector direction. The χ^2/dof of the secondary vertex fit is also a powerful discriminant. A sketch of the variables associated with SV is shown in Fig. 4.1. The maximum distance

of the closest approach between the two muon tracks d_{ca}^{max} is also used in the BDT. The decay time is given by the product of l_{3D} and the invariant mass of the B candidate, divided by the magnitude of the B candidate momentum. Similarly, the decay time uncertainty is calculated by propagating uncertainties of momentum and flight length. Another important

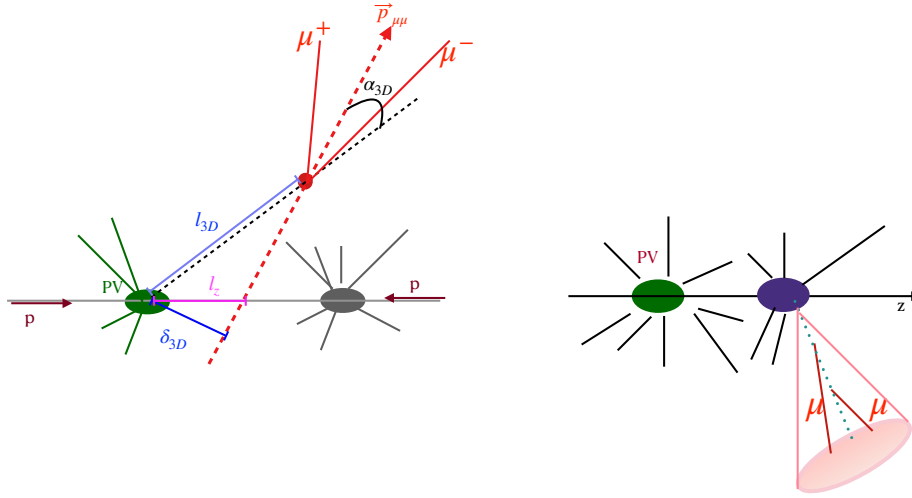


Figure 4.1: Visual representation of a B decay; the decay length l_{3D} , pointing angle α_{3D} , impact parameter δ_{3D} , the dimuon momentum $\vec{p}_{\mu\mu}$ and the longitudinal impact parameter l_z are shown (left figure). An isolated cone is shown in the right plot.

variable associated with the secondary vertex is isolation. The purpose of this variable is to separate the signal from the background events containing tracks that originated from the jets. It is determined from the B candidate transverse momentum and other charged tracks in a cone with radius $\Delta R = 0.7$ ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$, where φ is the azimuthal angle in radians) (Fig. 4.1) around the B momentum as follows,

$$I = \frac{p_T(B)}{p_T(B) + \sum_{trk} p_T} \quad (4.1)$$

where the sum includes tracks with $p_T > 0.9$ GeV that are

- not part of the B candidate but associated with the same PV as the B candidate or

- not associated to any PV but have a distance of the closest approach $d_{ca} < 500 \mu\text{m}$ to the B secondary vertex.

We use a few variables to reject candidates arising from partially reconstructed B decays:

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

In addition to the isolation variable, we also computed the single muon isolation I_μ , with other charged tracks in a cone with radius $\Delta R = 0.5$ around the muon:

$$I_\mu = \frac{p_T(\mu)}{p_T(\mu) + \sum_{trk} p_T} \quad (4.2)$$

where the sum includes tracks with $p_T > 0.5 \text{ GeV}$ that has a distance of closest approach $d_{ca} < 1 \text{ mm}$ 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 the level of the pile-up.

Since $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidates are analyzed with the same analysis BDT as the $B \rightarrow \mu^+ \mu^-$ candidates, the two muons from the J/ψ are refitted to a common vertex, and this fit χ^2/dof is used in the analysis BDT so as to have the same number of degrees of freedom as in the signal decay. The determination of the other variables is based on the complete B candidate secondary vertex, also including the additional kaon (s) in the fit.

4.2 Multivariate analysis for signal selection

The variables described in Section [4.1.2](#) are fed into a BDT discriminator. The preselection cuts are applied to the input variables to remove the outlier events, which could distort the

BDT architecture. The preselection also has a considerable influence on the BDT performance and characteristics. Table 4.1 summarizes the applied preselection requirements.

Table 4.1: Preselection cuts on variables for BDT training/testing.

Variable	Minimum	Maximum	unit
$p_{T,B}$	5.0	9999.0	GeV
$p_{T\mu}$	4.0	9999.0	GeV
l_{3D}	-	2.0	cm
$l_{3D}/\sigma(l_{3D})$	4.0	200.0	
$l_{xy}/\sigma(l_{xy})$	4.0	-	
l_z	-	1.0	cm
$l_z/\sigma(l_z)$	-	5.0	
χ^2/dof	-	5	-
δ_{3D}	-	0.1	cm
$\delta_{3D}/\sigma(\delta_{3D})$	-	5.0	
d_{ca}	-	0.08	
α_{3D}	-	0.20	
d_{ca}^0	-	2.50	
N_{trk}^{close}	-	21.0	
I	0	-	

Given the small cross-section of $b\bar{b}$ quark production, it is impossible to generate sufficient dimuon background MC samples. Thus, the data from the dimuon higher mass sideband is used as background for BDT training. We need to ensure that there is no possible bias in the training process. Therefore, the data and MC samples are split into three distinct subsets, depending upon the event type, to avoid possible systematic effects. We define the event type by the remainder of the event number divided by three, i.e.,

$$\text{type} = \text{iEvent.id().event()} \% 3. \quad (4.3)$$

To be very specific, three BDTs are obtained in following way,

- events of type 0 : analyzed by BDT0, trained on type-1 events, tested on type-2 events
- events of type 1 : analyzed by BDT1, trained on type-2 events, tested on type-0 events

- events of type 2 : analyzed by BDT2, trained on type-0 events, tested on type-1 events

4.2.1 Optimization, training, and characterization

The BDT optimization was achieved with about 10000 batch jobs where BDTs were trained. To finalize the best set of variables, a core set of variables was defined ($l_{3D}/\sigma(l_{3D})$, α_{3D} , $\delta_{3D}/\sigma(\delta_{3D})$, $d_{ca}^0/\chi^2/dof$, N_{trk}^{close} , I , $I_{\mu 1}$, $I_{\mu 2}$). Every BDT trained included these variables. In addition, a subset of variable list l_{3D} , $l_{xy}/\sigma(l_{xy})$, δ_{3D} , d_{ca}^{max} , p_T , η was added in the batch jobs. Each BDT response was tested with a decent Kolmogorov-Smirnov probability ($P > 0.05$) between the training and testing response distribution. Furthermore, the final analysis BDTs are chosen based on the maximum of figure of merit formula $S/\sqrt{(S+B)}$, where S is the expected $B_s^0 \rightarrow \mu^+ \mu^-$ yield in the region $5.30 < m_{\mu^+ \mu^-} < 5.45$ GeV from simulation and B is the expected combinatorial background in the same region, which is obtained by performing a fit to the higher sideband distribution and extrapolation. In the optimization process, the BDT settings are analyzed to obtain the final one by checking the optimum point of all three event types. The response required for each event type should be close. The BDT response from different event types is shown in Fig. 4.2. This approximate figure of merit was used only in the optimization procedure and not in the procedure used to obtain the final result. The final BDT working points for the result are obtained from the sensitivity of the toy study which will be discussed later.

4.3 Normalization ($B^+ \rightarrow J/\psi K^+$) and control ($B_s^0 \rightarrow J/\psi \phi$) channel fitting

The normalization channel $B^+ \rightarrow J/\psi K^+$ and control channel $B_s^0 \rightarrow J/\psi \phi$ were triggered with HLT paths subject to varying prescales for the data-taking year 2016. This must be accounted for in the analysis because the dimuon trigger (for signal) was never prescaled. The prescale is used as the candidate's weight when filling the mass histogram. Then a

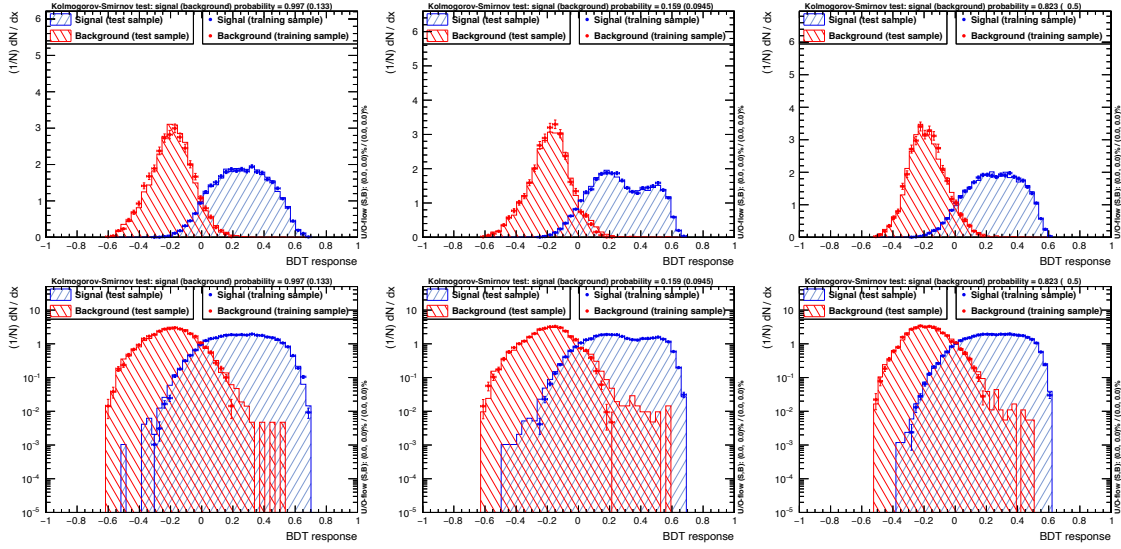


Figure 4.2: Overtraining checks for the three BDTs for 2016BF. The top row shows the BDT response in a linear scale, the bottom row shows the same plots in a logarithmic scale.

weighted binned maximum-likelihood fit is performed to determine the overall yield. The model used to fit the mass distribution is the sum of a double Gaussian for the signal component, an error function for partially reconstructed B decays, an exponential function for the combinatorial background, and finally, another triple Gaussian for the Cabibbo-suppressed component $B^+ \rightarrow J/\psi \pi^+$. The contribution of $B^+ \rightarrow J/\psi \pi^+$ is fixed at 3.8% of the signal component, based on the relative ratio of their branching fractions. Similarly, we fit the $B_s^0 \rightarrow J/\psi \phi$ mass distribution with the sum of a double Gaussian with a common mean for the signal component, an exponential function for the combinatorial background, and a sum of double Gaussian and a linear component for the satellite peak $B^0 \rightarrow J/\psi K^*$. Figures 4.3 and 4.4 show the projections of the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ invariant mass fits in different years.

The systematic uncertainty on the normalization yield is calculated by the difference between the results of the fit to the unconstrained B^+ and B_s^0 invariant mass distribution and the fit to the J/ψ mass-constrained invariant mass distribution.

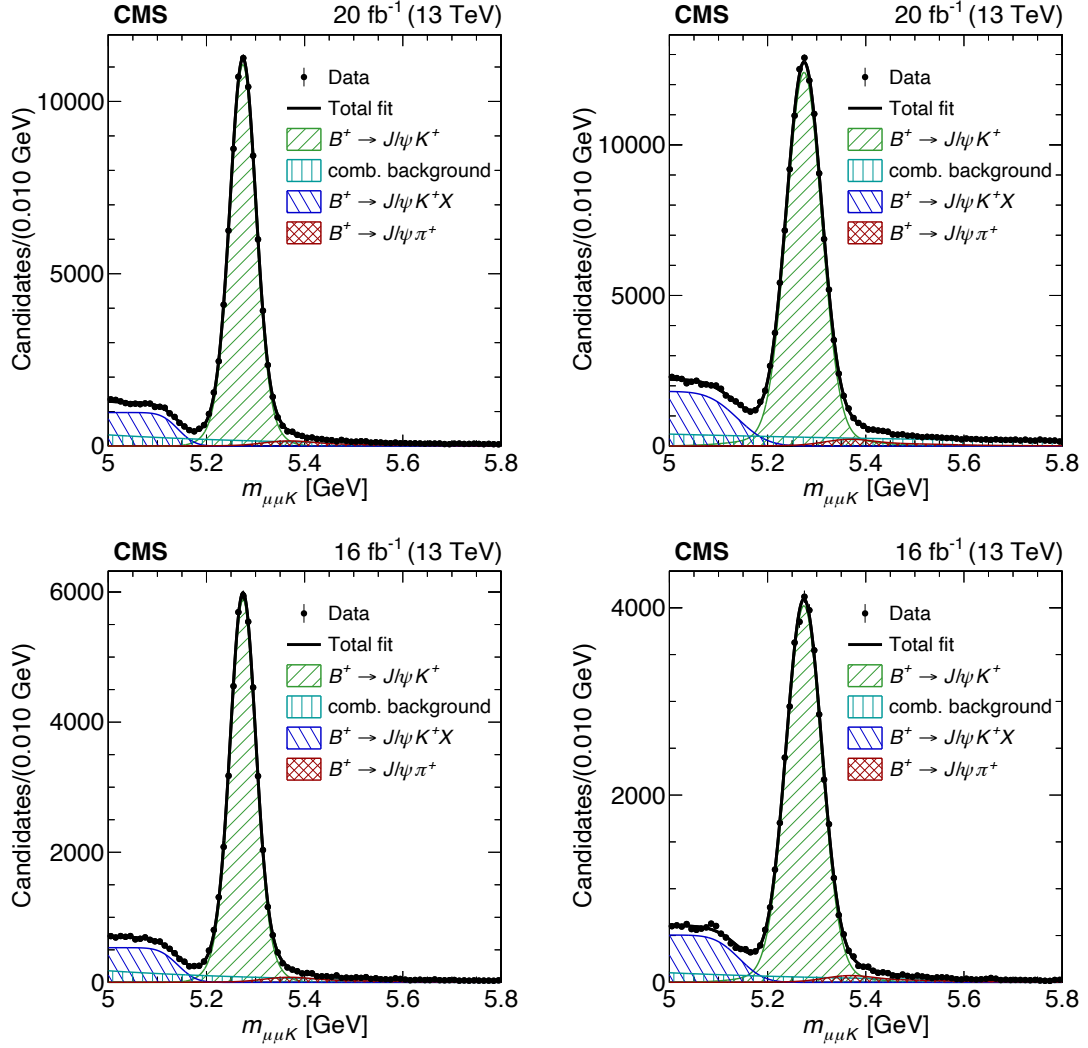


Figure 4.3: Invariant mass distributions of the selected $B^+ \rightarrow J/\psi K^+$ candidates for the year 2016BF (top), and 2016GH (bottom). The left and right column plots are for channel 0 and channel 1, respectively. The continuous curve (black) is for the combined fit and different hatched histograms are for signal and background components. The solid black points are data.

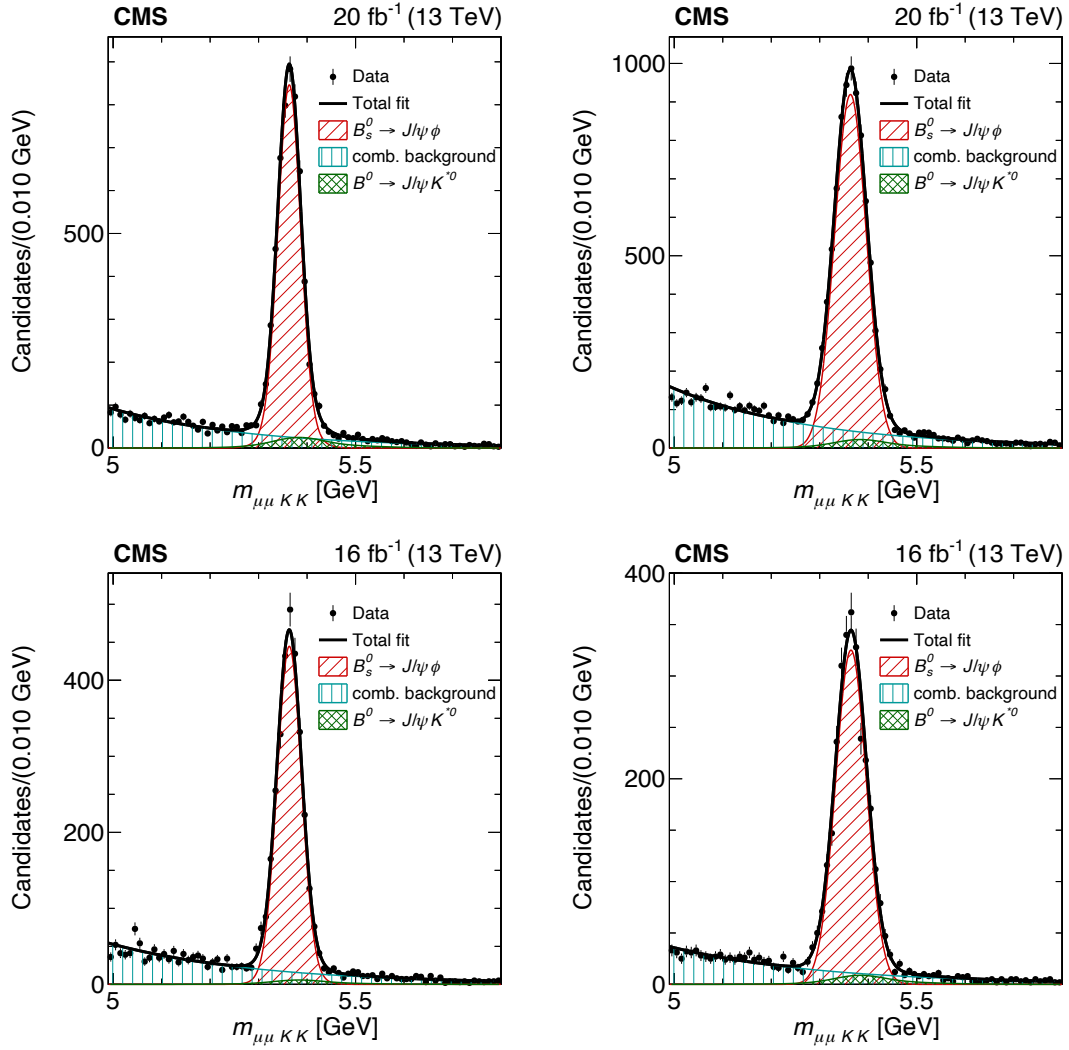


Figure 4.4: Invariant-mass distributions of the selected $B_s^0 \rightarrow J/\psi \phi$ candidates for the year 2016BF (top), and 2016GH (bottom). The left and right column plots are for channel 0 and channel 1, respectively. The continuous curve (black) is for the combined fit and different hatched histograms are for signal and background components. The solid black points are data.

4.4 MC corrections

During the initial period of 2016 data taking, the CMS detector was partially run with wrong detector configuration. In this section, I discussed the corrections used to counter that. The effect is first observed while looking at the $N(\mu\mu)$ yields from $B \rightarrow \mu\mu$ sideband, the number of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidates in full 2016 data taking year. The 2016 data year is separated with era like B, C, D, E, F, G, & H. The microstrip detector operational instabilities mostly affect the era B-F, later for era G-H it is fixed. The main difference of normalization trigger from the signal is the selection requirement on the vertex be well separated from the beamspot. This criteria affect the normalization trigger more strongly compared to signal trigger, which led to yield instability. This instability need to be corrected to get a sensible result. First a correction is derived using the the number of primary vertex because the major difference between the era is PU. This will allow to reduce the generic discrepancies between the eras or within the era. After PU-dependent correction, another set of weights is estimated to correct the decay time dependency. The variable used for this correction is the flight length significance, because it is directly used at the HLT. At the end, the corrections are applied event by event basis in the signal MC samples.

4.5 Data-MC simulation comparison

It is essential to verify that the MC sample used in this analysis describes the data accurately. We studied this by comparing (sideband-subtracted) data distributions with the corresponding distribution from MC simulation in $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ channels. The plots in Fig. 4.5 illustrate the concept of sideband subtraction. For a given selection, we fill (1) the mass distribution and (2) the variable distributions for the low sideband, the high sideband, and the signal region, respectively. Because the shapes of these distributions depend

on the relative sample composition in the different mass regions. The background composition differs in the low sideband from the high sideband. The high sideband is composed dominantly by combinatorial background (it can fit well with an exponential function, apart from the known and fixed small contribution from $B^+ \rightarrow J/\psi \pi^+$), while the low sideband shows a pronounced step at $m = 5.14$ GeV due to partially reconstructed B decays (dominated by $B^0 \rightarrow J/\psi K^*(K^+\pi^-)$ where one of the tracks from the K^* is not included in the reconstruction. The model used to fit the $B^+ \rightarrow J/\psi K^+$ distribution is already discussed in Section 4.3.

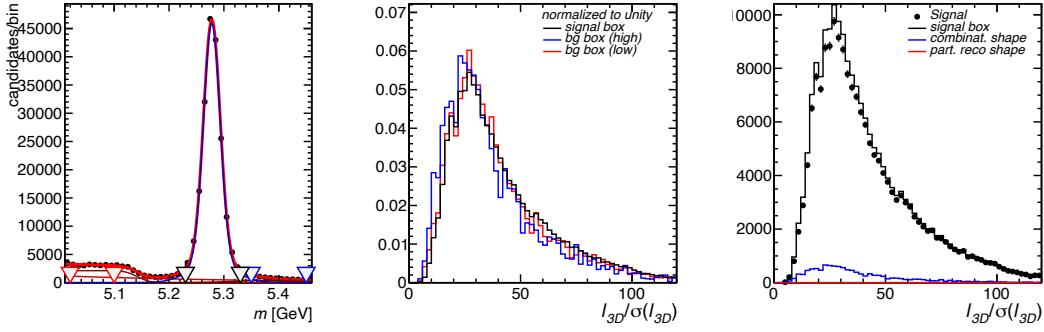


Figure 4.5: $B^+ \rightarrow J/\psi K^+$: Illustrations of the sideband subtraction methodology in 2016BF channel 0. The left column shows the fitted mass distribution indicating the low (red arrows) and high (blue arrows) sidebands together with the signal region (black arrows). The middle column shows the shape of the distributions for the three regions, each normalized to unity. The right column shows the signal distribution obtained after subtracting from the signal-region distribution the contributions from the combinatorial and partially-reconstructed background.

The variable distribution in the signal region is therefore determined after two corrections:

- the variable distribution corresponding to the high sideband is scaled by the ratio of the integrals of the exponential function in the signal region and the high sideband. This scaled histogram is then subtracted from the histogram corresponding to the signal region. We also determine the ratio of the integrals of the exponential function in

the low sideband and the high sideband and then subtracted the scaled high-sideband variable distribution from the low-sideband distribution to obtain the distribution for the partially reconstructed background.

- the variable distribution of the partially reconstructed background is then scaled by the ratio of the integral of the step function in the signal region to the low sideband and subtracted from the signal region distribution (obtained after subtraction of the combinatorial background).

In practice, the partially reconstructed background is always zero due to the choice of the signal region (indicated by the black arrows in Fig. 4.5).

After the subtraction, the data and MC are compared for all the variables. The most discriminating variables, along with decay time and the BDT discriminator are shown in Figs. 4.6 and 4.7. The comparison shows a good agreement between the variables in both normalization and control channels. In the BDT distribution, a visible difference of around $\pm 20\%$, in the peripheral part is observed. To account for this difference, a systematic uncertainty is assigned to the analysis efficiency ratio between the signal and normalization channel, which will be discussed in the systematic section.

4.6 Validation of muon BDT in the data and MC

As it is discussed in the previous chapter, the rate of rare peaking background made by two misidentified hadrons could be reduced to a significant level by using the muon identification algorithm. For the optimized muon BDT working point, the average misidentification rate is 6×10^{-4} and 10^{-3} for pions and kaons, respectively, for both Run 1 and 2016, together with a muon efficiency of about 70 (76)% for Run 1 (2016).

The performance of the muon BDT is validated by comparing its behaviour in simulation with that in data using event samples in which a kinematically selected two-body

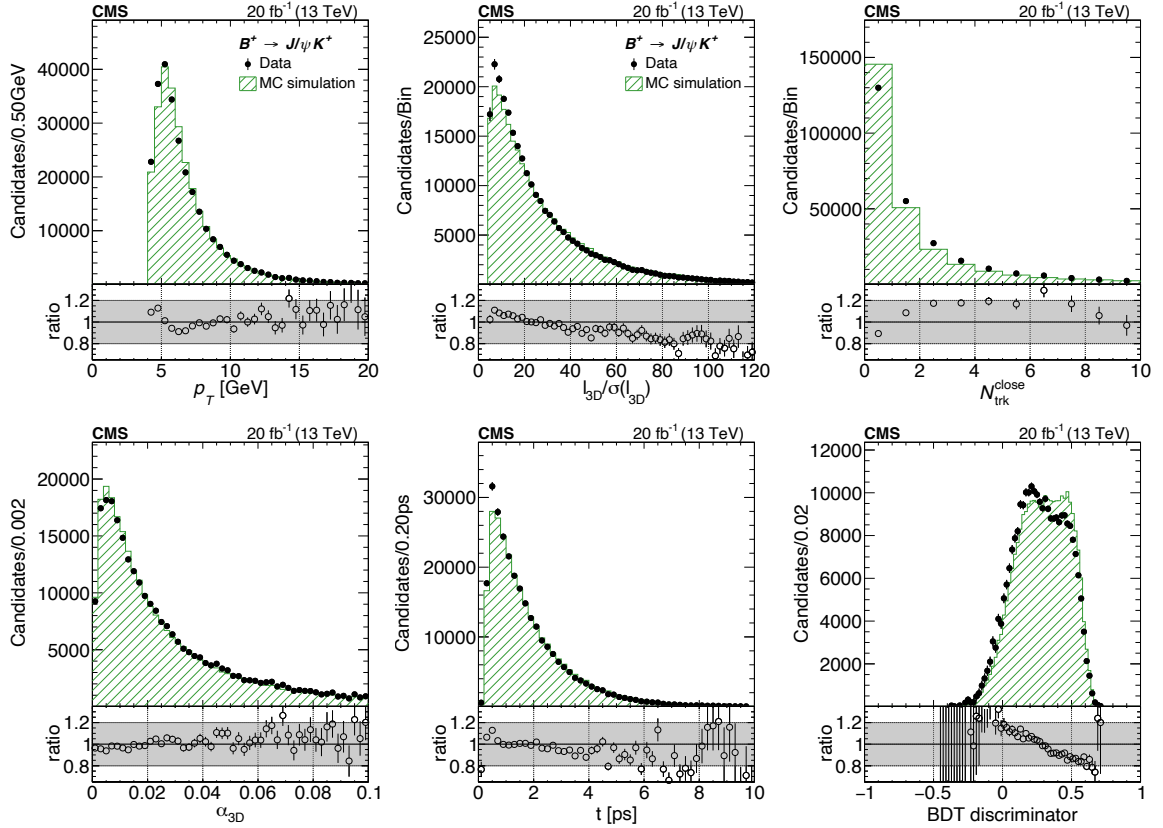


Figure 4.6: Illustrations of the sideband subtracted data (solid markers) and MC simulation (hatched histogram) comparison for few variables in $B^+ \rightarrow J/\psi K^+$ decay mode: the subleading muon p_T , the flight length significance ($l_{3D}/\sigma(l_{3D})$), N_{trk}^{close} , the pointing angle (α_{3D}), the decay time (t) and the analysis BDT discriminator. The comparison is for 2016BF channel 0. The MC simulation is normalized to the same number of events as the data. The shaded region in the ratio plot indicates a $\pm 20\%$ variation.

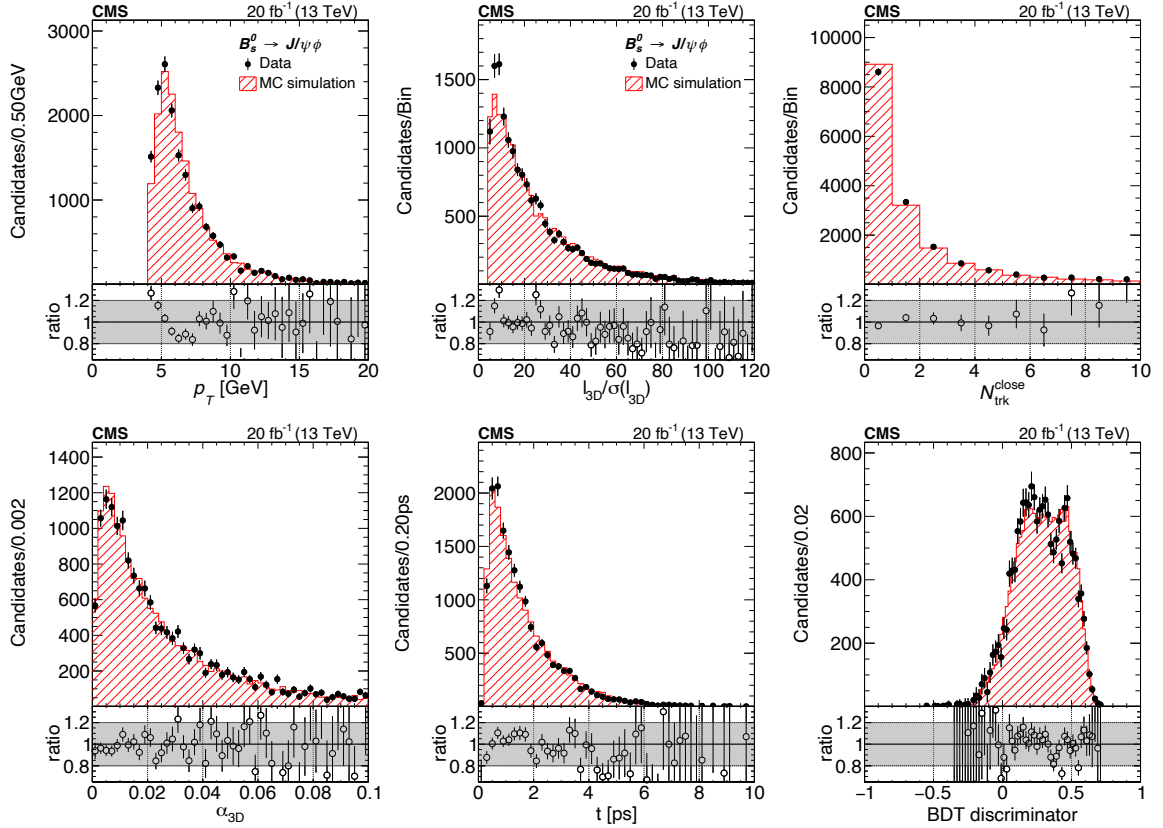


Figure 4.7: Illustrations of the sideband subtracted data (solid markers) and MC simulation (hatched histogram) comparison for few variables in $B_s^0 \rightarrow J/\psi \phi$ decay mode: the subleading muon p_T , the flight length significance ($l_{3D}/\sigma(l_{3D})$), N_{trk}^{close} , the pointing angle (α_{3D}), the decay time (t) and the analysis BDT discriminator. The comparison is for 2016BF channel 0. The MC simulation is normalized to the same number of events as the data. The shaded region in the ratio plot indicates a $\pm 20\%$ variation.

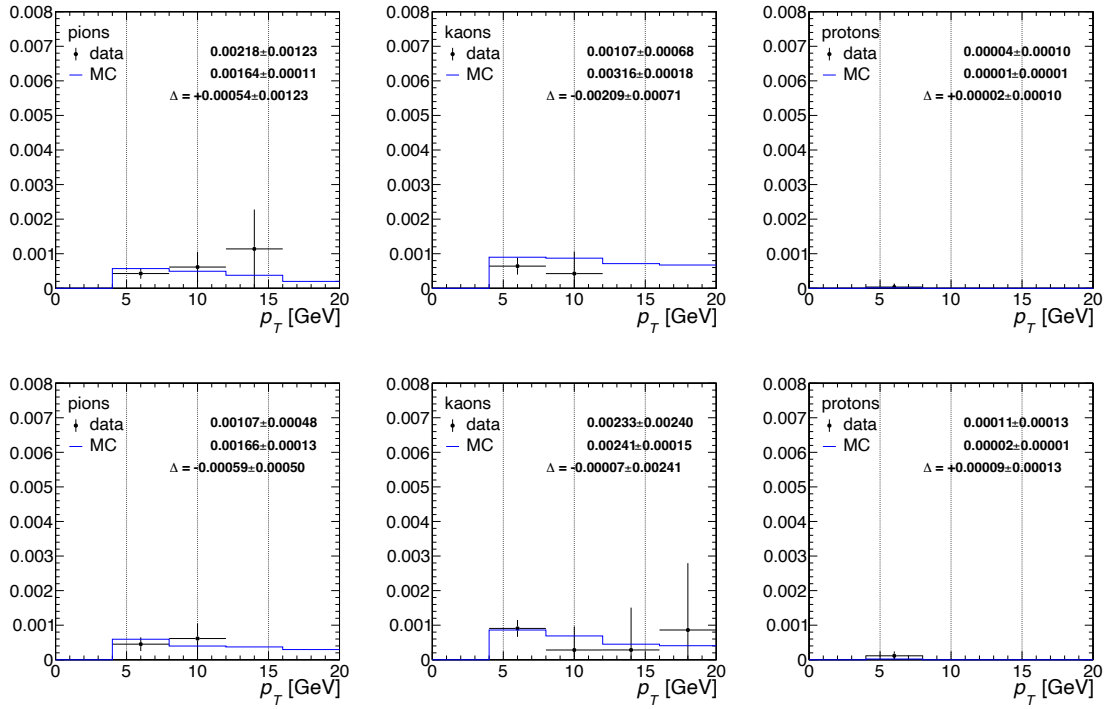


Figure 4.8: 2016 hadron misidentification probability in data and MC simulation. The top and bottom row shows the probabilities for channel 0, and channel 1, correspondingly. The systematic error is calculated from the error of the uncertainty-weighted average of data and MC simulation in the two regions.

decay provides a source of independently identified muons or hadrons. For muons, the decay $J/\psi \rightarrow \mu^+ \mu^-$ is used. Charged hadrons are selected with the decays $K_s^0 \rightarrow \pi^+ \pi^-$ for pions, $\phi \rightarrow K^+ K^-$ for kaons, and $\Lambda \rightarrow p^+ \pi^-$ for protons. These samples are used to compare the distributions of the variables used in the muon BDT in background-subtracted data and simulation, as well as the corresponding single-hadron misidentification probabilities. The distributions of all variables used in the muon BDT are found to be consistent between data and simulation.

Figure 4.8 illustrates the misidentification probabilities in the data and MC simulation for 2016. These comparisons give a feeling of agreement between the MC simulation-based misidentification rate and the one observed in the data. For proton, due to lack of statistics, it is hard to check the data-MC comparison. The numbers mentioned in the legend of the figures show the integrated (unnormalized) misidentification probability in the data and MC simulation. This comparison is used to assign 10% relative uncertainty in the pion and kaon misidentification probabilities and 60% relative systematic uncertainty for the proton.

4.7 Estimation of background and signal events

The expected number of signal events and rare background events are derived with respect to normalized channel $B^+ \rightarrow J/\psi K^+$ using the following formula: (for example, in the decay $Y \rightarrow X$)

$$N(X) = \frac{\mathcal{B}(Y \rightarrow X)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} \frac{f_Y}{f_u} \frac{A(Y)\epsilon_{ana}(Y)}{A(B^+)\epsilon_{ana}(B^+)} N_{B^+ \rightarrow J/\psi K^+}^{obs} \quad (4.4)$$

where $A(Y)$ and ϵ_{ana} are the acceptance and analysis efficiency which can be evaluated using the dedicated MC simulation, respectively. The acceptance efficiency is evaluated using the following formula,

$$A = \frac{N(\text{reco})}{N(\text{gen})} \quad (4.5)$$

where $N(\text{reco})$ is the number of event after the following criteria

- both the generated muons are requested to be within the geometric acceptance: $|\eta| < 2.5$ and $p_T > 3.5$ GeV.
- both muons must have a reconstructed high purity track with $|\eta| < 2.5$ and $p_T > 3.5$ GeV.
- For normalization and control channel, we have an additional selection criteria on the kaon track, at generator level, $p_{T,K} > 0.4$ GeV and $|\eta| < 2.5$ and to be reconstructed as a high purity track with $p_{T,K} > 0.5$ GeV and $|\eta| < 2.4$.

The denominator $N(\text{gen})$ is the total number of generated events in the signal or background decays. The analysis efficiency ϵ_{ana} can be written as a product of muon identification efficiency, trigger efficiency, and reconstruction efficiency. The reconstruction efficiency is defined as the ratio of the number of reconstructed candidates passing all selections to the number of events passing the acceptance criteria. Muon identification efficiency is defined

as the number of candidates, with two identified muons each satisfying the MVA muon selection criteria. The trigger efficiency is determined from those events which have two identified muons passing the trigger criteria.

During the evaluation of expected rare peaking and semileptonic yield, the trigger efficiency is taken as half of the $B_s^0 \rightarrow \mu^+\mu^-$ signal trigger efficiency. The choice of trigger efficiency is inherited from the anti-muon sample (AMS) study (Note that, In CMS, we don't have any dedicated hadronic trigger, so the trigger efficiency for hadronic decay mode is calculated from the signal by a conservative approach.) In AMS sample, the two muons are required to be global muon but fails the MVA muon selection.

Table 4.2: Branching fractions of all the decay modes used in this analysis.

Decays	Branching fractions
$B_s^0 \rightarrow \mu^+\mu^-$	$(3.57 \pm 0.17) \times 10^{-9}$ [81]
$B^0 \rightarrow \mu^+\mu^-$	$(1.06 \pm 0.09) \times 10^{-10}$ [81]
$B^+ \rightarrow J/\psi K^+$	$(6.02 \pm 0.18) \times 10^{-5}$ [82]
$B_s^0 \rightarrow J/\psi \phi$	$(3.17 \pm 0.24) \times 10^{-5}$ [82]
$B^0 \rightarrow J/\psi K^{*0}$	$(5.04 \pm 0.20) \times 10^{-5}$ [82]
$B^0 \rightarrow K^+\pi^-$	$(1.96 \pm 0.05) \times 10^{-5}$ [82]
$B^0 \rightarrow K^+K^-$	$(7.80 \pm 1.50) \times 10^{-8}$ [82]
$B^0 \rightarrow \pi^+\pi^-$	$(5.12 \pm 0.19) \times 10^{-6}$ [82]
$B^0 \rightarrow \pi^-\mu^+\nu$	$(1.50 \pm 0.06) \times 10^{-4}$ [82]
$B^0 \rightarrow \pi^0\mu^+\mu^-$	$(5.30 \pm 0.53) \times 10^{-8}$ [82]
$B^- \rightarrow \pi^-\mu^+\mu^-$	$(1.76 \pm 0.23) \times 10^{-8}$ [82]
$B_s^0 \rightarrow K^+\pi^-$	$(2.59 \pm 0.17) \times 10^{-5}$ [82]
$B_s^0 \rightarrow K^+K^-$	$(5.70 \pm 0.60) \times 10^{-6}$ [82]
$B_s^0 \rightarrow \pi^+\pi^-$	$(7.00 \pm 0.80) \times 10^{-7}$ [82]
$B_s^0 \rightarrow K^-\mu^+\nu$	$(9.40 \pm 2.40) \times 10^{-5}$ [83]
$\Lambda_b \rightarrow p\mu^-\nu$	$(4.10 \pm 1.00) \times 10^{-4}$ [82]
$\Lambda_b \rightarrow pK^-$	$(5.10 \pm 0.90) \times 10^{-6}$ [82]
$\Lambda_b \rightarrow p\pi^-$	$(4.20 \pm 0.80) \times 10^{-6}$ [82]

The branching fraction of different decay mode which goes into the Eq. 4.4 are given in Table 4.2. In Eq. 4.4, the normalization yield $N(B^+)$ from the data are directly used which are obtained from the fit to B^+ invariant mass. The expected yields for different rare decay

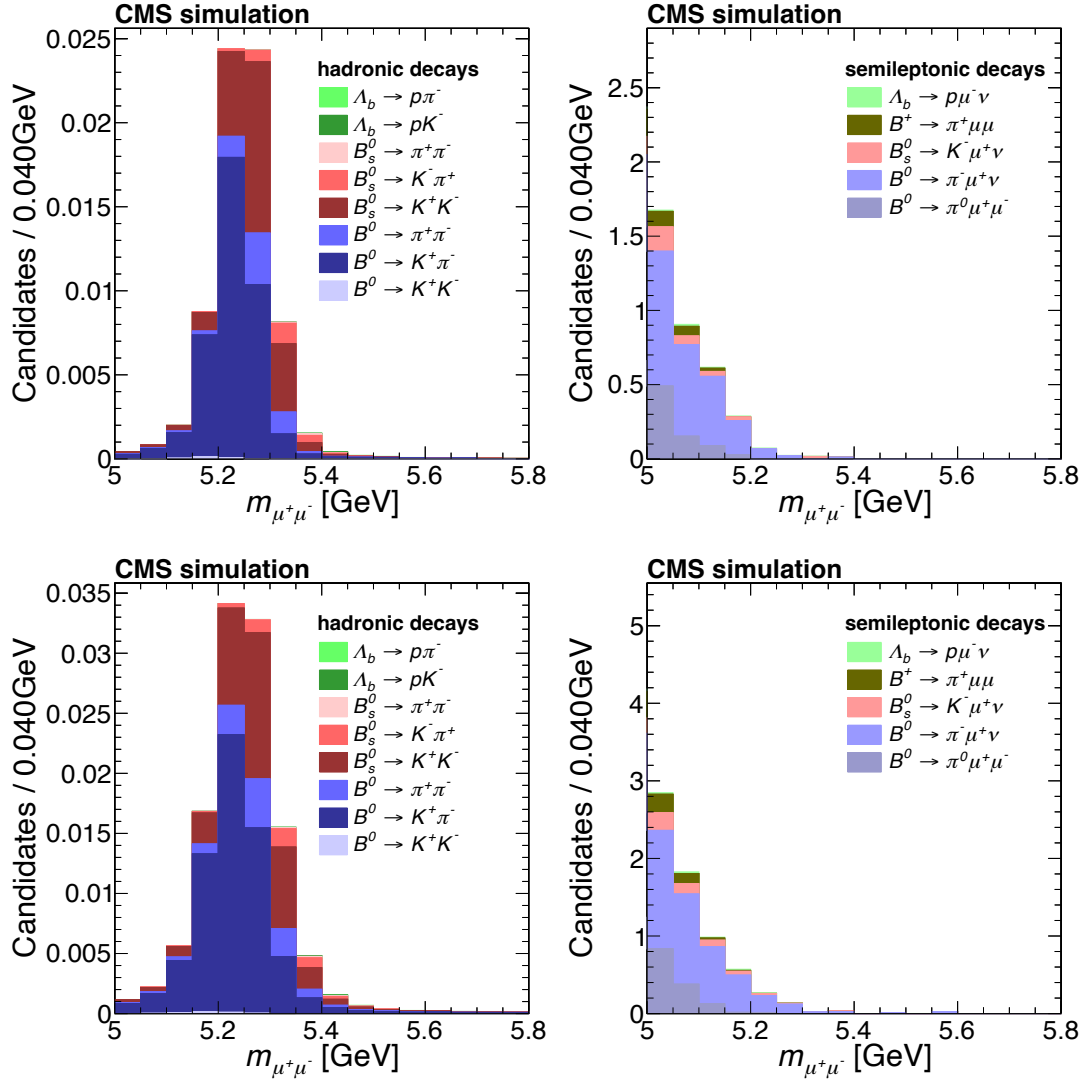


Figure 4.9: Invariant mass distribution of rare semileptonic (left column), and peaking (right column) backgrounds for 2016BF channel 0 (top row) and channel 1 (bottom row).

modes are plotted in Fig. 4.9.

The combinatorial background is formed with the random combination of two reconstructed muons from separate particle decays. It is possible to predict the number of events in the signal region by extrapolating the function, a zero-degree polynomial, used to fit the high sideband invariant mass distribution. The left sideband is not used for this estimation because it contains additional rare semileptonic background. An alternative Probability Distribution Function (PDF), an exponential function, is also used to check the expected combinatorial background yield. They are more or less consistent.

4.8 Unbinned maximum likelihood fit for the branching fraction

In this section, I will briefly describe the development of an unbinned maximum likelihood fitter to extract the branching fractions. The whole analysis strategy is developed and validated while keeping the signal region ($5.20 < m_{\mu^+\mu^-} < 5.45$ GeV) blinded, which is a good approach to avoid any possible bias by the experimenter due to incorrect discrimination of signal against background events. All the fitting procedures are validated on the pseudo-experiment, and after freezing the method, the signal region is uncovered.

4.8.1 Analysis strategy

To measure the branching fraction of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$, a UML fit setup is developed using invariant mass ($m_{\mu^+\mu^-}$), relative mass resolution ($\sigma(m_{\mu^+\mu^-})/m_{\mu^+\mu^-}$) and a binary distribution for the dimuon pairing configurations ($\mathcal{C}_{\mu^+\mu^-} = \text{seagull or cowboy}$). There are six components included in the UML: the B_s^0 signal, the B^0 signal, the semileptonic $B \rightarrow h\mu\nu$ background, the $B \rightarrow h\mu\mu$ background, the peaking background, and the combinatorial background. The likelihood PDF to describe the mass distribution for the

individual categories is

$$\mathcal{L} = N_{B_s^0} P_{B_s^0} + N_{B^0} P_{B^0} + N_{\text{peak}} P_{\text{peak}} + N_{h\mu\mu} P_{h\mu\mu} + N_{\text{semi}} P_{\text{semi}} + N_{\text{comb}} P_{\text{comb}} \quad (4.6)$$

where N_i is the number of events and P_i is the PDF, for each contribution i . The PDF for each component i is,

$$P_i(m_{\mu^+\mu^-}, \sigma(m_{\mu^+\mu^-}), \mathcal{C}_{\mu^+\mu^-}) = P_i(m_{\mu^+\mu^-}; \sigma(m_{\mu^+\mu^-})) P_i(\sigma(m_{\mu^+\mu^-})/m_{\mu^+\mu^-}) P_i(\mathcal{C}_{\mu^+\mu^-}) \quad (4.7)$$

The individual models introduced for each component are summarized below:

- **Signal B_s^0 and B^0 mass PDF:** the main mass model is a two-dimensional function produced by a Crystal-Ball (CB) line [84], for the dimuon invariant mass itself, times a kernel estimator [85] model for the relative mass resolution. A CB-line is a Gaussian function that has a power-law tail on the low mass side to model radiative energy loss in the final state.

$$P_s(m_i : \mu, \sigma, \alpha, n) = N \cdot \begin{cases} e^{-\frac{(m_i - \mu)^2}{2\sigma^2}}, & \text{for } \left(\frac{m_i - \mu}{\sigma}\right) > -\alpha \\ A \cdot (B - \frac{(m_i - \mu)}{\sigma})^{-n}, & \text{for } \left(\frac{m_i - \mu}{\sigma}\right) \leq -\alpha \end{cases} \quad (4.8)$$

where, $A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right)$ and $B = \frac{n}{|\alpha|} - |\alpha|$. N is a normalization factor. The width parameter of the CB-line is a conditional parameter linearly depending on the dimuon mass resolution. All of the parameters are fixed from fits to signal MC events. The dimuon mass scale is studied with $J/\psi \rightarrow \mu^+\mu^-$ and $\Upsilon(1S) \rightarrow \mu^+\mu^-$ decays, interpolated to $m_{B_s^0}$. In Run 1, the MC PDF is shifted by -6.0 (-7.0) MeV at the B_s^0 mass for the channel 0 (1), while in 2016 the shift is -4.4 (-3.1) MeV.

- **Peaking background mass PDF:** the model for dimuon invariant mass is an addition of Gaussian and a CB-line sharing a common mean. Similarly, the relative mass resolution is modeled with a kernel estimator function [85]. The Gaussian and CB-line widths for the invariant mass model are independent of the per-event mass resolution. They are fixed by fits to a mixture of weighted peaking background MC events,

together with the mean value and other parameters used in the CB-line. In the MC mixture, weights are calculated based on the expected yield for each process and are applied to the events.

- **Semileptonic $B \rightarrow h\mu\nu$ and $B \rightarrow h\mu\mu$ background mass PDFs:** the PDFs are simply modeled by a product of two kernel estimator models, one for the dimuon invariant mass and another one for the relative mass resolution. The models are extracted directly from a mixture of background MC samples with proper weights (calculated from the expected yields for each process).
- **Mass PDF for the combinatorial background:** the dimuon invariant mass is modeled with the first order, always positive Bernstein polynomial [86], while a kernel function also models the relative mass resolution. The slope of the polynomial is floated in the fits.
- **PDFs for dimuon pair configuration, $\mathcal{C}_{\mu^+\mu^-}$:** given this is just a binary observable for two muons bend towards (away from) each other in the magnetic field. The idea behind introducing this is to estimate the possible underestimation of $B \rightarrow h^+h^-$ background. The probability of getting two hadrons into two muons is significantly less, and one can assume them to be independent. However, this assumption is not necessarily correct if the two tracks overlap in the detector. A possible residual enhancement of the double-hadron misidentification probability exists relative to the assumption that this probability factorizes into the product of two separately measured misidentification probabilities. Potential bias from this source can be reduced by requiring that both muon candidates satisfy very strict quality criteria or by demanding that the tracks be spatially separated. The effect is different for two muon candidates that bend towards or away from each other, and thus the $\mathcal{C}_{\mu^+\mu^-}$ distribution is introduced. For most of the cases, such as signal B_s^0 and B^0 , combinatorial

background and $B \rightarrow h\mu\mu$ background, the distribution is flat. But for $B \rightarrow hh$ and $B \rightarrow h\mu\nu$ backgrounds, correction factors are introduced depending on the number of misidentified hadrons to the flat distribution.

The two signal yields, $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ are converted into branching fractions using the master formula (Eq. 4.4). The combinatorial background components, including the yields (N_{comb}) and the slopes in the PDF, are floated in the fit. The parameters which are treated as nuisance in the fit are described below.

- The branching fraction of $B^+ \rightarrow J/\psi K^+$ times the branching fraction of $J/\psi \rightarrow \mu\mu$ is constrained with a Gaussian prior.
- The fragmentation ratio f_s/f_u is constrained with a Gaussian prior.
- The ratios of the efficiencies between $B_{(s)}^0 \rightarrow \mu\mu$ and $B^+ \rightarrow J/\psi K^+$ in the branching fraction calculations are constrained with Gaussian priors.
- Yields of $B^+ \rightarrow J/\psi K^+$ events in the normalization of each category is constrained with Gaussian prior.
- Yields of semileptonic $B \rightarrow h\mu\nu$ background in each category are constrained with Gaussian prior.
- Yields of $B \rightarrow h\mu\mu$ background in each category are constrained with lognormal prior.
- Yields of peaking background in each category are constrained with lognormal prior.

4.8.2 BDT categories

As I mentioned in Section 3.3, the data are divided into two channels (channel 0 and channel 1) depending on the most forward muon pseudorapidity and into data collection running

periods (2011, 2012, 2016BF, and 2016GH). In total, there are eight channels: channel 0 and 1, in four data-taking running periods. To maximize the sensitivity and take full advantage of the BDT, channels with sufficient statistical precision are divided into mutually exclusive categories, low- and high-range categories in the analysis of the BDT discriminator. Table 4.3 summarizes the optimized BDT working points for different categories, channels, and data taking years. For 2011, only one category is considered because of small statistics. The low-range categories start from the first row of Table 4.3 to the boundary of the second row. The low-range categories are mostly dominated by combinatorial backgrounds compared to signal. The high-range categories boundary start from the second row and extends to +1. The boundaries are evaluated by maximizing the expected sensitivity using the full UML fit framework in Asimov data [87]. In summary, the UML fit is performed simultaneously in 14 categories. The main parameters of interest are the branching fractions of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$.

	2011		2012		2016BF		2016GH	
category\ channels	0	1	0	1	0	1	0	1
Low	0.28	0.21	0.27	0.23	0.19	0.19	0.18	0.23
High	n/a	n/a	0.35	0.32	0.30	0.30	0.31	0.38

Table 4.3: The final BDT selection used for the branching fraction determination in different channel, category, and data taking period. In 2011, only one category is selected because of low statistics.

Similarly, the final BDT selections in each channel are obtained from scanning the lifetime uncertainty using the Asimov data. A simplified mass times decay time model is used to estimate the uncertainty. For each BDT point, the expected event yields are calculated, and the Asimov data is generated using the simplified model. The sWeighted decay time distribution is then extracted, and a weighted binned fit is performed (similar to the method in Section 4.10). The optimized best B_s^0 lifetime thresholds are mentioned in Table 4.4.

Channel	2011	2012	2016BF	2016GH
0	0.22	0.32	0.22	0.22
1	0.19	0.32	0.20	0.29

Table 4.4: Optimized BDT discriminator boundaries per channel, and running period for the effective lifetime determination.

4.8.3 Fragmentation fraction ratio (f_s/f_u)

The fragmentation fractions, for example, f_s , f_u , and f_d , are the probabilities that a b quark will hadronize into a B_s^0 , B^+ , and B^0 meson. This factor arises in the master formula (Eq. 4.4) due to the use of $B^+ \rightarrow J/\psi K^+$ as a normalization channel. It is treated as an external parameter to this analysis. The $\frac{f_s}{f_u}$ value used in this analysis is

$$\frac{f_s}{f_u} = 0.252 \pm 0.012 (\text{exp}) \pm 0.015 (\text{CMS}). \quad (4.9)$$

where, exp stands for the experimental uncertainty and an additional uncertainty (labeled CMS) by adding in quadrature uncertainties evaluated from the consideration of $\frac{f_s}{f_u}$ obtained from $\sqrt{s} = 7$ TeV and $\sqrt{s} = 13$ TeV [88] and the difference of $\frac{f_s}{f_u}$ obtained from the transverse momentum (p_T) dependence [88] where the average p_T of $B_s^0 \rightarrow \mu^+\mu^-$ in this analysis is 17.2 GeV (Table 4.11).

4.8.4 Upper limit estimation for branching fraction

When computing the expected $B^0 \rightarrow \mu^+\mu^-$ signal yield, it is observed that the number of events is small. So, instead of quoting the branching fraction in the final result, the upper limit on the B^0 branching fraction is estimated using the CL_s criterion [89]. The chosen test statistic q is used to determine how signal- or background-like data are and is based on the profile likelihood ratio. The profile likelihood ratio is calculated by two fits, where one fit is carried out with a fixed B^0 rate to the scanning point and another fit is done with

a profiled B^0 branching fraction but with an upper bound set to the scanning B^0 rate (and with a lower bound of zero as well).

The ratio of probabilities to observe a value of the test statistic is at least as large as the one observed in the data, q_{obs} , under the signal+background (s+b) and background-only (b) hypotheses,

$$\frac{P(q < q_{\text{obs}}|\text{S} + \text{B})}{P(q < q_{\text{obs}}|\text{B}_{\text{only}})} = \text{CL}_s, \quad (4.10)$$

The $P(q < q_{\text{obs}}|\text{S} + \text{B})$ and $P(q < q_{\text{obs}}|\text{B}_{\text{only}})$ are the fraction of the events from the test statistic distributions ($-2\Delta NLL$) generated from a signal-plus-background hypothesis and a background-only hypothesis, respectively.

4.9 Unbinned maximum likelihood fit for the effective lifetime

This section discusses the effective lifetime of the B_s^0 mesons in the $B_s^0 \rightarrow \mu^+\mu^-$ final state, using the two-dimensional (2D) UML fit. The two dimensions used for the fit are the dimuon invariant mass ($m_{\mu^+\mu^-}$) and the decay time (t) of the B_s^0 meson. The decay time is defined as $t = m \frac{\ell_{3D}}{p}$, where m , ℓ_{3D} , and p are the mass, flight length, and momentum of the B_s^0 or B^+ meson. The fit method is validated on the $B^+ \rightarrow J/\psi K^+$ normalization channel because the lifetime of the B^+ meson is precisely measured. That is why the method validation on this channel will allow us to understand the various systematic uncertainties. The decay time below 1.0 ps and above 11.0 ps are not considered in the fit as the efficiency is very low below 1.0 ps and no signal events are found for a few channels in the MC after 11.0 ps.

4.9.1 Fit model

As I discussed previously, the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^+ \rightarrow J/\psi K^+$ decay channels have different background contributions. So the probability distribution function used for the 2D

UML fit to extract the lifetime can be written as,

$$L_{Tot} = N_{sig}\mathcal{F}_{sig}(m, t|\sigma_t) + N_{peak}\mathcal{F}_{peak}(m, t|\sigma_t) + N_{par}\mathcal{F}_{par}(m, t|\sigma_t) + N_{comb}\mathcal{F}_{comb}(m, t|\sigma_t) \quad (4.11)$$

where N_{sig} , N_{peak} , N_{par} , and N_{comb} are the signal and background (peaking, partially reconstructed, and combinatorial) yields. The variables $\mathcal{F}_{sig}$, $\mathcal{F}_{peak}$, $\mathcal{F}_{par}$ and $\mathcal{F}_{comb}$ are the corresponding 2D PDFs for signal and backgrounds. In $B_s^0 \rightarrow \mu^+\mu^-$ decay mode, $B_s^0 \rightarrow \mu^+\mu^-$ is treated as the signal, $B^0 \rightarrow \mu^+\mu^-$ is combined with $B \rightarrow h^+h^-$, collectively called as peaking background. The partial background consists of $B \rightarrow h\mu\mu$ and $B \rightarrow h\mu\nu$ decays. The combinatorial background consists of the events with a random combination of two muons. For $B^+ \rightarrow J/\psi K^+$ decay, the source of the peaking background is from $B^+ \rightarrow J/\psi \pi^+$ decay, and the partial background is from $B^+ \rightarrow J/\psi K^+ X$, where X can be pion or kaon, which is not reconstructed. The 2D PDF for different components are defined as,

$$\mathcal{F}_j = P_j(m)P_j(t|\sigma_t) \quad (4.12)$$

where index j is for signal, peaking, combinatorial, and partial reconstructed background. $P_j(m)$ and $P_j(t|\sigma_t)$ are the PDF for mass and decay time distribution.

4.9.2 Mass distribution model

The invariant mass of the B^+ hadron is modeled with double Gaussian function.

$$P_{sig}(m_i : \mu, \sigma_1, \sigma_2, f) = \frac{f}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(m_i - \mu)^2}{2\sigma_1^2}} + \frac{(1-f)}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(m_i - \mu)^2}{2\sigma_2^2}}, \quad (4.13)$$

where σ_1 , σ_2 and f are related mass resolution and fraction. The combinatorial background is modeled with the exponential function.

$$P_b(m_i : \lambda) = e^{(-\lambda m_i)}, \quad (4.14)$$

where the slope (λ) is floated in the fit. A triple Gaussian function is used to describe the peaking background distribution obtained after fitting the $B^+ \rightarrow J/\psi \pi^+$ simulation sample. The partial reconstructed background, which originated from the four body decay, for example, $B^0 \rightarrow J/\psi K^*$ where $K^* \rightarrow K^+ \pi^-$ and the pion is not reconstructed, is modeled with error function.

$$P_{par}(m_i : s, d) = \text{Erf}((-m_i + s)/d + 1), \quad (4.15)$$

In $B_s^0 \rightarrow \mu^+ \mu^-$ case: The signal B_s^0 is parametrized with a CB-line. The peaking background mass distribution is modeled with a sum of CB-line and Gaussian function. Then, a Gaussian function and Bernstein polynomial are used to model the semileptonic and combinatorial background, respectively.

4.9.3 Decay time distribution model

The signal decay time distribution is modeled by an exponential function convoluted with detector resolution.

$$P_s(t_i : \tau_B, \sigma_{t_i}) = [e^{(-t_i/\tau_B)} \otimes R(t_i, \sigma_{t_i})] \cdot \epsilon(t), \quad (4.16)$$

where $R(t_i, \sigma_{t_i})$ is the Gaussian resolution function which uses per-event decay time uncertainty (σ_t). The decay time uncertainty is calculated by propagating the uncertainty on the mass, flight length, and momentum of the B meson. τ_B is the lifetime of B meson which is floated in the fit. $\epsilon(t)$ is the efficiency as a function of decay time, more detail will be discussed in Section [4.9.4](#).

In $B^+ \rightarrow J/\psi K^+$ decay mode,

- decay time distribution for the combinatorial background is modeled with a sum of Gaussian and exponential functions,

- the peaking and partial reconstructed background are each modeled with exponential function convoluted with resolution function.

In $B_s^0 \rightarrow \mu^+ \mu^-$ decay mode,

- combinatorial background is modeled with a single exponential function with resolution,
- the peaking and partial reconstructed background are each modeled with exponential function convoluted with resolution function and efficiency correction.

4.9.4 Efficiency as a function of decay time

The exponential decay time for the signal events gets distorted due to the selection criteria (or efficiency effect). This signal efficiency (after all selection criteria) as a function of decay time distribution is determined using simulated samples. The efficiency is defined as the truth (generation) level decay time distribution of the reconstructed events after selection cuts divided by the exponential decay distribution generated with the fixed lifetime value (this value is fixed to the generator value of MC). Then the distribution is modeled with the function mentioned below,

$$\epsilon(t; p_0, p_1, p_2, p_3) = p_0 + p_1 t + \frac{p_2}{1 + \exp(-tp_3)} \quad (4.17)$$

The fitted efficiency plots for different channels in $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ are shown in Figs. [4.10](#) and [4.11](#).

4.9.5 Fit procedure

To extract the effective lifetime of B_s^0 meson, a simultaneous fit over the eight channels is performed in data for $B_s^0 \rightarrow \mu^+ \mu^-$ decay. Similarly, the validation is performed on $B^+ \rightarrow J/\psi K^+$ decay mode, and the fit procedure is described below.

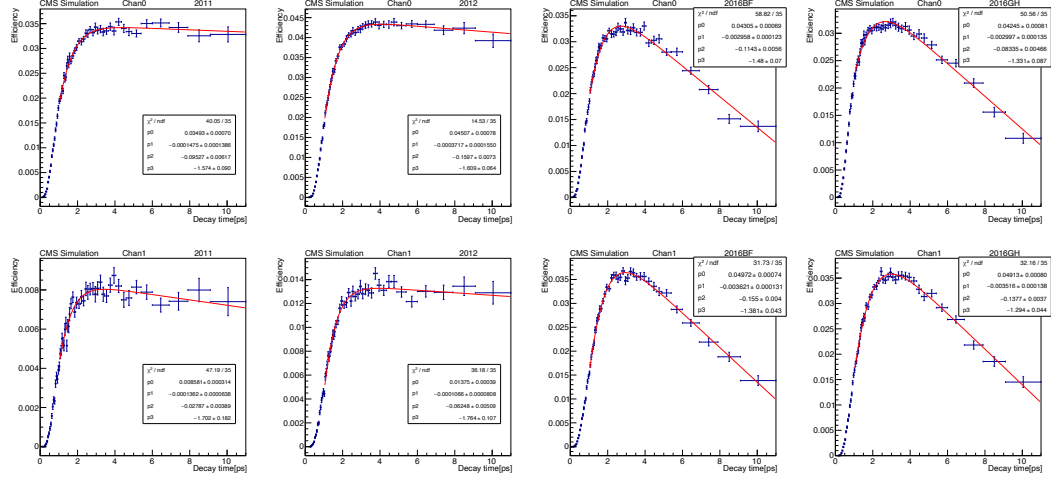


Figure 4.10: Efficiency plots for channel 0 (top row) and channel 1 (bottom row) in 2011, 2012, 2016BF and 2016GH $B^+ \rightarrow J/\psi K^+$ samples (from left to right). For 2011 and 2012, the MC is generated with lifetime 1.67 ps where as for 2016, it is 1.637 ps. The distributions are fitted with efficiency function, as mentioned in Eq. 4.17.

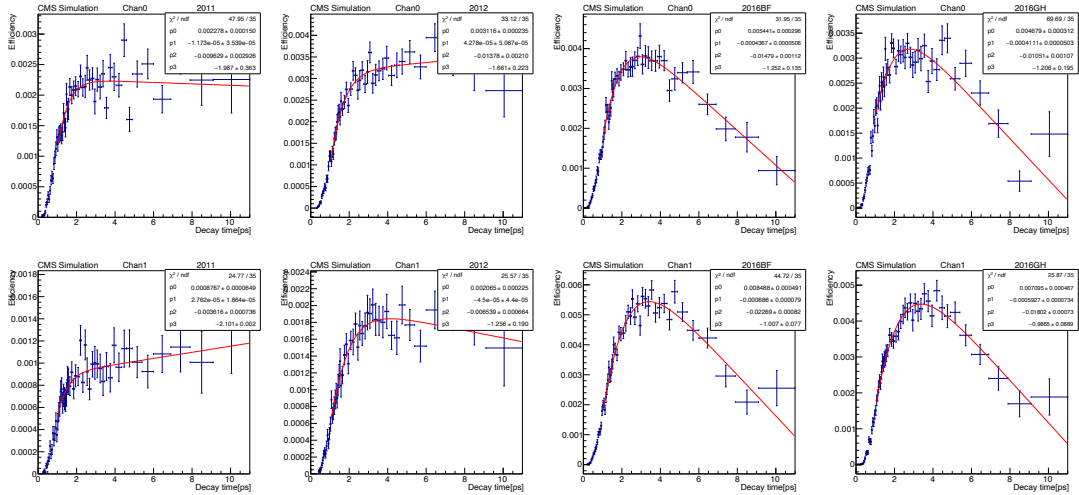


Figure 4.11: Efficiency plots for channel 0 (top row) and channel 1 (bottom row) in 2011, 2012, 2016BF and 2016GH $B_s^0 \rightarrow \mu^+\mu^-$ samples (from left to right). The efficiency distribution is fitted with the function mentioned in Eq. 4.17.

- The B^+ lifetime in each channel is extracted after going through a sequential fitting procedure:
 - Prefits are performed to determine the starting values of the parameters of the distribution for the next steps. A fit is performed to the $J/\psi K^+$ invariant mass in MC with the above-discussed PDF.
 - The peaking background, from $B^+ \rightarrow J/\psi \pi^+$, shape is obtained from the simulated sample and fixed.
 - The peaking background yield in the data fit is fixed to a branching fraction ratio of $B^+ \rightarrow J/\psi \pi^+$ to $B^+ \rightarrow J/\psi K^+$.
 - The signal efficiency as a function of decay time is measured by using a signal simulated sample.
 - Combinatorial background decay time shape is obtained from the mass higher sideband region ($5.36 < m_{J/\psi K^+} < 5.60$ GeV).
 - Now, using the initial value from the MC fit, a 2D fit between the mass and decay time is performed on MC and then on the collision data using the model described in Eq. 4.11.
 - The fit is performed with the decay time range [1.0, 11.0] ps and mass range [5.03, 5.6] GeV.
 - After performing the individual 2D fit in 8 different channels, a simultaneous fit to all the eight channels is done, where one single lifetime parameter for the signal is used to describe the eight channels. In simultaneous fit, we fix all the floated parameters obtained from a 2D fit to individual channels except the yields and the signal lifetime.
- After validating the method on the $B^+ \rightarrow J/\psi K^+$ decay mode, the procedure to mea-

sure the B_s^0 lifetime is developed. The procedure is developed a priori and validated on the pseudo-experiment, by keeping the B_s^0 signal window blinded in data. With the help of MC samples and the higher sideband data ($m_{\mu\mu} \in [5.45, 5.9]$), the mass shape and decay time efficiencies of the different components are obtained. Since the expected number of signal and background events are very small, it is necessary to fix the shape of different components to avoid a non-converging fit. The details of the procedure for each channel are discussed below.

- The dimuon invariant mass shape and the decay time efficiency function parameters for $B_s^0 \rightarrow \mu^+\mu^-$ signals are obtained from the MC sample. The invariant mass shape is fixed in the final data fit.
- The $B^0 \rightarrow \mu^+\mu^-$ and $B \rightarrow h^+h^-$ final state constitute our peaking background. The invariant mass shape, decay time shape, and efficiency functions are obtained from the weighted combined ($B \rightarrow h^+h^-$, and $B^0 \rightarrow \mu^+\mu^-$) MC sample and fixed in the fit. The weight is calculated based on the expected number of peaking background events.
- Similarly, the mass and decay time shape and efficiency as a function of decay time for the semileptonic backgrounds are taken from the weighted (based on expected numbers) MC sample and fixed in the fit.
- The combinatorial mass and decay time slope parameters are floated in the fit.
- The signal and combinatorial background yields are floated in the fit.
- The yields of the peaking and semileptonic backgrounds are constrained with lognormal priors to the expected yield from the MC.

4.9.6 Lifetime of B^+ from data

The 2D fit results obtained from the collision data and MC for different periods and channels are mentioned in Table 4.5. The BDT selection used for the measurement is the same as the final selection for B_s^0 . The lifetime values from the fits to MC samples are always close to the generated values as well as the PDG value (1.638 ps). Using the decay time efficiency from MC samples, the lifetime fit is performed on the data. The observed lifetime for a few channels shows a good agreement with the PDG value, whereas in 2016GH channel 1, the deviation is around 0.07 ps. During the initial period of the Run 2 data taking, the microstrip detector experienced operational instabilities resulting in a significant impact on the trigger efficiency as the level of pileup increased. Later in mid-2016, the issue was fixed. It is believed that the observed deviation is a result of imperfect MC simulation for detector conditions in different eras. The data fit projections for channel 0 and channel 1 are shown in Figs. 4.12 and 4.13.

Table 4.5: The 2D UML fit results to the individual channels, using the the signal MC as well as collision data. The events samples are taken after the BDT cut as mentioned in the 2nd column.

Era/channel	BDT	MC (in ps)	Data (in ps)	Diff. (PDG-Data) (in ps)
2011/ch0	> 0.22	1.671 ± 0.007	1.628 ± 0.006	+0.010
2011/ch1	> 0.19	1.666 ± 0.014	1.622 ± 0.014	+0.016
2012/ch0	> 0.32	1.667 ± 0.006	1.620 ± 0.004	+0.018
2012/ch1	> 0.32	1.669 ± 0.011	1.633 ± 0.007	+0.005
2016BF/ch0	> 0.22	1.635 ± 0.007	1.635 ± 0.006	+0.003
2016BF/ch1	> 0.30	1.633 ± 0.007	1.641 ± 0.006	-0.003
2016GH/ch0	> 0.22	1.631 ± 0.008	1.689 ± 0.008	-0.051
2016GH/ch1	> 0.29	1.630 ± 0.008	1.707 ± 0.009	-0.069

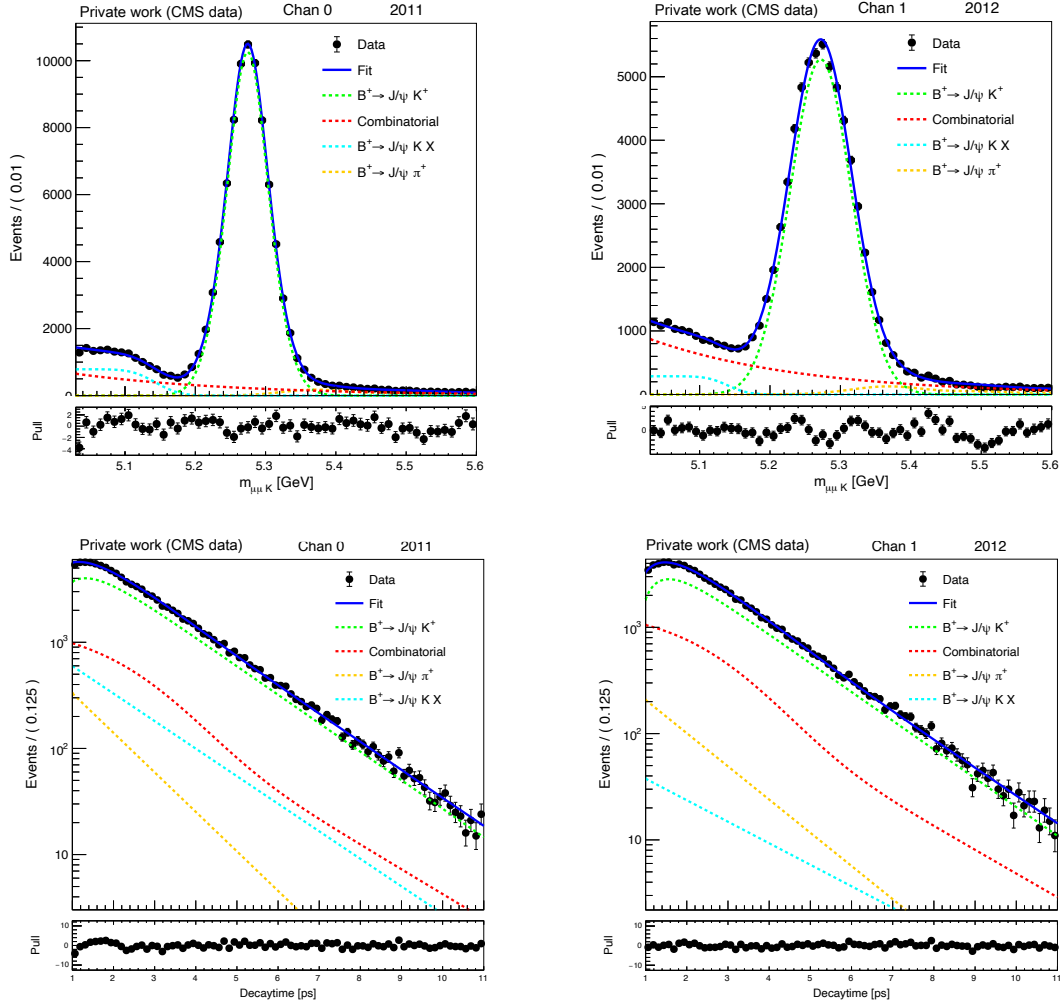


Figure 4.12: Invariant mass and decay time fit projections in $B^+ \rightarrow J/\psi K^+$ decay mode for 2011 channel 0 and 2012 channel 1. Solid blue curve is for the total PDF. The signal and background components are shown as dotted curves. Pull plots are shown in the bottom pad, which indicates the goodness of fit.

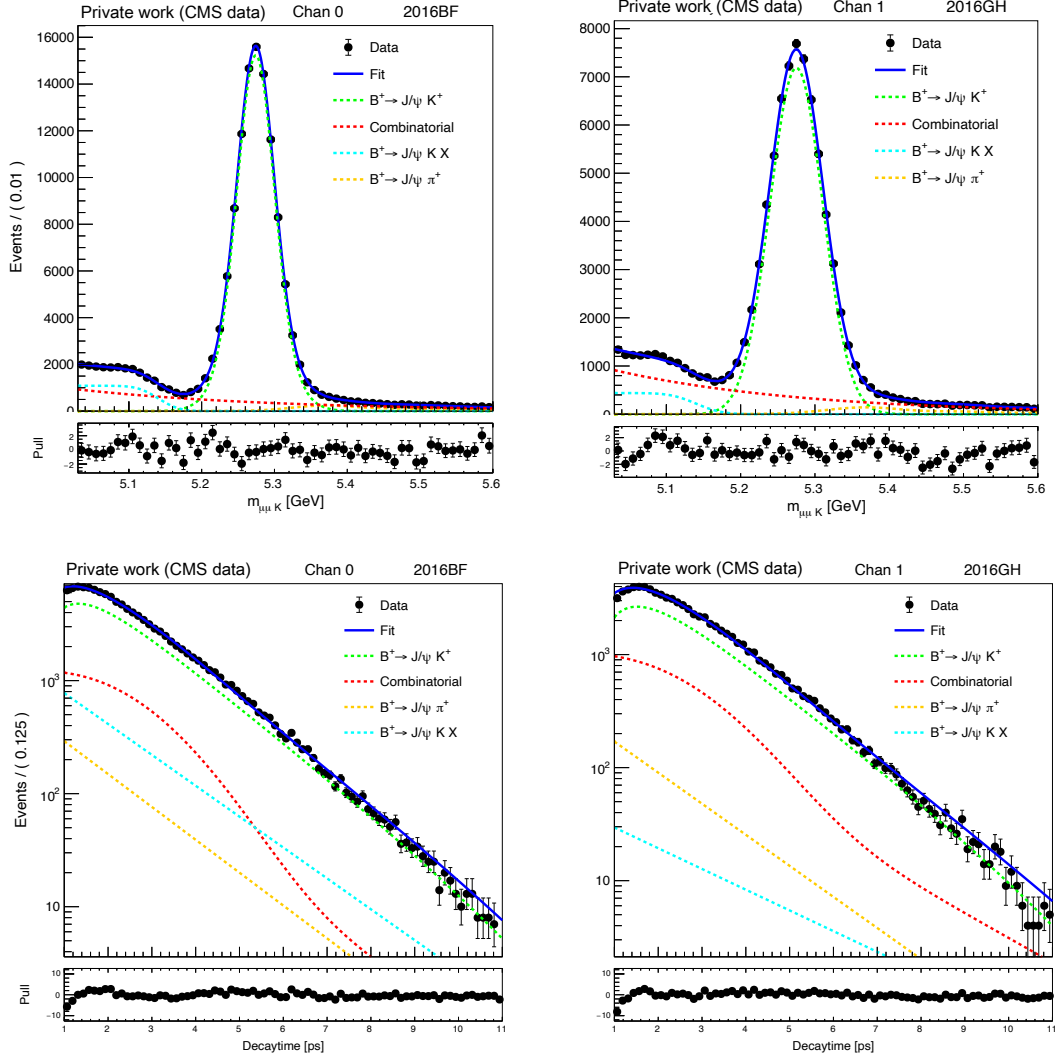


Figure 4.13: Invariant mass and decay time fit projections in $B^+ \rightarrow J/\psi K^+$ decay mode for 2016BF channel0 and 2016GH channel 1. Solid blue curve is for the total PDF. The signal and background components are shown as dotted curves. Pull plots are shown in the bottom pad, which indicates the goodness of fit.

4.10 sPlot fit method for effective lifetime

The second method to estimate the effective lifetime is discussed here. The lifetime fit is carried out in three steps. First, the branching fraction fit is performed to estimate the proper weights for each event. Then the weights are used to produce the sPlot in the decay time. At last, a weighted binned fit is performed to extract the effective lifetime. In the final lifetime fit, the fit PDF, which is the exponential function, includes the efficiency and resolution functions. This method is more straightforward than the UML fit. Like the UML fit, the validation is also performed on the $B^+ \rightarrow J/\psi K^+$ channel. Table 4.6 shows the observed lifetime from the sPlot fit, and the results are in very good agreement with the UML fit results (Table 4.5). This proves that the fit bias observed in the data is not from the fit model but from the detector miscalibration.

Table 4.6: A comparison of the effective lifetime from $B^+ \rightarrow J/\psi K^+$ data and the difference with PDG lifetime for different eras and channels.

era/channel	PDG	Data	Difference (PDG–Data)
2011/ch0	1.638 ± 0.004 ps	1.631 ± 0.006 ps	+0.007 ps
2011/ch1		1.612 ± 0.012 ps	+0.026 ps
2012/ch0		1.617 ± 0.003 ps	+0.021 ps
2012/ch1		1.620 ± 0.007 ps	+0.018 ps
2016BF/ch0		1.605 ± 0.006 ps	+0.033 ps
2016BF/ch1		1.616 ± 0.006 ps	+0.022 ps
2016GH/ch0		1.677 ± 0.009 ps	–0.039 ps
2016GH/ch1		1.709 ± 0.009 ps	–0.071 ps

The effective lifetime in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay is measured using the PDF described in Section 4.8. The full PDF is used to estimate the event weight, and using the event weight, the decay time histogram is produced. The number of the signal event is expected to be small in each category, so an integration is performed over all eras and channels to get a final decay time histogram. The effective lifetime is determined by doing a weighted binned

fit with a modified exponential function (which includes the channel-dependent resolution and efficiency effects). The uncertainty on the effective lifetime is obtained from a custom algorithm [90]. This algorithm has several features. First, the central value and the covariance matrix are calculated from the fit. Second, the bin-integrated PDF is calculated to reduce the bias due to the large histogram bin width. Third, it incorporates a resolution and efficiency model into the effective decay time PDF. Finally, it gives asymmetric uncertainties in the fit parameters. The fitter implementation is validated on the pseudo-experiments to estimate the expected uncertainties.

4.11 Systematic uncertainties

Systematic uncertainties affect the analysis in different ways. The lifetime error is dominated by lifetime-dependent uncertainties from trigger and selection variations that might not be well modeled in the MC simulation. The branching fraction uncertainty is limited by the badly modeled efficiencies in the MC simulation and the differences between the signal/control samples and the normalization sample. The systematic uncertainties for the branching fraction and lifetime are discussed below.

4.11.1 Systematic uncertainties on the branching fraction

All errors on branching fractions (see Table 4.2), f_s/f_u , and the kaon tracking efficiency uncertainty (4% and 2.3% for Run 1 and 2016, respectively) are treated as external systematic uncertainties. There is no systematic uncertainty from the integrated luminosity, as that number does not enter the analysis at any stage (all yields in the MC simulation are normalized to the number of $B^+ \rightarrow J/\psi K^+$ normalization events).

Inspection of the master formula for the branching fraction determination reveals that the systematic uncertainties come in different categories.

- Correlated errors on efficiency ratios.
- Single explicit uncertainties on specific observables (e.g. the normalization yield systematic error).
- Intrinsic uncertainties that do not appear explicitly in Eq. 3.1 (e.g. the systematic uncertainty on the hadron fake rate influences the number of signal events and its error).

The systematic uncertainty is determined by taking the control sample as a place holder for the signal. The double ratio is defined as

$$D = \frac{\left[\frac{\varepsilon(B^+ \rightarrow J/\psi K^+)}{\varepsilon(B_s^0 \rightarrow J/\psi \phi)} \right]_{data}}{\left[\frac{\varepsilon(B^+ \rightarrow J/\psi K^+)}{\varepsilon(B_s^0 \rightarrow J/\psi \phi)} \right]_{MC}} \quad (4.18)$$

and use $1 - D$ as systematic uncertainty for the ratio. The efficiency ratio, even though determined in MC simulation, should also describe the corresponding ratio in data. Therefore, the difference between the MC simulation-based ratio to the one obtained in data is taken as the systematic uncertainty.

4.11.1.1 Selection efficiency (data/MC simulation)

To obtain efficiency in data, loose preselection criteria are applied, and the number of candidates before and after the application of the selection criteria are determined with a binned χ^2 fit to the appropriate mass distributions. The invariant mass distributions of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ are fitted with the model described in Section 4.3. The deviation from unity is taken as systematic uncertainty. Systematic uncertainty due to analysis BDT requirement is calculated in a similar way. In 2016, it is approximately 5%, while in Run 1, it varies between 7 and 10% depending on the year and channel.

4.11.1.2 Signal efficiency dependence on lifetime

The effective lifetime of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay is predicted by the SM to be $\tau(B_{s,H}^0) = 1.615$ ps (this is the theoretical value at the time of publication). Since the selection favors signal decays that happen at longer proper decay times, the selection efficiency depends on the unknown effective lifetime: the selection efficiency is larger for longer effective lifetimes. As shown in Fig. 4.14, the efficiency is seen to depend fairly linearly on the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime. The value used in the CMS MC simulation is somewhat arbitrary (1.47 ps) and has no relation to the expected SM value. Therefore, a systematic uncertainty ranging from 1-3 % is assigned for different data-taking years.

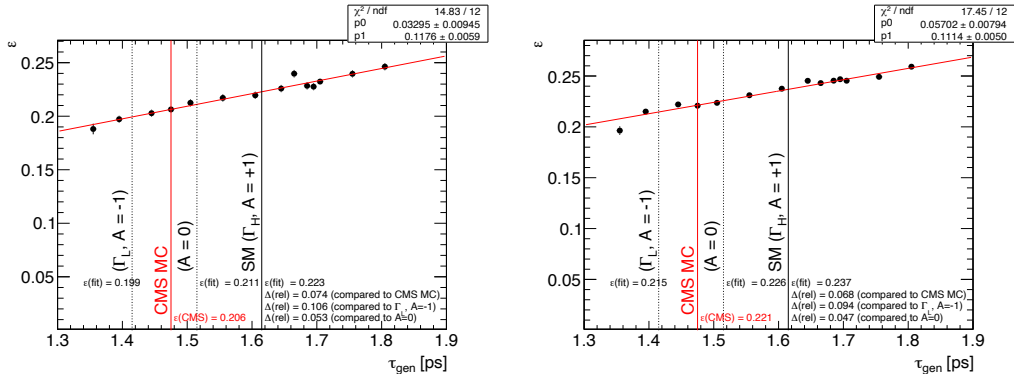


Figure 4.14: Dependence of the selection efficiency on the (unknown) effective lifetime of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay for 2016BF channel 0 (left column) and channel 1 (right column). The fit with a polynomial of first degree always passes through the point of the CMS MC simulation because that point has much more statistics than other points. Note that in the plots A is used as an abbreviation for $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$.

4.11.1.3 Production processes

The systematic uncertainty due to differences between data and simulation for the production mechanism mixture (introduced in subsection 2.5) is estimated as follows. In events with a $B^+ \rightarrow J/\psi K^+$ or $B_s^0 \rightarrow J/\psi \phi$ candidate and a third muon μ_3 , presumed to originate from the decay of the other b hadron in the event, the variable $\Delta R(B, \mu_3)$ provides discrimination between gluon splitting on the one hand and gluon fusion plus flavor excitation on

the other. The reason behind the addition of gluon fusion and flavor excitation is that they have very similar total efficiencies. Templates from MC simulation are fitted to the data distribution and used to determine the relative production mechanism fractions in the data. The sideband-subtracted data is used for only considering the signal for the fitting, and the fit projections for two BDT selections (loose and tight) are shown in Fig. 4.15. Reweighting the fractions in the MC simulation to the fractions determined in the data provides an estimate of 3% for the systematic uncertainty in the efficiency ratio.

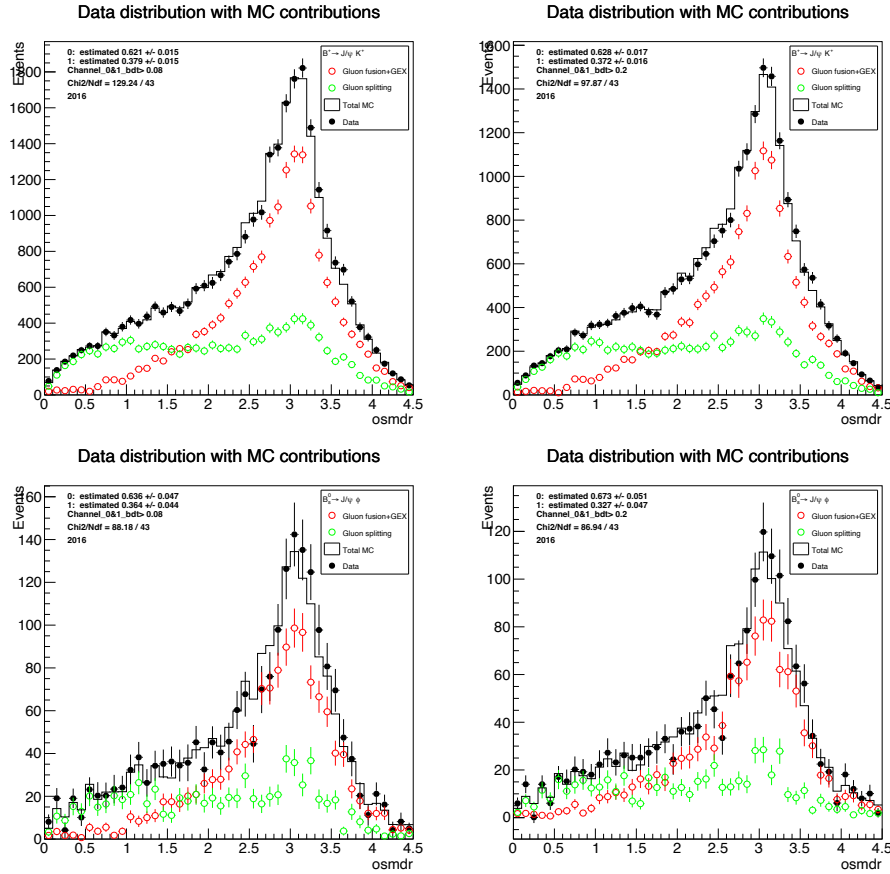


Figure 4.15: Data fit using the production process from the MC for 2016 $B^+ \rightarrow J/\psi K^+$ (top row) and $B^0 \rightarrow J/\psi \phi$ (bottom row) samples. The left and right columns are for loose and tight BDT selections. The red distribution is for gluon fusion plus flavor excitation, and the green distribution is for gluon splitting. The process fractions are mentioned in the plot legend.

4.11.1.4 Trigger efficiency

Unfortunately, the double ratio approach (Eq. 4.18) cannot determine the trigger efficiency systematics. The reason is that the control sample cannot be used as a placeholder for the signal, as the control sample uses the identical trigger as the normalization sample. We, therefore, use the variation over a wide BDT range of the trigger efficiency ratio between signal and normalization as determined in the MC simulation. In Fig. 4.16, the ratio of the signal and normalization trigger efficiencies is shown as a function of the BDT cut from 0 onwards. From the numbers provided on these plots, we assign a systematic uncertainty for the trigger efficiency ratio of 3%.

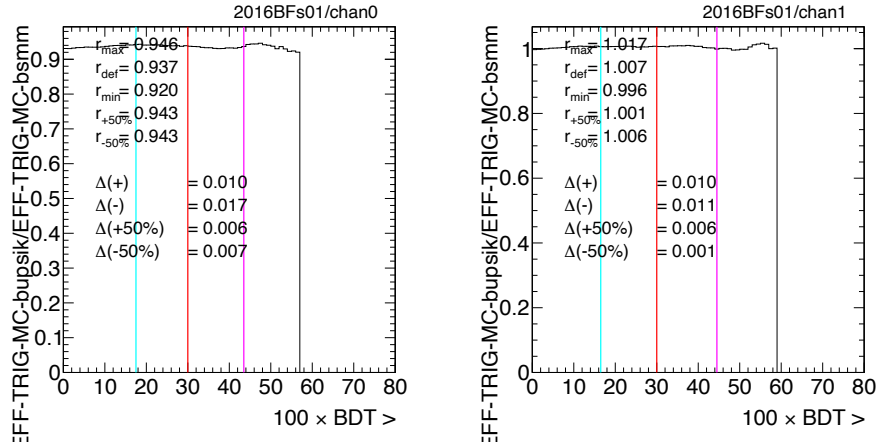


Figure 4.16: Ratio of trigger efficiency of the normalization sample to the signal sample as a function of the BDT threshold. These plots are used to estimate the systematic uncertainty on the trigger efficiency ratio by varying the BDT selection criteria. The red line corresponds to the default BDT cut, the magenta line is at the BDT value where the BDT-cut related efficiency has dropped by 50%, and the cyan line is displaced by the same distance from the red line. The trigger efficiency ratios corresponding to these lines correspond to the numbers labeled as $\pm 50\%$. The numbers labeled as $\Delta(+)$ or $\Delta(-)$ correspond to the variation of the trigger efficiency ratio over the entire BDT range shown.

Table 4.7: Summary of systematic uncertainty sources described in the text. The uncertainties quoted for the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ are relative uncertainties. The Era dependence systematic uncertainty is derived from the MC correction method for era-dependent variation on the ratios between $N(\mu\mu)$ from $B \rightarrow \mu^+ \mu^-$ sideband and $B^+ \rightarrow J/\psi K^+$ events.

Source	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ [%]
Kaon tracking	2.3-4
Normalization yield	4
Production processes	3
Muon identification	3
Trigger	3
Selection efficiency (data/MC simulation)	5-10
Efficiency lifetime dependence	1-3
Era dependence	5-6
Background parametrization	2.3
Total systematic uncertainty	0.3×10^{-9}

4.11.1.5 Muon identification

Muon identification criteria are applied to both signal and normalization samples. The systematic uncertainty due to muon selection is determined by comparing the muon identification efficiency ratio between the control sample and the normalization in data and MC simulation. The data and MC simulation agree to be better than 3% in all analysis channels.

4.11.1.6 Background parametrization

The nominal fit model uses a polynomial of degree one to describe the combinatorial background. An alternative PDF-like exponential function is used to check the bias introduced by using a different background shape. The change in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ value wrt the nominal result is 2.3%, which is considered as a systematic uncertainty.

In Table [4.7](#) all relative systematic uncertainties are summarized.

4.11.2 Systematic uncertainties on the effective lifetime

There are many source of systematic uncertainties involving our measurement of effective lifetime in the decay $B_s^0 \rightarrow \mu^+ \mu^-$. They are addressed below:

4.11.2.1 Modeling of decay time efficiency

The default efficiency function used, which is mentioned in Eq. 4.17, is not an accurate function to describe the efficiency. So we used two more efficiency functions (Eqs. 4.19 and 4.20) on the $B^+ \rightarrow J/\psi K^+$ sample to estimate the systematic error due to efficiency parametrisation. Figure 4.17 shows the three efficiency functions plotted on the efficiency histogram. The maximum deviation obtained from different channels after comparing two other efficiency functions with respect to the default efficiency function (Eq. 4.17), which is 0.01 ps, is taken as systematic error.

$$\epsilon(t; p_0, p_1, p_2, p_3, p_4) = p_0 \left(1 + p_1 t + \frac{p_2}{t^2} + \frac{p_3}{1 + \exp(-tp_4)} \right) \quad (4.19)$$

$$\epsilon(t; p_0, p_1, p_2, p_3, p_4) = p_0 + p_1 t^{p_2} \exp(tp_3 + t^2 p_4) \quad (4.20)$$

4.11.2.2 Scanning signal lifetimes using different generated samples

As the effective lifetime has not been measured precisely, its true value is not known. Eventually, one needs to check if the use of different efficiency functions introduces any type of bias or not. To perform this check, a set of simulated samples are generated with different effective lifetime values between 1.40 ps to 1.80 ps. The decay time efficiency for each sample is then estimated and later used to correct the decay time shape for the sample. Figure 5.24 shows the difference between the fitted lifetime values of different MC samples to the generated lifetime while using four different efficiency functions, obtained from samples 1.47 ps, 1.55 ps, 1.70 ps, and 1.75 ps. As we can see, the fitted results (or, in other words, the fitted-generated lifetime) fluctuates up and down while using the same

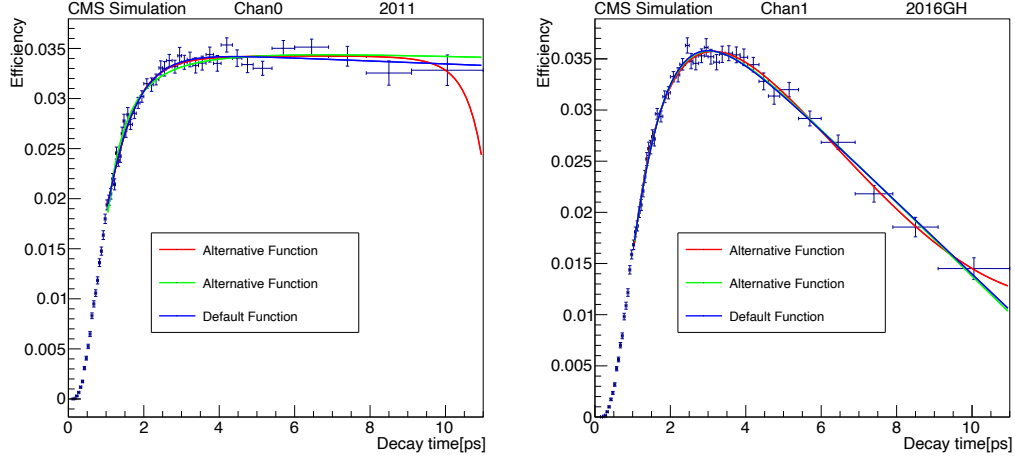


Figure 4.17: Three efficiency functions overlapped on the efficiency histogram for $B^+ \rightarrow J/\psi K^+$ sample in 2011 channel 0 and 2016GH channel 1. The blue, green, and red curves are from Eqs. [4.17](#), [4.19](#), and [4.20](#), respectively.

efficiency functions. The overall shift is estimated by fitting this distribution to a straight line. The shift obtained from different efficiency functions is not a real bias due to the lifetime value used in the MC generation but is due to sample-to-sample statistical fluctuations. We assign a systematic uncertainty of 0.03 ps from the shift due to the choice of a particular efficiency function from the statistically limited MC sample.

4.11.2.3 BDT dependency

The effective lifetime extracted from the fit may depend on the choice of the BDT cut value. So, by changing the cut values, the effect of BDT selections is studied and included as one of the sources of systematic error. The lifetimes obtained from the fit by changing the BDT efficiency by $\pm 10\%$ are shown in Table [4.8](#). After changing the cut threshold, the efficiency function has been re-evaluated, and the lifetime fitting procedure is performed again. The deviation of 0.02 ps in the lifetime is considered as a systematic uncertainty.

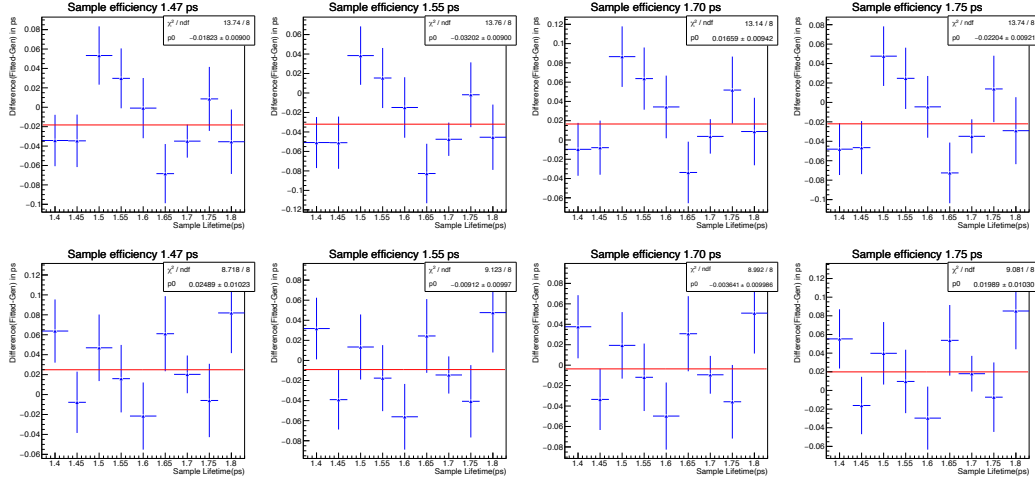


Figure 4.18: The top row shows results from 2016BF MC sample and bottom row plots are using 2016GH MC sample. The top-left plot shows the fitted lifetime from different samples (with a generated lifetime in the range of 1.40 ps - 1.80 ps), while using efficiency from 1.47 ps MC sample. Other similar plots are shown with different efficiency functions.

Table 4.8: The 2D UML fit results to the individual channels from $B^+ \rightarrow J/\psi K^+$ channel, using the collision data as well as the variation of BDT efficiency.

Era/channel	BDT	Data (in ps)	Difference at $+/- 10\%$ BDT eff.(in ps)
2011/ch0	> 0.22	1.628 ± 0.006	$-0.011/+0.013$
2011/ch1	> 0.19	1.622 ± 0.014	$-0.009/+0.014$
2012/ch0	> 0.32	1.620 ± 0.004	$+0.005/-0.008$
2012/ch1	> 0.32	1.633 ± 0.007	$+0.004/-0.007$
2016BF/ch0	> 0.22	1.635 ± 0.006	$-0.017/+0.019$
2016BF/ch1	> 0.30	1.641 ± 0.006	$-0.013/+0.015$
2016GH/ch0	> 0.22	1.689 ± 0.008	$-0.015/+0.017$
2016GH/ch1	> 0.29	1.707 ± 0.009	$-0.014/+0.020$

4.11.2.4 Era dependency

Table 4.5 shows the values of the effective lifetime with the BDT cut that is used in the analysis. The maximum deviation between the results in different channels and the PDG value is taken as systematic uncertainty, which is 0.069 ps.

4.11.2.5 Fit validation

Before examining the $B_s^0 \rightarrow \mu^+ \mu^-$ data, 1500 toys (each toy having similar to real data events) are generated by taking the fit model to check the bias on the lifetime. The signal PDFs and efficiencies are used from the signal $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample. The peaking and semileptonic background PDFs are obtained from the weighted simulated background samples as shown in Figs. 4.20 and 4.21. The shapes corresponding to combinatorial background are obtained from the upper sideband of the dimuon data, as shown in Fig. 4.22, after releasing the muon BDT cut (to increase the statistics). The shape of the event decay time error are obtained from MC or the high sideband data of dimuon invariant mass, for each channel, which are used to describe the per-event decay time error distribution in the toy. The signal and background expected yields are generated with Poisson fluctuation. After creating the eight different datasets for eight channels, a global simultaneous fit to eight channels is performed where a single lifetime for the signal is extracted. The fitting procedures are the same as mentioned in subsection 4.9.5. Figure 5.29 shows the results from the simultaneous fit to all channels in different eras. The fit bias is negligible. There is no systematic error assigned due to fit bias. From the toy studies, the expected lifetime uncertainty from the simultaneous fit to all channels is $(+0.39, -0.30)$ ps, whereas for the sPlot fit, the expected uncertainties are $(+0.49, -0.31)$ ps. Because of the smaller expected lifetime error, the UML fit is chosen as the primary method for the lifetime estimation. In the sPlot fit, we observe a bias of 0.09 ps on the lifetime, mainly coming from the low statistics.

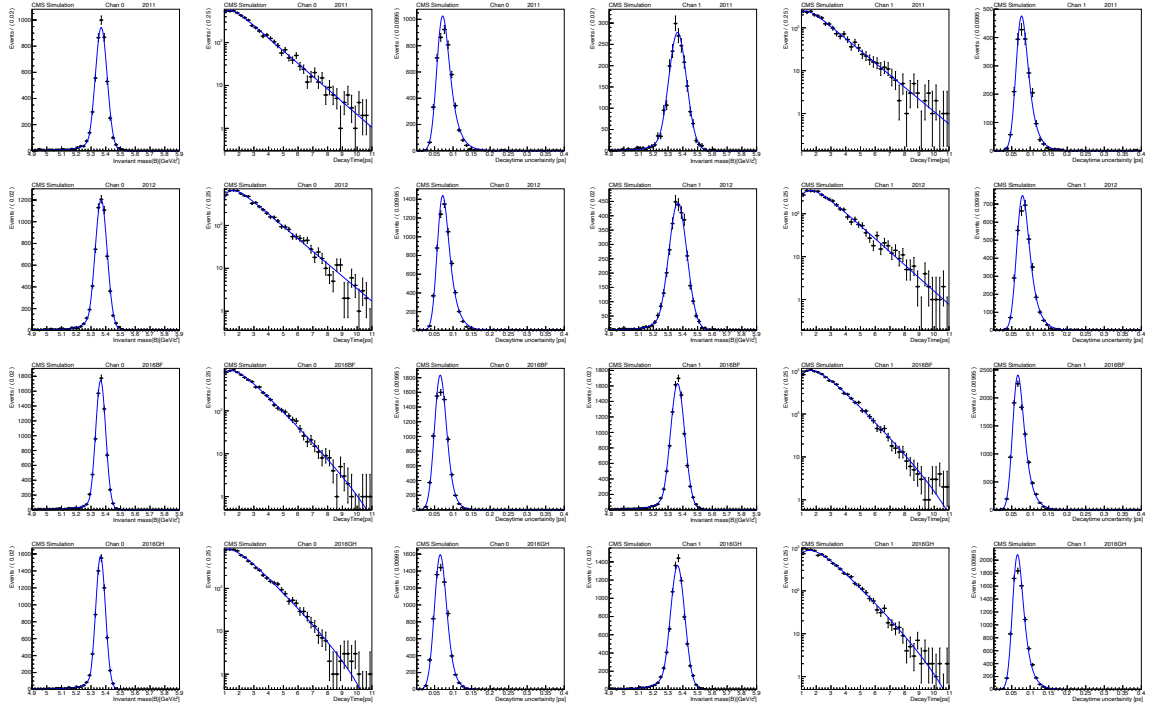


Figure 4.19: The invariant mass, decay time and decay time uncertainty fit projections from the 8 different channels obtained from the $B_s^0 \rightarrow \mu^+\mu^-$ signal MC samples (2011, 2012, 2016BF, and 2016GH).

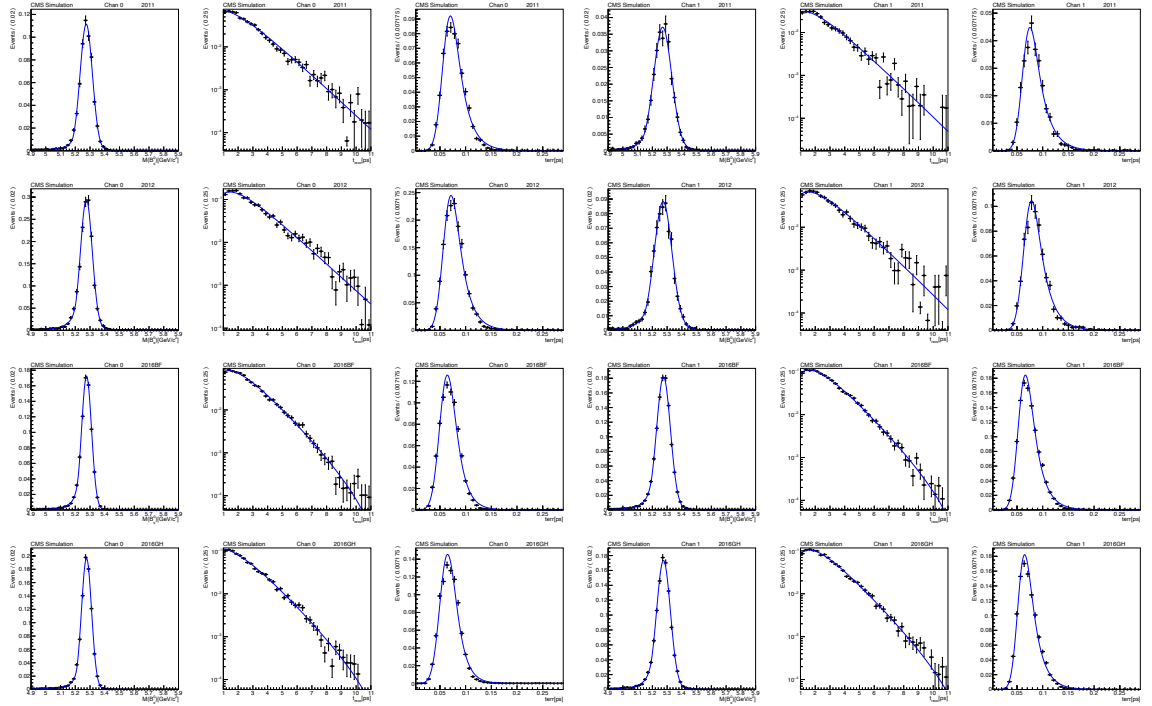


Figure 4.20: The invariant mass, decay time and decay time uncertainty fit projections from the 8 different channels obtained from the peaking MC samples (2011, 2012, 2016BF, and 2016GH).

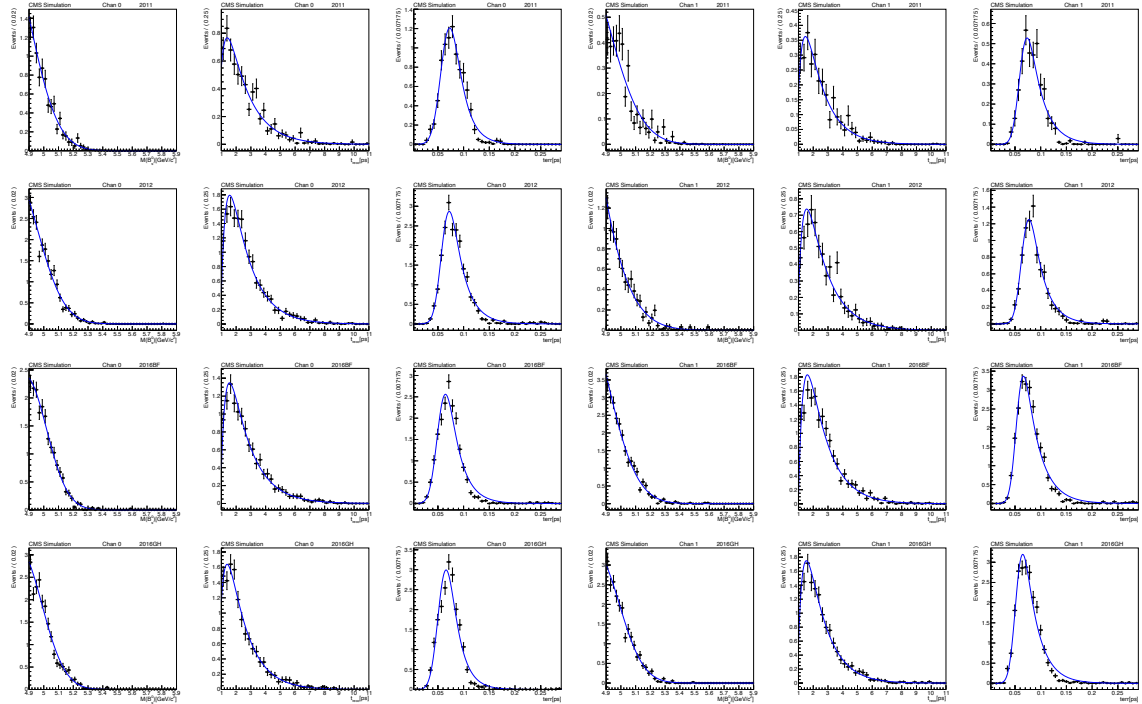


Figure 4.21: The invariant mass, decay time and decay time uncertainty fit projections from the 8 different channels obtained from the semileptonic MC samples (2011, 2012, 2016BF, and 2016GH).

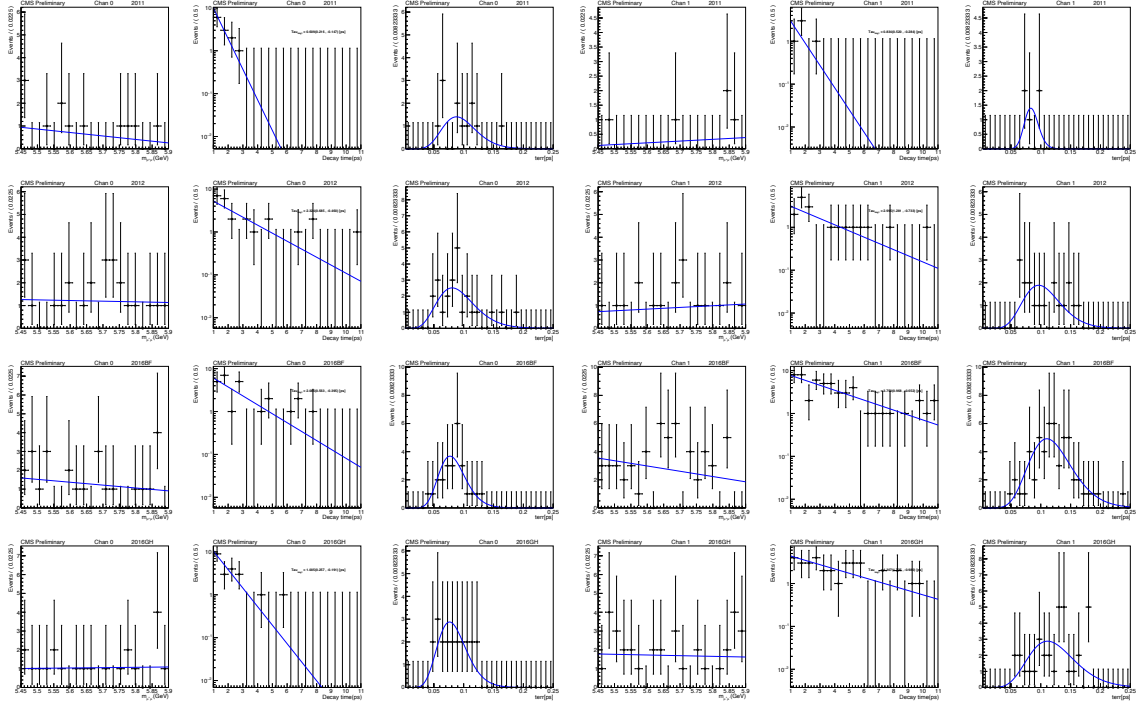


Figure 4.22: Combinatorial PDF shapes (mass, decay time, decay time uncertainty) for channel 0 and channel 1 from 2011, 2012, 2016BF and 2016GH high sideband data ($m \in [5.45, 5.9]$ GeV) without MVA muon BDT cut, used for toy generation.

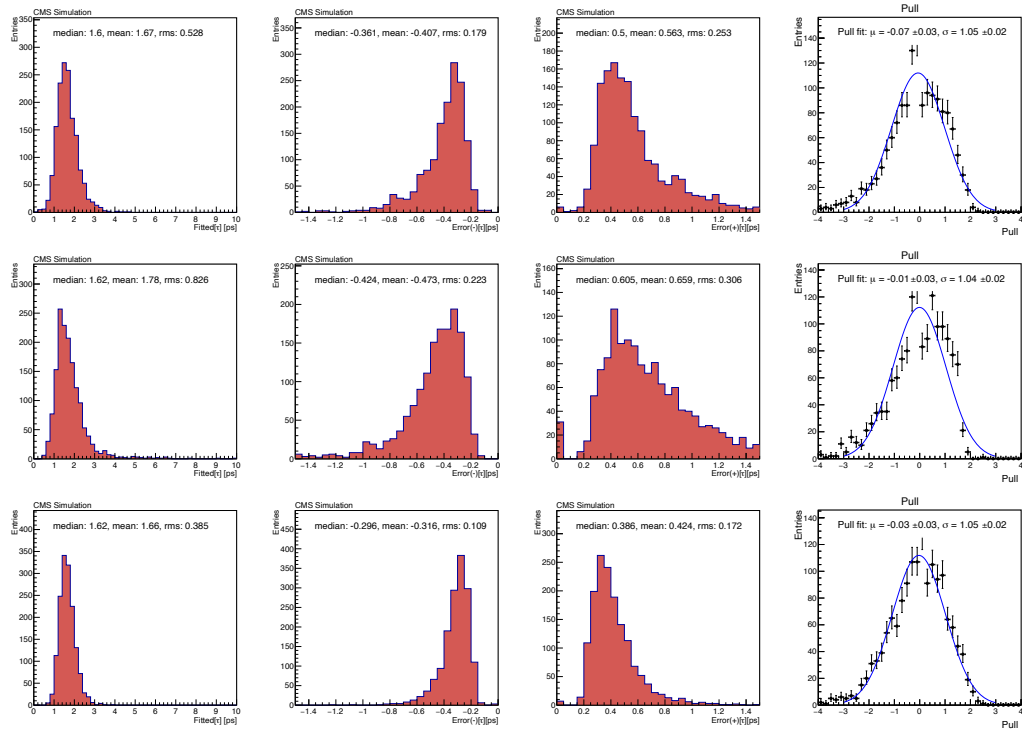


Figure 4.23: The above figure shows B_s^0 lifetime, B_s^0 lifetime positive and negative error and the pull distribution from the 1500 toy sample. The pull is calculated using the generation value of lifetime to be 1.615 ps. The top row is for 2011-2012 era, middle row is for the 2016 era and the bottom row is the combination of all era.

4.11.2.6 Tracker misalignment

Tracker misalignment systematics is estimated by splitting the 2016 data into different ϕ region of the B meson using $B^+ \rightarrow J/\psi K^+$ decay channel. The total data sample is splitted into four subsamples based on the azimuthal angle (ϕ) of the B meson. Our study is performed with a loose BDT selection. The lifetime is then extracted from each of the four subsamples separately, which are mentioned in Table 4.9. The deviation in lifetime which is the difference of simultaneous fit lifetime from individual ϕ region to total ϕ region lifetime (1.603 ± 0.004 ps) of B meson is calculated and the maximum deviation (0.02 ps) is used as a systematic error.

Table 4.9: Lifetime results in different ϕ region of B meson obtained from the total 2016 $B^+ \rightarrow J/\psi K^+$ sample in different channel. The last coloumn is the difference from the simultaneous fit to channel 0 and channel 1 result to the lifetime obtained from the total ϕ region (0 to 2π).

ϕ region	Channel 0	Channel 1	Simultaneous Fit	Deviation
0 to $\frac{\pi}{2}$	1.5757 ± 0.0123	1.5884 ± 0.0117	1.5824 ± 0.0085	0.021
0 to $-\frac{\pi}{2}$	1.6187 ± 0.0117	1.6087 ± 0.0113	1.6136 ± 0.0081	0.010
0 to π	1.5804 ± 0.0126	1.6185 ± 0.0122	1.6005 ± 0.0088	0.003
0 to $-\pi$	1.6048 ± 0.0114	1.6303 ± 0.0114	1.6177 ± 0.0081	0.014

4.11.2.7 Double muon correction

The double muon correction factor is extracted from the BF fit and injected into the fit after sccaling the expected yield. The study is performed in the $B_s^0 \rightarrow \mu^+ \mu^-$ unblinded data. The associated systematic uncertainty is found to be 0.01 ps.

4.11.2.8 Systematics from background yields and shape uncertainties

In the final $B_s^0 \rightarrow \mu^+ \mu^-$ data fit, the semileptonic and peaking background yields (nuisance parameters) for each channel are inside lognormal constraints. The systematic uncertainties are calculated by taking the variation of the nuisance parameter $\pm 20\%$ of the mean value

and $\pm 1\sigma$ of the semileptonic and peaking background shape parameter for each channel one by one. The maximum deviation from the nominal fit is found out to be 0.03 ps.

The total systematic uncertainties obtained from 2D UML and sPlot method are summarized in Table 4.10. In the sPlot fit, the “Limited MC simulation sample statistics”, “Tracker misalignment”, and “Background yields” are included in the efficiency modeling.

Table 4.10: Summary of the lifetime systematics. The contributions that are included in other items are indicated by (*).

Source	2D UML(in ps)	sPlot (in ps)
Efficiency modeling	0.01	0.04
Limited MC simulation sample statistics	0.03	(*)
BDT dependency	0.02	0.02
Era dependency	0.07	0.07
Tracker misalignment	0.02	(*)
Double muon correction	0.01	0.01
Fit validation	-	0.09
Background yields and shape uncertainty	0.03	(*)
Total	0.09	0.12

4.12 Results

4.12.1 Branching fraction

The signal region is “unblinded” after the finalization of selection and validation of procedure on pseudo-experiments. We observe (expect, in the SM) a total $B_s^0 \rightarrow \mu^+\mu^-$ yield of $60.8^{+14.5}_{-13.3}(74 \pm 3.5)$ candidates (summing over all categories). The result of the branching fraction fit is

$$\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (2.9 \pm 0.6(\text{stat}) \pm 0.3(\text{syst}) \pm 0.2(\text{frag}) \times 10^{-9}, \quad (4.21)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (0.8^{+1.4}_{-1.3}) \times 10^{-10} \quad (4.22)$$

where the first and the second uncertainties in $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ are the statistical and systematic uncertainty. The third uncertainty is due to f_s/f_u . In $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$ result, the

Table 4.11: Summary of the fitted yields (for $B_s^0 \rightarrow \mu^+\mu^-$, $B^0 \rightarrow \mu^+\mu^-$, the combinatorial background for $4.9 < m_{\mu\mu} < 5.9$ GeV, and the $B^+ \rightarrow J/\psi K^+$ normalization), the average p_T of the $B_s^0 \rightarrow \mu^+\mu^-$ signal, and the ratio of efficiencies between the normalization and the signal in all 14 categories of the 3D UML branching fraction fit. The size of peaking background is approximately 5–10% of the $B^0 \rightarrow \mu^+\mu^-$ signal. The average p_T (in GeV) is calculated from the MC simulation and has negligible uncertainties. The errors shown include the statistical and systematic components. It should be noted that the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ yield errors (1) are determined from the branching fraction fit and (2) also include the normalization errors.

Category	$N(B_s^0)$	$N(B^0)$	N_{comb}	$N_{\text{obs}}^{B^+}/100$	$\varepsilon_{\text{tot}}/\varepsilon_{\text{tot}}^{B^+}$	$\langle p_T(B_s^0) \rangle$
2011/ channel 0 / high	$3.6^{+0.9}_{-0.8}$	$0.4^{+0.7}_{-0.6}$	8.4 ± 3.8	750 ± 30	3.9 ± 0.5	16.4
2011/ channel 1 / high	$2.0^{+0.5}_{-0.4}$	$0.2^{+0.4}_{-0.3}$	3.2 ± 2.2	220 ± 12	7.5 ± 0.8	14.9
2012/ channel 0 / low	$3.7^{+0.9}_{-0.8}$	$0.4^{+0.6}_{-0.6}$	115.8 ± 11.3	790 ± 32	3.8 ± 0.5	16.1
2012/ channel 0 / high	$9.3^{+2.3}_{-2.1}$	$1.0^{+1.7}_{-1.6}$	30.2 ± 7.3	2360 ± 95	3.2 ± 0.4	17.3
2012/ channel 1 / low	$1.7^{+0.4}_{-0.4}$	$0.2^{+0.3}_{-0.3}$	116.7 ± 11.0	190 ± 9	7.3 ± 1.0	14.3
2012/ channel 1 / high	$4.7^{+1.2}_{-1.1}$	$1.0^{+0.9}_{-0.8}$	31.0 ± 6.5	660 ± 27	5.9 ± 0.8	15.5
2016A/ channel 0 / low	$2.2^{+0.5}_{-0.5}$	$0.2^{+0.4}_{-0.4}$	43.0 ± 7.1	580 ± 23	3.1 ± 0.4	17.5
2016A/ channel 0 / high	$4.0^{+1.0}_{-0.9}$	$0.4^{+0.8}_{-0.7}$	13.2 ± 4.7	1290 ± 57	2.5 ± 0.3	19.3
2016A/ channel 1 / low	$3.7^{+0.9}_{-0.8}$	$0.4^{+0.7}_{-0.7}$	168.8 ± 13.5	780 ± 31	3.9 ± 0.5	15.8
2016A/ channel 1 / high	$8.1^{+2.0}_{-1.8}$	$0.8^{+1.5}_{-1.4}$	64.2 ± 9.7	1920 ± 78	3.4 ± 0.4	17.3
2016B/ channel 0 / low	$4.1^{+1.0}_{-0.9}$	$0.4^{+0.8}_{-0.7}$	128.8 ± 12.0	1020 ± 44	3.3 ± 0.4	17.2
2016B/ channel 0 / high	$3.6^{+0.9}_{-0.8}$	$0.4^{+0.7}_{-0.6}$	7.8 ± 3.6	1320 ± 54	2.2 ± 0.2	20.8
2016B/ channel 1 / low	$6.1^{+1.5}_{-1.4}$	$0.6^{+1.1}_{-1.0}$	133.4 ± 12.5	1260 ± 51	3.9 ± 0.4	16.2
2016B/ channel 1 / high	$3.9^{+1.0}_{-0.9}$	$0.4^{+0.8}_{-0.7}$	14.2 ± 4.6	1180 ± 49	2.7 ± 0.3	19.5

uncertainty combines the statistical and systematic terms. In the $B_s^0 \rightarrow \mu^+\mu^-$ decay, the observed (expected) significance, determined using Wilks' theorem [91], is 5.6 (6.5) standard deviations. The significance is estimated by taking the difference in the likelihood in fits when the signal component is included and excluded from the fit.

In Fig. 4.24, the mass projections are shown for all categories. Figure 4.25 combines the mass projections for the high and low BDT-threshold categories. In Table 4.11, the fitted yields (for $B_s^0 \rightarrow \mu^+\mu^-$, $B^0 \rightarrow \mu^+\mu^-$, the $B^+ \rightarrow J/\psi K^+$ normalization and the combinatorial background for $4.9 < m_{\mu\mu} < 5.9$ GeV), the ratio of efficiencies between the normalization and the signal and the transverse B_s^0 momentum in all 14 categories of the 3D UML branching fraction fit are shown.

As it can be seen from Table 4.11, the $B^0 \rightarrow \mu^+\mu^-$ yield obtained from the fitter is not significant. The observed (expected) significance is 0.6 (0.8) standard deviations based on

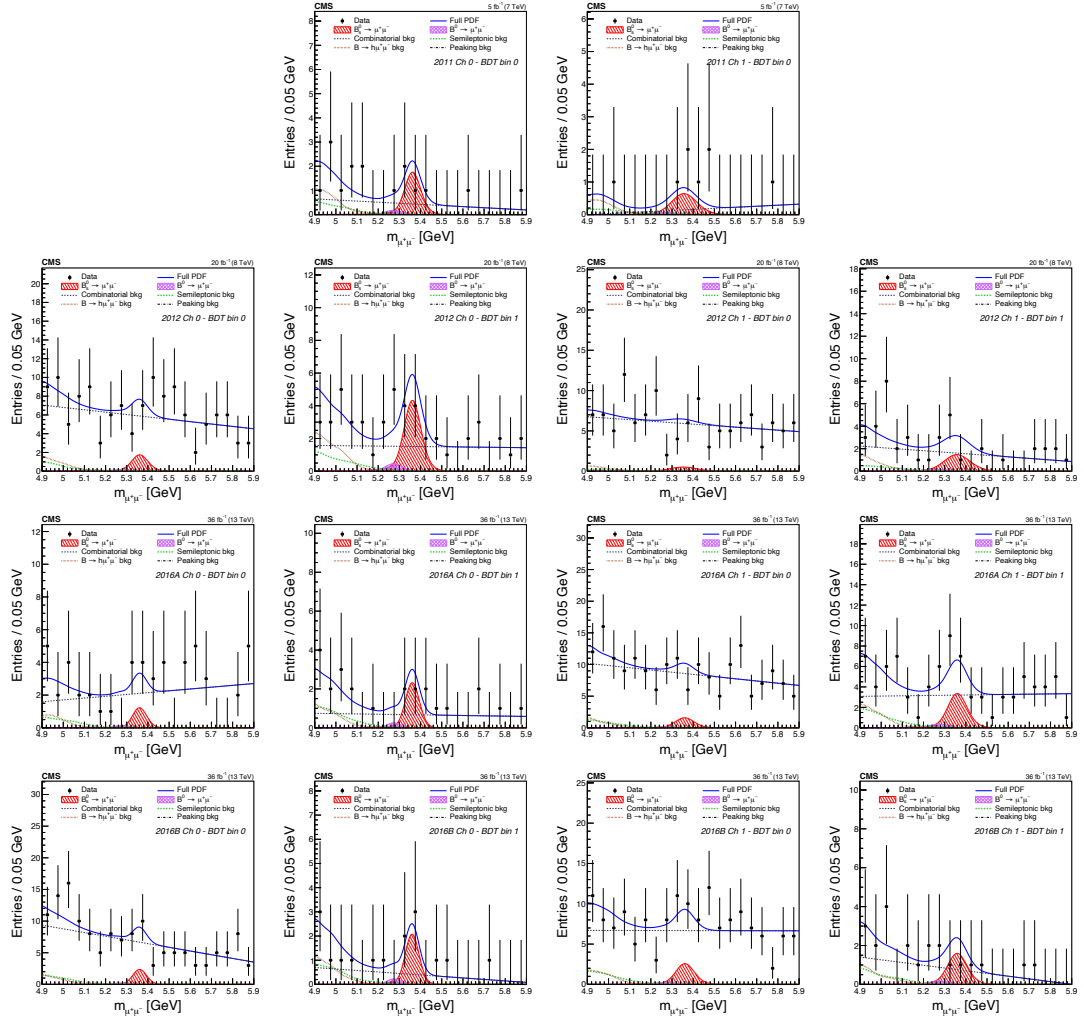


Figure 4.24: Mass projections from the final UML fits in different categories. The total fit is shown in solid blue curve, and the different background components are shown in dotted curves. The signal $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ components are shown in the red and violet hatched histograms.

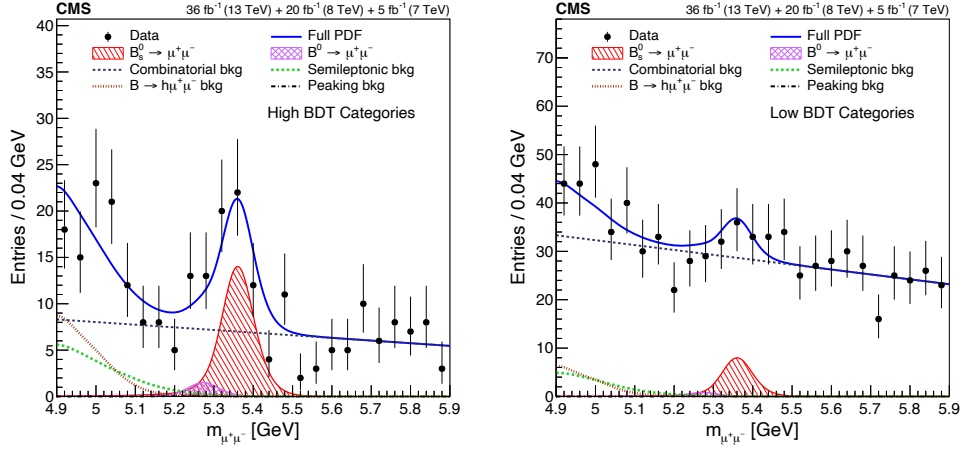


Figure 4.25: Combined mass projections for the high and low BDT-threshold categories. No event weighting is applied.

Wilks' theorem, treating $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ as a nuisance parameter. The upper limit is set on the the branching fraction of $B^0 \rightarrow \mu^+\mu^-$ using CL_s method.

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 3.6 \times 10^{-10} (95\% \text{ CL}) \quad (4.23)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 3.1 \times 10^{-10} (90\% \text{ CL}) \quad (4.24)$$

4.12.2 Effective lifetime using sPlot method

The sPlot-fit result of the effective lifetime of $B_s^0 \rightarrow \mu^+\mu^-$ is shown in Fig. 4.26 together with the merged mass projection. The result is

$$\tau(B_s^0 \rightarrow \mu^+\mu^-) = 1.55_{-0.33}^{+0.52} \text{ ps}, \quad (4.25)$$

where the error combines the statistical and systematic uncertainties.

4.12.3 Effective lifetime using UML method

The effective lifetime obtained from the 2D UML fit is

$$\tau(B_s^0 \rightarrow \mu^+\mu^-) = 1.70_{-0.44}^{+0.61} \text{ ps}, \quad (4.26)$$

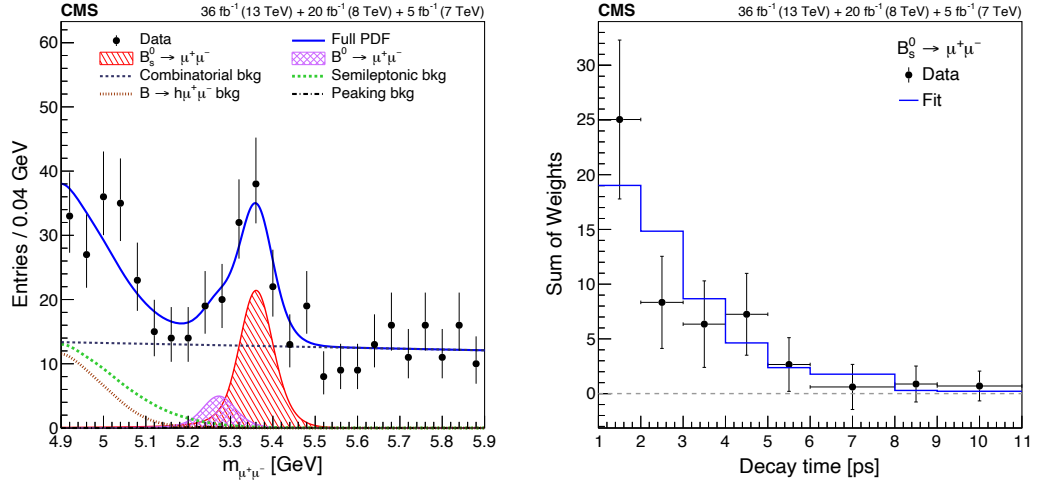


Figure 4.26: (Right) Decay time distribution and fit result for the sPlot lifetime fit setup. Note that no requirement on the decay time (> 1 ps) is applied. (Left) Merged mass projection of all the contributing data.

where the error combines the statistical and systematic uncertainties. The observed performance is compared to the expectation in Fig. 4.27.

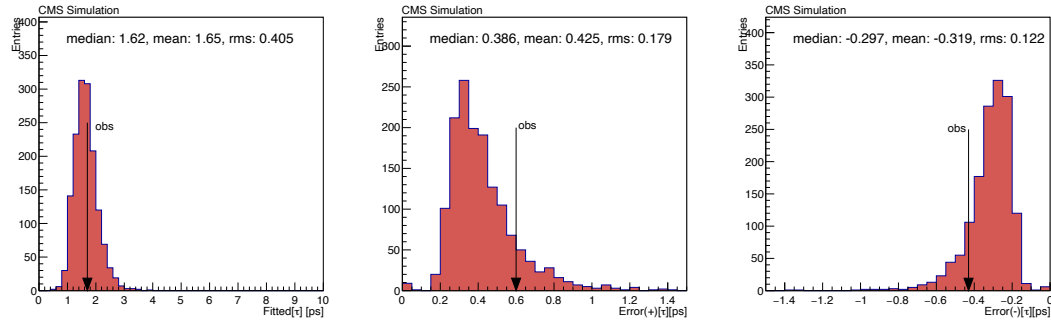


Figure 4.27: Comparison of the 2D UML fitted results to the expectations for the effective lifetime. From left to right the central value, the upper error, and the lower error are shown.

The merged mass and decay time projections are provided in Fig. 4.28 and the individual plots for each category in data are shown in Fig. 4.29.

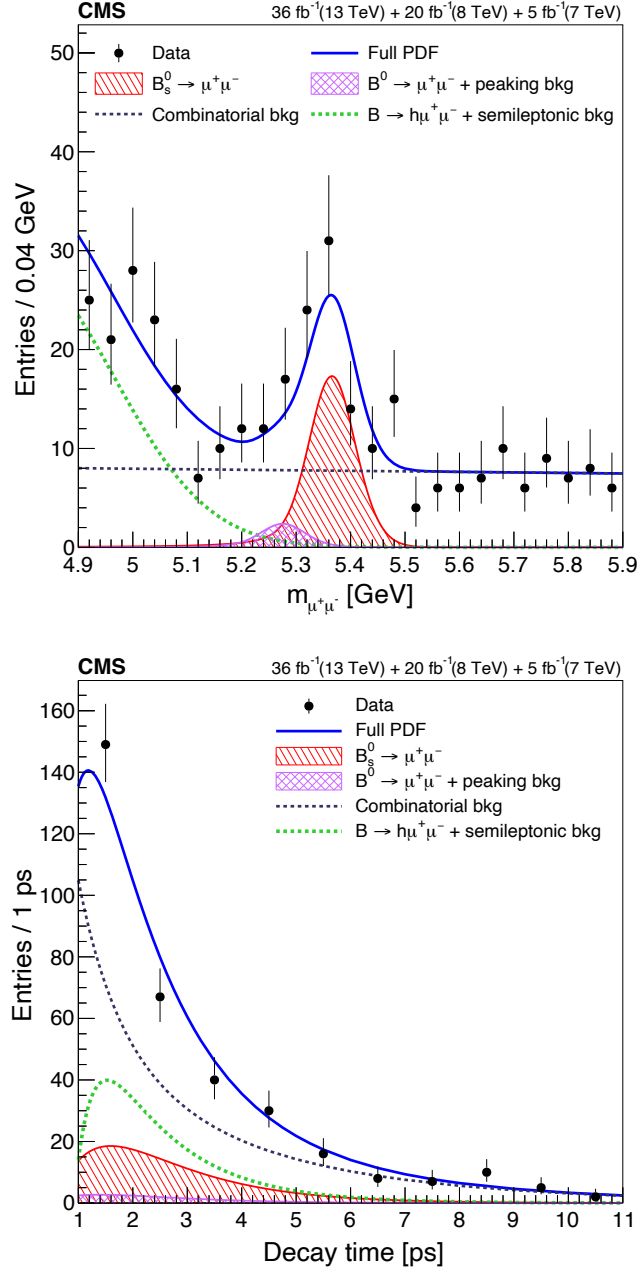


Figure 4.28: The data overlayed with fit projections for the dimuon invariant mass (top) and the proper decay time distribution distributions in linear (bottom) scale after merging all the channels. The total fit is shown in solid blue curve, and the different background components are shown in dotted curves and hatched histograms. The signal component is shown in the red hatched histogram.

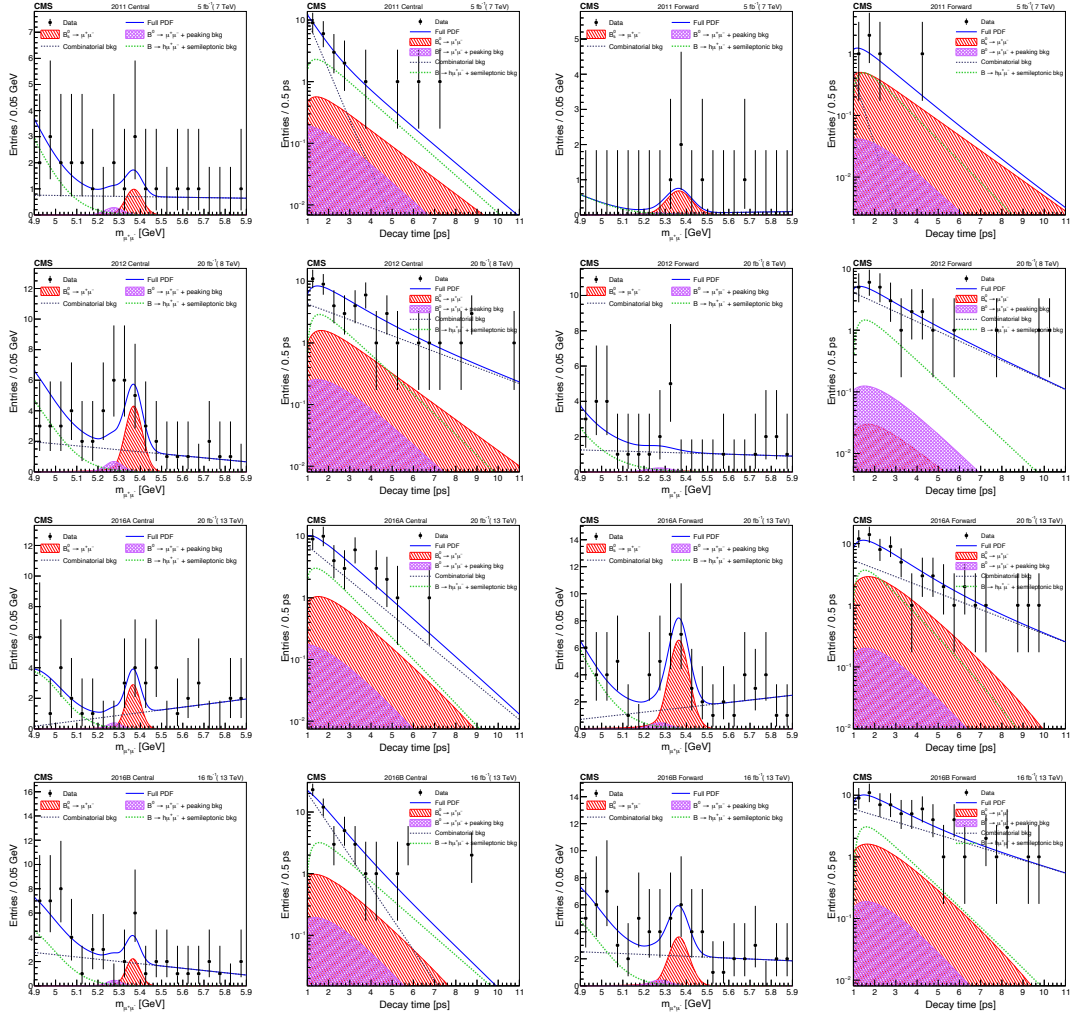


Figure 4.29: Fit projection on the dimuon mass and decaytime distributions for 2011 (first column from the left), 2012 (second column from the left), 2016BF (second column from the right), and 2016GH (first column from the right). The top two rows show channel 0, the bottom two rows show channel 1. The slope of the Bernstein polynomial for 2011 chan1 is fixed to zero because of the low statistics.

4.13 Summary

The measurement of $B_s^0 \rightarrow \mu^+ \mu^-$ properties and search for $B^0 \rightarrow \mu^+ \mu^-$ decay performed by the CMS experiment have been presented. The data used in the analysis were collected in 2011, 2012, and 2016 and correspond to an integrated luminosity of 5 fb^{-1} (7 TeV), 20 fb^{-1} (8 TeV), and 36 fb^{-1} (13 TeV), respectively. The $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction measured from the simultaneous fit to 14 BDT categories is $(2.9 \pm 0.7(\text{exp}) \pm 0.2(\text{frag})) \times 10^{-9}$, where the first uncertainty is the combination of statistical and systematic uncertainty, and the second is due to the fragmentation fraction. The observed significance for $B_s^0 \rightarrow \mu^+ \mu^-$ is 5.6 standard deviations. This is the first observation of $B_s^0 \rightarrow \mu^+ \mu^-$ by the CMS experiment. There is no evidence of $B^0 \rightarrow \mu^+ \mu^-$ decay. Therefore, an upper limit on the branching fraction is assigned, $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 3.1 \times 10^{-10}$, at 90% confidence level. The B_s^0 meson effective lifetime in $B_s^0 \rightarrow \mu^+ \mu^-$ decay mode, for the first time by the CMS experiment, is $\tau(B_s^0 \rightarrow \mu^+ \mu^-) = 1.70_{-0.44}^{+0.61} \text{ ps}$, where the uncertainty includes both statistical and systematic contributions. All the measurements are consistent with the theoretical predictions and other experimental results.

Chapter 5

Run 2 data analysis

5.1 Introduction

The discovery of $B_s^0 \rightarrow \mu^+ \mu^-$ decay has paved the way for more precise measurements of properties to observe the effect beyond the SM. In this chapter, the rare leptonic decay of B_s^0 and B^0 meson properties are carried out using full Run 2 data. The data were collected by the CMS experiment in 2016, 2017, and 2018 at COM energy of 13 TeV. The total integrated luminosity corresponds to 140 fb^{-1} . Since this analysis is similar to the previous one, this chapter focuses on the new changes and validations. I have contributed significantly to the development and validation of different MC corrections, measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ properties such as the branching fraction and the effective lifetime and associated systematics.

This chapter is organized as follows. Section 5.2 discusses the dataset and triggers used in this analysis. The requirement imposed to select the signal and normalization modes are discussed in Section 5.3. Similar to previous analysis, the dominant backgrounds are reduced by the use of a new multivariate classifier. The classifier is validated on the normalization channel. To make the normalization channel useful, a restrictive phase space is considered. The detail of the study is discussed in Section 5.4. Section 5.5 discusses the procedure to estimate the normalization and control channel yields. The fit model used for the measurement of branching fractions and lifetime are discussed in Sections 5.6 and 5.7, respectively. Different data-driven corrections used on the signal simulation sample are discussed in Section 5.8. The systematic uncertainty associated with the branching fraction and the lifetime are explored in Section 5.9. The expected performance of the branching fraction and effective lifetime are discussed in Sections 5.10 and 5.11. Section 5.12 discusses the treatment of fragmentation fraction in this analysis. Finally, the results are summarized

in Section [5.13](#).

5.2 Dataset and trigger

The dataset used in this analysis were collected in 2016, 2017, and 2018. But, all these years, the performance of the detector was not stable. Several issues related to different sub-detector were reported. In early 2016, the APV chip used in the silicon microstrips detector had some operational instability, leading to lower tracker efficiency. This was fixed in the later part of the year. The pixel detector was replaced at the beginning of 2017, and because of it, high acceptance and better resolution were achieved. At the end of 2017, many pixel detectors had DC-DC converter issues, resulting in almost 11% of the detector being unresponsive. The DC-DC converters were replaced in early 2018, and later, the reason behind the failure was understood, and a workaround was developed. Different sub-detectors were calibrated to resolve most of the issues, and the data and MC samples were reprocessed, commonly known as ultralegacy data.

The dataset, called Charmonium, used in this analysis is split into 2016BF, 2016GH, 2017, and 2018 with corresponding luminosity 20.0, 16.6, 42.0, and 61.3 fb^{-1} , respectively. To select good quality data, muon certification is chosen, in which only the tracker and muon chamber performance is checked.

The trigger used to collect the signal decay is HLT_DoubleMu4_3_Bs, whereas the normalization decay mode uses HLT_DoubleMu4_3_Jpsi_Displaced for the years 2016 and HLT_DoubleMu4_3_Jpsi for the year 2017, and 2018. The details of the trigger were discussed in Table 3.2.

5.3 Selection requirement

The selections between signal and control channel are kept as similar as possible among all candidates to be able to benefit from the cancelation of systematic effects in ratios of different observables. In order to reject muon fakes originating from the decay in the flight of pions and kaons, the same soft muon BDT identification is employed, which was discussed previously in Section 3.5. After a thorough analysis of the muon fakes, the muon BDT criterion is lowered from 0.58 to 0.45 (to be discussed in Appendix A). Similarly, a multivariate classifier output is used to increase the signal to background separation, referred to as d_{MVA} . The classifier is built using the kinematic and topological variables. The definitions of the variables are discussed in the next section.

5.3.1 Vertexing variables

The analysis has been developed by using the previous analysis as a baseline. Many of the variable definitions are similar. One major difference is the selection of primary vertex. The B decay vertex for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$, and $B_s^0 \rightarrow J/\psi \phi$ is determined with a kinematic vertex fit using two, three, and four daughter particle tracks, respectively. A refitted transient track representing the B candidate is constructed from the candidate's decay vertex and its momentum. For each primary vertex, in 3D space, the track is then extrapolated to the closest point to that vertex. The distance between the closest point and the primary vertex is named the absolute impact parameter 3D (δ_{3D}). The best primary vertex, at which the B candidate is produced, is chosen based on the smallest δ_{3D} . The variables associated with candidate vertex, α_{3D} , $\delta_{3D}/\sigma(\delta_{3D})$, δ_z , $\delta_z/\sigma(\delta_z)$, ℓ_{3D} , $\ell_{3D}/\sigma(\ell_{3D})$, I and I_μ are then computed. Since the analysis is developed using many ideas from the previous analysis, the above variable definitions are similar to previous analysis.

In addition to the isolation variables, a few more variables are computed which are

designed to suppress partial and combinatorial backgrounds.

- Number of tracks compatible with the $\mu\mu$ vertex based on $d_{\text{ca}} < 300 \mu\text{m}$ and smaller impact parameter significance with respect to $\mu\mu$ vertex than with PV
- Maximum vertexing probability of a candidate muon with a charged track for tracks with $p_T > 1.0 \text{ GeV}$ and $p_T > 2.0 \text{ GeV}$

5.3.2 Offline selection requirement

The previous analysis used a tighter preselection of the variables, and to gain the signal efficiency, the cut has been loosened. Table 5.1 shows a summary of B candidate selection used as input to the analysis d_{MVA} and the ML fitter. Furthermore, the final result is measured after applying the d_{MVA} selection, which will be discussed in Section 5.6.

Selection	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^+ \rightarrow J/\psi K^+$	$B_s^0 \rightarrow J/\psi \phi$
B candidate mass [GeV]	[4.90, 5.90]	[4.90, 5.90]	[4.90, 5.90]
Blinding window [GeV]	[5.15, 5.50]		
$p_{T\mu}$ [GeV]	> 4	> 4	> 4
$ \eta_\mu $	< 1.4	< 1.4	< 1.4
3D SV displacement significance	> 6	> 4	> 4
$p_{T\mu\mu}$ [GeV]	> 5	> 7	> 7
$\mu\mu$ SV probability	> 0.025	> 0.1	> 0.1
$\mu\mu$ invariant mass [GeV]		[2.9, 3.3]	[2.9, 3.3]
Kaon p_T [GeV]		> 1	> 1
Mass-constrained vertex fit probability		> 0.025	> 0.025
2D $\mu\mu$ pointing angle [rad]		< 0.4	< 0.4
ϕ candidate mass [GeV]			[1.01, 1.03]

Table 5.1: Selection summary without analysis d_{MVA} .

The 2D $\mu\mu$ pointing angle is defined as the angle between the dimuon momentum and dimuon flight direction. The angle is calculated in the transverse plane with respect to the beam spot. The J/ψ trigger has a cut on the pointing angle, $\cos(\alpha) > 0.9$ at the online level. In order to match the same requirement, the 2D $\mu\mu$ pointing angle cut is introduced in the offline selection.

5.4 Background suppression with multivariate analysis

The most challenging and dominant backgrounds in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, which limits the analysis sensitivity, are the combinatorial and partially reconstructed backgrounds. The background distribution is compared with the signal distribution in different variables, and based on that, the discriminating variables are selected. Most of the discriminating variables have relatively weak discrimination power on their own. Therefore a multivariate analysis to combine their inputs into a single more powerful discriminator is formed. Here the XGBoost library [92] package is used, which implements an advanced gradient boosting algorithm. After a few iterations, the final set of variables is obtained, and the score from the final training is shown in Table 5.2. A high score signifies the increased importance of the variable in training. The d_{MVA} is trained using the signal $B_s^0 \rightarrow \mu^+ \mu^-$

Feature	Score
Pointing angle of the B candidate with respect to the primary vertex in 3D	1364.1
Pointing angle of the B candidate with respect to the beam spot in the transverse plane	221.4
3D impact parameter of the B candidate with respect to the primary vertex	40.4
Significance of the 3D impact parameter of the B candidate with respect to the primary vertex	71.0
B candidate isolation	211.3
Leading muon isolation	30.4
Trailing muon isolation	27.1
Flight length significance of the B candidate	38.0
$\mu\mu$ vertex χ^2/dof	13.1
Number of tracks compatible with the $\mu\mu$ vertex using tracks with $d_{\text{ca}} < 300\mu$ and smaller impact parameter significance with respect to $\mu\mu$ vertex than with PV	10.8
Maximum vertexing probability of a candidate muon with a charged track for tracks with $p_T > 1.0\text{GeV}$	7.2
Maximum vertexing probability of a candidate muon with a charged track for tracks with $p_T > 2.0\text{GeV}$	18.4

Table 5.2: $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ d_{MVA} input variables and their importance scores.

MC and sideband background events. The data sideband consists of a low sideband region ($m \in [4.9, 5.1]$ GeV), which has both a partially reconstructed background and combi-

natorial background, and a high sideband region ($m \in [5.5, 5.9]$ GeV) containing only combinatorial background. To avoid any possible bias, the events are split into three categories using the event number modulo 3. Since the observed number of sideband events is small, the training and testing event ratio is in a 2:1 proportion. Figure 5.1 shows ROC curves and the d_{MVA} distributions for simulated signal $B_s^0 \rightarrow \mu^+ \mu^-$ and data background events. The ROC curve shows the effectiveness of the d_{MVA} at distinguishing the signal and background classes. In this case, it represents the signal efficiency against the background rejection. There is some mild degree of over-training resulting from low statistics. The final optimized d_{MVA} working points for branching fraction and lifetime will be discussed in Section 5.6.3.

A comparison between the previous analysis of BDT and this new d_{MVA} shows that the signal efficiency increased by a factor of 2 for the same background rejection. This is one of the reasons to reanalyze the 2016 dataset in this analysis.

5.4.1 Validation of MVA using $B^+ \rightarrow J/\psi K^+$ decay

$B^+ \rightarrow J/\psi K^+$ decays are more favorable than $B_s^0 \rightarrow J/\psi \phi$ for all validation processes because there is one less track in the final state. But there is one caveat in the $B^+ \rightarrow J/\psi K^+$ decay mode. It has an additional charge track in the final state which makes the $B \rightarrow \mu^+ \mu^-$ d_{MVA} classifier like the background as shown in Fig. 5.2 (left).

This decay channel will be only useful if one can find suitable alternative variables to train the d_{MVA} . Let's discuss the differences between the two decay modes. First, the dimuon mass in $B^+ \rightarrow J/\psi K^+$ decay mode is smaller compared to the $B_s^0 \rightarrow \mu^+ \mu^-$ (i.e., 3.1 vs. 5.37 GeV). This mass difference significantly impacts the opening angle of the two muons. Again the smaller opening angle leads to sizeable uncertainty on the dimuon vertex. Since the flight length significance variable uses the vertex position, the ratio of mass (1.6) can be used as a scale factor to match the $B^+ \rightarrow J/\psi K^+$ distribution with

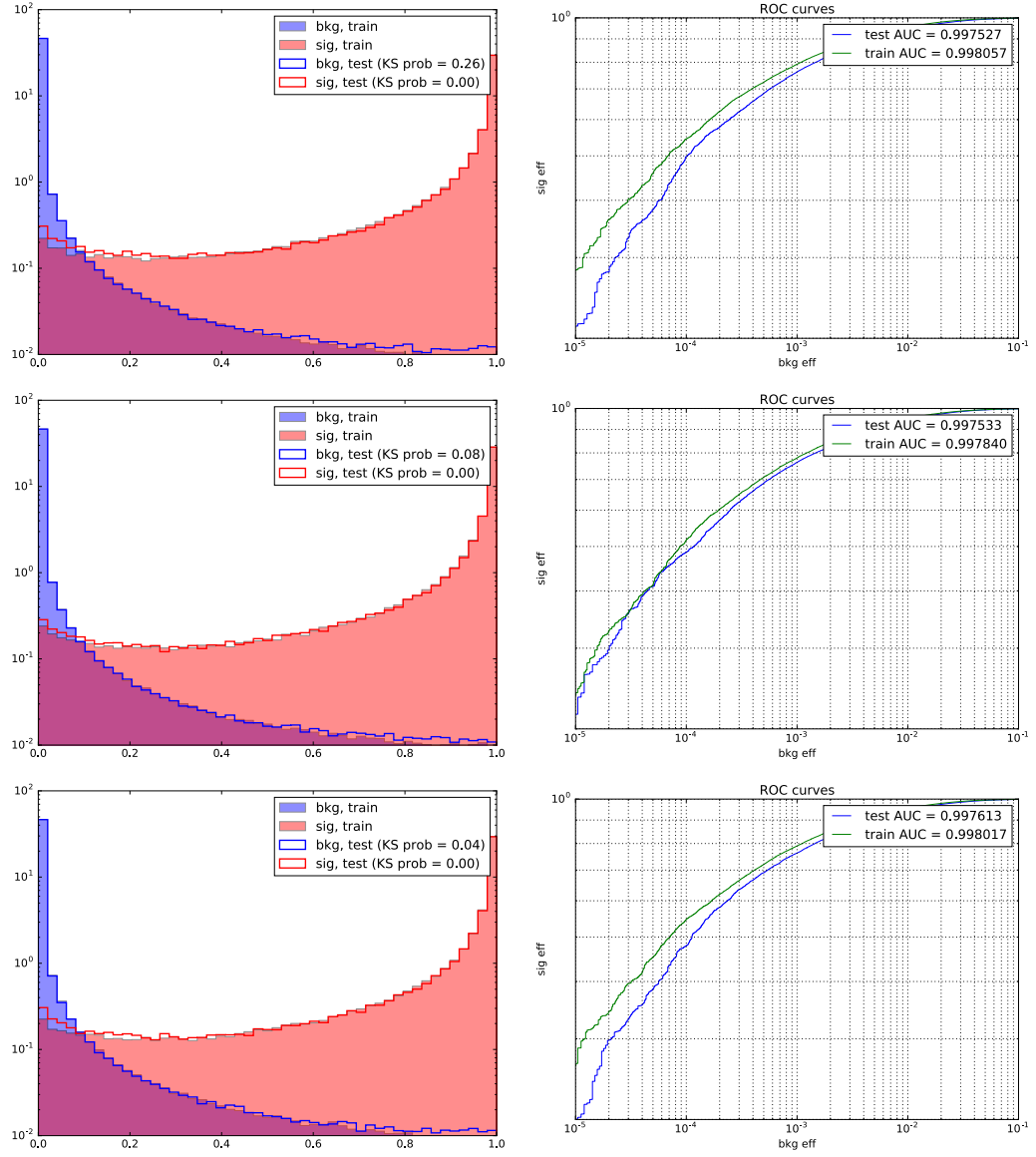


Figure 5.1: d_{MVA} distribution for $B_s^0 \rightarrow \mu^+\mu^-$ MC simulation and data sideband background (left) and ROC curves (right) for three different event splittings.

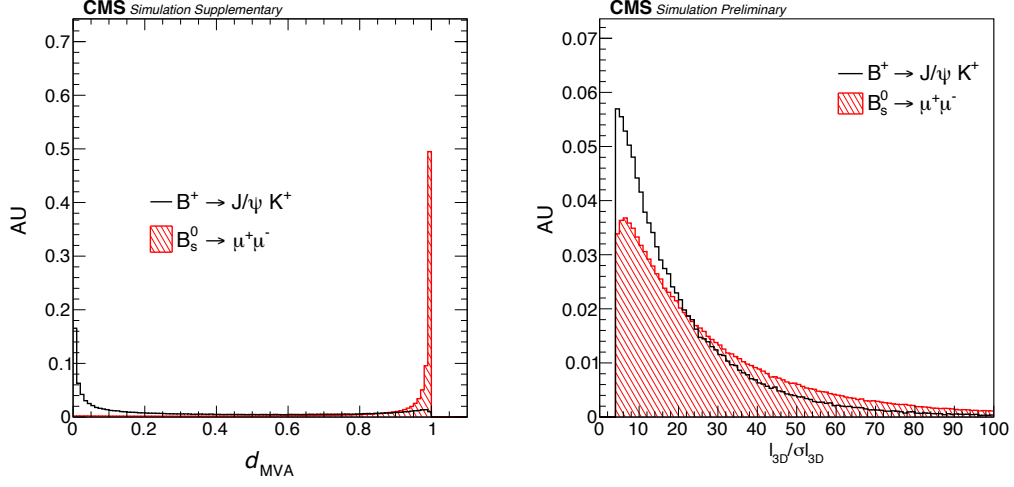


Figure 5.2: d_{MVA} and flight length significance distributions for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ events.

$B_s^0 \rightarrow \mu^+ \mu^-$. Even though the exact mass ratio is 1.7, 1.6 works perfectly fine. The flight length significance distribution with and without the scale factor is shown in Figs. 5.2 and 5.4, respectively.

The presence of kaon in the final state of $B^+ \rightarrow J/\psi K^+$ decay is one of the major differences between the decay modes. It is observed that the choice of kaon momentum has a significant impact on the various observables. For example, the comparison of pointing angle distribution, one of the most discriminating variables, for several different kaon p_T ranges is shown in Fig. 5.3. As it can be seen, the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ distribution agree well for kaon $p_T < 1.5$ GeV. For tighter cuts, the difference is quite significant.

The final inputs for the $B^+ \rightarrow J/\psi K^+$ d_{MVA} to make similar $B_s^0 \rightarrow \mu^+ \mu^-$ d_{MVA} , are summarized in Table 5.3. The isolation and track counting variables in $B^+ \rightarrow J/\psi K^+$ decay mode do not include the kaon track in order to match with $B_s^0 \rightarrow \mu^+ \mu^-$. Figure 5.4 shows the comparison of few variables after the new selection and modified d_{MVA} inputs. The agreement is reasonably close.

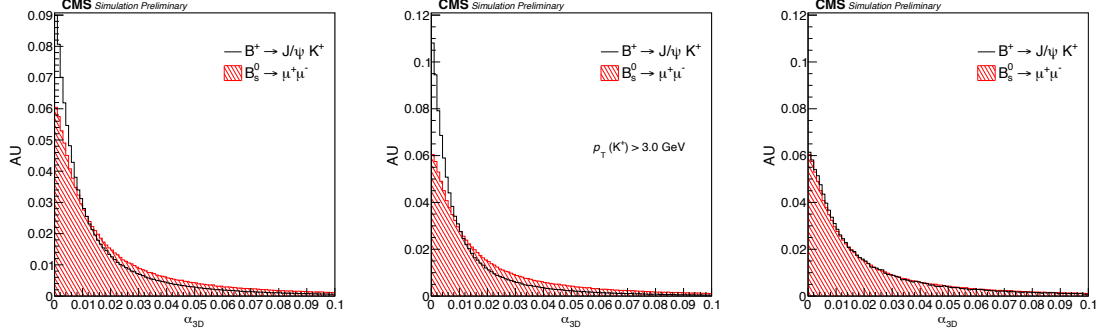


Figure 5.3: Point angle distributions for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ events in MC. All events (left), kaon $p_T > 3.0$ GeV (center), kaon $p_T < 1.5$ GeV (right) are shown. Pointing angle of $B^+ \rightarrow J/\psi K^+$ is measured using B^+ momentum.

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ variable	$B^+ \rightarrow J/\psi K^+$ variable
Pointing angle 3D ($\mu\mu$)	Pointing angle 3D ($\mu\mu K$)
Pointing angle 2D ($\mu\mu$)	Pointing angle 2D ($\mu\mu K$)
Impact parameter 3D ($\mu\mu$)	Impact parameter 3D ($\mu\mu K$)
Impact parameter significance ($\mu\mu$)	Impact parameter significance ($\mu\mu K$)
Flight length significance ($\mu\mu$)	Flight length significance ($\mu\mu$)
$\mu\mu$ vertex χ^2/dof	$\mu\mu$ vertex χ^2/dof
B candidate isolation	B candidate isolation excluding kaon
Muon isolation	Muon isolation excluding kaon
Number of extra tracks	Number of extra tracks excluding kaon
Vertexing with other tracks	Vertexing with other tracks excluding kaon

Table 5.3: d_{MVA} input variables used to match $B^+ \rightarrow J/\psi K^+$ to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$.

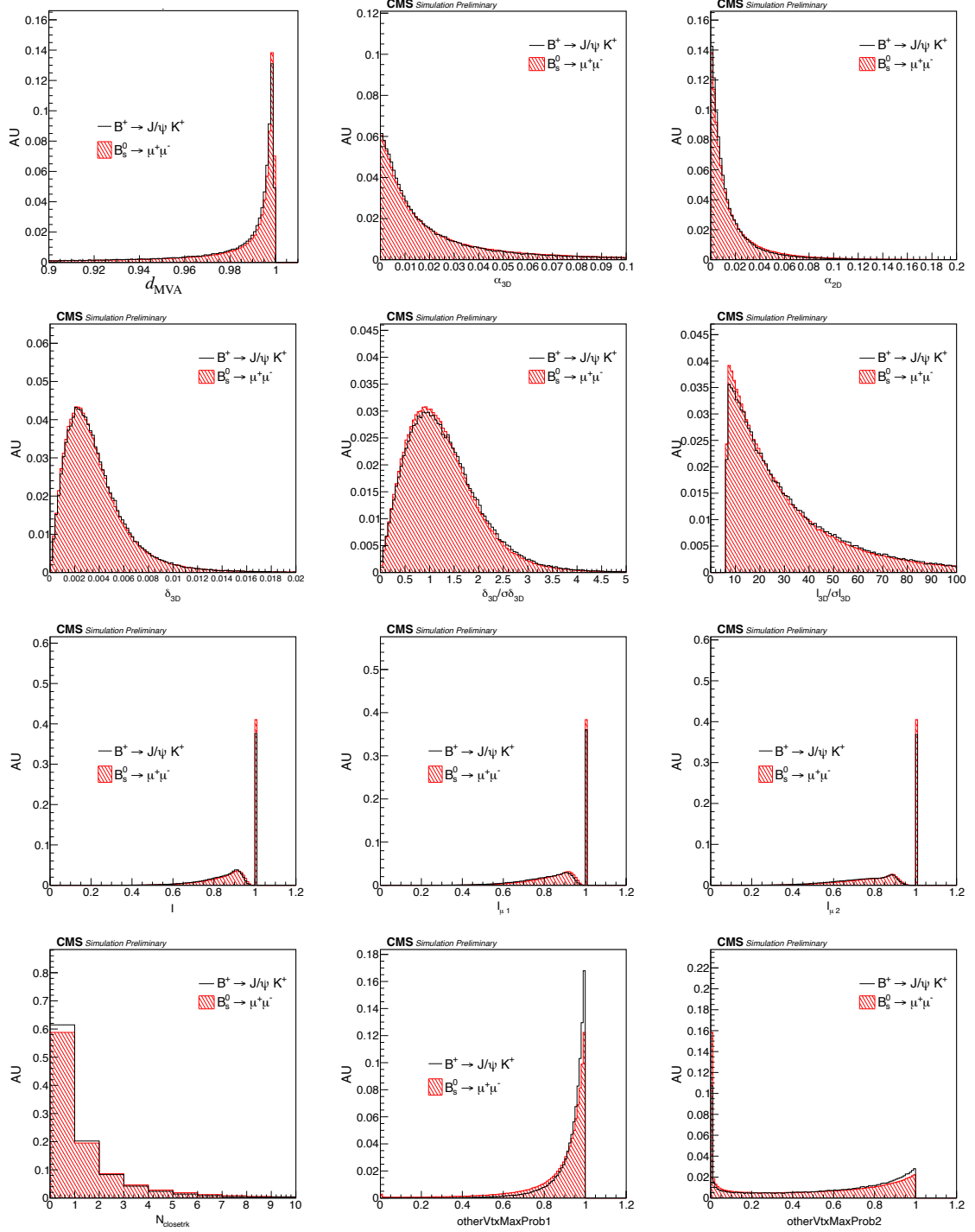


Figure 5.4: Comparison of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ MC distribution for different variables.

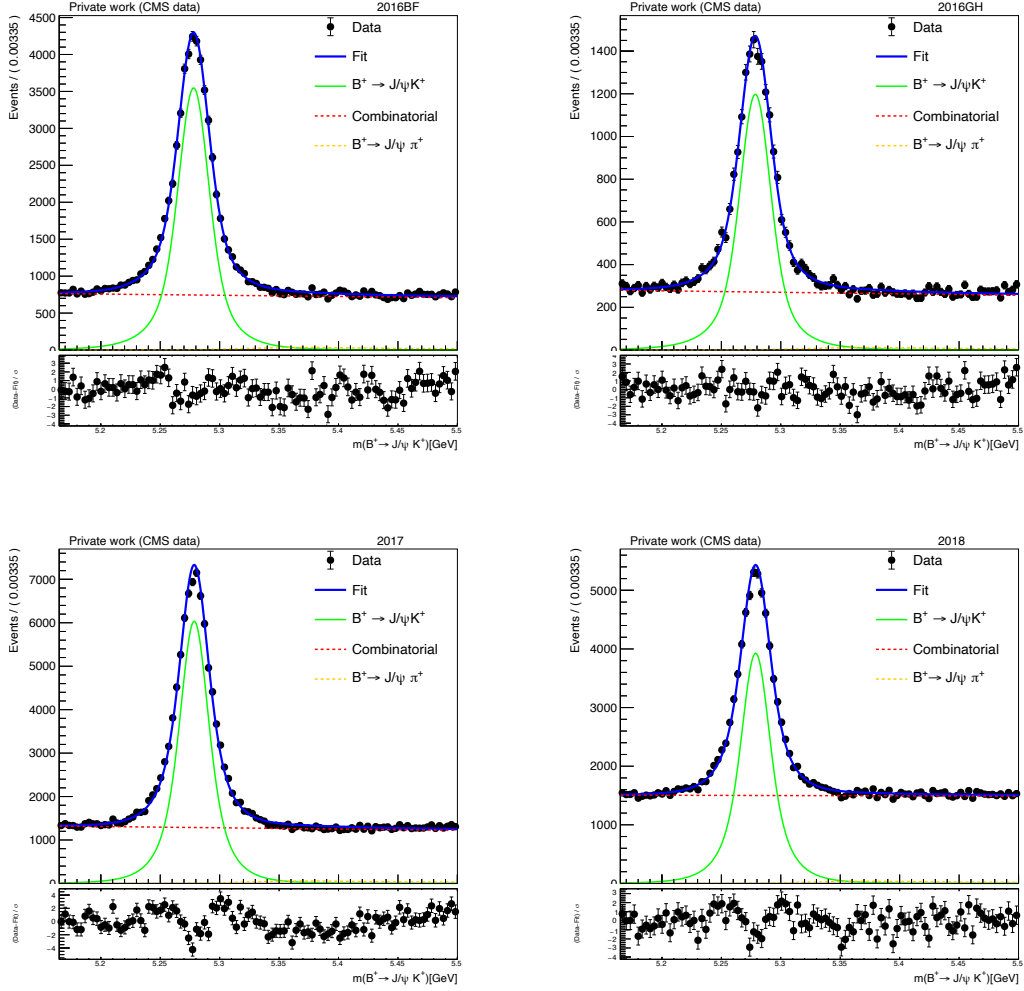


Figure 5.5: Invariant mass distributions along with UML fit projections for $B^+ \rightarrow J/\psi K^+$ events in 2016BF, 2016GH, 2017, and 2018 data. The result of the fit is overlaid (blue curve) and the different components detailed in the legend.

5.4.2 MVA validation in data with $B^+ \rightarrow J/\psi K^+$ events

Various observables in $B^+ \rightarrow J/\psi K^+$ between data and MC are compared because if there is any discrepancy between the data and MC, the correction can be calculated in the $B^+ \rightarrow J/\psi K^+$ decay mode, and it will then be applied on $B_s^0 \rightarrow \mu^+ \mu^-$ MC. To compare the data distribution with MC, the background distribution in the data needs to be subtracted. The best and easy solution is to use the sPlot technique to estimate the sWeight for each event by performing an unbinned maximum likelihood fit to the $B^+ \rightarrow J/\psi K^+$ invariant mass distribution. The signal component is modeled with double Crystal-Ball function, the peaking background is modeled with a triple Gaussian function, and the combinatorial background is modeled with a Bernstein polynomial function. Figure 5.5 shows the invariant mass UML fit and the Fig. 5.6 shows the d_{MVA} distributions and a few most problematic observables for all years.

The data and MC d_{MVA} distributions are closely aligned in the 2016BF and 2016GH samples, but in 2017 and 2018 samples, there is a significant discrepancy between the two. The discrepancies in d_{MVA} are investigated by looking at input variables to the d_{MVA} classifier. We observe variables like impact parameter 3D, and its significance shows an apparent disagreement in 2017 and 2018. A correction is estimated based on the discrepancy and applied to the MC to mitigate this issue. A more detailed study will be discussed in Section 5.8.2.

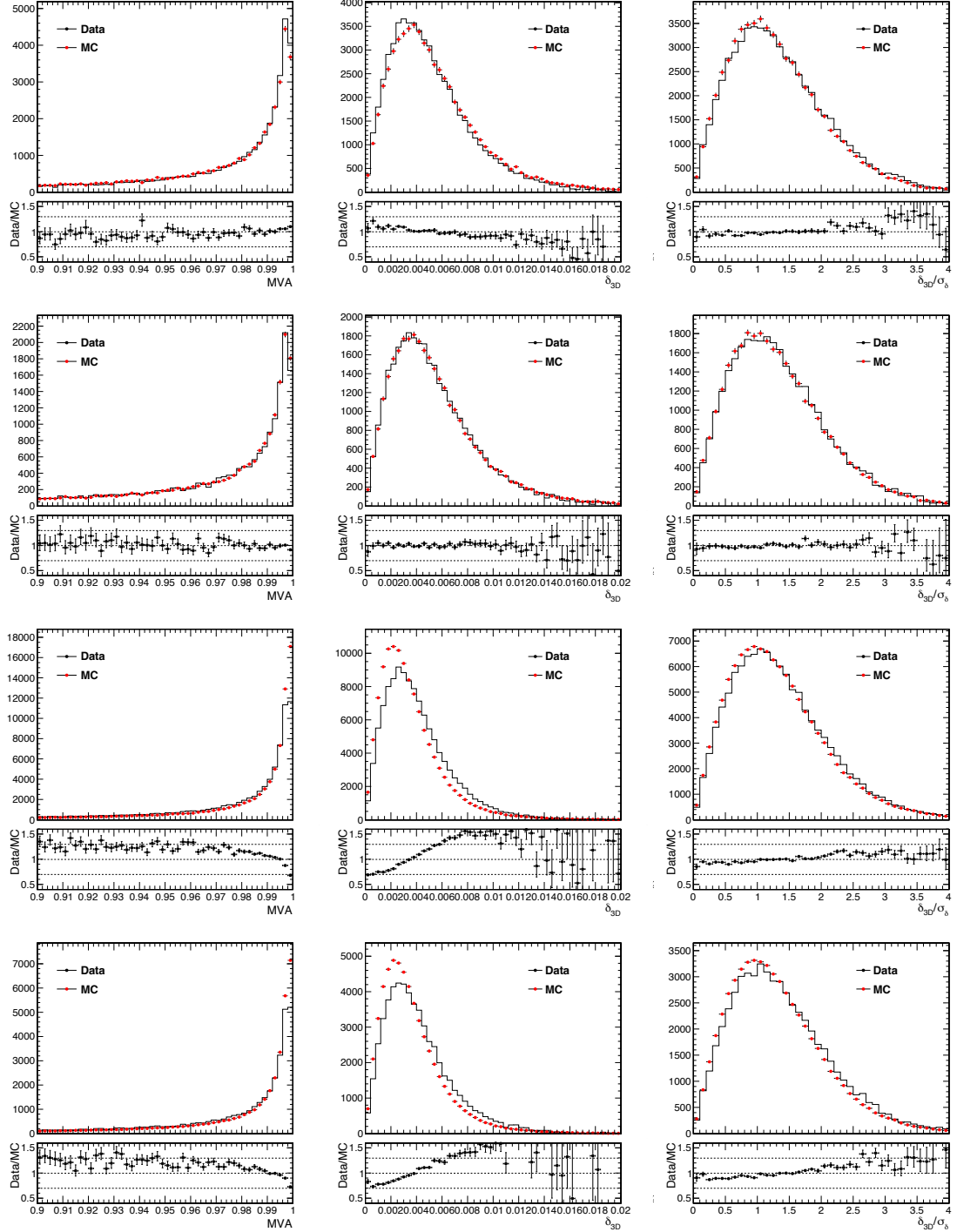


Figure 5.6: Distributions of the d_{MVA} output (left), the impact parameter with respect to PV in 3D (center) and its significance (right) in 2016BF, 2016GH, 2017, and 2018 data samples (from top to bottom).

5.5 Normalization and control channel yields

The subject of this section is the measurement of the normalization and control sample yields. As done in the previous chapter, the normalization channel and control channel invariant mass fit is performed after considering the trigger prescale. The prescale is used as the candidate's weight, and then a weighted extended UML fit is performed to determine the overall yield. Here we decide not to use the d_{MVA} selection, and the reason will be discussed later.

The PDF used to fit the data distribution is described below

$$\mathcal{L} = N_{B^+/B_s^0} P_{B^+/B_s^0}(m) + N_{\text{peak}} P_{\text{peak}}(m) + N_{\text{comb}} P_{\text{comb}}(m) \quad (5.1)$$

where N_i corresponds to the yield for B^+ or B_s^0 , peaking background and combinatorial background. The mass PDF for B^+ or B_s^0 is,

$$P_{B^+/B_s^0}(m) = [f \cdot \text{CB} + (1 - f) [f1 \cdot G1 + (1 - f1) \cdot G2]] \quad (5.2)$$

which is a sum of double Gaussian ($G1, G2$) and a Crystal-Ball (CB) function having a common mean. To achieve stable fits, the MC distributions are fitted using the PDF described in Eq. 5.2. Then the parameters of CB-line and double Gaussian functions are fixed in the data fit except for the μ , and σ . The peaking background in the $B^+ \rightarrow J/\psi K^+$ decay mode, $B^+ \rightarrow J/\psi \pi^+$, is modeled with triple Gaussian function, and in $B_s^0 \rightarrow J/\psi \phi$ decay mode, $B^0 \rightarrow J/\psi K^*$, is modeled with a kernel estimator. The shape of the peaking background distribution is fixed in the data fit, and the normalization is fixed to the ratio of peaking and signal branching fractions.

The PDF for the combinatorial background is modeled with a Bernstein polynomial. The floated parameters in the data fit are $\mu, \sigma, N_{B^+/B_s^0}, N_{\text{comb}}$ and the slope of the Bernstein polynomial.

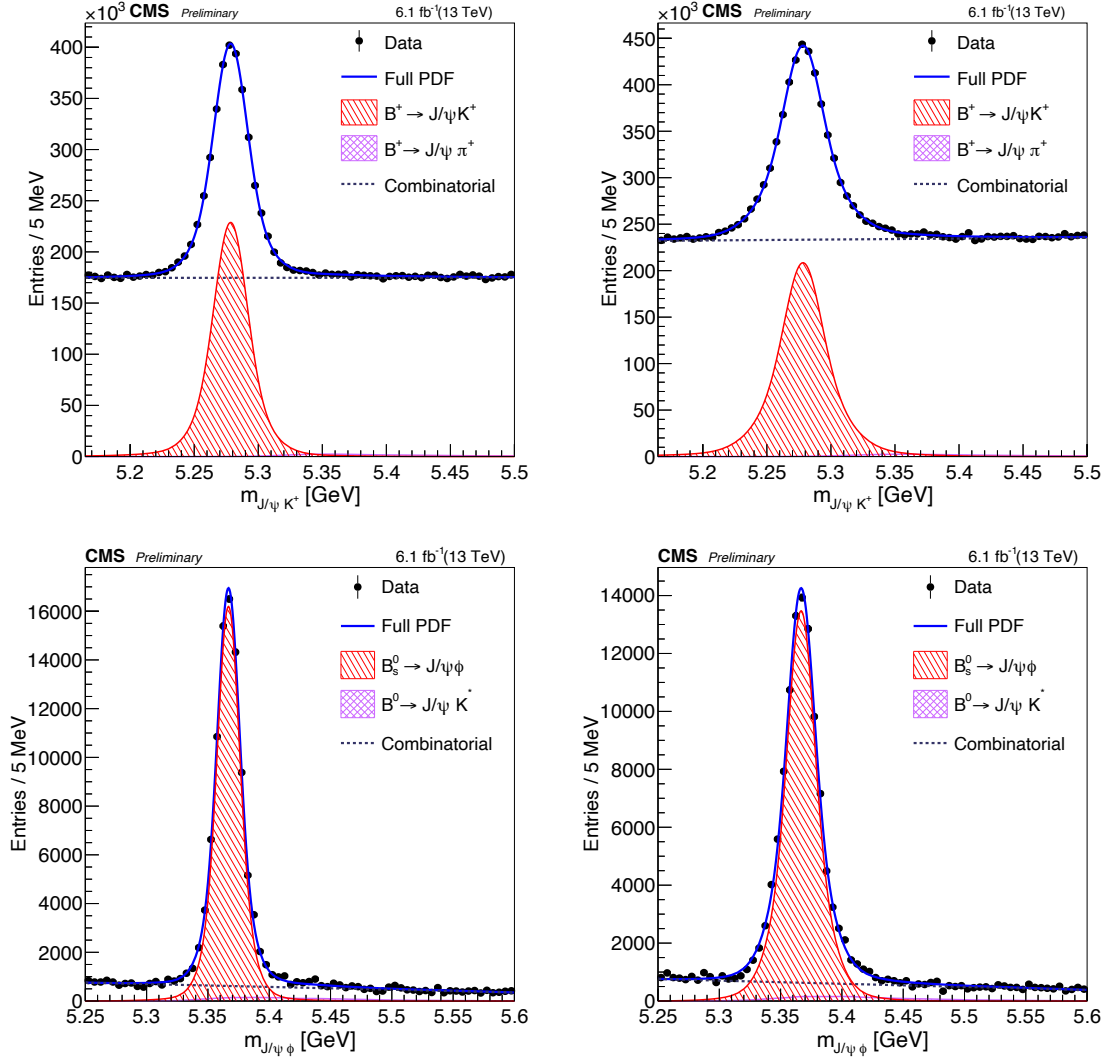


Figure 5.7: The $B^+ \rightarrow J/\psi K^+$ (top row) and $B_s^0 \rightarrow J/\psi \phi$ (bottom row) invariant mass distributions from 2018 channel 0 (left column) and channel 1 (right column) are shown, along with the fit projections.

The fits to the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidate mass distributions in data for 2018 are illustrated in Fig. 5.7. The systematic uncertainty of the yield is estimated by comparing the yield obtained from the use of different signal and parametric background models. The fit results for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ are shown in Tables 5.4 and 5.5. Based on the results, 1% and 1.5% systematic uncertainty on the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ yield are assigned.

Period	$B^+ \rightarrow J/\psi K^+$ yield			
	Channel 0		Channel 1	
	Nominal	Alternative	Nominal	Alternative
2016BF	$299,881 \pm 1,185$	$299,202 \pm 1,057$	$439,299 \pm 1,838$	$440,086 \pm 1,632$
2016GH	$378,825 \pm 1,324$	$378,115 \pm 1,431$	$476,369 \pm 1,987$	$478,703 \pm 1,810$
2017	$1,084,880 \pm 2,151$	$1,082,589 \pm 2,407$	$1,233,267 \pm 3,005$	$1,224,996 \pm 3,576$
2018	$1,764,249 \pm 2,779$	$1,748,627 \pm 3,213$	$2,093,494 \pm 4,029$	$2,069,176 \pm 3,576$

Table 5.4: $B^+ \rightarrow J/\psi K^+$ event yield weighted by the trigger prescale for the nominal and alternative fits.

Period	$B_s^0 \rightarrow J/\psi \phi$ yield			
	Channel 0		Channel 1	
	Nominal	Alternative	Nominal	Alternative
2016BF	$14,242 \pm 129$	$14,218 \pm 128$	$19,345 \pm 157$	$19,485 \pm 154$
2016GH	$17,685 \pm 145$	$17,623 \pm 143$	$21,716 \pm 165$	$21,599 \pm 162$
2017	$52,068 \pm 248$	$51,306 \pm 245$	$56,125 \pm 268$	$55,227 \pm 261$
2018	$85,665 \pm 325$	$85,282 \pm 316$	$96,148 \pm 349$	$94,690 \pm 343$

Table 5.5: $B_s^0 \rightarrow J/\psi \phi$ event yield weighted by the trigger prescale for the nominal and alternative fits.

5.6 Branching fraction measurement with UML fit

One of the main observables of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis is the branching fraction of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decay mode, which is obtained from an UML fit to the data. The UML fit setup is identical to the analysis's previous iteration with minor changes. A simultaneous fit to the dimuon invariant mass ($m_{\mu^+ \mu^-}$) and the relative mass resolution ($\sigma(m_{\mu^+ \mu^-})/m_{\mu^+ \mu^-}$) distributions with multiple event categories is performed. The events are categorized as follows:

- Data taking period: 2016BF, 2016GH, 2017 and 2018
- Signal purity based on the analysis d_{MVA} : [0.90, 0.99], [0.99, 1.00]
- η of the most forward muon: [0.0, 0.7] (Channel 0), [0.7, 1.4] (Channel 1)

In the data fit, the likelihood to describe the event has five components: B_s^0 signal, B^0 signal, partial-reconstructed semileptonic 3-body $B \rightarrow h\mu\nu$ and $B \rightarrow h\mu\mu$ background, peaking $B \rightarrow h^+ h^-$ background, and the combinatorial background.

The complete model introduced for each signal and background component is the product of the PDF for each dimension:

$$P(m_{\mu^+ \mu^-}, \sigma(m_{\mu^+ \mu^-})) = P(m_{\mu^+ \mu^-}; \sigma(m_{\mu^+ \mu^-})) \times P(\sigma(m_{\mu^+ \mu^-})/m_{\mu^+ \mu^-}), \quad (5.3)$$

To extract signal yields the fitter uses the extended likelihood:

$$\mathcal{L} = e^{-(N_{\text{bs}} + N_{\text{bd}} + N_{\text{semi}} + N_{\text{peak}} + N_{\text{comb}})} \prod_{\text{events}} \left[N_{\text{bs}} P_{\text{bs}}(x) + N_{\text{bd}} P_{\text{bd}}(x) + N_{\text{semi}} P_{\text{semi}}(x) + N_{\text{peak}} P_{\text{peak}}(x) + N_{\text{comb}} P_{\text{comb}}(x) \right] \quad (5.4)$$

The likelihood model for each of the component is described in Table [5.6](#). The shape of each component is derived from dedicated MC sample. In the final fit, two Gaussians are used

Component	Mass	Width	Mass resolution
Signal	CB	cond., $\sigma_{CB} = \kappa \times \sigma_{m_{\mu\mu}}$	Keys
background, hh	CB+G	fixed to MC	Keys
background, $h\mu\mu$, $h\mu\nu$	G	floated	Keys
Combinatorial background	Bernstein pol	N/A	Keys

Table 5.6: Models used to describe the signal background mass and mass resolution. Here the CB, G and Keys stands for Crystal-Ball, Gaussian and kernel estimator function. The width parameter (σ_{CB}) of the CB-line is a conditional parameter linearly depending on the dimuon mass resolution. Here κ is the multiplicative factor.

to model the semileptonic background, where one Gaussian is used for the low purity bin ($d_{MVA} \in [0.90, 0.99]$) correlated over the different eras and channels and another Gaussian for the high purity bin ($d_{MVA} \in [0.99, 1.0]$). The mean and sigma of the Gaussians are floated. All parameters of the CB and G are fixed to MC values unless otherwise stated. Similarly, the parameter (the slope) of the Bernstein polynomial is floated in the fits and the kernel models are obtained from the data sideband events with lower event d_{MVA} scores ($0.6 < d_{MVA} < 0.9$).

5.6.1 Estimation of the signal and background components

The signal yields N_{bs} and N_{bd} in each category are derived from the common branching fractions $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, which are the main parameters of interest across all the categories:

$$N_{bs}^i = \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \times \frac{\varepsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^i}{\varepsilon_{B^+ \rightarrow J/\psi K^+}^i} \times \frac{f_s}{f_u} \times \frac{N_{B^+ \rightarrow J/\psi K^+}^i}{\mathcal{B}_{B^+ \rightarrow J/\psi K^+} \mathcal{B}_{J/\psi \rightarrow \mu^+ \mu^-}} \quad (5.5)$$

$$N_{bd}^i = \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) \times \frac{\varepsilon_{B^0 \rightarrow \mu^+ \mu^-}^i}{\varepsilon_{B^+ \rightarrow J/\psi K^+}^i} \times \frac{N_{B^+ \rightarrow J/\psi K^+}^i}{\mathcal{B}_{B^+ \rightarrow J/\psi K^+} \mathcal{B}_{J/\psi \rightarrow \mu^+ \mu^-}} \quad (5.6)$$

where the superscript i indicates the parameter for the specific category i in the fit; the factors $\varepsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^i$, $\varepsilon_{B^0 \rightarrow \mu^+ \mu^-}^i$ and $\varepsilon_{B^+ \rightarrow J/\psi K^+}^i$ are the total efficiencies (including the effects of detector acceptance) for $B_s^0 \rightarrow \mu^+ \mu^-$, $B^0 \rightarrow \mu^+ \mu^-$, and $B^+ \rightarrow J/\psi K^+$ decays,

respectively. The ratio f_s/f_u corrects the fragmentation rate between B_s^0 and B^+ mesons.

The $B^+ \rightarrow J/\psi K^+$ and $B_{(s)}^0 \rightarrow \mu^+\mu^-$ branching fractions are:

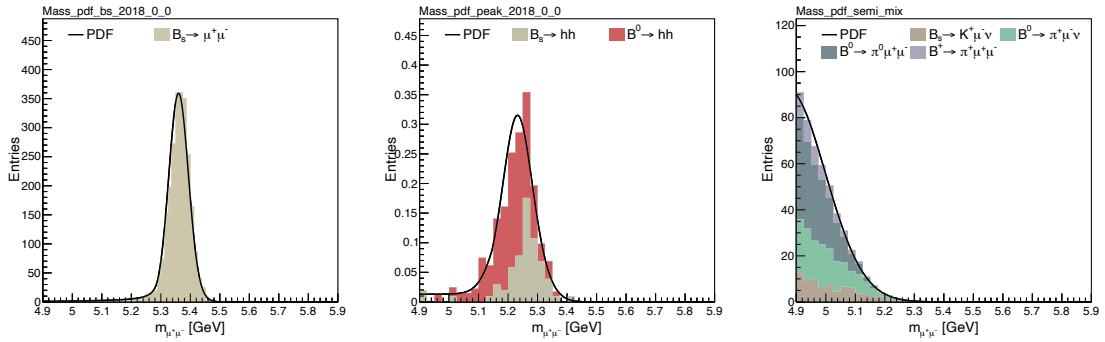
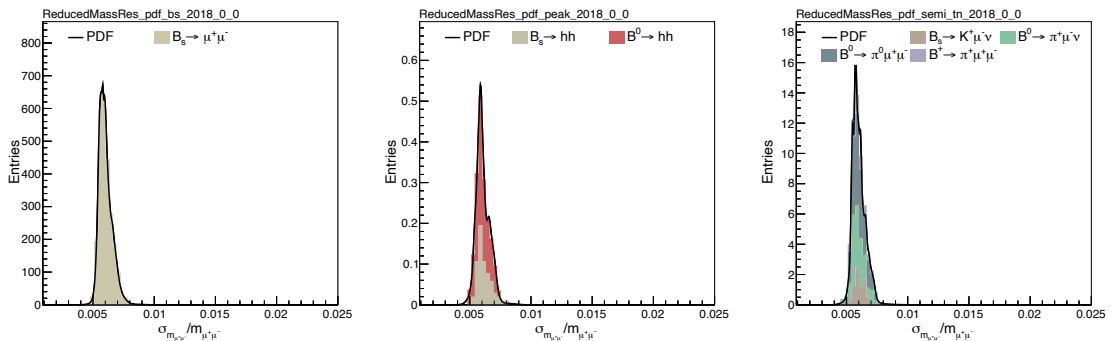
- $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.020 \pm 0.019) \times 10^{-3}$ [94]
- $\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-) = (5.961 \pm 0.033) \times 10^{-2}$ [94]
- $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.66 \pm 0.14) \times 10^{-9}$
- $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (1.03 \pm 0.05) \times 10^{-10}$

The combinatorial background arises from $b\bar{b} \rightarrow \mu^+\mu^-X$ process can not be entirely removed by the d_{MVA} selection. The expected yields in the signal region are obtained from the extrapolation of events from the high mass sideband region $5.5 < m_{\mu\mu} < 5.9$ GeV in each category. In the full range, the combinatorial distribution is assumed to be flat distribution.

The peaking and semileptonic backgrounds are estimated using a dedicated MC simulation. The background yields are normalized to the $B^+ \rightarrow J/\psi K^+$ event yields in the same way as it is done for the signal components. In the case of two body Λ_b decays such as $\Lambda_b \rightarrow p\pi$ and $\Lambda_b \rightarrow pK$ one may expect a distribution similar to $B \rightarrow h^+h^-$ backgrounds. However, after the analysis selection requirements the number of expected events is negligible. The simulated distributions of dimuon invariant mass of the 3-body B decays background share the same PDF between different channels and eras. The branching fraction of two and three-body B decays backgrounds used for the estimation of the background yields are mentioned in Table 5.7.

The simulated distributions of dimuon invariant mass of the signal, peaking and semileptonic components and their PDFs are shown in Fig. 5.8 with the selection soft muon BDT > 0.45 and $d_{\text{MVA}} > 0.99$. The simulated distributions of relative mass resolution and their PDFs are shown in Fig. 5.9, with the same selection.

Processes	Branching fractions
$\mathcal{B}(B^0 \rightarrow K\pi)$	$(1.96 \pm 0.05) \times 10^{-5}$ [94]
$\mathcal{B}(B^0 \rightarrow KK)$	$(7.80 \pm 1.50) \times 10^{-8}$ [94]
$\mathcal{B}(B^0 \rightarrow \pi\pi)$	$(5.12 \pm 0.19) \times 10^{-6}$ [94]
$\mathcal{B}(B_s^0 \rightarrow K\pi)$	$(5.80 \pm 0.70) \times 10^{-6}$ [94]
$\mathcal{B}(B_s^0 \rightarrow KK)$	$(2.66 \pm 0.22) \times 10^{-5}$ [94]
$\mathcal{B}(B_s^0 \rightarrow \pi\pi)$	$(7.00 \pm 1.00) \times 10^{-7}$ [94]
$\mathcal{B}(B_s^0 \rightarrow K^-\mu^+\nu)$	$(1.06 \pm 0.09) \times 10^{-4}$ [96]
$\mathcal{B}(B^0 \rightarrow \pi^-\mu^+\nu)$	$(1.50 \pm 0.06) \times 10^{-5}$ [94]
$\mathcal{B}(B^0 \rightarrow \pi^0\mu^+\mu^-)$	$(6.9 \pm 0.69) \times 10^{-8}$ [94]
$\mathcal{B}(B^+ \rightarrow \pi^+\mu^+\mu^-)$	$(1.78 \pm 0.23) \times 10^{-8}$ [94]

Table 5.7: Branching fractions of charmless two body B decays and 3-body B decays.Figure 5.8: The simulated distribution of dimuon invariant mass of $B_s^0 \rightarrow \mu^+\mu^-$ (left), $B \rightarrow h^+h^-$ (middle) and semileptonic (right) backgrounds and its PDF for 2018 channel 0. All channels and eras share the same semileptonic PDF.Figure 5.9: The simulated distribution of the relative mass resolution of $B_s^0 \rightarrow \mu^+\mu^-$ (left), $B \rightarrow h^+h^-$ (middle) and semileptonic (right) backgrounds and its PDF for 2018 channel 0.

The expected (or the so-called “pre-fit”) yields and parameters are used to construct the fit model, which is used to calculate the expected performance with the toy-based pseudo-experiments. The outcome of the performance will be discussed in Section 5.10.

5.6.2 Nuisance parameters

The main parameters of interest are the branching fractions of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$. The corresponding yields (N_{bs} and N_{bd}) are derived from them. In addition to the signal branching fraction, many nuisance parameters are floated in the fit. Table 5.8 provides a complete list of nuisance parameters with some details about their purpose and implementation. Here the constrained means the uncertainty on the parameter included in the fit as a Gaussian/lognormal constraint. Lognormal constraints are applied to N_{semi} and N_{peak} because the expected numbers are close to zero.

Uncertainty Type	Status in fit	Categories	Comments
BF($B^+ \rightarrow J/\psi K^+$)	Constrained	Global	PDG value
Combined systematics on BF	Constrained	Global	Kaon efficiency systematics
$N(B^+ \rightarrow J/\psi K^+)$, $\varepsilon(B^+ \rightarrow J/\psi K^+)$	Constrained	4 eras x 2 channels (ch)	yield, systematic on efficiency on $B^+ \rightarrow J/\psi K^+$ (stats, J/ψ trigger etc)
$\varepsilon(B_s^0 \rightarrow \mu^+ \mu^-)$, $\varepsilon(B^0 \rightarrow \mu^+ \mu^-)$	Constrained	4 eras x 2 ch x 2 d_{MVA}	MC stat
Combined systematics on $\varepsilon(\mu\mu)$	Constrained	4 eras x 2 ch	$\mu\mu$ trigger efficiency
d_{MVA} scale factor uncertainties	Constrained	3 years x 2 d_{MVA}	
N_{peak} , N_{semi}	Constrained	4 eras x 2 ch x 2 d_{MVA}	Combined systematics (MC stat, BFs, muon fake rate, normalization etc)
N_{comb} , slope	Floated	4 eras x 2 ch	

Table 5.8: Detail list of nuisance parameters and their purpose and implementation.

5.6.3 Optimization of analysis d_{MVA} and muon BDT selection

The analysis sensitivity varies with the choice of analysis d_{MVA} selection. In order to achieve the best sensitivity, a two-dimensional optimization of analysis d_{MVA} and muon BDT is performed. In this optimization, the significance of the B_s^0 and B^0 is measured for each d_{MVA} and muon BDT selection with the Asimov data. The input branching fraction and lifetime are set to be the SM expectation, and the expected yields for the different components are based on the MC simulation. The fit model is described in the above section.

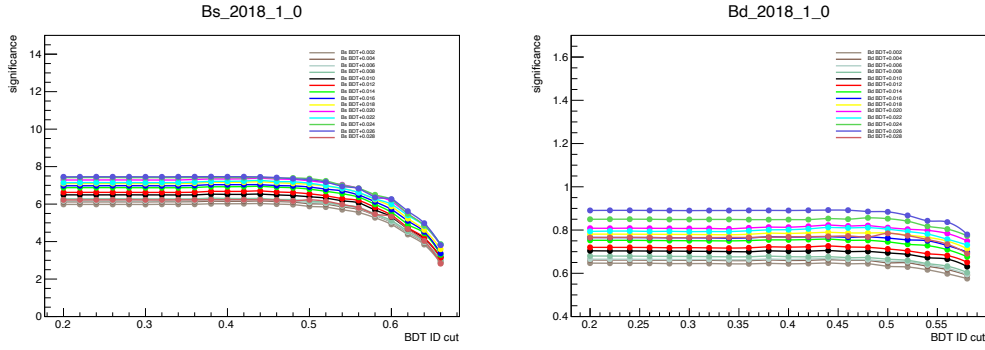


Figure 5.10: Significance of B_s^0 (left) and B^0 (right) in terms of different analysis d_{MVA} and muon BDT selections in 2018 channel 0. Other years and channel plots have similar trends which are not shown here. Here X-axis represents the muon BDT requirement and different labels represents the analysis d_{MVA} requirement with benchmark $d_{\text{MVA}} > 0.97 + X$ (where X is mentioned in the legend.)

The scan result, in Fig. 5.10, indicates a looser muon BDT selection requirement compared to the old value (0.58) to gain the sensitivity. We can not use a very low muon BDT value which will give more fake muons (more details in Appendix A). To be in the safe side, muon BDT > 0.45 is chosen for the final analysis. Figure 5.11 shows the significance of B_s^0 and B^0 for different d_{MVA} selections. The tight d_{MVA} selection should be avoided because of low statistics and bad mismodeling. In the end, the final selection is chosen to be $d_{\text{MVA}} > 0.99$ to simplify the systematic evaluation.

Another test is performed to improve the sensitivity, where the analysis d_{MVA} is sepa-

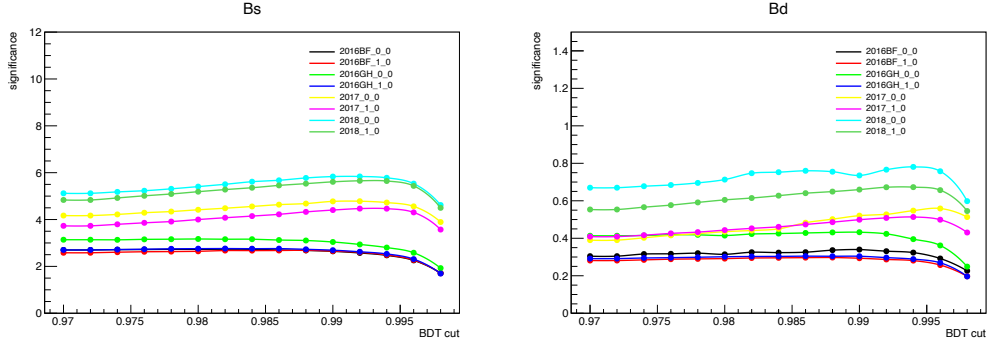


Figure 5.11: Significance of B_s^0 (left) and B^0 (right) in terms of different analysis d_{MVA} in different years and channels.

rated in to two categories (“2-bin”) representing events with low d_{MVA} [$d_{\text{MVA}}^{\text{low}}$, $d_{\text{MVA}}^{\text{high}}$] scores and high d_{MVA} [$d_{\text{MVA}}^{\text{high}}$, 1.0] scores. We find the choice of different $d_{\text{MVA}}^{\text{low}}$ points do not affect

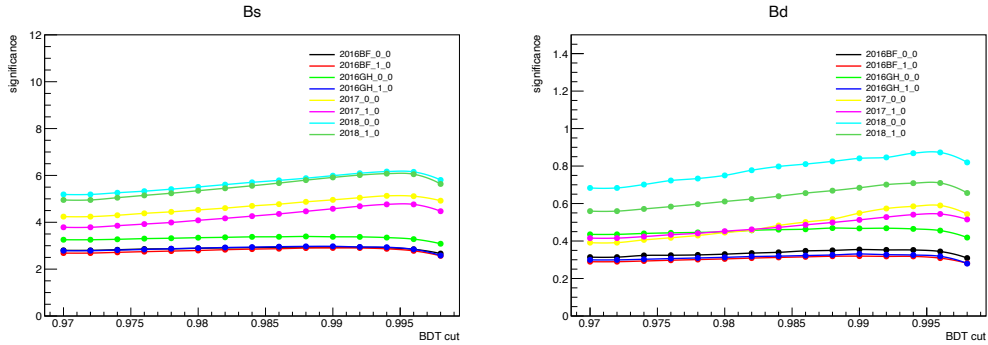


Figure 5.12: The significance of B_s^0 (left) and B^0 (right) signal in terms of different $d_{\text{MVA}}^{\text{high}}$ with muon BDT > 0.45 , $d_{\text{MVA}}^{\text{low}} = 0.93$.

the significance much. Keeping $d_{\text{MVA}}^{\text{low}} = 0.93$, the scan again is performed to find the $d_{\text{MVA}}^{\text{high}}$ point. The scan results are shown in Fig. 5.12. The results are similar to the 1-bin setup, and the final selection for 2-bin is chosen to be $d_{\text{MVA}}^{\text{low}} = 0.9$ and $d_{\text{MVA}}^{\text{high}} = 0.99$.

The relative error on the lifetime is calculated in terms of different requirements on analysis d_{MVA} and muon BDT, using Asimov datasets. The lifetime UML uses the model described in Section 5.7. To estimate the relative error on the lifetime, first the invariant mass fit is performed and the sWeight’s for each events are calculated. The sWeighted decay time distribution are then fitted using the exponential function corrected by decay

time efficiency and resolution. Figure 5.13 shows the scanning of relative error for different muon BDT and analysis d_{MVA} in 2018 channel 0 and 1. We conclude that the same analysis d_{MVA} selection ($d_{\text{MVA}} > 0.99$) can be used to obtain the highest precision in terms of the lifetime measurement.

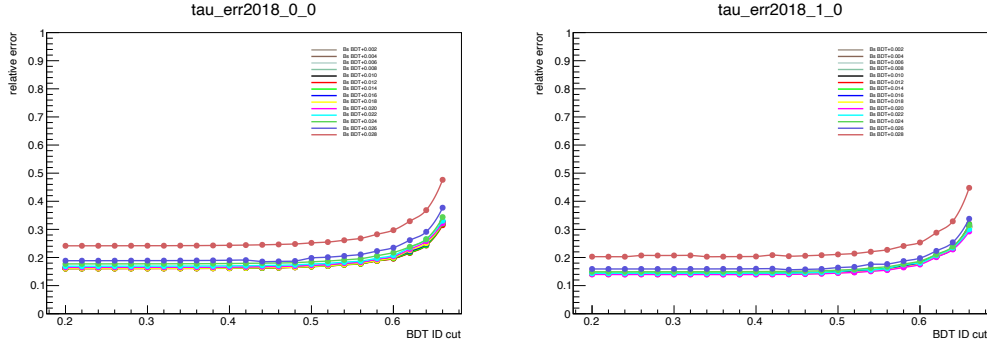


Figure 5.13: The relative lifetime error in terms of different requirements of analysis d_{MVA} and muon BDT in 2018 channel 0 (left) and channel 1 (right). Other eras and channels have similar trend. Here x-axis represents the muon BDT requirement and different labels represents the analysis d_{MVA} requirement with benchmark $d_{\text{MVA}} > 0.97 + X$ (where X is mentioned in the legend).

5.7 Lifetime measurement with UML fit

The second parameter of interest is the effective lifetime of B_s^0 meson in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay process. The lifetime is extracted from a simultaneous UML fit to mass, decay time, and decay time uncertainty in multiple event categories. The current fitter setup is developed using the old fitter discussed in the previous chapter, as a baseline, aiming to improve the stability of the fitter and decrease the statistical uncertainty on the lifetime. Many lifetime analyses have been using the similar UML fit to extract the lifetime, but we still need to validate our extraction procedure to trust our method. In this analysis, the best possible choice to measure the lifetime is the B^+ meson lifetime in $B^+ \rightarrow J/\psi K^+$ decay mode. Even though the two decay channels differ in the background composition, a fit component included in the likelihood and the selection requirements, they are very similar in the

lifetime measurement part. In order to minimize the difference, $B^+ \rightarrow J/\psi K^+$ events are selected as similar as possible to $B_s^0 \rightarrow \mu^+ \mu^-$ decays, including the analysis d_{MVA} , constructed in such a way that it matches d_{MVA} of $B_s^0 \rightarrow \mu^+ \mu^-$ decays (a detailed discussion is given in Section 5.4.2). The proper decay time (t) in 3D space is defined as, $t = m \frac{\ell_{3D}}{p}$ where, m and p are the invariant mass and momentum of the reconstructed $B_s^0(B^+)$ meson calculated for each event. ℓ_{3D} is the distance between the primary to the secondary vertex in 3D space. The proper decay time uncertainty (σ_t) for each event is calculated by using the covariance matrices. This takes the uncertainty of mass, vertex positions, and momentum into account.

The fit procedure is somewhat different for $B^+ \rightarrow J/\psi K^+$, and $B_s^0 \rightarrow \mu^+ \mu^-$ fits. The lifetime in the normalization channel is investigated in four different samples, i.e., 2016BF, 2016GH, 2017, and 2018. At the same time, the signal $B_s^0 \rightarrow \mu^+ \mu^-$ lifetime is extracted from a simultaneous fit over the four eras, and two channels with the signal yield constrained between different categories based on the MC simulation prediction. The benefit of adding constraint conditions is to allow the fitter to give sensible results in very low statistics regions. In both types of fits, the decay time is restricted to a window of [1, 14] ps as the signal efficiency is very low outside this window. Compared to previous analysis, a PDF for the decay time uncertainty is introduced in the fitter to avoid the Punzi effect [95].

5.7.0.1 Model overview

Complete event probability functions are defined as:

$$\begin{aligned}
 P_{B^+ \rightarrow J/\psi K^+} = & N_{sig} P_{sig}(m_{B^+}) \epsilon(t) [f(t, \tau_{sig}) \otimes G(t, S\sigma_t)] P_{sig}(\sigma_t) + \\
 & N_{bkg} P_{bkg}(m) P_{bkg}(t) P_{bkg}(\sigma_t) + \\
 & N_{peak} P_{peak}(m) P_{peak}(t) P_{peak}(\sigma_t)
 \end{aligned} \tag{5.7}$$

$$\begin{aligned}
 P_{B_s^0 \rightarrow \mu^+ \mu^-} = & N_{sig} \epsilon(t) [f(t, \tau_{sig}) \otimes G(t, S\sigma_t)] P_{sig}(m_{B_s}) P_{sig}(\sigma_t) + \\
 & N_{bkg} P_{bkg}(m) P_{bkg}(t) P_{bkg}(\sigma_t) + \\
 & N_{peak} \epsilon_{peak}(t) [f(t, \tau_{peak}) \otimes G(t, S\sigma_t)] P_{peak}(m) P_{peak}(\sigma_t) + \\
 & N_{semi} \epsilon_{semi}(t) [f(t, \tau_{semi}) \otimes G(t, S\sigma_t)] P_{semi}(m) P_{semi}(\sigma_t)
 \end{aligned} \tag{5.8}$$

where N_{sig} , N_{bkg} , N_{semi} , and N_{peak} are the yields assigned to different components. A complete list of component PDFs are shown below:

- Signal
 - $P_{sig}(m)$ is the mass PDF modeled by a sum of CB-line and a double Gaussian with a common mean for $B^+ \rightarrow J/\psi K^+$ and a sum of CB-line and a Gaussian function for $B_s^0 \rightarrow \mu^+ \mu^-$. For $B^+ \rightarrow J/\psi K^+$ fits, the mass model parameters are floated, while for the $B_s^0 \rightarrow \mu^+ \mu^-$ fit, the signal model is fixed to a corrected mass distribution model used in the branching fraction fit.
 - $P_{sig}(\sigma_t)$ is the decay time uncertainty PDF.
 - $f(t, \tau_{sig})$ is the decay time PDF.
 - $G(t, \sigma_t)$ is the Gaussian resolution function with per event decay time uncertainty.
 - $\epsilon(t)$ is the efficiency as a function of decay time, described in Section [5.7.0.2](#).
- Combinatorial background
 - $P_{bkg}(m)$ - mass PDF modeled by the 1st order Bernstein polynomial function, where the parameters are floated in the data fit.
 - $P_{bkg}(t)$ & $P_{bkg}(\sigma_t)$ - decay time and decay time uncertainty PDF.
- Peaking backgrounds

- $B^+ \rightarrow J/\psi \pi^+$ background for $B^+ \rightarrow J/\psi K^+$ fits
 - * $P_{peak}(m)$ - mass PDF modeled with a triple Gaussian function. The parameters of the model are fixed in the final data fit and the yield is fixed to the ratio of branching fraction w.r.t to $B^+ \rightarrow J/\psi K^+$ (which is 4%).
 - * $P_{peak}(t)$ & $P_{peak}(\sigma_t)$ - decay time and decay time uncertainty PDF.
- $B \rightarrow h^+ h^-$ and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ backgrounds for $B_s^0 \rightarrow \mu^+ \mu^-$ fits
 - * $P_{peak}(m)$ - mass PDF modeled by a sum of the Crystal-Ball function and a Gaussian function.
 - * $P_{peak}(\sigma_t)$ - decay time uncertainty PDF.
 - * $f(t, \tau_{peak})$ - decay time PDF.
 - * $\epsilon_{peak}(t)$ - the efficiency as a function of decay time.
- Semileptonic partially reconstructed backgrounds for $B_s^0 \rightarrow \mu^+ \mu^-$
 - $P_{semi}(m)$ - mass PDF modeled by a Gaussian function.
 - $P_{semi}(\sigma_t)$ - decay time uncertainty PDF.
 - $f(t, \tau_{semi})$ - decay time PDF.
 - $\epsilon_{semi}(t)$ - the efficiency as a function of decay time.
- Normalization yield
 - For $B^+ \rightarrow J/\psi K^+$ data fits, all the normalization yields are floated.
 - For $B_s^0 \rightarrow \mu^+ \mu^-$ fit,
 - * N_{sig} - is converted using the master formula (Eq. 5.5). The master formula uses the efficiency ratio, fragmentation fraction ($\frac{f_s}{f_u}$), branching fraction of $B^+ \rightarrow J/\psi K^+$ and $B^+ \rightarrow J/\psi K^+$ yield from the data fit. A

common branching fraction for all eight channels is floated in the fit. The $B^+ \rightarrow J/\psi K^+$ branching fraction and yields for eight channels are floated inside a Gaussian constraint.

- * N_{peak} and N_{semi} - peaking and semileptonic background yields are floated inside Lognormal constraint.
- * N_{bkg} - combinatorial background yields for eight channels are floated.

5.7.0.2 Efficiency as a function of decay time

The decay time efficiency is a relative efficiency derived by comparing the reconstructed decay time distribution and the ideal exponential one with the SM lifetime. The efficiency is given by:

$$\epsilon(t) = \frac{t_{reco}}{t_{gen} \otimes \delta(t)} \quad (5.9)$$

where,

- t_{reco} : is the reconstructed decay time distribution
- t_{gen} : is an exponential generated decay time distribution with the lifetime $\tau = 1.638$ and $\tau = 1.609$ ps for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$, correspondingly.
- $\delta(t) = t_{gen} - t_{reco}$ is the decay time resolution function. The resolution is included in order to keep the efficiency independent from the effect of the detector reconstruction.

The efficiency distribution is modeled in the range 1-14 ps with the following function,

$$\epsilon(t, p_0, p_1, p_2, p_3, p_4) = p_0 \left(1 + p_1 t + \frac{p_2}{t^2} + \frac{p_3}{(1 + \exp(-tp_4))} \right) \quad (5.10)$$

5.7.0.3 Decay time model

The decay time distribution for the signal is modeled by a single exponential function convoluted with a Gaussian resolution function, which uses the event-by-event decay time uncertainty. Furthermore, the signal component of the decay time is corrected by the decay

time efficiency. Similarly, the combinatorial background and peaking background distributions are modeled with exponential functions with Gaussian resolution.

The decay time uncertainty is computed by propagating the uncertainties of the vertex position, mass, and the B meson momentum. For $B^+ \rightarrow J/\psi K^+$, the distribution of the signal decay time uncertainty is modeled by three Gamma functions. In data, we perform a prefit to obtain the decay time uncertainty distribution for signal and backgrounds. The signal distribution is obtained by subtracting the combinatorial background from the signal region (5.22-5.32) GeV. The combinatorial background decay time uncertainty distribution is taken from the mass range (5.4-5.5) GeV and is modeled with two Gamma functions. Since the peaking background is similar to decay as the signal, we use the same (three Gamma) function as the signal. Gamma distribution is defined as,

$$f(\sigma_t, \mu, \gamma, \beta) = \frac{(\sigma_t - \mu)^{\gamma-1} \exp^{-(\sigma_t - \mu)/\beta}}{\Gamma(\gamma) \beta^\gamma} \quad (5.11)$$

where, $\Gamma(\gamma)$ is the gamma function $\Gamma(\gamma) = (\gamma - 1)!$, μ , γ and β are three shape parameters.

For $B_s^0 \rightarrow \mu^+ \mu^-$, each signal and background decay time uncertainty distributions are modeled with the Gamma function, which are obtained from MC simulation and high sideband data.

5.7.1 Lifetime fit validation on $B^+ \rightarrow J/\psi K^+$ decay using d_{MVA}

The analysis d_{MVA} has non-trivial correlations with the decay time. One of the main contributions to the correlation comes from the fact that the pointing angle of the B meson is the most powerful discriminator against the combinatorial background in d_{MVA} . The pointing angle resolution is strongly correlated with the decay distance, which means that d_{MVA} gives a higher score to events with a larger decay distance. Figure 5.14 shows the correlation between the pointing angle and the decay distance, and the decay time.

The strong correlation between the decay time and d_{MVA} may lead to a potential bias in

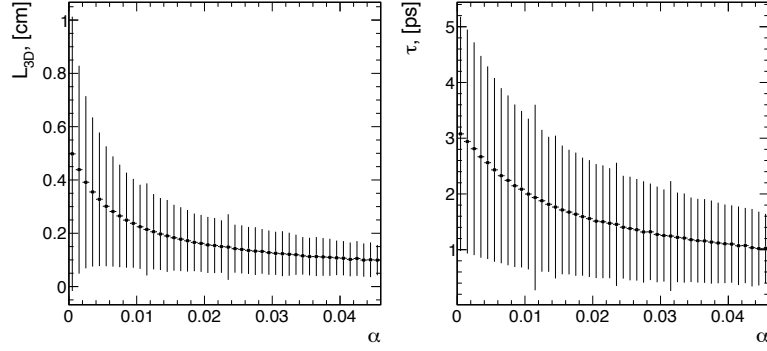


Figure 5.14: Correlation between the B meson point angle and its decay distance (left) and decay time (right) for $B_s^0 \rightarrow \mu^+ \mu^-$ events.

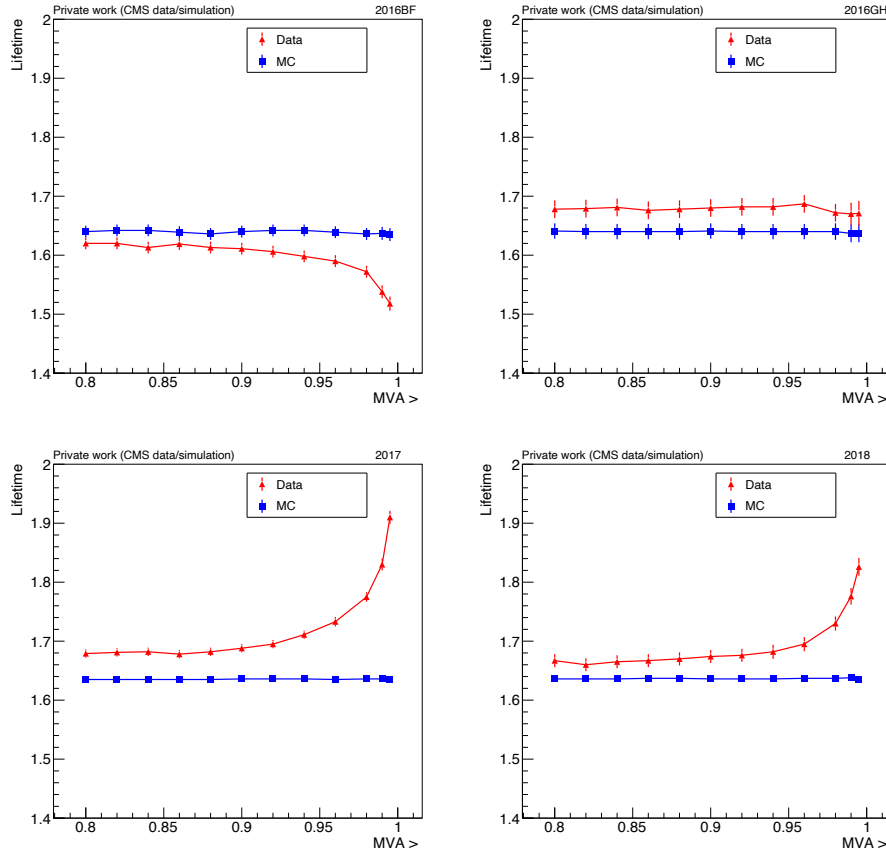


Figure 5.15: Measured lifetime from the data (red marker) and MC (blue marker) using different d_{MVA} points for 2016BF, 2016GH, 2017, and 2018 data samples.

the lifetime measurement. For the tight d_{MVA} selection requirements, we observe a sizable deviation between MC simulation predictions and data. To study this effect, we select the $B^+ \rightarrow J/\psi K^+$ events as similar as possible to $B_s^0 \rightarrow \mu^+ \mu^-$ decays as discussed in Section 5.4.1. We then performed a scan on the different analysis d_{MVA} points to measure the lifetime in data and MC. For each d_{MVA} point, the efficiency as a function of decay time from the MC is evaluated and the signal decay time distribution is corrected. Figure 5.15 shows the lifetime obtained from the fit to MC and data sample.

As expected, the B^+ meson lifetime measured in fits on MC simulated events reproduces the SM value of the lifetime (1.638 ps) for different d_{MVA} cut. For data, the fit results show a significant bias in the lifetime for tighter d_{MVA} selection. The bias in the lifetime distribution has a constant value for loose d_{MVA} selection requirements and shows a strong correlation with d_{MVA} for $d_{\text{MVA}} > 0.90$. In order to address this problem, we derive a data-driven correction, which is described in detail in Section 5.8.3. Figure 5.16 shows the fit projection for a typical $B^+ \rightarrow J/\psi K^+$ fit after $d_{\text{MVA}} > 0.90$.

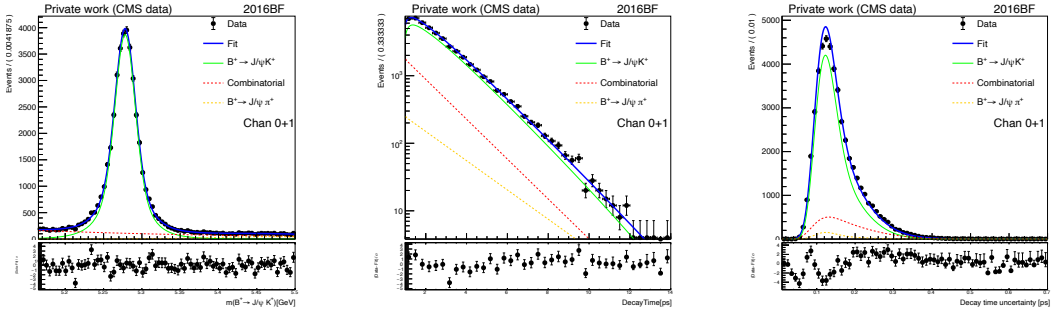


Figure 5.16: Invariant mass, decay time and decay time uncertainty fit projections in data after $d_{\text{MVA}} > 0.90$ selection for the year 2016BF. The final fit PDF is shown as the solid blue curve. The signal and other background components are shown as the dotted curves. Pull plots are shown in the bottom pads to visualize the goodness of fit.

5.8 MC corrections

5.8.1 Mass scale and resolution

The dimuon mass ($m_{\mu\mu}$) and its resolution (σ_m) are the main input observables in the UML fit to measure the branching fraction and lifetime. As the expected number of $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ events are small, it will be difficult to determine the shape from the UML fit. So, the shapes of the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ distribution are necessary to perform the fit. Therefore, the $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ MC samples are used to estimate the mean value of the dimuon mass distribution and the resolution scale. Now to avoid any fit bias due to data and MC differences, the mass shapes are calibrated using the two control channels, $J/\psi \rightarrow \mu^+\mu^-$, and $\Upsilon(1S) \rightarrow \mu^+\mu^-$.

The mass scale correction is defined as

$$\delta S_m = \frac{m_{\mu\mu}(Data) - m_{\mu\mu}(MC)}{m_{\mu\mu}(PDG)} \quad (5.12)$$

We use it to compute the mass offset for $B_s^0 \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow \mu^+\mu^-$ defined as

$$\delta m = m_{\mu\mu}(Data) - m_{\mu\mu}(MC) \quad (5.13)$$

In the fitter, the mass offset (δm) is used to correct the mean. Whereas the more appropriate observable for this study is the scale difference (δS_m) because it has a weak correlation with the invariant mass, which simplifies parameterization of the corrections and it can be easily converted into the offset for any mass point.

A ratio of the resolution scales in data and MC simulations is then measured to make sure that the event by event mass uncertainty matches the actual mass resolution.

The $J/\psi \rightarrow \mu^+\mu^-$ and $\Upsilon(1S) \rightarrow \mu^+\mu^-$ events are selected using following requirements to match with the signal decay. The event selection requirements are:

- $p_{T\mu} > 4$, $|\eta_\mu| < 1.4$, muon BDT > 0.45 , two opposite charges and vtx_prob > 0.1

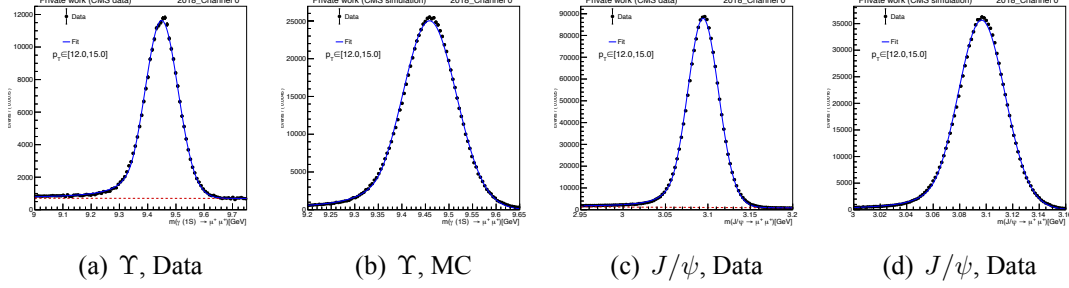


Figure 5.17: Fit projection in $\Upsilon(1S) \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ channel from MC and Data for $p_T \in [12.0, 15.0]$ in 2018 channel 0.

- triggers for $J/\psi \rightarrow \mu^+\mu^-$: HLT_DoubleMu4_3_Jpsi(_Displaced) for 2018 (2016, and 2017)

For $\Upsilon(1S) \rightarrow \mu^+\mu^-$ events, no trigger requirement is required and used all events in the corresponding datasets.

The signal model used in this study is identical to the model used in the main fit. It is a two-dimensional PDF consisting of the CB-line for the dimuon invariant mass itself and a kernel estimator model for the relative mass resolution, as described in Section 5.6. The width parameter of the CB-line is a conditional parameter linearly depending on the dimuon mass resolution, $\sigma_{CB} = \kappa \times \sigma_m$. The background is modelled by the first order Bernstein polynomial function. A few examples of the fits are shown in Fig. 5.17.

The model fits the data reasonably well, but not perfectly. This is not an issue though. The small mismatches are only visible because the control samples have much larger number of events than the expected $B_s^0 \rightarrow \mu^+\mu^-$ signal. The goal is not to find a perfect description of the mass line shape in the control samples, but to correct the mass scale and resolution of the model used for the branching fraction and lifetime measurements.

Figure 5.18 shows the p_T dependence of the mass scale correction for different eras and channels. The mean shift distribution for $J/\psi \rightarrow \mu^+\mu^-$ for the year 2018 is different from 2016 and 2017, because of the change in the trigger. The observed p_T dependence has a

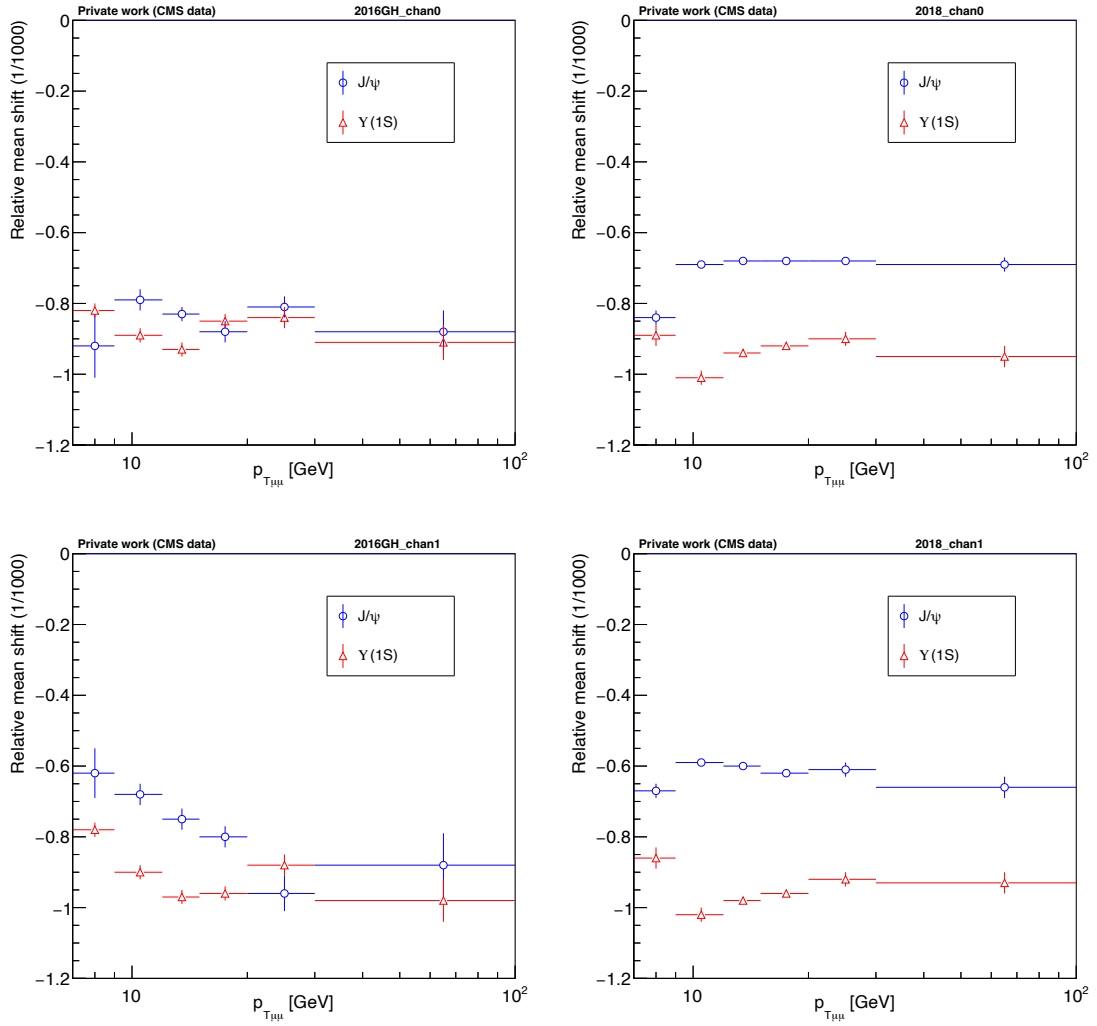


Figure 5.18: Relative mean shift from J/ψ and $\Upsilon(1S)$ for channel 0 (top row) and channel 1 (bottom row) in 2016GH (left column) and 2018 (right column) data samples.

negligible impact on the analysis, which is concluded from the toy study. A simple fit is performed to invariant mass distribution in the toy sample having the total event close to the expected number (300). The observed mass uncertainty is found to be 2.3 MeV which is 0.4‰ of the mass scale. Corrections smaller than that will not make a significant impact on the analysis.

Channel	Mass shift (GeV)			
	2016BF	2016GH	2017	2018
0	−0.0047	−0.0046	−0.0048	−0.0042
1	−0.0046	−0.0045	−0.0051	−0.0040

Table 5.9: Summary of the mean shift between data and MC for B_s^0 in different years and channels.

The ratio of mass resolution scale factors are extracted in similar way and the results are shown in Table 5.10.

Channel	$\frac{\kappa(Data)}{\kappa(MC)}$			
	2016BF	2016GH	2017	2018
0	1.040	1.027	1.023	1.032
1	1.043	1.038	1.040	1.045

Table 5.10: Ratio of the mass resolution scales in data and MC for B_s^0 for different years and channels.

5.8.2 MC corrections for analysis d_{MVA}

As discussed in Section 5.4.2, an evident discrepancy has been seen in the d_{MVA} distribution of data and MC simulation for 2017 and 2018. As a result, the tight d_{MVA} selection will result in different efficiencies in data and MC. This difference must be corrected to obtain the correct final results.

To minimize the effect, two possible ways can be possible, which are

- apply the same d_{MVA} selection on $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ and rely on the cancellation of effect in the efficiency ratio
- derive corrections for data and MC discrepancies

The first approach cannot be carried out easily because the d_{MVA} distributions do not match in these two decay modes. However, they do match for a very restrictive kaon momentum requirement ($p_T \in [1.0, 1.5]$ GeV). This p_T restriction will bring sizeable systematics.

The second solution, correction for data and MC discrepancies, is obtained from two methods.

- Take the ratio of sWeighted data and MC efficiency for a particular d_{MVA} selection in $B^+ \rightarrow J/\psi K^+$ channel, and apply it to the $B_s^0 \rightarrow \mu^+ \mu^-$ sample for that selection. The efficiencies and the corresponding corrections can be found in Tables 5.11 and 5.12 respectively (termed as “Efficiency ratio” in Table).
- Use advanced machine learning techniques to derive the correction from $B^+ \rightarrow J/\psi K^+$ and reweight $B_s^0 \rightarrow \mu^+ \mu^-$ samples.

The details of method 2 are discussed below.

5.8.2.1 MC reweighting with XGBoost

The second way to correct the MVA selection efficiency is to reweight MC simulation samples to match the data. To start with the deriving correction, several input variables to the d_{MVA} are checked. There is no single variable found, which would allow for the compensation of the discrepancy. An alternative approach using the XGBoost algorithm [93] is adopted to train a classifier on the $B^+ \rightarrow J/\psi K^+$ samples and use it to reweight $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ MC samples.

	Type	Efficiency					
		$\frac{N(d_{MVA} > 0.90)}{N(d_{MVA} > 0.00)}$			$\frac{N(d_{MVA} > 0.99)}{N(d_{MVA} > 0.90)}$		
		2016	2017	2018	2016	2017	2018
$\mu\mu K$	MC	0.701 ± 0.006	0.849 ± 0.002	0.826 ± 0.002	0.500 ± 0.006	0.677 ± 0.002	0.660 ± 0.002
	MC _{corr}	0.694 ± 0.006	0.807 ± 0.002	0.764 ± 0.002	0.513 ± 0.006	0.621 ± 0.002	0.603 ± 0.002
	Data	0.709 ± 0.007	0.797 ± 0.006	0.746 ± 0.006	0.529 ± 0.007	0.603 ± 0.005	0.584 ± 0.006
$\mu\mu$	MC	0.705 ± 0.004	0.846 ± 0.002	0.840 ± 0.001	0.531 ± 0.004	0.674 ± 0.002	0.660 ± 0.001
	MC _{corr}	0.699 ± 0.004	0.803 ± 0.002	0.770 ± 0.001	0.535 ± 0.004	0.610 ± 0.002	0.599 ± 0.001

Table 5.11: Efficiencies obtained from MC samples with (labelled as MC_{corr}) and without (labelled as MC) XGBoost reweighting. The efficiencies obtained from the sWeighted data are labelled as Data.

The XGBoost classifier is trained with sWeighted data and MC simulation using the same inputs as the analysis d_{MVA} . The sWeights are obtained from the UML fit to the $B^+ \rightarrow J/\psi K^+$ invariant mass distribution.

The classifier returns the probability of each event to be data and MC. Using the the ratio of the two probabilities, the weight is defined as:

$$w(x) \simeq \frac{p_{data}(x)}{p_{mc}(x)} \quad (5.14)$$

For 2017 and 2018, the most important and therefore the most mis-modeled variables are the 3D impact parameter wrt PV, its significance, and the maximum vertexing probability of a candidate muon with a charged track.

Most discriminating variable comparisons between sWeighted data and weighted MC, for the year 2017 + 2018, are shown in Fig. 5.19.

Figure 5.20 compares the sWeighted data and weighted MC for the analysis d_{MVA} . The reweighted MC distribution shows a better agreement with data, especially for 2017 and 2018.

After validation of the reweighting procedure, the weight is applied to the $B_s^0 \rightarrow \mu^+ \mu^-$ MC samples and efficiencies of the d_{MVA} cuts are computed. The results are shown in Table 5.11.

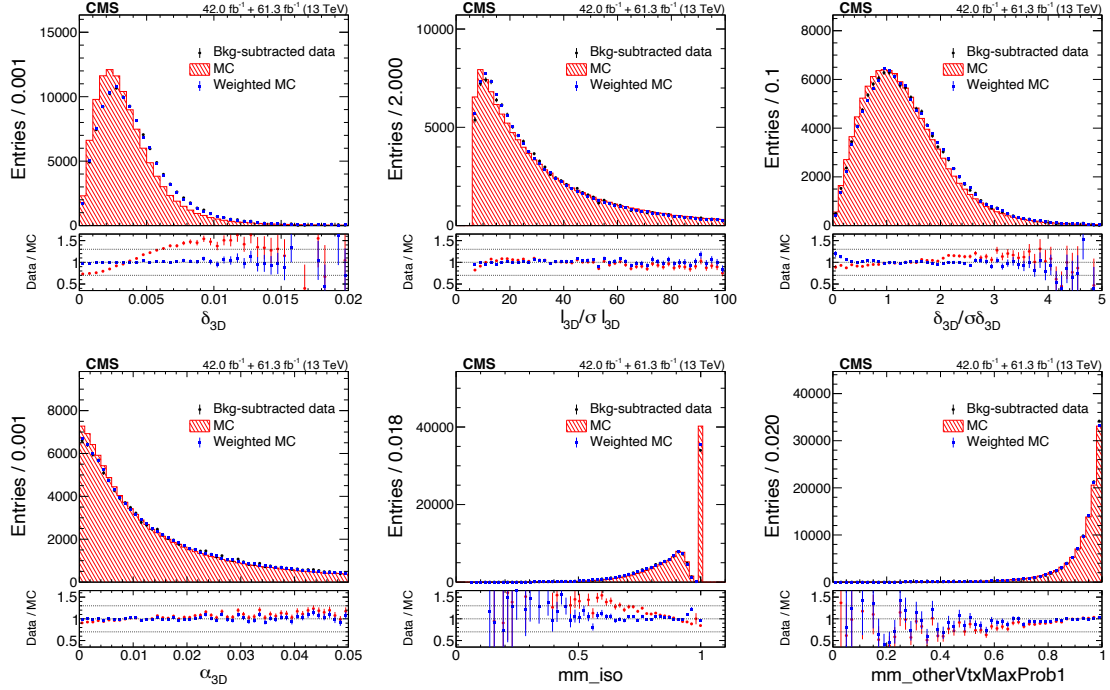


Figure 5.19: Data and MC comparison before and after reweighting for few important variables in the 2017 + 2018 sample.

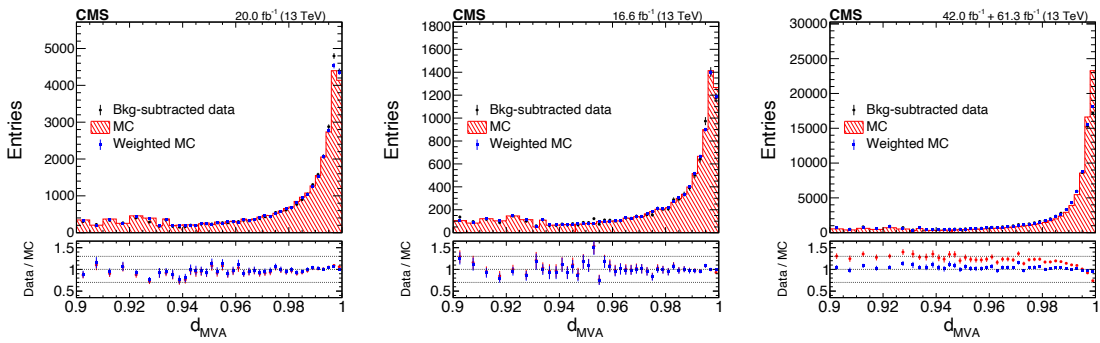


Figure 5.20: Analysis d_{MVA} comparison (in plot it is written as d_{MVA}) in data and MC for both weighted and unweighted distributions for 2016BF (left), 2016GH (center) and 2017 + 2018 (right). The bottom plots represent Data/MC ratios for weighted MC (blue marker) and unweighted MC (red marker). The MC distributions are normalized to the total number of events in data.

Method	Scale Factors		
	$d_{\text{MVA}} \in [0.90, 1.00]$		
	2016	2017	2018
XGBoost	0.991 ± 0.008	0.949 ± 0.003	0.917 ± 0.002
Ratio	1.011 ± 0.013	0.939 ± 0.007	0.903 ± 0.008
Method	$d_{\text{MVA}} \in [0.99, 1.00]$		
	2016	2017	2018
	2016	2017	2018
XGBoost	1.008 ± 0.011	0.905 ± 0.004	0.908 ± 0.002
Ratio	1.058 ± 0.019	0.891 ± 0.008	0.885 ± 0.010

Table 5.12: Normalization corrections for $B_s^0 \rightarrow \mu^+ \mu^-$ derived using two different methods: XGBoost reweighting and the efficiency ratio between data and MC in $B^+ \rightarrow J/\psi K^+$ events.

5.8.2.2 Summary

The normalization corrections from the two methods are summarized in Table 5.12. In general, the two methods give results compatible with each other at the $1\text{-}2\sigma$ level. The correction from XGBoost method is chosen to correct the MC, and the choice is made by flipping a coin. Based on the difference between the two estimations, 2% systematics for the corrections for $d_{\text{MVA}} > 0.90$ and 3% for the corrections for $d_{\text{MVA}} > 0.99$ are assigned.

5.8.3 Decay time corrections

The decay time is correlated with many selection requirements. Most of them are fairly well simulated, leading to residual systematics of 0.03-0.04 ps. The situation changes when we introduce d_{MVA} . Figure 5.15 shows that starting from $d_{\text{MVA}} > 0.90$, a bias on the lifetime is observed in the data.

Before we dive into the details of the correction extraction, it is important to understand how the decay time efficiency works and its meaning. Figure 5.21 shows the impact of the decay time efficiency corrections when applied to the decay time distribution in MC and data. By dividing the decay time distribution by the efficiency, one can effectively unfold

the reconstructed decay time distribution to the ideal shape, i.e., the exponential distribution with the Standard Model lifetime as the decay constant. The figure clearly shows that with the $d_{\text{MVA}} > 0.99$ selection requirement, the efficiency correction derived from MC modifies the decay time distribution to a distribution with a slightly different decay constant and some deviation from the exponential shape. All this leads to a bias in the lifetime measurement.

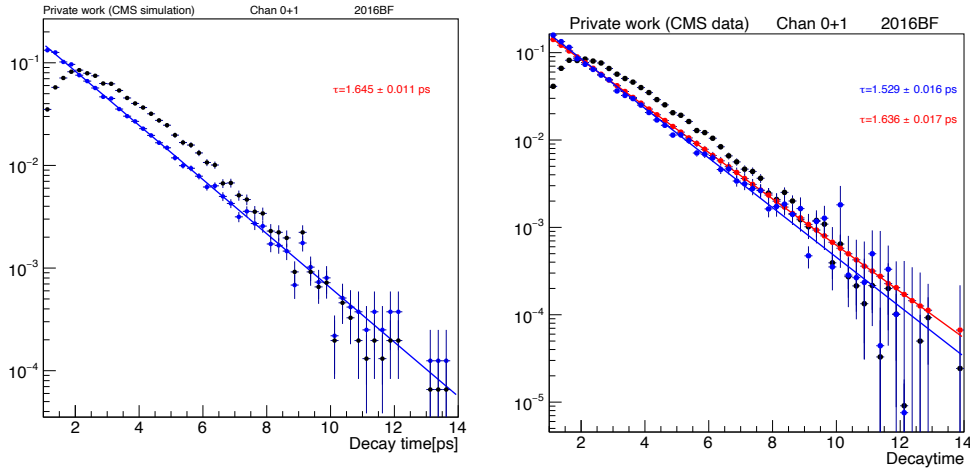


Figure 5.21: Impact of the decay time efficiency correction on MC(left) and data(right) with the $d_{\text{MVA}} > 0.99$ for the year 2016BF. Black markers are the original distributions without correction. Blue markers are the distribution after the MC efficiency correction. Red markers in the right plot are after the data efficiency correction. The distributions are unit normalized.

To compensate for the detector mismodeling of the decay time distribution in MC simulations, we need to derive a correction from $B^+ \rightarrow J/\psi K^+$ events in data. Due to potential differences between $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ efficiencies, the correction is derived for the d_{MVA} selection only. In other words, the correction for decay time distribution for $d_{\text{MVA}} > 0.90$ selection is obtained from the MC simulation, and for the transition from $d_{\text{MVA}} > 0.90$ to $d_{\text{MVA}} > 0.99$, the correction is derived from $B^+ \rightarrow J/\psi K^+$ channel.

Before we derive the correction, we need to show that the method works on MC. For that, we will compare the lifetime measurement performed on $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample

using $B_s^0 \rightarrow \mu^+ \mu^-$ only efficiency corrections vs. corrections derived on $B_s^0 \rightarrow \mu^+ \mu^-$ MC for $d_{\text{MVA}} > 0.90$ selection and $B^+ \rightarrow J/\psi K^+$ MC for $d_{\text{MVA}} > 0.99$ with respect to $d_{\text{MVA}} > 0.90$.

First, let us define corrections and their meaning. In general, the decay time PDF is defined as

$$P(t) = \frac{f(t)}{\varepsilon(t)} \quad (5.15)$$

where $f(t)$ is the decay time distribution for a given selection and $\varepsilon(t)$ is the efficiency with $1/\varepsilon(t)$ is effectively a correction function that transforms the distribution in the exponential with the decay constant equal to the expected lifetime.

For $d_{\text{MVA}} > 0.90$, the decay time PDFs are

$$P_{0.90}^u(t) = \frac{f_{0.90}^u(t)}{\varepsilon_{0.90}^u(t)}$$

$$P_{0.90}^s(t) = \frac{f_{0.90}^s(t)}{\varepsilon_{0.90}^s(t)}$$

where $P_{0.90}^u(t)$ is the PDF for $B^+ \rightarrow J/\psi K^+$ and where $P_{0.90}^s(t)$ is the PDF for $B_s^0 \rightarrow \mu^+ \mu^-$.

For $d_{\text{MVA}} > 0.99$, the PDFs are

$$P_{0.99}^u(t) = \frac{f_{0.99}^u(t)}{\varepsilon_{0.99}^u(t)}$$

$$P_{0.99}^s(t) = \frac{f_{0.99}^s(t)}{\varepsilon_{0.99}^s(t)}$$

We know that MC simulations match data reasonably well for a loose selection such as $\text{MVA}_B > 0.90$. Therefore, we will rely on simulated samples to derive the first level of corrections, i.e. $\varepsilon_{0.90}^u(t)$ and $\varepsilon_{0.90}^s(t)$. For the tighter $\text{MVA}_B > 0.99$ selection, we want to use $B^+ \rightarrow J/\psi K^+$ control sample. In other words:

$$P_{0.99}^u(t) = \frac{f_{0.99}^u(t)}{\varepsilon_{0.99}^u(t)} = \frac{f_{0.99}^u(t)}{\varepsilon_{0.90}^u(t)} \cdot \frac{\varepsilon_{0.90}^u(t)}{\varepsilon_{0.99}^u(t)} = \frac{1}{\alpha^u(t)} \frac{f_{0.99}^u(t)}{\varepsilon_{0.90}^u(t)} \quad (5.16)$$

	2016BF	2016GH	2017	2018
τ_s	1.601 ± 0.008	1.598 ± 0.007	1.606 ± 0.004	1.607 ± 0.003
$\tau_{s'}$	1.505 ± 0.008	1.542 ± 0.007	1.591 ± 0.004	1.591 ± 0.003
Bias	0.096	0.057	0.015	0.016

Table 5.13: Lifetime obtained in $B_s^0 \rightarrow \mu^+ \mu^-$ MC using $B_s^0 \rightarrow \mu^+ \mu^-$ only corrections (τ_s) and a mix of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ corrections for $d_{\text{MVA}} > 0.99$ ($\tau_{s'}$). The generated lifetime for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ are 1.609 ps and 1.638 ps.

where the efficiency correction $\alpha(t)$ is defined as

$$\alpha(t) = \frac{\varepsilon_{0.99}(t)}{\varepsilon_{0.90}(t)} \quad (5.17)$$

In other words, $1/\alpha(t)$ is the correction that one applies to make $f_{0.99}^s(t)/\varepsilon_{0.90}^s(t)$ distribution look like a perfect exponential distribution with the target lifetime.

To verify the method, we need to compare lifetime measurements extracted using the following efficiencies with the decay time distribution $f_{0.99}^s(t)$ extracted from $B_s^0 \rightarrow \mu^+ \mu^-$ MC simulations.

$$\varepsilon(t) = \alpha^s(t) \cdot \varepsilon_{0.90}^s(t) \quad (5.18)$$

$$\varepsilon'(t) = \alpha^u(t) \cdot \varepsilon_{0.90}^s(t) \quad (5.19)$$

Table 5.13 summarizes the lifetime obtained from the $B_s^0 \rightarrow \mu^+ \mu^-$ MC using $B_s^0 \rightarrow \mu^+ \mu^-$ only corrections and a mix of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ corrections. The differences between the two estimates is considered as a systematic uncertainty of the method.

After performing the validation on the MC, the correction is measured in data using $B^+ \rightarrow J/\psi K^+$ decay mode. The corrections as a function of decay time are shown in Fig. 5.22. We observe a significant inefficiency for low decay time values. This is expected since it is hard to measure the pointing angle precisely for short decay time, making such events harder to separate from the combinatorial background. The corrections are applied bin-by-bin to the efficiency obtained from $d_{\text{MVA}} > 0.90$.

	2016BF	2016GH	2017	2018
$\tau_{B^+}^{Data}$	1.601 ± 0.012	1.666 ± 0.018	1.668 ± 0.008	1.645 ± 0.012
Bias	-0.04	0.03	0.04	0.01

Table 5.14: Lifetime obtained in $B^+ \rightarrow J/\psi K^+$ data using the correction from $B^+ \rightarrow J/\psi K^+$ data for $d_{MVA} > 0.99$. Here the MC value is 1.638 ps.

Effect	2016BF	2016GH	2017	2018
Method systematic	0.10	0.06	0.02	0.02
Lifetime bias for $d_{MVA} > 0.90$	0.03	0.04	0.05	0.04
Lifetime bias for $d_{MVA} > 0.99$	0.04	0.03	0.04	0.01
Total	0.11	0.07	0.05	0.04

Table 5.15: Systematics obtained from the correction for $d_{MVA} > 0.99$. Here the PDG value is 1.638 ps.

As a final cross-check, we measure the lifetime of B^+ meson in data using the corrected efficiency function. The results are shown in Table 5.14. The bias is expected to be consistent with the bias observed for $d_{MVA} > 0.90$ selection. We take the maximum of the two by absolute value as the control sample measurement systematics.

Table 5.15 shows various systematic effects and the final systematic that we assign for the lifetime measurement for each data-taking period. We introduce these systematic effects as nuisance parameters in the lifetime fit. The difference between the data and MC result is the lifetime bias for $d_{MVA} > 0.90$, which is obtained from Fig. 5.15.

5.8.4 Muon fake rate correction

As shown in the optimization study in Section 5.6.3, the muon BDT working point can be released to a lower value (0.45) to gain sensitivity without compromising the control over the misidentified muon backgrounds. A dedicated study on the muon fake rate is discussed in Appendix A. From the study, the kaon and pion scale factors (scale factor = data fake

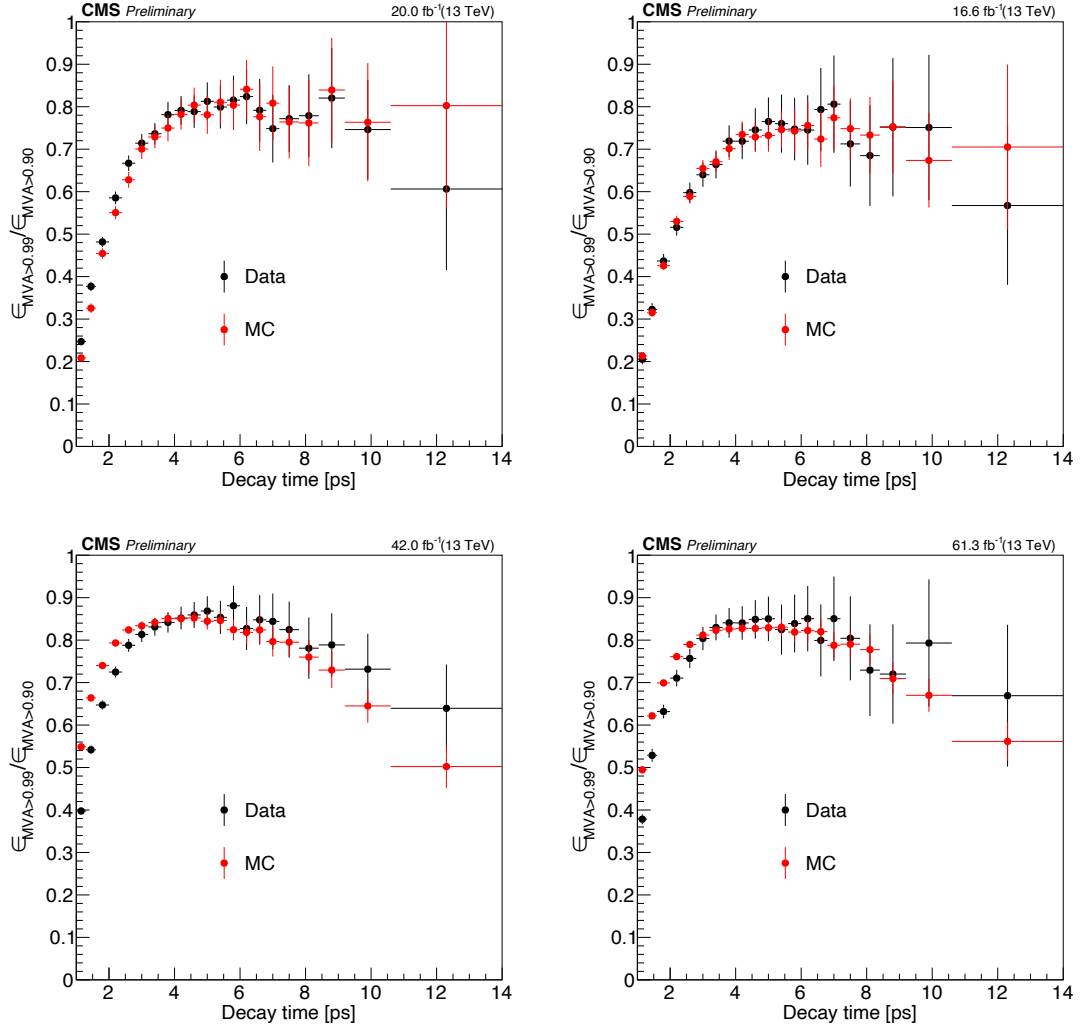


Figure 5.22: Ratios of the decay time efficiencies for the tight ($d_{\text{MVA}} > 0.99$) and loose ($d_{\text{MVA}} > 0.90$) selections observed in 2016BF, 2016GH, 2017, and 2018 data and MC simulations for $B^+ \rightarrow J/\psi K^+$ decays. The difference between the two ratios is used to correct the decay time efficiency for $B_s^0 \rightarrow \mu^+ \mu^-$ decays.

Selection	4-8 GeV	8-12 GeV	12-16 GeV	16-20 GeV	20-30 GeV
pion 2016	0.98 ± 0.01	0.98 ± 0.03	0.86 ± 0.09	0.53 ± 0.07	0.72 ± 0.20
pion 2017	0.97 ± 0.01	0.92 ± 0.03	0.84 ± 0.09	0.99 ± 0.15	1.15 ± 0.70
pion 2018	1.07 ± 0.01	1.16 ± 0.04	0.93 ± 0.08	1.08 ± 0.22	0.62 ± 0.18
kaon 2016	1.21 ± 0.02	1.29 ± 0.03	1.31 ± 0.08	0.82 ± 0.04	2.39 ± 0.66
kaon 2017	1.12 ± 0.04	1.17 ± 0.11	1.22 ± 0.20	3.71 ± 0.82	0.59 ± 0.09
kaon 2018	1.13 ± 0.03	1.35 ± 0.09	1.19 ± 0.17	1.43 ± 0.28	0.89 ± 0.22

Table 5.16: Pion and kaon fake rate ratios for data over MC simulation for 2016, 2017, and 2018.

rate/MC fake rate) are summarized in Table 5.16 for muon $\text{BDT} > 0.45$. The pion scale factor is consistent with one, whereas kaon scale factors are 20% higher. The reason could be that kaon decay-in-flight may have larger decay angles, and exact details of such decay identification may be harder to simulate correctly.

The scale factors of the fake rate are applied to the peaking and semileptonic background, and a 25% systematic uncertainty is assigned to each fake muon.

5.8.5 Trigger efficiency

Events in this analysis used high-level trigger whose absolute efficiency is of the order 50-60%. Since the analysis dealt with low p_T muons, a significant fraction of the event fall into the turn-on part of the trigger efficiency. The inefficiency collectively comes from the L1 trigger requirement, which is responsible for the 10% of inefficiency, and the HLT failures to reconstruct the low p_T muons.

Since the trigger efficiency from MC is directly used as input to the final branching fraction measurements, the HLT and L1 efficiency between the data and MC are checked to ensure everything cancels out in ratio.

HLT trigger efficiency: The signal and normalization triggers are essentially identical, as shown in Table 3.2, with a couple of minor differences in the vertex probability and the

displacement requirement.

Despite the similarity of the triggers and the phase space of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ decays, one needs to make sure that the rapid change in the efficiency at the low momentum is not causing a significant bias in the branching fraction measurement. For that, the HLT trigger efficiency as a function of the muon momentum and psuedo-rapidity is measured using orthogonal datasets, i.e., the datasets that are defined without muon triggers. A special method, known as Tag-n-Probe method [97], is used to measure the HLT trigger efficiency. In this method, the tag muon is fired by an unprescaled muon trigger for the dimuon events, and the probe muon is required to pass the signal trigger (prescaled trigger will reduce the usable set of events to a minimum leading to large statistical uncertainties.)

Figure 5.23 shows the HLT trigger efficiency as a function of the muon p_T for two channels in era 2018.

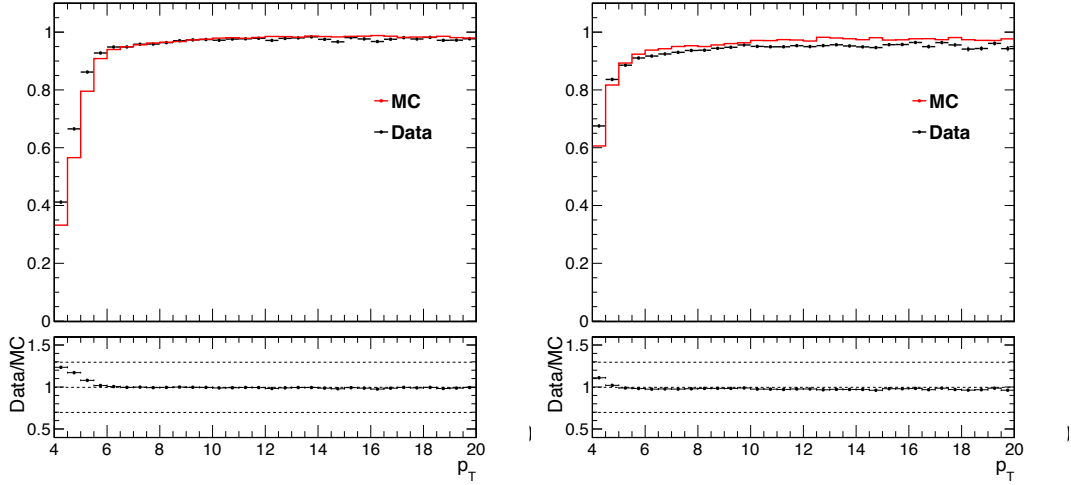


Figure 5.23: HLT trigger efficiency in data and simulated $B_s^0 \rightarrow \mu^+ \mu^-$ events as a function of the muon momentum for channel 0 (left) and channel 1 (right) for 2018.

At low momentum, a slightly higher trigger efficiency in data than MC is observed. Assuming the signal and normalization trigger are identical, the ratio of trigger efficiencies

in data and MC are calculated. The ratios are close to one, and a 1% systematic is assigned due to this effect.

Finally, the difference between the signal and the normalization triggers is considered. The main differences are the invariant mass selection, dimuon vertex requirement, and pointing angle requirement. An alternate reference trigger, without these selections, allows for the measurement of the trigger efficiency in signal and normalization trigger in data and MC simulations. Based on the ratio between $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ in data and MC, 1% systematic is assigned for 2017 and 2018 and 3% for 2016.

L1 trigger efficiency: The L1 trigger efficiency of the signal and normalization triggers is measured using an orthogonal data sample. The efficiency measured with respect to offline selections matches well between data and MC simulation for 2017 and 2018. The minor differences are considered a systematic uncertainty. The systematic uncertainty assigned for 2017 and 2018 is 2%, and for 2016, it is 3%.

5.9 Systematic uncertainty

5.9.1 Systematic uncertainties on the effective lifetime

A number of possible systematics checks are performed for the effective lifetime measurement. In this section, only the dominant contributions are discussed.

5.9.1.1 Efficiency modeling

Two alternative parametric functions, Eqs. [5.20](#) and [5.21](#), are tested on the normalization decay mode ($B^+ \rightarrow J/\psi K^+$) to estimate the contribution of systematic uncertainty due to the choice of efficiency function. The maximum deviation (0.01 ps) obtained from different channels after using two other efficiency functions with respect to default efficiency

function (Eq. 5.10) is taken as a systematic uncertainty.

$$\epsilon(t; p_0, p_1, p_2, p_3, p_4, p_5) = \exp(p_0 + p_1 t) \cdot \text{Chebychev4}(t) \quad (5.20)$$

$$\epsilon(t; p_0, p_1, p_2, p_3, p_4) = p_0 + p_1 t^{p_2} \exp(tp_3 + t^2 p_4) \quad (5.21)$$

5.9.1.2 Scanning lifetimes from different generated signal samples

As the effective lifetime has not been measured precisely, its true value is not known. A dedicated study has been performed to check the effect of the choice of input lifetime value while generating the MC sample. A set of MC samples are generated using the lifetime value ranging from 1.40 ps to 1.80 ps (with 0.05 ps increments). Then efficiency is calculated for each sample and used as correction on the other samples. Figure 5.24 shows the difference between fitted and generated lifetime values for a set of samples after using a particular efficiency function. For example, the left plot shows the fitted to generated lifetime difference in sample range 1.40 -1.80 ps for particular efficiency function taken from 1.50 ps sample. As can be seen, the fitted results (or, in other words, the fitted-generated lifetime (y-axis)) fluctuate up and down while using the same efficiency functions. The overall shifts are estimated by fitting this distribution to a straight line. The maximum shift is 0.01 ps which will be taken as a systematic uncertainty.

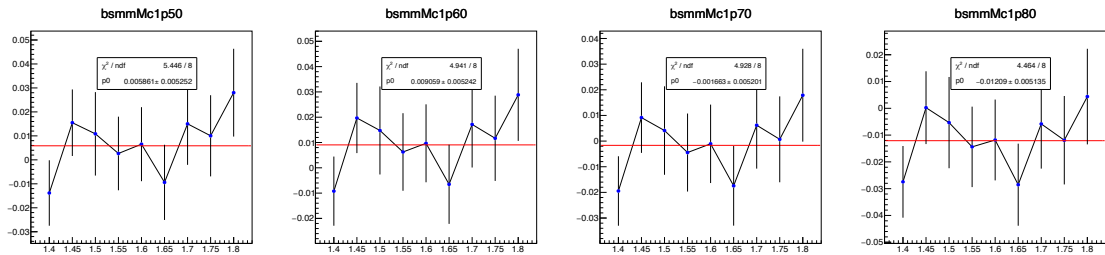


Figure 5.24: The results are from 2018 MC sample. The left plot shows the fitted-generated (y-axis) lifetime from different samples (with a generated lifetime in the range of 1.40 ps - 1.80 ps), while using the efficiency from the 1.50 ps MC sample. Similar plots are shown with different efficiency functions.

5.9.1.3 Combinatorial background decay time shape

The dominant background contribution under the signal is the combinatorial background. Therefore, a good knowledge of the combinatorial decay time distribution is necessary to obtain the correct signal lifetime. In the lifetime fitter, each channel uses a single exponential function to describe the combinatorial decay time shape. So, it is necessary to check if the function is sufficient to describe the combinatorial background.

The combinatorial background shape may vary for different d_{MVA} selections. The test has been performed in three d_{MVA} regions, one is the low d_{MVA} region, (0.2, 0.8) of data where the signal contribution is negligible. The second and third regions are the high sideband ($m \in [5.5, 5.9]$) of data for d_{MVA} region (0.90, 1.0) and d_{MVA} region (0.99, 1.0). The looser d_{MVA} regions have a larger number of events, and it is clear that a double exponential distribution describes data better, as seen in Fig. 5.25. With an increase in the d_{MVA} selection, the difference between the two functions becomes smaller. In addition to that, the shape of the distribution changes itself as a function of the d_{MVA} . To gain more insight from the limited statistics in $d_{\text{MVA}} > 0.99$ region, all the channels are merged to give a single distribution. The fit projections for different d_{MVA} selections are shown in Fig. 5.25.

The combined fits for $d_{\text{MVA}} > 0.99$ show that both models describe data reasonably well. By keeping in mind that the level of the background in the signal box is smaller than in the sideband, we expect a small difference between the two models for the signal region. Nevertheless, a set of pseudo-experiments generating data with single (hypothesis 1) and double (hypothesis 2) exponential models using parameters from the combined fits are performed. The pseudo-data is fitted with our nominal fitter using a single exponential model for the background.

The distributions for the fitted mean values, positive and negative uncertainties, and the pulls for the lifetime are shown in Fig. 5.26 (hypothesis 1). We found the median of the

fitted B_s^0 lifetime between the two hypotheses is consistent, indicating that the lifetime bias is negligible. At the same time, the lifetime uncertainty is slightly larger for the second hypothesis. From this study, we conclude that in the real data, if the combinatorial background decay time has some different shape, then we may see a larger lifetime error but the mean value will not change.

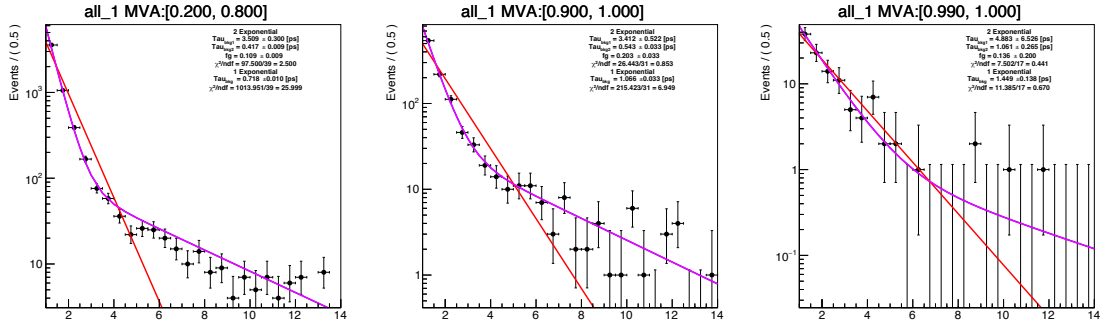


Figure 5.25: Combined fit projections for d_{MVA} (0.2, 0.8), d_{MVA} (0.90, 1.0), and d_{MVA} (0.99, 1.0). Single and double exponential functions are shown in red and magenta color. The combined fit is performed after merging data in different channels and eras.

5.9.2 Systematic uncertainties on branching fraction

The tracking efficiency difference between data and MC is estimated by using the D^{*+} decay mode, where $D^{*+} \rightarrow D^0 \pi^+$. The method calculates the ratio of event yields in $D^0 \rightarrow K \pi$ and $D^0 \rightarrow K \pi \pi \pi$. A systematic uncertainty of 2.3% is assigned on the track efficiency from the study [98]. Similarly, the data and MC simulation pileup distribution differences are expected to cancel out in the efficiency ratio of $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ efficiencies. The residual effect is taken as systematic uncertainty, which is 1%.

The triggers used in the normalization and signal channel are different. The dimuon vertex probability requirement in the trigger are 0.025 and 0.1 for normalization and signal channel respectively. The effect of this requirement may differ in data and MC, which leads to a systematic uncertainty. The sWeighted data and MC simulations as a function of vertex

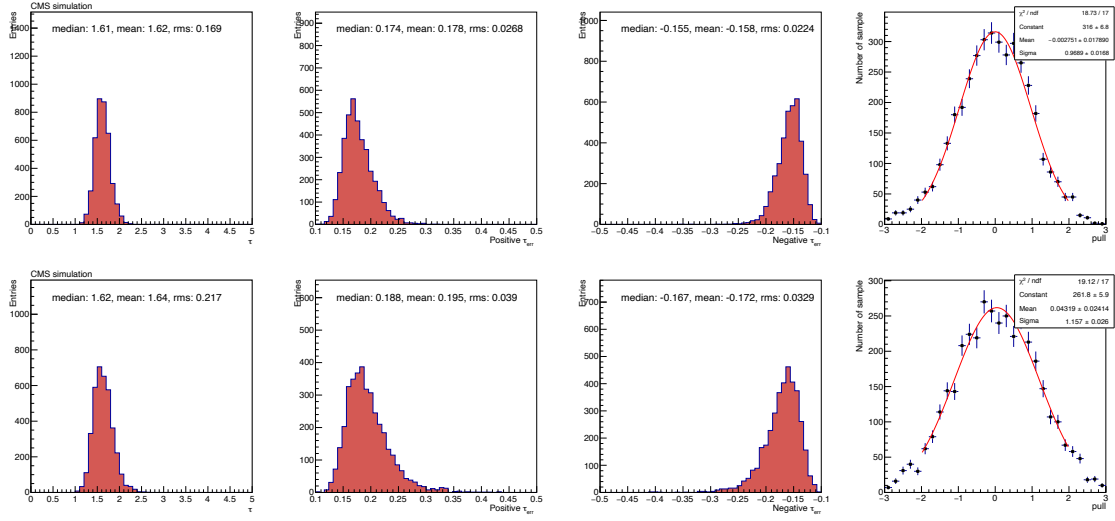


Figure 5.26: The top row (hypothesis 1) and bottom row (hypothesis 2) of the figure show the B_s^0 lifetime, B_s^0 lifetime positive and negative error and the pull distribution from the 3920 toy sample. The pull is calculated using the generation value of lifetime to be 1.615 ps.

probability in $B^+ \rightarrow J/\psi K^+$ channel are compared, and the efficiency difference (1%) in data and MC are taken as a systematic uncertainty.

The detail of trigger correction are discussed in Section 5.8.5. Since the study has been performed in three parts: L1 trigger, common HLT trigger, and the different selection in the HLT trigger, the systematics is derived for three parts. A 2% systematic uncertainty is assigned to the L1 trigger, a 1% systematic uncertainty to the common part of the HLT trigger, and 3 (1)% to the different part of the HLT trigger in 2016 (2017, 2018). Therefore, the total systematic uncertainty is 3.74% for the 2016 trigger and 2.44% for the 2017 and 2018 triggers.

The potential fit bias on the branching fraction extraction with the UML fit is checked using pseudo-experiments. A non-trivial bias on the $B^0 \rightarrow \mu^+ \mu^-$ branching fractions is observed, will be shown in Fig. 5.28. The fit bias is measured using median expected branching fractions and the truth input values used in the pseudo-experiment. A 2% (8%)

bias is observed for the $B_s^0 \rightarrow \mu^+ \mu^-$ ($B^0 \rightarrow \mu^+ \mu^-$) branching fraction. The issue was investigated by using several PDF models for different background components. In the end, it was found that the main contribution is from the semileptonic background component in the fit and partially due to limited statistics in each component.

5.9.3 Signal efficiency dependence on the lifetime

The lifetime of B_s^0 mesons has a significant impact on the signal efficiency of $B_s^0 \rightarrow \mu^+ \mu^-$. The total efficiency obtained from different lifetime samples is shown in Fig. 5.27 (left). The difference between the efficiency corresponds to a longer lifetime sample (1.62 ps) and a shorter lifetime (1.42 ps [99]) sample is $\sim 10\%$. Other eras and channels have similar distributions. Thus the effect of signal efficiency on the branching fraction is tested using different efficiency values in the pseudo-experiments (toys). The toys are generated with different signal efficiency and fitted assuming SM lifetime sample efficiency. The result of the toy study is plotted and shown in Fig. 5.27 (right). The branching fraction dependence for different lifetime samples is fitted with the polynomial function. The fitted function will be useful to estimate the branching fraction for any other lifetime value. The correction factor for the branching fraction is $1.577 - 0.358\tau$, where τ is the B_s^0 meson lifetime in ps. Here one thing to note that the signal efficiency is taken from the sample having the SM lifetime, so no systematic uncertainty is assigned for this test.

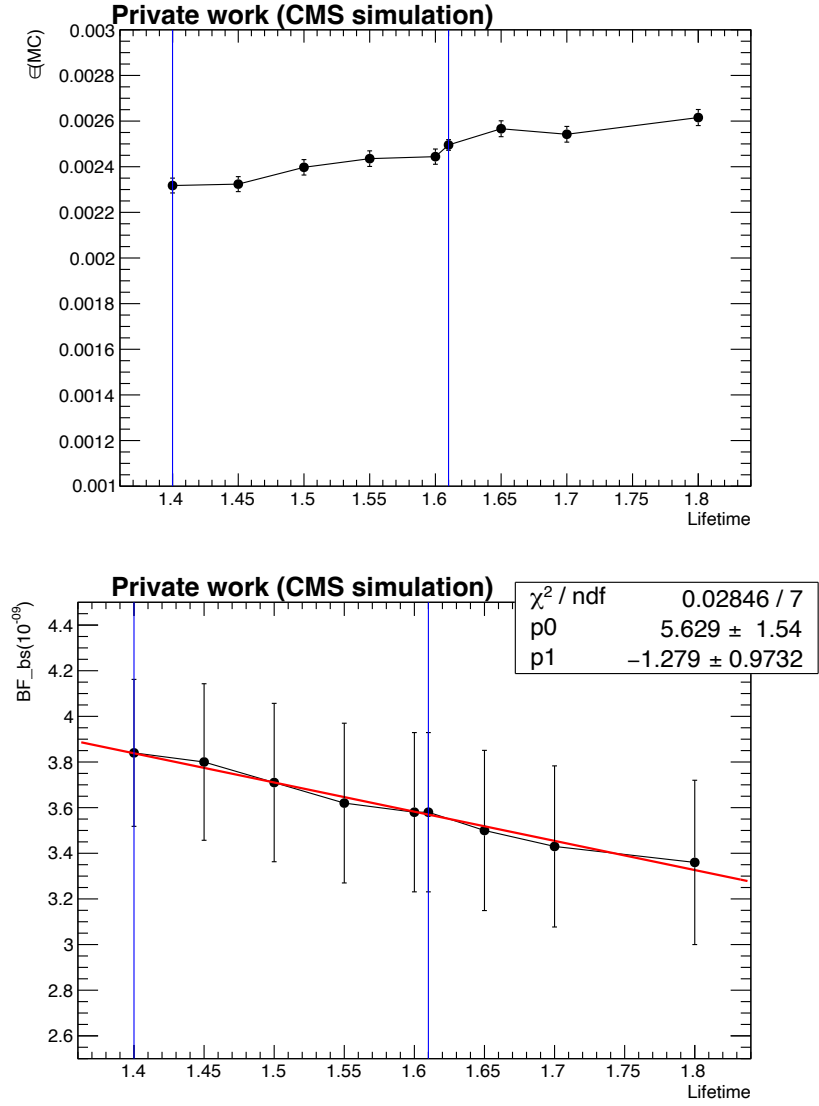


Figure 5.27: Signal efficiency from 2018 channel 0 (top) and the median of $B_s^0 \rightarrow \mu^+ \mu^-$ BF obtained from the toy MC simulation for different lifetime assumptions (bottom) are shown. Error on each point (in the bottom plot) corresponds to the median of the BF error.

5.9.4 Systematics summary

Table 5.17: Summary of the systematic uncertainties for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements.

Effect	$B_s^0 \rightarrow \mu^+ \mu^-$	$B^0 \rightarrow \mu^+ \mu^-$
Trigger efficiency		2–4%
Pileup		1%
Vertex quality requirement		1%
d_{MVA} correction		2–3%
Tracking efficiency (per kaon)		2.3%
$B^+ \rightarrow J/\psi K^+$ shape uncertainty		1%
$B_s^0 \rightarrow J/\psi \phi$ shape uncertainty		1.5%
Fit bias	2.2%	4.5%
f_s/f_u ratio of the B meson production fractions	3.5%	-

Effect	2016BF	2016GH	2017	2018
Efficiency modeling		0.01		
Scanning lifetime from different generated signal sample		0.01		
Correction method systematic	0.10	0.06	0.02	0.02
Lifetime bias	0.04	0.04	0.05	0.04
Total	0.11	0.07	0.05	0.04

Table 5.18: Summary of systematic uncertainties for $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime measurement (ps).

5.10 Expected performance for branching fraction measurement

In this section, the expected performance of the branching fraction analysis is estimated based on the pseudo-experiments using $B^+ \rightarrow J/\psi K^+$ as normalization. Each pseudo-experiment consists of the expected number of event for signal and backgrounds in the full

mass region. The performance has been estimated for two different configurations: the “1-bin” setup introduces a single threshold for each channel and data-taking era; the “2-bin” setup includes the events with lower analysis d_{MVA} score, as the secondary bins in each channel. Hence, the “1-bin” (“2-bin”) setup contains 8 (16) categories in the UML fits.

The input yields used in the toy MC event generation are summarized in Table 5.19. Using several thousands toy MC’s, the performance is estimated and summarized in Table 5.20. The “2-bin” configuration has lower relative uncertainty on branching fraction and larger significance than the “1-bin” setup. The distributions for the fitted mean values, positive and negative uncertainties, and the pulls, as well as the expected significances, are shown in Fig. 5.28 for “2-bin” configurations.

Era	Channel	N_{bs}	N_{bd}	N_{comb}	N_{peak}	N_{semi}
$d_{\text{MVA}} > 0.99$						
2016BF	0	5.4 ± 0.4	0.62 ± 0.07	5.0 ± 3.5	0.24 ± 0.12	6.0 ± 1.5
2016BF	1	9.6 ± 0.8	1.08 ± 0.12	10.0 ± 5.0	0.49 ± 0.25	10.9 ± 2.7
2016GH	0	6.5 ± 0.5	0.74 ± 0.08	1.0 ± 1.0	0.27 ± 0.14	8.7 ± 2.2
2016GH	1	10.1 ± 0.8	1.14 ± 0.12	12.5 ± 5.6	0.51 ± 0.26	13.5 ± 3.4
2017	0	23.7 ± 1.8	2.78 ± 0.29	45.0 ± 10.6	0.91 ± 0.46	27.7 ± 6.9
2017	1	34.4 ± 2.6	3.93 ± 0.35	92.5 ± 15.2	1.69 ± 0.85	45.9 ± 11.5
2018	0	35.1 ± 2.7	4.05 ± 0.42	50.0 ± 11.2	1.53 ± 0.77	45.0 ± 11.3
2018	1	51.0 ± 3.9	5.86 ± 0.61	127.5 ± 17.9	2.83 ± 1.42	62.8 ± 15.7
$0.99 > d_{\text{MVA}} > 0.9$						
2016BF	0	4.8 ± 0.4	0.56 ± 0.06	112.5 ± 16.8	0.21 ± 0.11	8.5 ± 2.1
2016BF	1	9.0 ± 0.7	1.04 ± 0.11	287.5 ± 26.8	0.49 ± 0.25	16.0 ± 4.0
2016GH	0	5.6 ± 0.5	0.67 ± 0.07	82.5 ± 14.4	0.21 ± 0.11	10.0 ± 2.5
2016GH	1	9.3 ± 0.8	1.08 ± 0.12	235.0 ± 24.4	0.51 ± 0.26	17.8 ± 4.5
2017	0	15.3 ± 1.2	1.80 ± 0.19	535.0 ± 36.6	0.81 ± 0.41	26.9 ± 6.7
2017	1	22.0 ± 1.7	2.60 ± 0.27	1297.5 ± 57.0	1.26 ± 0.63	45.4 ± 11.4
2018	0	23.3 ± 1.8	2.76 ± 0.29	715.0 ± 42.3	1.39 ± 0.65	44.4 ± 11.1
2018	1	34.3 ± 2.6	4.06 ± 0.42	1962.5 ± 70.0	2.14 ± 1.07	78.4 ± 19.6

Table 5.19: Expected yields for each component and for each category included in the UML implementation. These yields are used in the toy event generation for the expected performance estimations.

Observable	relative uncertainties	expected significance
“1-bin” configuration		
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	+11.2%/-10.5%	13.1σ
$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	+67.6%/-62.9%	1.64σ
“2-bin” configuration		
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	+10.5%/-9.9%	13.7σ
$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	+65.2%/-61.0%	1.75σ

Table 5.20: Summary of expected relative uncertainties on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, as well as the expected significance from the pseudo-experiments in “1-bin” and “2-bin” configurations.

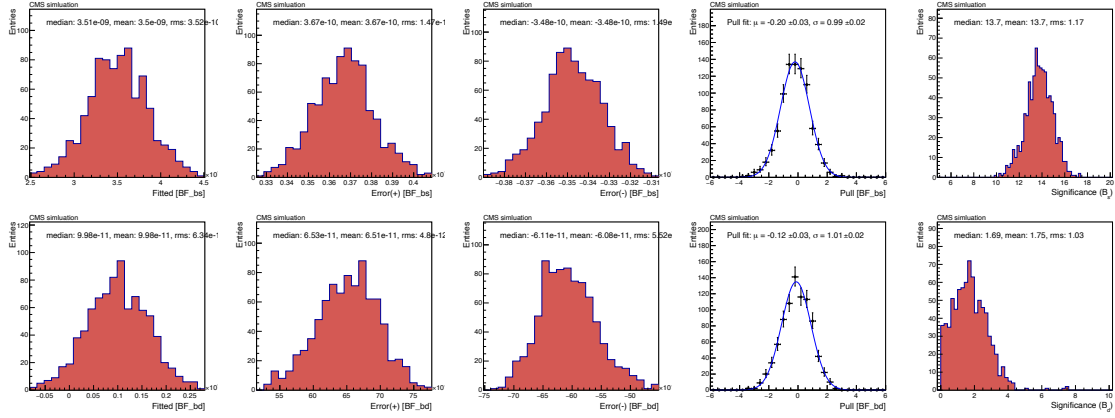


Figure 5.28: From left to right, the distributions of fitted mean, hi/low uncertainties, pull, and expected significance for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (upper row) and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ (bottom row) from the pseudo-experiments. The results are obtained from the “2-bin” configuration with fits to 16 categories.

5.10.1 Expected performance using $B_s^0 \rightarrow J/\psi \phi$ as normalization

The second set of $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction results is measured by replacing the normalization $B^+ \rightarrow J/\psi K^+$ to $B_s^0 \rightarrow J/\psi \phi$. This will allow excluding the fragmentation factor f_s/f_u in the final equation. The yield calculations in Eq. 5.5 can be replaced as,

$$N_{\text{bs}}^i = \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \times \frac{\varepsilon_{B_s^0 \rightarrow \mu^+ \mu^-}^i}{\varepsilon_{B_s^0 \rightarrow J/\psi \phi}^i} \times \frac{N_{B^+ \rightarrow J/\psi K^+}^i}{\mathcal{B}_{B_s^0 \rightarrow J/\psi \phi} \mathcal{B}_{J/\psi \rightarrow \mu^+ \mu^-} \mathcal{B}_{\phi \rightarrow K K}} \quad (5.22)$$

where $\varepsilon_{B_s^0 \rightarrow J/\psi \phi}^i$ are the total efficiencies (including the effects of detector acceptance) for $B_s^0 \rightarrow J/\psi \phi$ decays. $\mathcal{B}_{B_s^0 \rightarrow J/\psi \phi}$ is the branching fraction of $B_s^0 \rightarrow J/\psi \phi$. To date, the

most precise $B_s^0 \rightarrow J/\psi \phi$ branching fraction is by the LHCb experiment, which uses $B^+ \rightarrow J/\psi K^+$ as normalization.

- $\mathcal{B}(B_s^0 \rightarrow J/\psi \phi) = (1.018 \pm 0.05) \times 10^{-3}$ [100],
- $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033) \times 10^{-2}$ [94],
- $\mathcal{B}(\phi \rightarrow K^+ K^-) = (0.492 \pm 0.005)$ [94].

Similarly, other background yield calculations will be modified by replacing $B^+ \rightarrow J/\psi K^+$ with $B_s^0 \rightarrow J/\psi \phi$. But this decay mode will add two additional systematics, one on the $\mathcal{B}_{B_s^0 \rightarrow J/\psi \phi}$ and another on the extra kaon reconstruction in $\varepsilon_{B_s^0 \rightarrow J/\psi \phi}$. The expected performance is summarized in Table 5.21. The $B_s^0 \rightarrow \mu^+ \mu^-$ expected significance is lower compared to result from $B^+ \rightarrow J/\psi K^+$ normalization, mainly due to yield ratio $N_{B_s^0}/N_{B^+}$ in data is smaller compared to f_s/f_u ($f_s/f_u = 0.252$ is used in the toy study). The $B^0 \rightarrow \mu^+ \mu^-$ signal has the same yield but slightly lower significances due to the overall larger systematic uncertainty and decorrelated uncertainties between branching fractions of $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$.

Observable	relative uncertainties	expected significance
“1-bin” configuration		
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	+12.9%/-11.9%	11.8σ
$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	+73.9%/-68.9%	1.51σ
“2-bin” configuration		
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	+12.2%/-11.4%	12.2σ
$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	+71.1%/-66.8%	1.58σ

Table 5.21: Summary of expected relative uncertainties on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$, as well as the expected significance from the pseudo-experiments in “1-bin” and “2-bin” configurations using $B_s^0 \rightarrow J/\psi \phi$ normalization.

5.11 Expected performance for effective lifetime measurement

The expected performance of the lifetime measurement is evaluated with a set of toy MC studies. Using the prefit model, 5000 toy samples are generated, and the fit is performed on them to get the lifetime, lifetime error, and pull. The following steps are adopted to generate the toy MC sample.

- The correction derived from the data (in Section 5.8.3) are applied to the signal efficiency obtained from $d_{\text{MVA}} > 0.90$. The mass, decay time, and decay time uncertainties for the signal, semileptonic and peaking PDFs are obtained from the weighted (weighting wrt to expected yields) simulated samples. The shapes corresponding to combinatorial background are obtained from di-muon high sideband of the data.
- The signal and background expected yields are generated with Poisson fluctuations in each toy.
- Global constraints such as the branching fraction of $B^+ \rightarrow J/\psi K^+$ are randomized (generate randomly) once for eight channels.
- The lifetime systematics, as mentioned in Table 5.18, are treated as correlated nuisance parameters in the fit and randomized in each toy.
- We then randomize the constraint on the peaking background yield and semileptonic yield, for each channel.

After creating the eight different datasets for eight channels, a simultaneous fit to all channels is performed where a single lifetime and a single branching fraction for signal are extracted.

Figure 5.29 shows the fitted mean values, positive and negative uncertainties, and the pulls for the effective lifetime and branching fraction. The fit bias in the lifetime is negligible, and no systematic is assigned due to fit bias. The expected uncertainty on the lifetime is (+0.18, -0.16) ps.

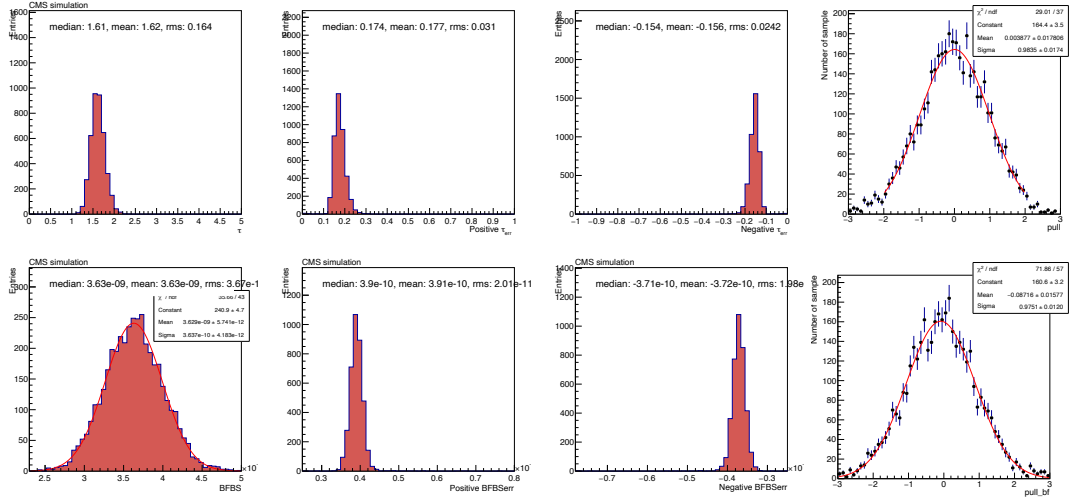


Figure 5.29: The top row of the figure shows B_s^0 lifetime, B_s^0 lifetime positive and negative error, and the pull distribution from the 5000 toy sample. The pull is calculated using the generation lifetime value of 1.62 ps. The bottom row shows the B_s^0 branching fraction, branching fraction positive and negative error, and the branching fraction pull distributions. The fitted means of the pull distributions are at zero.

5.12 Fragmentation fraction treatment

The fragmentation fraction ratio is an essential external input for the branching fraction measurement. Experimental precise measurement of this ratio is necessary to limit the systematics on the branching fraction. Such measurements are performed in the fiducial phase space and may differ from one experiment to another. In particular, the ratio may depend on the momentum and psuedo-rapidity of the B meson. Therefore, it is important to consider the effective phase space of $B_s^0 \rightarrow \mu^+ \mu^-$ decays to obtain a reliable result.

The results from the LHCb experiment cover the low momentum and high pseudo-rapidity region, but our analysis is for the high momentum ($p_T > 10$ GeV) and low pseudo-rapidity ($|\eta| < 1.4$) region. The p_T distribution of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ decay in data and MC are shown in Fig. 5.30.

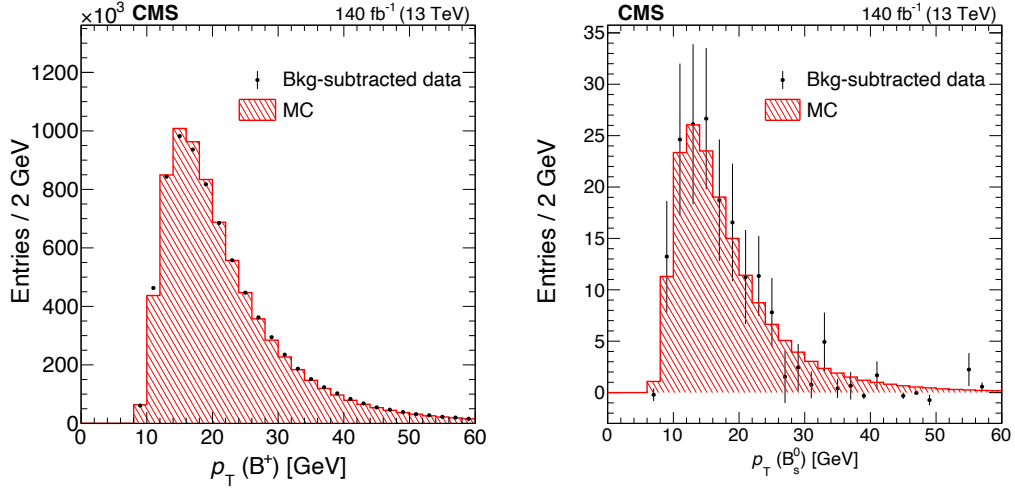


Figure 5.30: The distribution of the B meson p_T from the sPlot fit in data is compared with simulated B meson p_T in both $B^+ \rightarrow J/\psi K^+$, and $B_s^0 \rightarrow \mu^+ \mu^-$ region.

Since there are no measurements from the CMS experiment, the final value used in this analysis is obtained by integrating the ratio from the LHCb experiment in p_T bins. The p_T dependent LHCb value is [100],

$$f_s/f_u(p_T) = (0.263 \pm 0.008) + ((-17.6 \pm 2.1) \times 10^{-4}) \cdot p_T. \quad (5.23)$$

The integrated f_s/f_u after using the B_s^0 meson p_T at generator level is 0.231 ± 0.008 . The total uncertainty includes the variation for different d_{MVA} selections and channels. The integrated f_s/f_u results are almost identical when calculated with the reconstructed B meson p_T instead of generator-level p_T from the B_s^0 signal samples.

5.13 Results

5.13.1 Branching fraction results with $B^+ \rightarrow J/\psi K^+$ as normalization

In this section, the results of the unblinded fits on data are discussed. The branching fractions are found to be:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = [3.83_{-0.36}^{+0.38}(\text{stat.})_{-0.16}^{+0.19}(\text{syst.})_{-0.13}^{+0.14}(\text{frag.})] \times 10^{-9}, \quad (5.24)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = [0.37_{-0.67}^{+0.75}(\text{stat.})_{-0.09}^{+0.08}(\text{syst.})] \times 10^{-10}. \quad (5.25)$$

The mass projections for each era and channel are available in Appendix (Figs. [B.1](#) and [B.2](#)). Figure [5.31](#) displays the combined mass projections obtained by merging the projections of mass in two MVA bins.

The observed significances evaluated from the log-likelihood difference are 12.5σ and 0.51σ for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays, respectively. Since the observed $B^0 \rightarrow \mu^+ \mu^-$ yields are statistically insignificant, an upper limit on the $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ is evaluated using the full toy-based CLs technique. The values are

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.5 \times 10^{-10} \text{ at } 90\% \text{ C.L.} \quad (5.26)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) < 1.9 \times 10^{-10} \text{ at } 95\% \text{ C.L.} \quad (5.27)$$

Figure [5.32](#) shows the upper limit on $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ from the full toy-based CLs technique and two-dimensional profile likelihood scan as a function of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ branching fractions.

The impacts of nuisance parameters on the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction are checked. The uncertainty on $B^+ \rightarrow J/\psi K^+$ branching fraction and the kaon tracking systematic uncertainty have the largest impact at about 20% of the total uncertainty each. Whereas the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction measurement is statistically limited, and all nuisances have negligible impact.

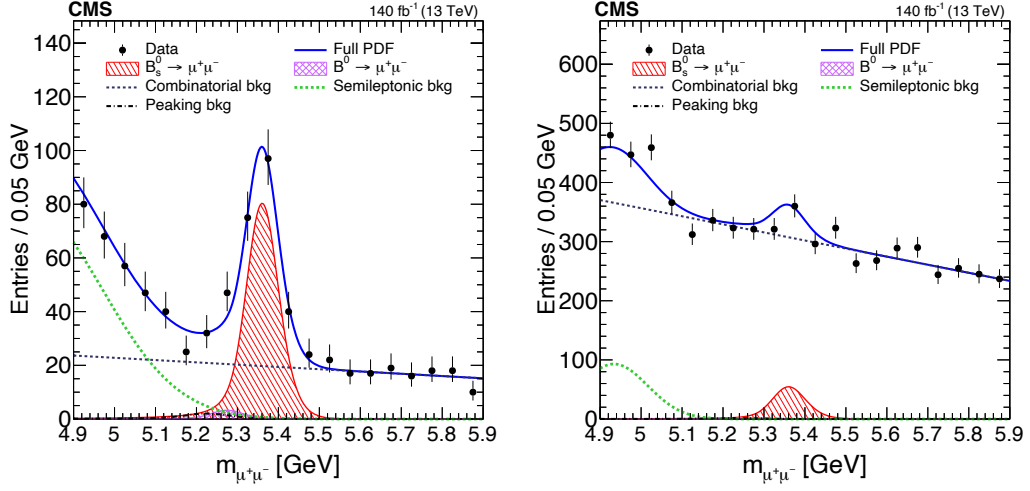


Figure 5.31: The merged dimuon invariant mass distributions of the unblinded fit in the $d_{\text{MVA}} > 0.99$ category (left) and $0.99 > d_{\text{MVA}} > 0.90$ category (right), with the PDF models used in the generation overlaid.

5.13.1.1 Branching fraction results with $B_s^0 \rightarrow J/\psi \phi$ as normalization

The second set of results on the branching fractions with $B_s^0 \rightarrow J/\psi \phi$ normalization are found to be:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = [3.95_{-0.37}^{+0.39}(\text{stat.})_{-0.22}^{+0.27}(\text{syst.})_{-0.19}^{+0.21}(\text{BF})] \times 10^{-9}, \quad (5.28)$$

$$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = [0.34_{-0.66}^{+0.75}(\text{stat.})_{-0.08}^{+0.09}(\text{syst.})] \times 10^{-10}. \quad (5.29)$$

As it can be seen, the systematic uncertainty on the branching fraction is dominated by the uncertainty on $\mathcal{B}(B_s^0 \rightarrow J/\psi \phi)$.

5.13.2 Lifetime measurement

The fitted decay branching fraction and lifetime for $B_s^0 \rightarrow \mu^+ \mu^-$ obtained from the lifetime unblinded fit are

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 3.89_{-0.40}^{+0.42} \times 10^{-9} \quad (5.30)$$

$$\tau_{B_s^0} = 1.83_{-0.20}^{+0.23}(\text{stat})_{-0.04}^{+0.04}(\text{syst}) \text{ ps} \quad (5.31)$$

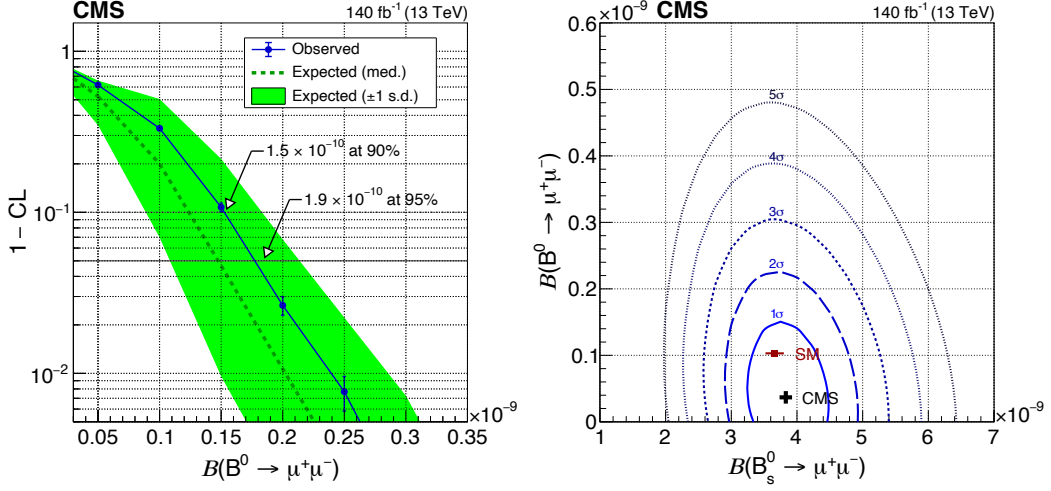


Figure 5.32: The left plot is the upper limits on $B^0 \rightarrow \mu^+\mu^-$ decay branching fraction from the full toy-based CLs technique, and the right plot is the two-dimensional profile likelihood scan for $B(B_s^0 \rightarrow \mu^+\mu^-)$ and $B(B^0 \rightarrow \mu^+\mu^-)$. The 2D contours enclose the regions with 1–5 standard deviation coverage, where the 1 σ , 2 σ , and 3 σ regions correspond to 68.3%, 95.4%, and 99.7% the confidence levels, respectively.

The fitted branching fraction is in agreement with the value obtained from the main fitter (Eq. 5.24). The uncertainties are only statistical and larger because the lifetime fit restricts the decay time to the [1, 14] ps window, and the branching fraction fit uses the mass uncertainty as an observable.

The lifetime of B_s^0 is consistent with SM value at 1 σ level and LHCb published result. The observed lifetime error is larger compared to the expected median error from the toy study, which is the effect of larger central value. The merged mass, decay time and decay time uncertainty projections are shown in Fig. 5.33. The decay time in the signal window [5.28, 5.48] GeV is shown in Fig. 5.34.

The impacts of nuisance parameters are checked and the effect on the lifetime variable is tiny compared to statistical uncertainty.

The result of branching fraction and the lifetime in individual year and channels are mentioned in Tables B.1 and B.3. A simple goodness of fit test is performed by measuring

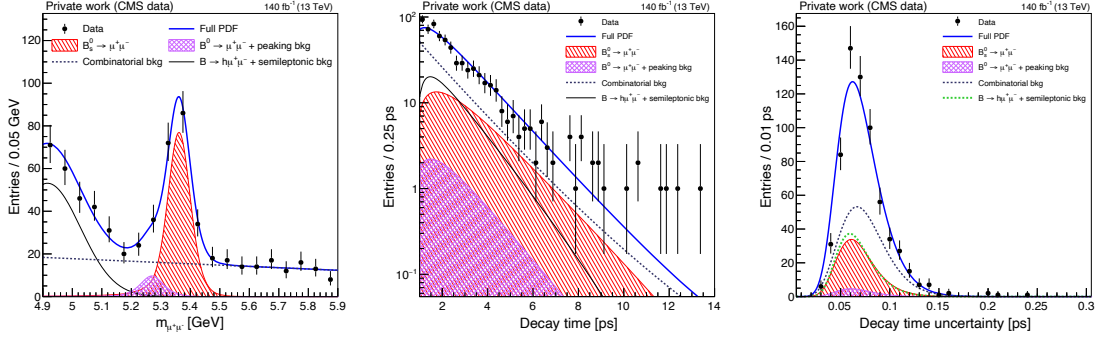


Figure 5.33: The merged dimuon invariant mass, decay time and decay time uncertainty distribution in the $d_{MVA} > 0.99$ categories obtained from the lifetime fit in the data. The channel-wise mass, decay time and decay time uncertainty are projected in Figs. B.3-B.5.

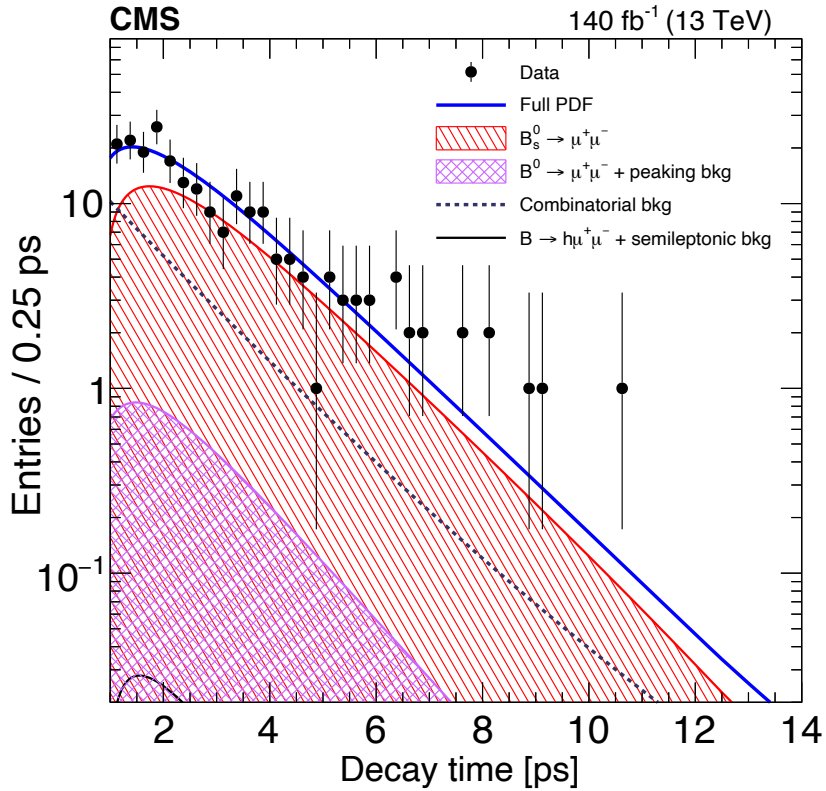


Figure 5.34: The merged decay time distribution in the signal region [5.28, 5.48] GeV is shown. The combined fit PDF along with signal and different background components are projected.

$\chi^2/\text{number-of-bins}$ and the results for individual year and channels from branching fraction and lifetime fit are mentioned in Tables B.2 and B.4. We found that the $\chi^2/\text{number-of-bins}$ values are acceptable. Finally, a beautiful visualization of $B_s^0 \rightarrow \mu^+ \mu^-$ candidate is shown in Fig. B.6.

B

5.14 Summary

This chapter presents the recent studies of $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and the B_s^0 meson effective lifetime and the search for the rare $B^0 \rightarrow \mu^+ \mu^-$ decay with the data collected at 13 TeV center-of-mass energy by the CMS experiment. The total integrated luminosity corresponds to 140 fb^{-1} . The measured $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = [3.83_{-0.36}^{+0.38}(\text{stat.})_{-0.16}^{+0.19}(\text{syst.})_{-0.13}^{+0.14}(\text{frag.})] \times 10^{-9}$ where the first, second, and third uncertainties are related to the statistical, systematic, and fragmentation fractions. No signal $B^0 \rightarrow \mu^+ \mu^-$ is observed and the branching fraction $\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$ is observed to be less than 1.9×10^{-10} at 95 % confidence level. The effective lifetime of the B_s^0 meson is found to be $\tau_{B_s^0} = 1.83_{-0.20}^{+0.23}(\text{stat.})_{-0.04}^{+0.04}(\text{syst.}) \text{ ps}$, where the first error is the statistical uncertainty and the second one systematic. The branching fraction and effective lifetime results are the most precise to date and in agreement with the Standard Model predictions and other experimental results.

Chapter 6

Conclusion and Outlook

Rare B meson decays to dimuon final states allow a unique insight into the $b \rightarrow sl^+l^-$ decays, and measurement of their properties provides a testing ground for the validation of the new physics models. Thanks to CERN LHC data, the observation of the rare $B_s^0 \rightarrow \mu^+\mu^-$ decay was made possible with the combination of CMS and LHCb data collected at COM energies of 7 TeV and 8 TeV in 2011 and 2012 [48], respectively. Later, the LHC underwent a major upgrade to cope with higher pp collision COM energy (13 TeV) and instantaneous luminosity. This allowed us to study the $B_s^0 \rightarrow \mu^+\mu^-$ properties with more events.

In this thesis, I have presented two analyses measuring the branching fraction and the effective lifetime of the B_s^0 meson in the $B_s^0 \rightarrow \mu^+\mu^-$ decay and searched for the $B^0 \rightarrow \mu^+\mu^-$ decay using the CMS data. The first analysis uses the data sample collected by CMS experiments in 2011 at 7 TeV, 2012 at 8 TeV, and 2016 at 13 TeV, which correspond to 5 fb⁻¹, 20 fb⁻¹, and 36 fb⁻¹, respectively. With this data, CMS observed the $B_s^0 \rightarrow \mu^+\mu^-$ decay with more than 5σ significance for the first time. The second analysis uses the data collected in 2016, 2017, and 2018 at 13 TeV. The total integrated luminosity corresponds to 140 fb⁻¹.

In the analysis, the signal events are reconstructed by considering the two opposite sign muons from a common vertex. The events are selected by applying preselection criterias on the different kinematic and vertex variables. The analysis dealt with large combinatorial and semileptonic backgrounds, which limits the analysis sensitivity. In the first analysis, a BDT algorithm in the TMVA package is trained using B candidate variables as input. The algorithm uses various combinations of the input variables, resulting in a single powerful

variable that can be used to suppress the background. In the second analysis, the preselection requirements on the variables are released, and a popular gradient boost algorithm using the XGBoost library is adopted to improve the sensitivity. Both the algorithms are strong enough to suppress the dominant backgrounds to a reasonable level. Similarly, another important achievement of the analysis is to select good muon candidates using the advanced muon identification algorithm with the help of BDT. The algorithm is able to suppress the fake hadrons to a controllable level.

Figure 6.1 shows the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction, the B_s^0 meson effective lifetime, and the upper limit on the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction obtained from the different experiments. All the individual experimental measurements are consistent with the SM prediction, whereas the combined analysis measurements show a slight deviation. A combined analysis was performed using the results from CMS (Run 1 + 2016 data), ATLAS (Run 1 + 2015 + 2016 data), and LHCb (Run 1 data). The combined branching fraction results are consistent with the SM prediction at 2.5σ . Similarly, the effective lifetime is consistent with the CP-odd state (SM) at 0.8σ and the CP-even state at 1.4σ .

Figure 6.1 also includes the CMS measurements with the Run 2 data of integrated luminosity 140 fb^{-1} . The relative error on the branching fraction and effective lifetime of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay are the most precise results obtained by a single experiment. However, the analysis is still statistically limited; more data is needed to significantly constrain the new physics parameter space. Using the effective lifetime value in Eq. 1.32, the $\mathcal{A}_{\Delta\Gamma}$ value is consistent with the SM expectations $+1$.

The largest systematic uncertainty contributing to the branching fraction measurement is the fragmentation fraction ratio of B_s^0 and B^0 mesons, f_s/f_u . The systematic related to this is $\sim 3.5\%$ and is extrapolated from the LHCb measurements. A couple of measurements are ongoing in CMS that will clarify the p_T dependence of f_s/f_u and also be useful for this analysis because of the same phase space.

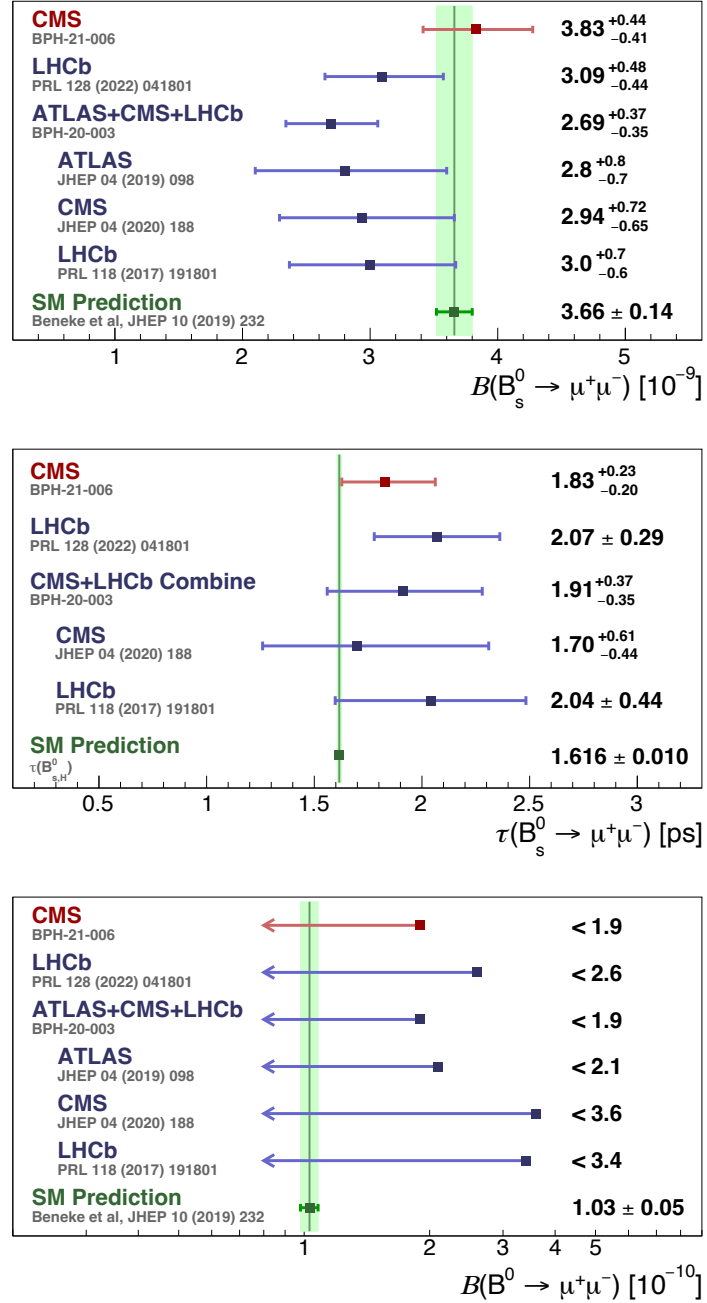


Figure 6.1: Summary of the $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction (top), the B_s^0 meson effective lifetime (middle) in the $B_s^0 \rightarrow \mu^+\mu^-$ mode and the $B^0 \rightarrow \mu^+\mu^-$ branching fraction upper limit (bottom) from different experiments.

No evidence of the $B^0 \rightarrow \mu^+\mu^-$ decay has been observed by any experiment. Only an upper limit on the branching fraction is set with 95% statistical confidence level, that is, two parts in a billion. This value is consistent with the combined analysis result [101]. The CMS detector is planning to undergo a phase-2 upgrade to cope with high luminosity. With this upgrade, the mass resolution is expected to improve significantly of the order 40-50%, and the semileptonic background leakage to $B^0 \rightarrow \mu^+\mu^-$ signal region can be reduced by 30%. Accounting for all these improvements and with an integrated luminosity of 3000 fb^{-1} , the expected performance is estimated using pseudo-experiments. The result infers that CMS will be able to observe the $B^0 \rightarrow \mu^+\mu^-$ decay with more than 5σ significance.

Over the last decade, several B meson decays mediated by $b \rightarrow s$ quark-level transition indicated tantalizing hints of new physics. For example, R_{K^*} anomalies (before December 2022) in $B \rightarrow K^{(*)}l^+l^-$ decay where $l = e, \mu$, anomaly in angular observable P'_5 of $B \rightarrow K^*\mu^+\mu^-$ decay, and branching ratio of $B_s^0 \rightarrow \phi\mu^+\mu^-$ decay. Since all the anomalies are observed in $b \rightarrow sl^+l^-$ processes, more evidently in $b \rightarrow s\mu^+\mu^-$ decay, the simultaneous anomalies can not be ignored. A global fit is performed prior to 2022, using the above anomalies along with $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ and the $b \rightarrow s\gamma$ branching fraction is used to interpret the result in terms of effective Wilson coefficients. The two Wilson coefficients sensitive to NP are C_{10}^μ and C_9^μ , and these coefficients are directly influenced by the branching fraction of $B_s^0 \rightarrow \mu^+\mu^-$ and angular observables in $B^0 \rightarrow K^{*0}\mu^+\mu^-$, respectively. From the global fit, the observed result was 5.8σ away from the SM value [102] and consistent with other experimental data results. In December 2022, a new publication regarding $R_{K^{(*)}}$ [103] was public by LHCb experiment where the deviation from the SM was found to be absent. It is expected that the current measurement of $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ and $R_{K^{(*)}}$ will push the average value to a large value. This implies that the anomalies observed in several $b \rightarrow sl^+l^-$ decays are more likely constrained only by operator $\mathcal{O}_9$.

Another interesting study, which could be possible with the HL-LHC dataset, is CP

violation in $B_s^0 \rightarrow \mu^+ \mu^-$ decays. The observable related to the CP violation study is the CP asymmetry, where the production rate of $B_s^0 \rightarrow \mu^+ \mu^-$ and $\bar{B}_s^0 \rightarrow \mu^+ \mu^-$ are measured. As the two decay modes have the same final state, flavor tagging is necessary. The flavor of the B_s^0 meson is deduced by tagging the opposite-sign muon in an event produced from the semileptonic b quark decay. The tagging power and efficiency achieved by CMS in $B_s^0 \rightarrow J/\psi \phi$ decay are $\sim 10\%$ and $\sim 50\%$ [104], respectively, which should be similar in the $B_s^0 \rightarrow \mu^+ \mu^-$ case. This implies a large dataset is needed to improve the tagging power and perform the CP violation study in the $B_s^0 \rightarrow \mu^+ \mu^-$ decay.

The $B_s^0 \rightarrow \mu^+ \mu^-$ decay has been observed by different LHC experiments with the Run 1 + Run 2 data. Now the focus is to measure the properties precisely and compare them with the SM predictions. In 2022, the LHC has started delivering pp collision data with record high energy, at 13.6 TeV, and will continue for the next four years. In addition, several upgrades in the detector components, data readout, computing, and selection systems were performed. The changes will improve the performance of the CMS detector and allow for the collection of more high-quality data. With the large dataset and advanced techniques, particle physicists will perform precise measurements of known processes to find the effects of new physics and also search for previously inaccessible processes.

Appendix A

Fake rate measurement

Misidentified low momentum muons are the dominant source of background for the rare $B^0 \rightarrow \mu^+ \mu^-$ decay. The contribution of these backgrounds is so large that the signal sensitivity will be biased without proper control. The background such as $B^0 \rightarrow K^+ \pi^-$ is one of the backgrounds that can mimic the $B^0 \rightarrow \mu^+ \mu^-$ decay. The source of hadron misidentification can be of any type. When the muon momentum has a significantly different value than the original hadron momentum, these fakes can be easily removed by applying criteria on the track matching in the muon system and tracker.

To preserve the momentum, the original hadron should either decay deep in the detector so that the inner tracker track contains mostly hits from the hadron or decay to a muon emitting a soft neutrino. In the first case, the hadron will not penetrate very deeply due to the strong interaction with the material, and such a fake can be easily rejected due to the poor quality of the muon matching. In the second case, the requirement to preserve the momentum limits the phase space of the relevant decays, but the event yield is still fairly large for kaons and pions. These fakes are very hard to suppress since they have a muon with momentum close to the original hadron. Now I will discuss the primary source of muon fakes, i.e., hadron decay in flight. The probability of a particle to decay follow the decay law equation,

$$P \propto e^{-\frac{t}{\tau}} = e^{-\frac{L}{c\tau} \frac{m}{p}} \quad (\text{A.1})$$

where L is the flight length of the hadron and p is the momentum of the particle. The average distance ($c\tau$) traveled by pion is $7.8m$ and for kaon is $3.7m$. This signifies that only a small fraction of the pion and kaon survive to the CMS calorimeters.

The above decay probability can be approximated for the flight length significantly shorter than the medium flight length as

$$P = \frac{L}{c\tau} \frac{m}{p} \quad (\text{A.2})$$

In other words, the decay rate is proportional to the flight length of the hadron and inversely proportional to its boost and, therefore, its momentum (p).

The fake rate depends not only on the decay probability but also on the fraction of hadron decays reconstructed as muon tracks. This fraction depends on the decay angle and the decay vertex location in the detector.

A.1 Fake rate measurement procedure

To measure the pion, kaon, and proton fake rates, the suitable decay modes are $K_s^0 \rightarrow \pi^+\pi^-$, $\phi \rightarrow K^+K^-$, and $\Lambda \rightarrow p^+\pi^-$. The sample used for this study is independent of muon trigger to avoid any possible bias due to trigger. As discussed in the previous chapter, the analysis uses an advanced muon identification algorithm to control the peaking and semileptonic background. The algorithm uses Boosted Decision Tree (BDT) from the TMVA package. The fake rates are studied for different muon BDT points, and the data-MC

comparison is performed to check the compatibility.

The list of selection requirements for this study are discussed below.

- For the pion control sample
 - $p_T^\pi > 1\text{GeV}$, $|\eta_\pi| < 1.4$, pion track is high purity
 - $m_{K_s^0} \in [0.45, 0.55]$
 - $\frac{L_{xy}}{\sigma_L} > 3$
 - vertex probability greater than 0.001
 - vertex displacement in XY plane wrt Beam Spot less than 8 (*)
 - cosine of pointing angle in XY wrt BS greater than 0.999
 - impact parameter significance of the candidate trajectory in 3D wrt PV less than 3
 - 2D impact parameter significance for Track 1 and 2 wrt Beam Spot greater than 5
 - kinematic vertex fit χ^2/dof of the two track less than 3
 - distance of closest approach of the tracks less than 0.004
- For the kaon control sample
 - $p_T^K > 3\text{GeV}$, $|\eta_K| < 1.4$, kaon track is highPurity
 - $m_\phi \in [1.00, 1.04]$
 - vertex displacement in XY plane wrt Beam Spot less than 4

- vertex probability greater than 0.3
- impact parameter significance of the candidate trajectory in 3D wrt PV less than 1
- distance of closest approach of tracks less than 0.004
- For the proton control sample
 - $p_T^p > 4\text{GeV}$, proton track is highPurity
 - $m_\Lambda \in [1.095, 1.14]$
 - $\frac{L_{xy}}{\sigma_L} > 3$
 - impact parameter significance of the candidate trajectory in 3D wrt PV less than 1
 - 2D impact parameter significance for Track 1 and 2 wrt Beam Spot greater than 2

The requirement on the flight length (marked as (*)) in the K_s^0 meson selection is motivated from the fake rate dependence as a function of flight length, as shown in Fig. [A.1](#).

The fake rate dramatically decreases for larger flight lengths with the muon BDT selection, which could be the effect of the requirement on the pixel and silicon strip hits for the inner track in muon BDT. This selection also allows us to reject K_s^0 events whose flight lengths are larger than the B meson. In this way, the K_s^0 events are restricted to be with the first few layers of the pixel layers. Since the ϕ meson has a shorter lifetime, the selection cut on the displacement is not required.

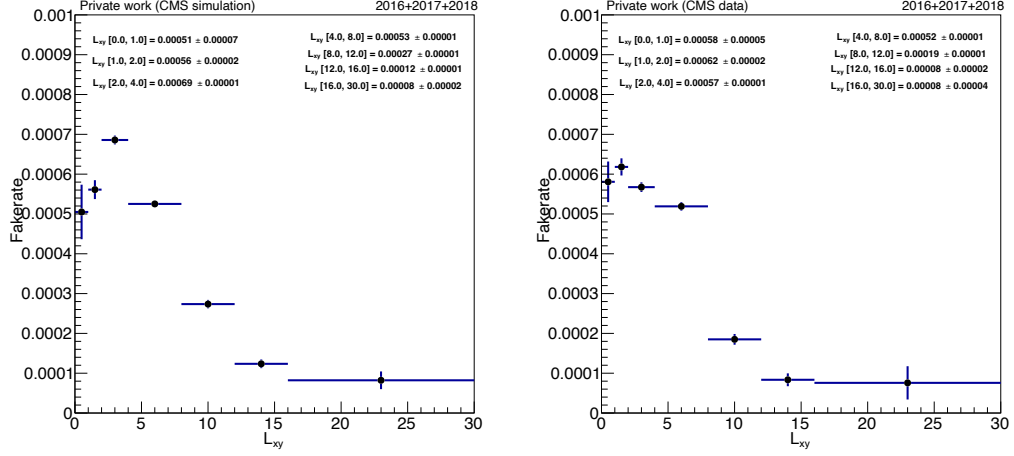


Figure A.1: Muon fake rates from $K_s^0 \rightarrow \pi\pi$ decays as a function of the transverse flight length from MC simulation (left) and data (right) for all the year.

The fake rate for pions is measured by taking the ratio of event yield in $K_s^0 \rightarrow \mu\pi$ and $K_s^0 \rightarrow \pi\pi$ decay. The event yields are calculated from the extended binned maximum likelihood fit to $\pi\pi$ and $\mu\pi$ invariant mass distribution. The models used to describe the signal and combinatorial background component are a triple Gaussian function and a 2nd order Bernstein polynomial function, respectively. The signal shape in the $K_s^0 \rightarrow \mu\pi$ distribution is fixed from the $K_s^0 \rightarrow \pi\pi$ distribution to avoid the different signal shape due to the low statistics.

The fake rate is estimated in bins of muon p_T . For each $K_s^0 \rightarrow \pi\pi$ candidate we check p_T and acceptance for both pions and add the K_s^0 candidate to the appropriate p_T bin. The same K_s^0 candidate may enter the fit twice if both pions fall within one muon p_T bin - it is just a reflection of the fact that we have two pions and, therefore, the probability of observing a muon decay increases by a factor of two.

A similar procedure is followed for the measurement of the kaon fake rate. The fit models for signal and combinatorial background are the Voigtian function and Bernstein function, respectively.

The proton fake rate is measured using $\Lambda \rightarrow p\pi$ and $\Lambda \rightarrow \mu\pi$ decay mode. Due to low statistics in $\Lambda \rightarrow \mu\pi$, the fake rate is measured in the p_T range $[0., 50.]$ GeV.

Figure [A.2](#) shows the fit projections for $K_s^0 \rightarrow \pi\pi$, $\phi \rightarrow KK$ and $\Lambda \rightarrow p\pi$.

A.2 Muon selector

The fake rate measurement is performed for the following muon selectors

- Loose (used as a preselection for all other selections)
 - $p_T > 4\text{GeV}$, $|\eta| < 1.4$, innerTrack is highPurity
 - isTrackerMuon, isGlobalMuon, and isLooseMuon
- Medium
 - Loose selection, and CutBasedIdMedium
- BDT
 - Loose selection
 - Soft muon BDT with the following thresholds
 - * 0.2, 0.3, 0.4, 0.45, 0.5, 0.55, default (0.58), and 0.6

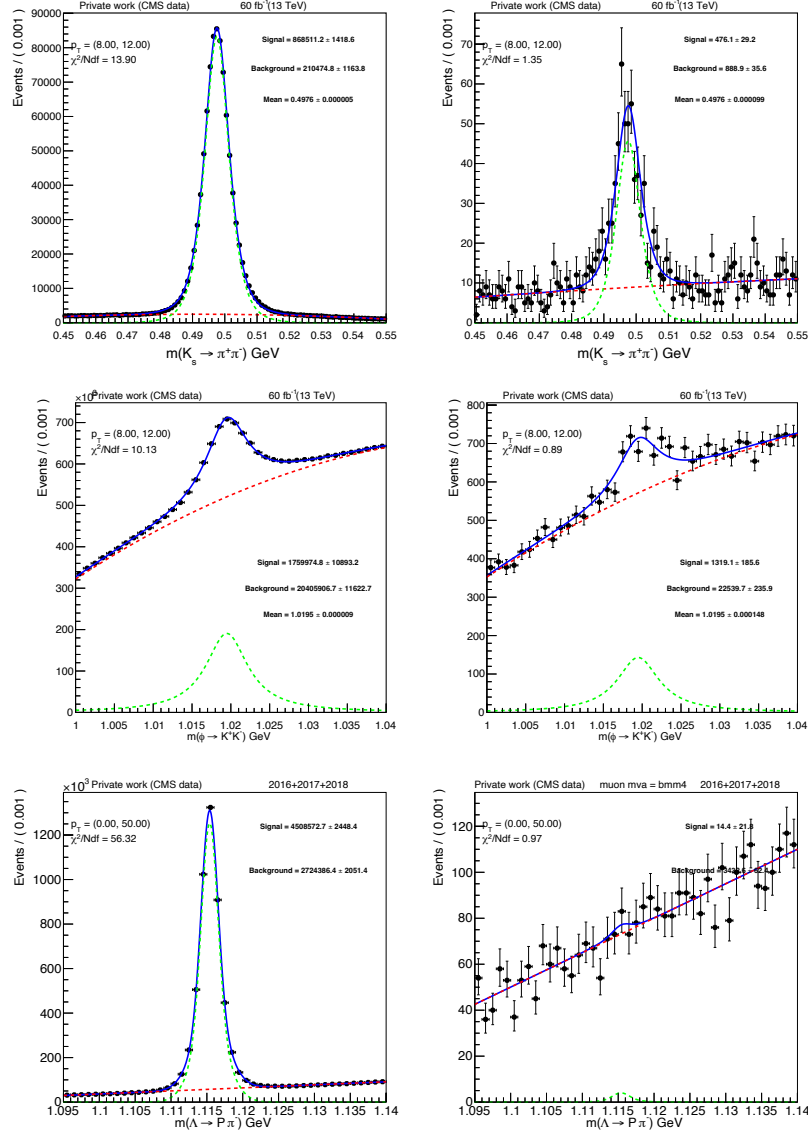


Figure A.2: Invariant mass projections from the $K_s^0 \rightarrow \pi\pi$ (top row), $\phi \rightarrow KK$ (middle row) and $\Lambda \rightarrow \mu\pi$ (bottom row) decay. The fit projections for K_s^0 and ϕ are in the μ p_T range 8-12.0 GeV, where as projection with full p_T range is shown for Λ . The left and right side plots are before and after soft muon BDT cut, respectively. The green dashed curve is for the signal distribution, the red dotted curve is for combinatorial background and the result of the fit projection is the solid blue curve. The data are shown as the solid black circles with error bars.

A.3 Results

The pion and kaon fake rate for data and MC, using Loose, Medium and muon BDT selections with different thresholds are shown in Figs. [A.3](#) to [A.4](#). The ratio of data to MC fake rates for kaon and pion are shown in Tables [A.1](#) to [A.2](#). The figures and the tables are only for 2018, other years are not shown but have similar trend.

The proton fake rate for data and MC and ratio of data to MC fake rates, using Loose, Medium and BDT selections with different thresholds, are shown in Table [A.3](#). For the tight muon BDT selection, the observed $\Lambda \rightarrow p\pi$ yields both in data and MC are not significant. Therefore an upper limit on the proton fake rate is estimated by taking 95% of the area of the probability density.

The agreement between the data and MC proton fake rate is reasonably good for loose selections. For tighter selection, data fake rate is smaller than MC. It also indicates that the tighter selection will fairly reject the fakes originating from proton.

A.3.0.0.1 Systematics The systematic effects related to fake rate measurements are studied thoroughly and categorized as follows:

- Method systematics
- Vertex displacement selection (for $K_s^0 \rightarrow \pi\pi$)
- Monte Carlo sample selection
- Pile up

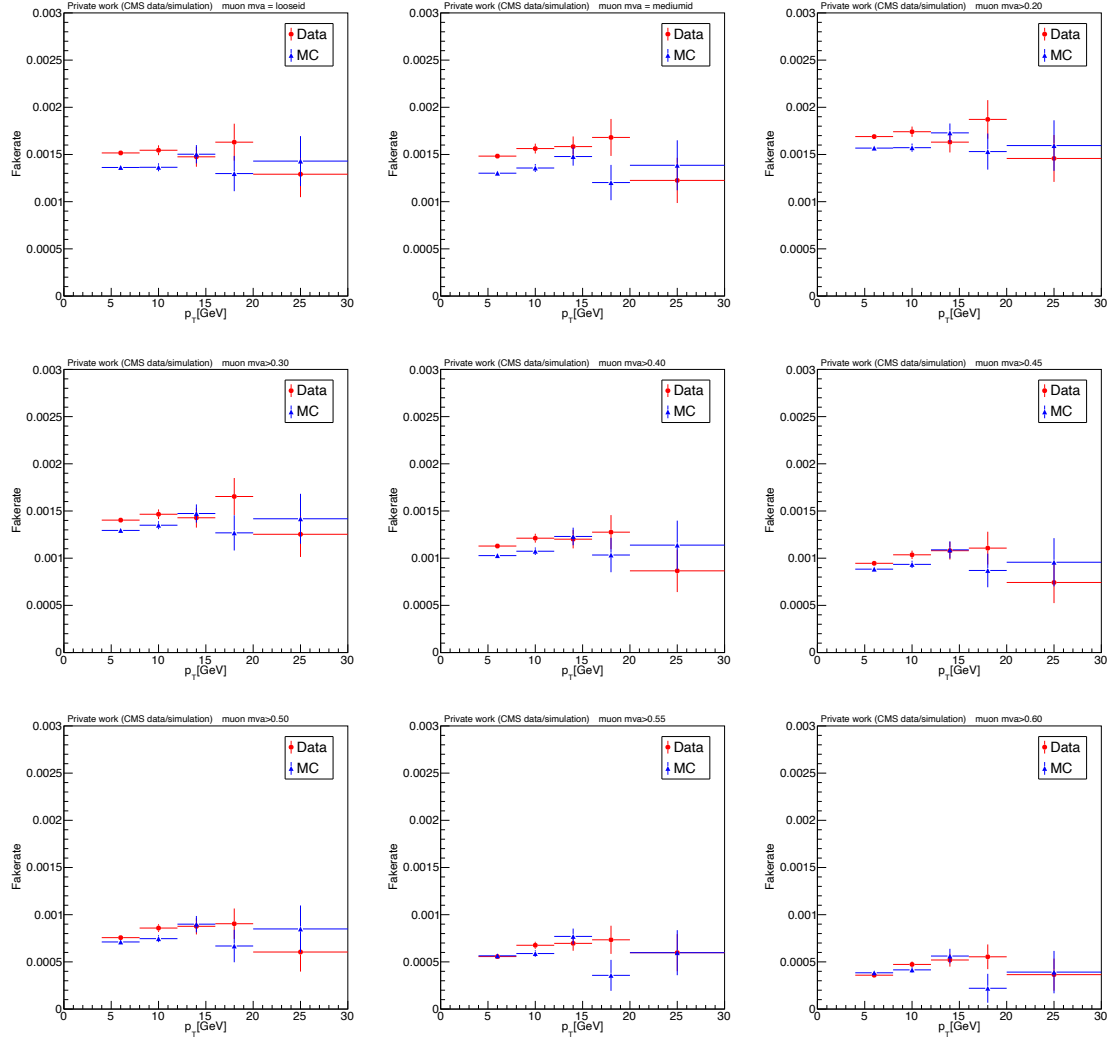


Figure A.3: Pion fake rates from 2018 data and MC as a function muon p_T for various selections.

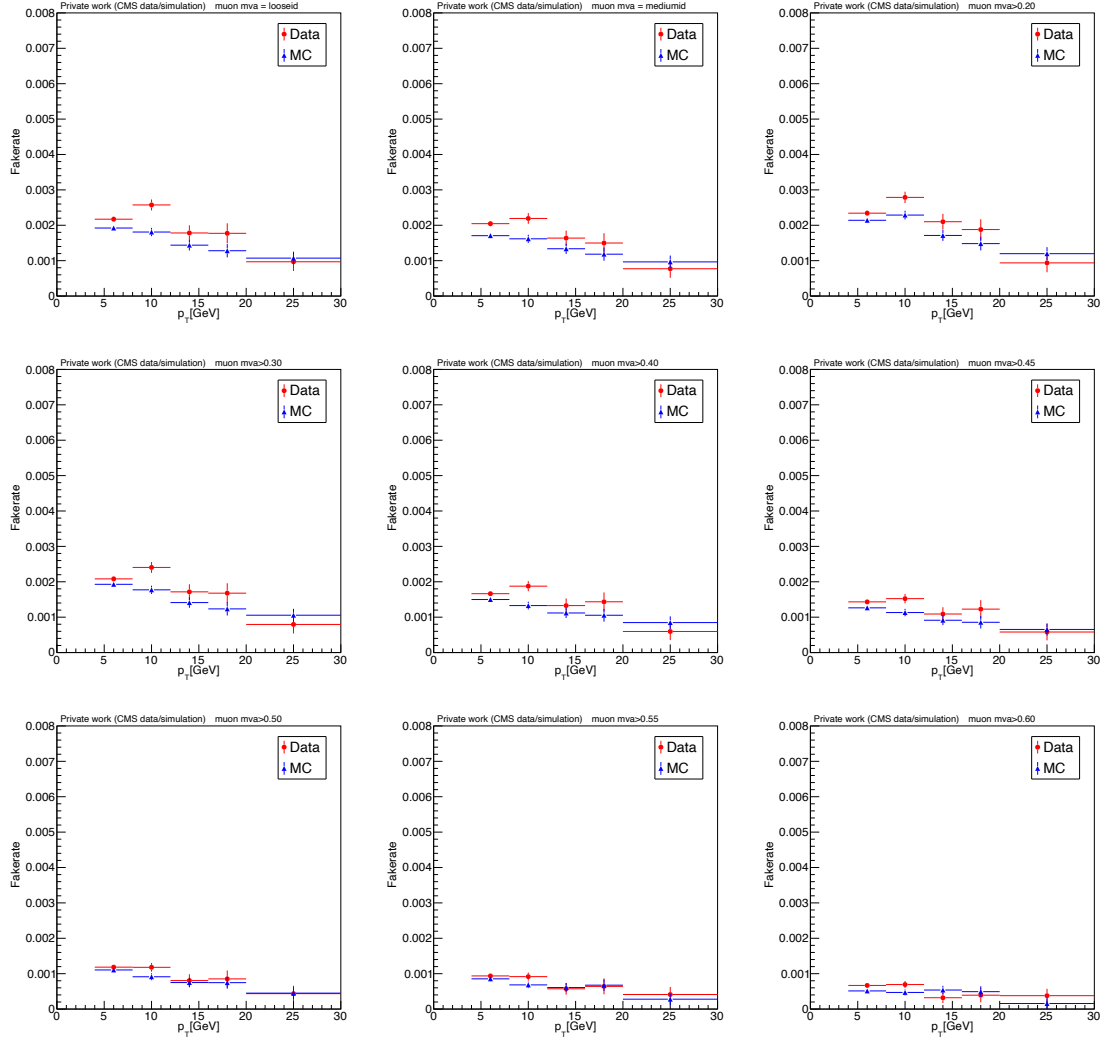


Figure A.4: Kaon fake rates from 2018 data and MC as a function muon p_T for various selections.

Selection	4-8 GeV	8-12 GeV	12-16 GeV	16-20 GeV	20-30 GeV
default_2018	0.93 ± 0.01	1.31 ± 0.10	0.80 ± 0.10	1.31 ± 0.55	0.53 ± 0.22
looseid_2018	1.10 ± 0.00	1.18 ± 0.03	0.94 ± 0.06	1.17 ± 0.17	0.79 ± 0.16
mediumid_2018	1.15 ± 0.00	1.21 ± 0.03	1.00 ± 0.06	1.35 ± 0.22	0.69 ± 0.13
mvabdt20_2018	1.07 ± 0.00	1.13 ± 0.02	0.90 ± 0.05	1.17 ± 0.15	0.82 ± 0.15
mvabdt30_2018	1.08 ± 0.00	1.13 ± 0.03	0.93 ± 0.06	1.25 ± 0.20	0.78 ± 0.16
mvabdt40_2018	1.10 ± 0.01	1.18 ± 0.04	0.92 ± 0.07	1.14 ± 0.21	0.61 ± 0.15
mvabdt45_2018	1.07 ± 0.01	1.16 ± 0.04	0.93 ± 0.08	1.08 ± 0.22	0.62 ± 0.18
mvabdt50_2018	1.05 ± 0.01	1.22 ± 0.06	0.93 ± 0.09	1.06 ± 0.26	0.56 ± 0.19
mvabdt55_2018	0.96 ± 0.01	1.27 ± 0.08	0.83 ± 0.10	1.18 ± 0.39	0.71 ± 0.27
mvabdt60_2018	0.90 ± 0.02	1.34 ± 0.12	0.77 ± 0.10	1.28 ± 0.58	0.50 ± 0.24

Table A.1: Pion fake rate ratio between data and MC simulation for various selections in 2018.

Selection	4-8 GeV	8-12 GeV	12-16 GeV	16-20 GeV	20-30 GeV
default_2018	1.23 ± 0.04	1.45 ± 0.14	0.78 ± 0.13	0.82 ± 0.20	1.76 ± 0.70
looseid_2018	1.13 ± 0.03	1.42 ± 0.07	1.24 ± 0.13	1.38 ± 0.21	0.90 ± 0.16
mediumid_2018	1.20 ± 0.03	1.36 ± 0.08	1.23 ± 0.14	1.27 ± 0.21	0.80 ± 0.14
mvabdt20_2018	1.09 ± 0.03	1.22 ± 0.05	1.23 ± 0.11	1.27 ± 0.18	0.78 ± 0.12
mvabdt30_2018	1.08 ± 0.03	1.36 ± 0.07	1.21 ± 0.13	1.36 ± 0.22	0.75 ± 0.11
mvabdt40_2018	1.11 ± 0.03	1.41 ± 0.09	1.19 ± 0.15	1.36 ± 0.23	0.70 ± 0.11
mvabdt45_2018	1.13 ± 0.03	1.35 ± 0.09	1.19 ± 0.17	1.43 ± 0.28	0.89 ± 0.22
mvabdt50_2018	1.07 ± 0.03	1.29 ± 0.10	1.07 ± 0.17	1.14 ± 0.25	0.97 ± 0.33
mvabdt55_2018	1.10 ± 0.03	1.34 ± 0.12	0.94 ± 0.17	0.95 ± 0.21	1.47 ± 0.64
mvabdt60_2018	1.30 ± 0.04	1.47 ± 0.13	0.60 ± 0.05	0.80 ± 0.20	2.41 ± 0.63

Table A.2: Kaon fake rate ratio between data and MC simulation for various selections in 2018.

Selection	MC fake rate ($\times 10^{-4}$)	Data fake rate ($\times 10^{-4}$)	Ratio (Data/MC)
looseid	2.377 ± 0.097	2.860 ± 0.110	1.203 ± 0.067
mediumid	2.362 ± 0.096	3.060 ± 0.109	1.295 ± 0.070
mvabdt20	2.508 ± 0.098	2.556 ± 0.106	1.019 ± 0.058
mvabdt30	1.491 ± 0.090	1.512 ± 0.091	1.014 ± 0.087
mvabdt40	0.825 ± 0.084	0.741 ± 0.076	0.898 ± 0.130
mvabdt45	0.610 ± 0.081	0.453 ± 0.068	0.743 ± 0.149
mvabdt50	0.419 ± 0.077	0.213 ± 0.060	0.508 ± 0.170
mvabdt55	0.286 ± 0.073	0.086 ± 0.053	0.299 ± 0.200
Bmm4 default	0.287	0.093	0.325
mvabdt60	0.251	0.062	0.248

Table A.3: Proton fake rates in MC, data and the ratio between data and MC simulation for various thresholds with all available data. For the default and BDT > 0.60 selection, we estimate the fake rate at 95% CL.

The fake rate measurements use the $K_s^0 \rightarrow \pi\pi$ and $\phi \rightarrow KK$ yields from the fit to their invariant mass distributions. An alternative approach is tried in which the fake rate is directly measured by matching muons to pions and kaons in the $B^0 \rightarrow K^+\pi^-$ MC sample. The second method uses geometrical matching and may select different fakes. Since the final $B_s^0 \rightarrow \mu^+\mu^-$ branching fraction is obtained from the fit, the first approach is more appropriate for this study. Averaging over all years, the fake rate is found to be off by 25% for pion and 20% for kaons. To be conservative, these numbers are taken as systematic uncertainties. The second dominant systematic effect comes from the vertex displacement cut on the K_s^0 which is less than 8 cm. The variation of fake rate in different flight length bins is estimated and found to be 16% for MC and 10% for data. The 25% systematic uncertainty assigned above for the fitting method should cover this effect; that is why it is not necessary to increase the value.

In order to assess the potential sample dependence systematics, the fake rate is evaluated

in different MC samples. Similarly, the fake rate is measured as a function of the different number of primary vertices in an event. Ultimately, both contributions are neglected because the deviations are insignificant with respect to the average.

Appendix B

More results on Run 2 analysis

Table [B.1](#) summarizes the branching fraction results obtained from the branching fraction fitter with subsets of data.

Configuration	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	$\mathcal{B}(B^0 \rightarrow \mu^+ \mu^-)$	χ^2/dof	
			$d_{\text{MVA}} > 0.99$	$d_{\text{MVA}} \in [0.9, 0.99]$
Full data	$3.83^{+0.42}_{-0.40} \times 10^{-9}$	$0.37^{+0.75}_{-0.68} \times 10^{-10}$	$6.95 / 20 = 0.35$	$23.9 / 20 = 1.20$
2016 only	$3.18^{+0.81}_{-0.73} \times 10^{-9}$	$-0.62^{+1.50}_{-1.32} \times 10^{-10}$	$5.99 / 19 = 0.32$	$19.3 / 20 = 0.97$
2017 only	$4.11^{+0.75}_{-0.69} \times 10^{-9}$	$-0.10^{+1.15}_{-1.04} \times 10^{-10}$	$11.3 / 20 = 0.57$	$10.1 / 20 = 0.51$
2018 only	$3.95^{+0.61}_{-0.57} \times 10^{-9}$	$0.66^{+1.10}_{-1.06} \times 10^{-10}$	$9.23 / 20 = 0.46$	$20.6 / 20 = 1.03$
channel 0 only	$4.35^{+0.61}_{-0.56} \times 10^{-9}$	$0.06^{+0.86}_{-0.82} \times 10^{-10}$	$5.23 / 20 = 0.26$	$26.3 / 20 = 1.32$
channel 1 only	$3.28^{+0.56}_{-0.51} \times 10^{-9}$	$0.93^{+1.28}_{-1.08} \times 10^{-10}$	$6.73 / 20 = 0.34$	$11.2 / 20 = 0.56$

Table B.1: A comparison of branching fraction fits with subsets of data.

Era	Channel	$\chi^2 / \text{number of bins}$	
		$d_{\text{MVA}} > 0.99$	$d_{\text{MVA}} \in [0.9, 0.99]$
2016BF	0	$6.31 / 20 = 0.32$	$10.4 / 20 = 0.52$
2016BF	1	$11.3 / 20 = 0.57$	$11.6 / 20 = 0.58$
2016GH	0	$5.49 / 20 = 0.27$	$14.0 / 20 = 0.70$
2016GH	1	$7.14 / 20 = 0.36$	$15.8 / 20 = 0.79$
2017	0	$28 / 20 = 1.40$	$24.4 / 20 = 1.22$
2017	1	$12.5 / 20 = 0.63$	$6.90 / 20 = 0.35$
2018	0	$4.96 / 20 = 0.25$	$26.7 / 20 = 1.34$
2018	1	$13.6 / 20 = 0.68$	$10.1 / 20 = 0.51$
Combined	-	$6.95 / 20 = 0.35$	$23.9 / 20 = 1.20$

Table B.2: The χ^2 test for the mass projection for each category after the unblinded fit.

Table [B.3](#) summarizes the branching fraction and lifetime result obtained from lifetime fitter with subsets of data.

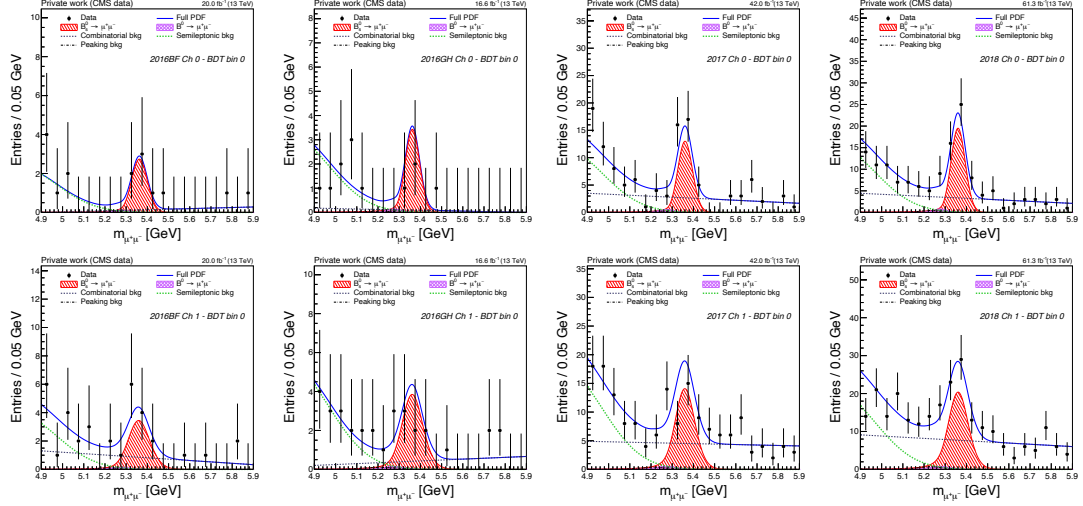


Figure B.1: The dimuon invariant mass distributions of the unblinded fit in the $d_{\text{MVA}} > 0.99$ categories, with the PDF models used in the generation overlaid. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively.

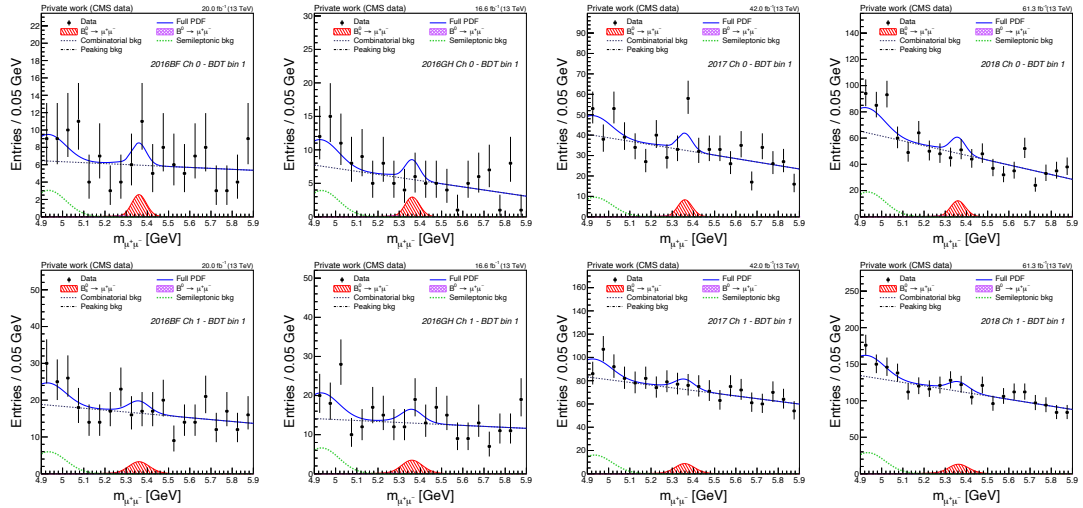


Figure B.2: The dimuon invariant mass distributions of the unblinded fit in the $0.99 > d_{\text{MVA}} > 0.90$ categories, with the PDF models used in the generation overlaid. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively.

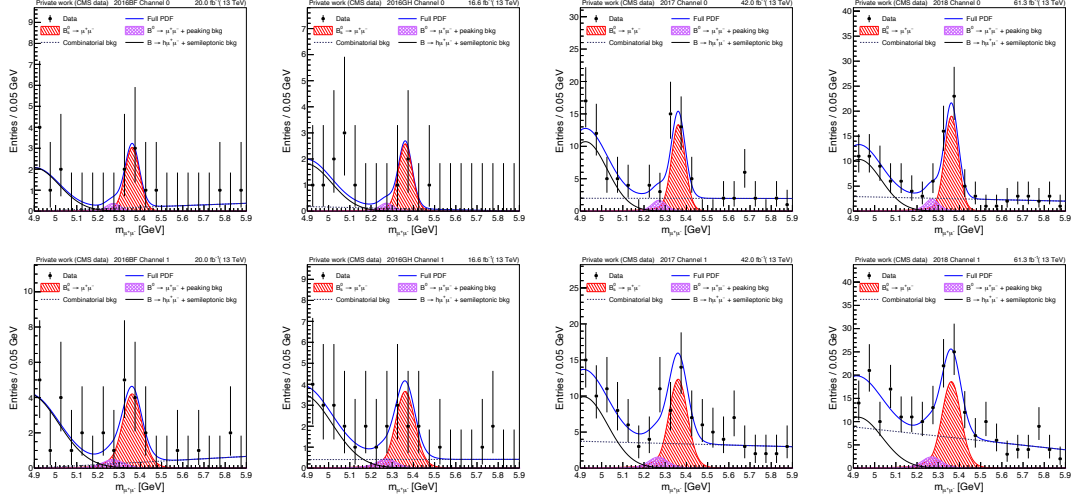


Figure B.3: The dimuon invariant mass distributions of the unblinded fit in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit. The mass distributions are obtained from the decay time range $[1, 14]$ ps. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively. The result of the fit is overlaid (blue line) and the different components detailed in the legend.

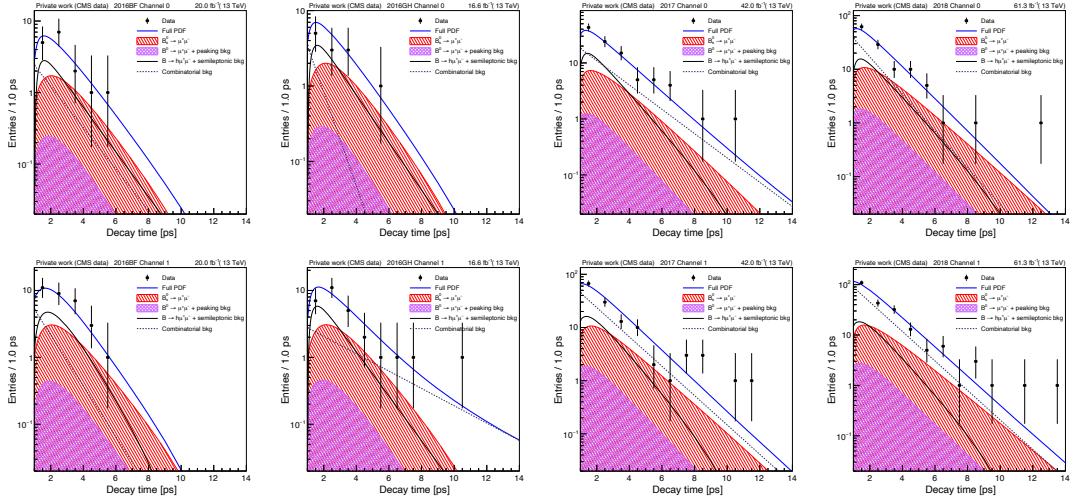


Figure B.4: The decay time distributions (on a log scale) in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively. The result of the fit is overlaid (blue line) and the different components detailed in the legend.

Configuration	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	$\tau_{B_s^0}$ ps
Full data	$3.89^{+0.42}_{-0.40} \times 10^{-9}$	$1.83^{+0.23}_{-0.20}$
2016 only	$3.07^{+0.79}_{-0.70} \times 10^{-9}$	$2.06^{+0.84}_{-0.51}$
2017 only	$3.86^{+0.75}_{-0.70} \times 10^{-9}$	$1.86^{+0.45}_{-0.37}$
2018 only	$4.33^{+0.62}_{-0.59} \times 10^{-9}$	$1.74^{+0.31}_{-0.24}$
channel 0 only	$4.57^{+0.65}_{-0.61} \times 10^{-9}$	$1.66^{+0.26}_{-0.21}$
channel 1 only	$3.30^{+0.53}_{-0.49} \times 10^{-9}$	$2.10^{+0.46}_{-0.35}$

Table B.3: A comparison of branching fraction and lifetime with subsets of data.

Era	Channel	χ^2 / number of bins		
		mass	decay time	decay time uncertainty
2016BF	0	4.19 / 20 = 0.21	1.29 / 13 = 0.10	4.06 / 30 = 0.14
2016BF	1	6.42 / 20 = 0.32	1.46 / 13 = 0.11	6.46 / 30 = 0.22
2016GH	0	4.72 / 20 = 0.24	1.21 / 13 = 0.10	2.05 / 30 = 0.07
2016GH	1	4.22 / 20 = 0.21	2.82 / 13 = 0.22	9.73 / 30 = 0.32
2017	0	13.81 / 20 = 0.69	4.76 / 13 = 0.36	7.41 / 30 = 0.24
2017	1	8.13 / 20 = 0.41	10.66 / 13 = 0.82	11.27 / 30 = 0.37
2018	0	4.09 / 20 = 0.20	6.33 / 13 = 0.49	6.77 / 30 = 0.23
2018	1	10.69 / 20 = 0.53	11.46 / 13 = 0.88	20.21 / 30 = 0.67
$d_{\text{MVA}} > 0.99$ combined	-	5.37 / 20 = 0.27	33.47 / 52 = 0.64	32.35 / 30 = 1.08

Table B.4: The χ^2 test for the mass and decay time projection of each category from the unblinded lifetime fit.

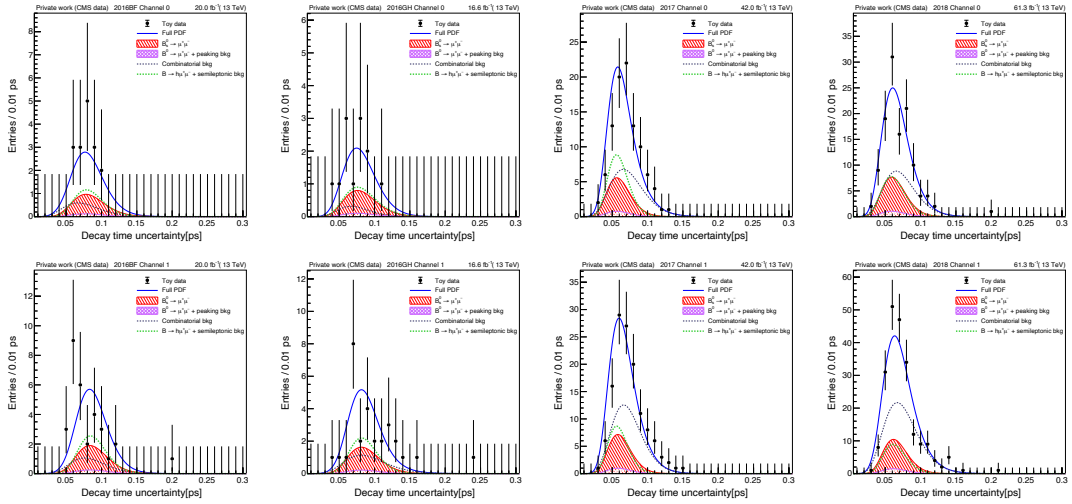


Figure B.5: The decay time uncertainty distributions (on a log scale) in the $d_{\text{MVA}} > 0.99$ categories obtained from the lifetime fit. The events for the categories in the channel 0 (channel 1) categories are shown in the upper (lower) row. From left to right, the plots show the events for 2016BF, 2016GH, 2017, and 2018 eras, respectively. The result of the fit is overlaid (blue line) and the different components detailed in the legend.

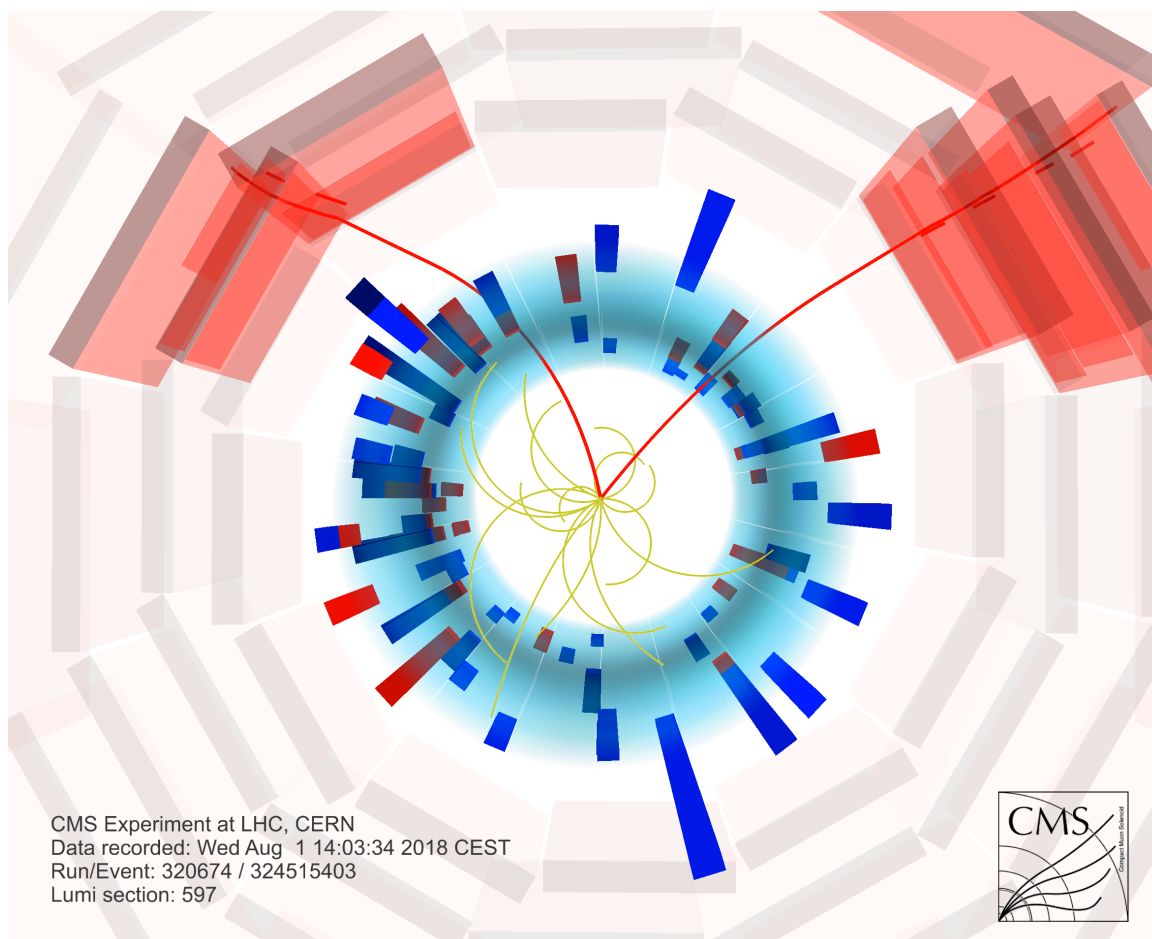


Figure B.6: Visualization of $B_s^0 \rightarrow \mu^+ \mu^-$ candidate in 2018 data. The two muons produced in the disintegration of the B_s^0 meson are represented by the red lines. Additional curved lines depict charged-particle tracks that also originate from the same collision point. The calorimeter's blue and red towers represent the deposits of hadrons, electrons, and photons, respectively. Tracks originating from other interactions during the event have been eliminated to enhance clarity [105].

References

- [1] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 122 (2019) 222001
- [2] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 125 (2020) 242001
- [3] P. A. M. Dirac, Proc. Roy. Soc. Lond. A117 (1928) 610-624
- [4] J. S. Schwinger, Phys. Rev. 73 (1948) 416-417
- [5] R. P. Feynman, Phys. Rev. 76 (1949) 769-789
- [6] R. P. Feynman, Phys. Rev. 76 (1949) 749-759
- [7] P. W. Higgs, Phys. Rev. Lett. 13 (1964) 508-509
- [8] F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964) 321
- [9] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, Phys. Rev. Lett. 13 (1964) 585-587
- [10] S.L. Glashow, Nucl.Phys. 22 (1961), pp. 579-588
- [11] Abdus Salam and John Clive Ward, Phys.Lett. 13 (1964), pp. 168-171
- [12] Steven Weinberg. Phys. Rev. Lett. 19 (1967), pp. 1264-1266
- [13] O.W. Greenberg, American Journal of Physics 50, 1074 (1982)
- [14] S.L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D 2(1970), pp. 1285-1292
- [15] J.J. Aubert et al., Phys. Rev. Lett. 33 (1974), pp. 1404-1406
- [16] J.E. Augustin et al., Phys. Rev. Lett. 33 (1974), pp. 1406-1408

- [17] D.J. Gross and Frank Wilczek, Phys. Rev. Lett. 30 (1973), pp. 1343-1346
- [18] H. David Politzer, Phys. Rev. Lett. 30 (1973), pp. 1346-1349
- [19] Martin L. Perl et al., Phys. Rev. Lett. 35 (1975), pp. 1489-1492
- [20] S.W. Herb et al., Phys. Rev. Lett. 39 (1977), pp. 252-255
- [21] F.Abe et al., Phys. Rev. Lett. 74 (1995), pp. 2626-2631
- [22] CMS Collaboration, S. Chatrchyan et al., Phys. Lett. B 716 (2012) 30-61
- [23] ATLAS Collaboration, Georges Aad et al., Phys. Lett. B 716 (2012) 1-29
- [24] https://en.wikipedia.org/wiki/Standard_Model
- [25] Lisa Randall and Raman Sundrum, Phys. Rev. Lett. 83 (1999) 3370-3373
- [26] Martin Schmaltz and David Tucker-Smith, Ann. Rev. Nucl. Part. Sci. 55 (2005) 229-270
- [27] Super-Kamiokande Collaboration, Yoshiyuki Fukuda et al., Phys. Rev. Lett. 81 (1998) 1562
- [28] Planck Collaboration, Nabila Aghanim et al., A&A 641, A6 (2020)
- [29] Rabindra Mohapatra and Alexei Smirnov, Ann. Rev. Nucl. Part. Sci. 56 (2006) 569
- [30] D. I. Kazakov, arXiv:hep-ph/0012288v2
- [31] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 127 (2021) 151801

- [32] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 125 (2020) 011802
- [33] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 126 (2021) 161802
- [34] CMS Collaboration, A. M. Sirunyan et al., Phys. Lett. B 781 (2018) 517
- [35] CMS Collaboration, A. M. Sirunyan et al., JHEP 04 (2021) 124
- [36] ATLAS Collaboration, M. Aaboud et al., JHEP 10 (2018) 047
- [37] A. Arbey, M. Battaglia, F. Mahmoudi, and D. Martinez Santos, Phys. Rev. D 87 (2013) 035026
- [38] N. Cabibbo, Phys. Rev. Lett. 10 (1963) 531
- [39] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. 49, 652 (1973)
- [40] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk and N. Tuning, Phys. Rev. D 86, 014027 (2012)
- [41] R. Fleischer and R. Knegjens, Eur. Phys. J. C 71, 1789 (2011)
- [42] Martin Beneke, Christoph Bobeth and Robert Szafron, JHEP 10 (2019) 232
- [43] Andrzej J. Buras, R. Fleischer, J. Girrbach and R. Knegjens, JHEP 1307 (2013) 77
- [44] G. D'Ambrosio, G.F. Giudice, G. Isidori, and A. Strumia, Nucl. Phys. B 645:155-187, 2002
- [45] K. Babu and C. F. Kolda, Phys. Rev. Lett. 84 (2000) 228-231

- [46] CMS Collaboration, S. Chatrchyan et al., Phys. Rev. Lett. 111 (2013) 101804
- [47] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 111 (2013) 101805
- [48] CMS and LHCb Collaboration, V. Khachatryan et al., Nature 522, 68-72 (2015)
- [49] http://www.scholarpedia.org/article/Rare_decays_of_b_hadrons
- [50] A. L. Read, J. Phys. G 28 (2002) 2693
- [51] Lyndon Evans and Philip Bryant, JINST 3 (2008) S08001
- [52] https://hilumilhc.web.cern.ch/sites/default/files/HL-LHC_Janvier2022.pdf
- [53] ATLAS Collaboration, G. Aad et al., JINST 3 (2008) S08003
- [54] ALICE Collaboration, K. Aamodt et al., JINST 3 (2008) S08002
- [55] CMS Collaboration, S. Chatrchyan et al., JINST 3 (2008) S08004
- [56] LHCb Collaboration, Jr. A. Augusto Alves et al., JINST 3 (2008) S08005
- [57] <https://www.home.cern/science/accelerators/accelerator-complex>
- [58] https://cmslumi.web.cern.ch/publicplots/pileup_allYears_stack.pdf
- [59] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>
- [60] https://tikz.net/axis3d_cms/
- [61] L.Viliani for the CMS Collaboration, Nucl. Instrum. Methods Phys. Res., A 824 (2016) 67-69

- [62] CMS Collaboration, W. Adam et al., JINST 16 (2021) P02027
- [63] Giovanni Mocellin on behalf of the CMS Muon Group 2019 J. Phys. Conf. Ser. 1390 012116
- [64] M. Tosi on behalf of CMS Collaboration, PoS (EPS-HEP2017) 523
- [65] CMS Collaboration, A. M. Sirunyan et al., JINST 13 (2018) P06015
- [66] CMS Collaboration, S. Davis, URL:<https://cds.cern.ch/record/2205172>
- [67] CMS Collaboration, A. M. Sirunyan et al., JINST 12 (2017) P10003
- [68] CMS Collaboration, Paolo Azzurri, Nucl. Phys. Proc. Suppl. 197 (2009) 275-278
- [69] P. Billoir and S. Qian, Nucl. Instrum. Meth. A 294 (1990) 219
- [70] J M Campbell et al., Rep. Prog. Phys. 70 (2007) 89
- [71] E. Norrbin and T. Sjostrand., Eur. Phys. J. C 17 (2000) 137-161
- [72] LHCb Collaboration, R. Aaij et. al., Phys. Rev. D 104, 032005 (2021)
- [73] CMS Collaboration, A. M. Sirunyan et al., JINST 15 (2020) P10017
- [74] T. Sjöstrand et al., 191:159177, 2015
- [75] D.J. Lange. Nucl. Instrum. Meth. A 462 (2001), pp. 152-155
- [76] S. Agostinelli et al. Nucl. Instrum. Meth. A, 506(3):250303, 2003
- [77] CMS Collaboration, S. Chatrchyan et al., JINST 7 (2012) P10002

- [78] H. Voss, A. Höcker, J. Stelzer and F. Tegenfeldt, PoS(ACAT)040
- [79] CMS Collaboration, S. Chatrchyan et al., Phys. Rev. Lett. 111 (2013) 101804
- [80] CMS Collaboration, A. M. Sirunyan et al., JINST 9 (2014) P10009
- [81] C. Bobeth et al., Phys. Rev. Lett. 112 (2014) 101801
- [82] PDG Collaboration, M. Tanabashi et al., Phys. Rev. D 98 (2018) 030001
- [83] J. M. Flynn, T. Izubuchi, T. Kawanai, C. Lehner, A. Soni, R. S. Van de Water and O. Witzel, Phys. Rev. D 91, (2015) 074510
- [84] T. Skwarnicki, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.
- [85] K. S. Cranmer, Comput. Phys. Commun. 136 (2001) 198
- [86] S. Bernstein(1912), Comm. Kharkov Math. Soc., 13: 1-2
- [87] G. Cowan, K. Cranmer, E. Gross et al., Cowan, Eur. Phys. J. C 71 (2011) 1554, [Erratum: Eur. Phys. J. C 73 (2013) 2501]
- [88] LHCb Collaboration, R. Aaij et al., Phys. Rev. D 100 (2019) 031102
- [89] T. Junk, Nucl. Instrum. Meth. A 434 (1999) 435
- [90] G. J. Feldman and R. D. Cousins, Phys. Rev. D 57 (1998) 3873
- [91] S. S. Wilks, Annals Math. Statist. 9 (1938) 60

- [92] T. Chen and C. Guestrin, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'166, pp. 785-794
- [93] Beijiang Liu, Xian Xiong, Guoyi Hou, Shiming Song, Lin Shen, arXiv:1811.06131
- [94] Particle Data Group Collaboration, PTEP 2020 (2020), no.8, 083C01
- [95] Giovanni Punzi, ECONF C030908:WELT002,2003
- [96] LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett. 126 (2021) 081804
- [97] CMS Collaboration, A. M. Sirunyan et al., JINST 16 (2021) P07001
- [98] CMS Collaboration, CMS Detector Performance Summary, CMS-DP-2022-012, 2022
- [99] Heavy Flavor Averaging Group, Y. Amhis et al., arXiv:1412.7515
- [100] LHCb Collaboration, R Aaij et al., Phys. Rev. D 104 (2021) 032005
- [101] ATLAS, CMS and LHCb Collaborations, CMS-PAS-BPH-20-003, URL:
<https://cds.cern.ch/record/2727216>
- [102] Li-Sheng Geng et al., Phys. Rev. D 104 (2021) 035029
- [103] LHCb Collaboration, R. Aaij et al., arXiv:2212.09152v1
- [104] CMS Collaboration, A. M. Sirunyan et al., Phys. Lett. B 816 (2021) 136188
- [105] CMS Collaboration, A. M. Sirunyan et al., <https://cds.cern.ch/record/2815125>